PRIFYSGOL ABERYSTWYTH UNIVERSITY

Accessing and Dissecting Useful Genetic Variations for Iron and Zinc Content in Pearl Millet (*Pennisetum glaucum*) Using Genome Wide Association Studies

by

Hanna Rose Manwaring

A thesis submitted at the Institute of Biological, Environmental and Rural Sciences (Aberystwyth University), for the degree of Doctor of Philosophy.

2018

.

Declaration

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed…… (Candidate)

Date……

STATEMENT 1

This thesis is the result of my own investigations, except where otherwise stated. Where correction services have been used, the extent and nature of the correction is clearly marked in a footnote(s).

Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed…… (Candidate)

Date……

STATEMENT 2

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed…… (Candidate)

Date……

Word Count of Thesis: 79,443

.

Acknowledgements

This thesis would not have been possible without the funding supplied by the iCASE studentship funded by the Biotechnology and Biological Sciences Research Council (BBSRC), with additional support from Unilever.

There are so many people who have made this journey possibly.

First and foremost, I would like to thank my supervisory team: Dr Rattan Yadav, Dr Matthew Hegarty, Dr Frances Bligh and Dr Ana Winters for their time, guidance, contributions and for placing many wonderful opportunities in front of me. My appreciation also extends to those who have supported my laboratory work including: Dr Charly Morgan, Dr Barbara Huack, Dr David Messenger, Dr Ruth Sanderson, Ms Caron Evans and Dr Sarah Beynon. Also, a special thanks to Mrs Andi Lovatt, your kindness and support over the years has meant the world to me.

For all your love, I would like to thank my soul sisters Charly, Amy and Cat. Thank you for all the amazing memories we have shared together. In you all I have found lifelong friendship, which I will cherish forever. I have also enjoyed being Auntie Hanna to 3 incredible children; Jack, Alfie and Mari.

Special thanks to my Mum and Dad for their continued support and for always asking me how my PCRs were/how many words I have written so far. Also, thanks to my Nan, for saying our prayer to keep me safe and for occasionally sending the dog some steak money.

Last but definitely not least, I would like to thank my boyfriend Andrew (Andypeep), who has always been there for me. I will love you forever and always, in all ways xxxxxx.

.

Table of Contents

.

.

.

.

Summary

Iron/zinc deficiencies cause extensive health problems in developing countries and contribute to a loss of human potential. Many who suffer from micronutrient deficiencies are dependent on staple crops to meet their dietary requirements. Thus, the biofortification of crop cultivars with elevated levels of grain micronutrients is becoming increasingly attractive and is largely facilitated by genetics/genomic platforms. Pearl millet is an excellent candidate for biofortification as it is accessible to many populations suffering with micronutrient malnutrition. It contains naturally high levels of micronutrients and thrives in dry, semi-arid regions where farming conditions are often unfavourable; therefore it is considered climate change ready.

The aim of this project was to utilise natural genetic variations, present in a germplasm diversity panel of pearl millet to identify genes associated with iron/zinc uptake, with potential to contribute to the development of micronutrient-rich varieties. Iron/zinc levels were quantified in 230 lines by Inductively Coupled Plasma-Atomic Emission Spectroscopy (ICP-AES) and ranged between 29.18–135.27mg/kg and 22.07–93.28mg/kg, respectively. Anti-nutritional factors affecting mineral bioavailability were also considered, including phytate and metal-chelating soluble phenolics. STRUCTURE analysis using Single Nucleotide Polymorphisms (SNPs), generated by Genotyping by Sequencing (GBS) revealed insignificant population structure, further supported by principal component analysis. The extent of Linkage Disequilibrium (LD) was also assessed among all pairs of loci and was found to be prominent on chromosomes 3 and 5. Genome Wide Association Studies (GWAS) resulted in hundreds of significant marker-trait-associations for grain iron/zinc content with p-values ranging from 3.99 E-06–7.54 E-07. Using the 4kb region surrounding the 35 most significant SNPs, a BLAST search of the NCBI database revealed 6 candidate genes associated with iron/zinc uptake. The most significant was YUCCA11, which drives zinc efficiency via auxin biosynthesis. Additionally, haplotypes covering the YUCCA11 gene were identified and their association with trait data was assessed.

List of Abbreviations

ABA abscisic acid
AFLP amplified fragment length polymorphism
Ag aglycone
ANOVA analysis of variance
AQC analytical quality control
Asp aspartic proteinase
ATP adenosine triphosphate
AU[1] Aberdeen University
AU[2] Aberystwyth University
AU[3] absorbance units
Aug August
BAC bacterial artificial chromosome
BBL Bangor breeding line
bHLH basic helix-loop-helix
BLAST basic local alignment search tool
bp base pair
br broad range
BTB broad-complex tramtrack and bric-a-brac
BU Bangor University
C carbon
Ca calcium
CA California (US State)
CAX cation exchanger
CC core collection
Cd cadmium
CDK cyclin-dependent kinase
CG candidate gene
CGT *c*-glycosyl transferase
CHI chalcone isomerase
CISP conserved intron-scanning primer
CO_2 carbon dioxide
CRISPR clustered regularly interspaced short palindromic repeats
Cu copper
CXE carboxylesterase
CXIP4 (CAX)-interacting protein 4
Da daltons
DF degrees of freedom
DM downy mildew
DMSO dimethyl sulfoxide
DNA deoxyribonucleicacid
DS double strand
EC entire collection
ESI electrospray ionisation
EtOH ethanol
E-Value expect value
EXP experiment
EXT extension time
F field

FAO Food and Agriculture Organisation
FDR false discovery rate
Fe iron
FM foxtail millet
FRO2 ferric chelate reductase
FWD forward
F2H flavanone-2-hydroxylase
G glasshouse
g/100g grams per 100grams
Ga Georgia (US State)
Gb gigabyte
GBS genotyping by sequencing
GEI genotype environment interactions
GFP green fluorescent protein
GLM general linear model
GM genetic modification
GS growth stage
GWAS genome wide association studies
Hb haemoglobin
HCL hydrochloric acid
HEXBP hexamer-binding protein
HMA4 heavy metal ATPase 4
HNO_3 nitric acid
HOX homeobox
HL hyperLadder
HPC high performance computer
HPLC high performance liquid chromatography
HPLC-ESR-MS HPLC-electron spin resonance- MS
HPLC-MS HPLC-mass spectrometry
HPLC-PDA HPLC with photodiode array
HPS high pressure sodium
HTA haplotype trait association
H_2O water
H_2O_2 hydrogen peroxide
H_0 null hypothesis
H_1 alternative hypothesis
H^+ proton
I ICRISAT
IAA indole acetic acid
IBERS Institute of Biological, Environmental and Rural Sciences
ICPAES inductively coupled plasma atomic emission spectroscopy
ICRISAT International Crops Research Institute for the Semi-Arid Tropics
IP no. accession number
IR infrared
IRK inwardly rectifying potassium
IS internal standard
ITP iron transport protein
IUPAC International Union of Pure and Applied Chemistry
K^1 potassium
K^2 kinship

Kb kilobyte
kb kilobase
L litre
LC liquid chromatography
LC-PDA-MS^n HPLC with PDA and tandem mass spectrometry detection
LD linkage disequilibrium
LG linkage group
LOD logarithm of the odds
LRR leucine rich repeat
MAF minor allele frequency
MAS marker assisted selection
Mb megabyte
MCC mini core collection
MCMC Markov chain Monte Carlo
MD method development
MDS multi-dimensional scaling
MeOH methanol
Mg magnesium
mg milligram
mGWAS metabolic GWAS
mg/kg milligram/kilogram
mg/100g milligram/100grams
min minute
mL millilitre
MLM mixed linear model
Mn manganese
MND micronutrient deficiency
M_r nominal mass
MRM multiple reaction monitoring
mRNA messenger RNA
MS^1 mass spectrometry
MS^2 mature seed
MTA marker trait association
MTP metal tolerance protein
n nucleotide
N nitrogen
Na sodium
NA nicotianamine
NaCl sodium chloride
NAN not A number
NaOH sodium hydroxide
NAC *No Apical Meristem*
NAS *Nicotianamine Synthase*
NCBI National Center for Biotechnology Information
NGS next generation sequencing
NH_3 ammonia
Ni nickel
nm nanometre
NMR nuclear magnetic resonance
No. number

O oxygen atom
Oct October
OP open pollinated
OPV open pollinated variety
P^1 phosphorus
P^2 phytate assays
PCA principal component analysis
PCR polymerase chain reaction
PDA photodiode array
PDI protein disulfide isomerase
pGWAS phenotypic GWAS
pH potential hydrogen
PHD plant homeodomain
PMiGAP Pearl Millet inbred Germplasm Association Panel
POZ poxvirus and zinc finger protein
Ppm parts per million
p value probability value
QC quality control
QQ quantile-quantile
QTL quantitative trait loci
R Pearson's product moment correlation
RAD-seq restriction site-associated DNA sequencing
Rep replicate
REV reverse
RFLP restriction fragment length polymorphism
RIL recombinant inbred line
RLK receptor-like kinase
RNA ribonucleic acid
ROS reactive oxygen species
RP reverse phase
rpm revolutions per minute
S^1 sulphur
S^2 selfed
s second
SMT seed multiplication trial
SNP single nucleotide polymorphism
SSR single sequence repeat
s.e.m standard error of the mean
Sept September
SPE solid phase extraction
STD standard
TA targeted analysis
TASSEL Trait Analysis by aSSociation, Evolution and Linkage
Tb terabyte
TE tris EDTA
TF transcription factor
TL total length
T_r retention time
TMC total mineral content
TOI trait of interest

TPC two-pore channel
UA untargeted analysis
UDP uridine diphosphate
UPLC-MS/MS ultra performance liquid chromatography –tandem MS
USA United States of America
USD United States dollar
USDA United States Department of Agriculture
UV ultraviolet
V volts
WGS whole genome shotgun
WHO World Health Organisation
x protein
YL young leaf
YSL yellow stripe like
ZIP zrt- and irt-like protein
Zn zinc
µg microgram
µL microlitre
(T)-DNA transfer DNA
°C degree celsius
2-Log 2-Log DNA ladder
3D three-dimensional

List of Figures

List of Tables

Chapter 1: Introduction

Disclaimer: Significant portions of this chapter were used in the publication of:

Manwaring, H.R., Bligh, H.F.J. and Yadav, R., 2016. The challenges and opportunities associated with biofortification of pearl millet (Pennisetum glaucum) with elevated levels of grain iron and zinc. Frontiers in Plant Science, 7, p.1944.

1.1 Micronutrient Malnutrition of Iron and Zinc

On a global scale, over three billion people suffer from chronic micronutrient malnutrition (Chasapis *et al.,* 2012), which refers to extended durations of inadequate intake of food /micronutrients. This is a vast and persistent challenge for global development (Maestre *et al.,* 2017), the consequences of which mostly apply to nutritionally vulnerable populations within developing countries. The economic cost of micronutrient malnutrition due to human capital losses is estimated to be around USD 3.5 trillion per year (Hoddinott, 2013), therefore prevention and treatment is perceived as a desirable worldwide goal.

Even though micronutrient malnutrition encompasses the lack of a wide range of essential vitamins and minerals, deficiencies in iron (Fe) and zinc (Zn) are two of the most common and widespread micronutrient deficiencies (MNDs). These result in; poor health, increased mortality, low work productivity, learning disabilities in children and poor national economic development (Welch & Graham, 2004). At an individual level, children require up to 10 mg of Fe and Zn per day (8 mg for adults). This amount is essential to sustain life and to ensure optimal physiological function. Dietary Fe is available in two forms: haem and non-haem Fe. Fe is complexed as Fe^{2+} (ferrous Fe) to haemoglobin in the haem form, which is available from flesh-food sources (Hyder *et al.,* 2004) and non-haem Fe (Fe^{3+} or ferric Fe) is avialiable from plant-based foods (dark leafy vegetables, brown rice, beans, nuts, and seeds, etc). Haem Fe contributes approximately 15% of total Fe intake in meat-eating populations, but because it has a better absorption rate, estimated to be up to 35% better than non-haem Fe, it can account for up to 40% of total absorbed Fe (Hurrel & Egli, 2010). In developing countries within Africa and Asia, the intake of flesh foods, which are abundant in readily available haem Fe and Zn is at best limited due to economic, cultural, or religious constraints. Instead, staple diets are primarily plant based (Gibson *et al.,* 2000), from which non haem Fe and Zn are obtained.

These are more widely available and are the source of most Fe and Zn in the diet world-wide. People living in poverty stricken areas have limited access to even these foods. Instead, they largely depend on grain from staple crops, which they eat on a daily basis (Manwaring *et al.,* 2016).

Anaemia affects almost one third of the global population and nearly half of cases are due to Fe deficiency (Lopez *et al.,* 2016). Fe deficiency anaemia is a major and global health problem that can lead to several chronic diseases, most notably chronic kidney disease, chronic heart failure, cancer, and inflammatory bowel disease (Lopez *et al.,* 2016). The World Health Organisation (WHO) estimates that the prevalence of Fe deficiency varies hugely around the world and mostly affects children aged 0–5 years, women of childbearing age and pregnant women in resource poor areas (Bailey *et al.,* 2015). Severe Fe deficiency is associated with increased maternal mortality, premature delivery and miscarriage (Carriaga *et al.,* 1991). Fe deficiency also contributes substantially to maternal deaths from dystocia (obstructed labour) and is associated with low offspring birth weight (<5.5lb). These deficits lead to poor health and low productivity in adulthood as the detrimental effects often persist from childhood. On a global scale, the lowest rate of anaemia associated with Fe deficiency was observed in the USA and Canada (2.9% of the population). However, in several poorer regions, including Central Asia (64.7%), South Asia (54.8%), and Andean Latin America (62.3%) a significantly higher proportion of anaemia is caused by Fe deficiency (Kassebaun *et al.,* 2014). Some biological components of diet directly affect Fe bioavailability. For example, phytate (a phosphorus storage compound found in cereals and vegetables), polyphenols (most notably certain flavonoids due to their metal chelating activity) and some proteins inhibit Fe absorption (Tako *et al.,* 2015). By contrast, ascorbic acid and muscle tissue enhance Fe absorption.

Zn is another key micronutrient of particular importance during childhood and pregnancy. Moderate Zn deficiency has been linked to an increased rate of infection, growth stunting in children and is a risk factor for adverse pregnancy outcomes and premature delivery (Lamberti *et al.,* 2016), poor immune function and increased incidence of diarrhoeal diseases and acute respiratory infections, which are major causes of mortality (Hambidge, 2000; Bailey *et al.,* 2015). Organ systems clinically affected by advanced Zn deficiency include the epidermal, gastrointestinal, central

nervous, immune, skeletal and reproductive systems (Hambridge, 2000). Zn deficiency is most prevalent in Africa and South-East Asia (Caulfield and Black, 2004) and many studies have demonstrated a relationship between low plasma Zn levels during pregnancy and low birth weight within these areas (Neggers *et al.,* 1990; Black *et al.,* 2013). This leads to stunted growth and impaired physical/neural development, which is often irreversible. Despite the serious health implications, Zn deficiency has received much less attention than other MNDs (Manwaring *et al.,* 2016).

1.2 Acute and Long-term Management

The aim of MND treatment is to supply enough Fe and Zn to replenish and maintain Fe and Zn stores. This would have a positive impact on quality of life, symptoms, and the prognosis of many chronic disorders (Lopez *et al.,* 2016). Two mitigation strategies include; prevention plans targeted at the populations at risk and Fe and Zn supplementation approaches in confirmed Fe and Zn deficient patients.

On a global level, food-based approaches including the promotion of access to and consumption of Fe and Zn rich foods are strongly recommended by the WHO. In addition to this, nutritional dietary supplementation is intended to provide adequate nutrients that may otherwise not be consumed in sufficient quantities through diet alone. Combined Fe and Zn supplementation is perceived to be a logical MND prevention strategy (Lind *et al.,* 2004) and is used to prevent Zn/Fe deficiency in at-risk populations, or as treatment for those with proven deficiency (Lopez *et al.,* 2016). Even though studies show significant improvements to human health including weight gain, improved development and a reduction in the incidence of diarrheal disease and lower respiratory infection in children, as measured by the Bayley Scales of Infant Development, these positive effects are often not experienced without adverse side effects such as vomiting and fever (Lind *et al.,* 2004). For example, Lind *et al.* (2004) investigated the effects of Fe/Zn supplementation in Indonesian infants over a period of 6 months. The supplements were administered orally as sugary syrup containing either 10mg of Fe as ferrous sulphate, 10mg of Zn as Zn sulphate or both (20mg in total). Results showed that severe vomiting as a side effect was observed in 33% of children who took the combined treatment of Fe and Zn, whereas when Fe was taken as an individual

supplement, only 18% reported vomiting and 21% reported vomiting for Zn individual supplementation (Lind *et al.*, 2004). The study also considers the possibility of negative biochemical interactions between Fe/Zn in the human body. For example, it has been proven that high concentrations of inorganic Fe may inhibit the absorption of Zn (Solomans & Jacob, 1981) and Zn given with water may also inhibit Fe absorption (Rossander-Hulten *et al.,* 1991). In terms of critical levels where one mineral would affect the other, Solomans and Jacob (1981) found that 25mg Fe given in water with 25mg Zn decreased total plasma Zn levels. However, lower amounts of minerals (10mg Fe + 5mg Zn, ratio 2:1) had no effect on Zn concentration. Other studies including one conducted by Sandstrom *et al.,* (1985) reported no effect on Zn absorption when the ratio was 1:1, however Zn absorption was affected when the ratio of Fe:Zn was 25:1.

The process of food fortification involves the addition of micronutrients and vitamins to food sources to improve nutritional quality with minimal risk to health (WHO, Food and Agricultural Organization of the United Nations, 2006). However, limiting drawbacks are often experienced as the fortified compounds at elevated levels may alter taste, appearance and have a negative effect on shelf-life, thus making the product unacceptable to the consumer. For example, the most bioavailable Fe compounds, which are freely water soluble are the most likely to cause adverse effects to colour and taste (Hurrell, 1997). For example, a noticeable precipitation reaction, can be observed when Fe fortified sugar is added to beverages, such as tea (Hurrell, 1997). Additionally, many Fe compounds are coloured and cause a noticeable change in appearance, especially when added to lighter coloured foods. For example, infant cereals have been found to turn grey/green upon addition of a ferrous sulphate and salt fortified with Fe will turn yellow/brown in colour. Negative effects on taste are often reported as a 'metallic' taste when Fe is used to fortify foods, which is more noticeable in beverages (Hurrell, 1997).

Enrichment of food with micronutrients is effective in improving human nutritional status (Lopez *et al.,* 2016). For example, many governments in the west have taken the decision to fortify white wheat flour, for example in the United Kingdom the Bread and Flour regulations act of 1998 make fortification of white flour with Ca, Fe and thiamine compulsory. While this option is available to countries with a population dependent on purchased and convenience foods, in those countries where

a far larger population is more directly dependent on the land, this is neither practicable nor affordable (Manwaring *et al.*, 2016).

Agronomic practices that positively influence the nutritional status of farm produce show huge potential via better fertilisation (to improve soil quality) and watering systems. However, systems reliant on water may not be available to regions which often suffer with severe drought. There are also a wide range of factors that influence the soil and in turn nutrient uptake, such as light intensity, temperature and rainfall. These are difficult to control in an agricultural environment; therefore control of nutrient uptake may not be reliable and reproducible. There are other variables relating to the soil that are difficult to control, such as the proportion of sand, silt, clay, and organic matter, which in turn has a direct effect on mineral composition and uptake by the plant (Hornick, 1992).

While many of these interventions are proven to be successful in the short term, for the individuals reached by them, they have proved to be unsustainable and incapable of reaching all of the people afflicted by micronutrient malnutrition; and are unlikely to reach those most at risk, namely resource-poor women, infants and children that live in remote communities that are either far away from a clinic or that do not have ready access to processed and fortified foods (Graham *et al.*, 2001). Improving Fe and Zn content in staple foods through biofortification is considered to be the most cost effective, sustainable and consumer friendly solution in meeting target levels of Fe and Zn in human populations (Velu *et al.*, 2007). The aim of biofortification is to make crops more nutritious as they grow, rather than adding nutrients when processing them into foods. This is largely facilitated by natural cross breeding and genetic improvement methods which rely on natural genetic variation, the use of modern tools for selection and the identification of new genes and gene combinations that can be used to improve levels of Zn and Fe. These techniques can be utilised for the production of elite crop varieties with increased micronutrient densities (Figure 1.1) for those even in remote areas as the problem of MND is tackled at the source.

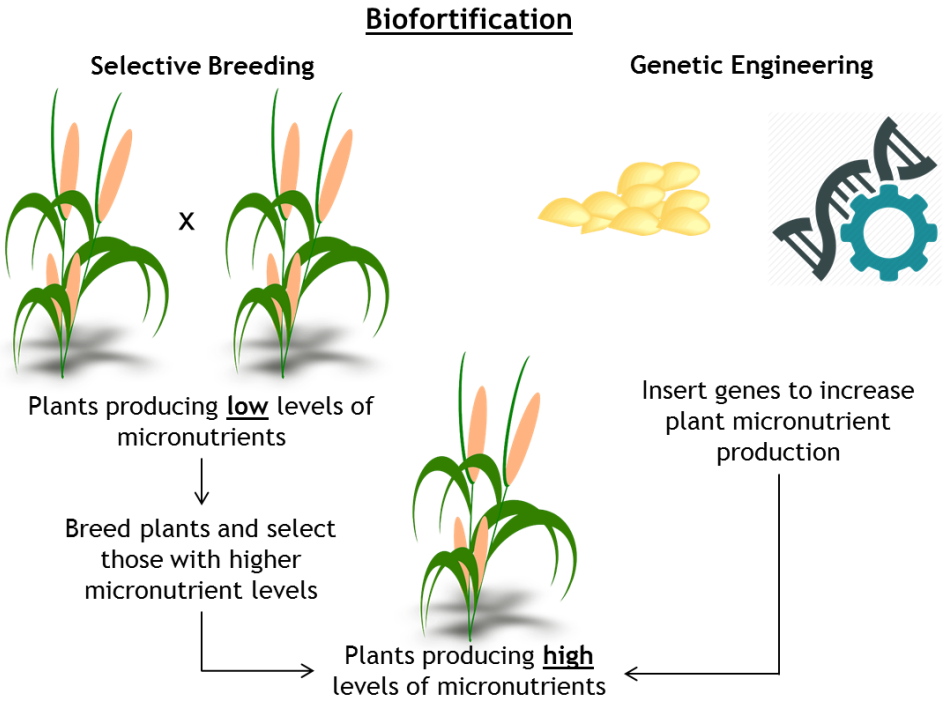


Figure 1.1: Biofortified crops are bred or engineered to produce increased levels of micronutrients (derived by author).

1.3 Pearl Millet (*Pennisetum glaucum*) ~ A Suitable Candidate for Biofortification

Pearl millet (family: *Poaceae*, subfamily: *Panicoideae*) is a multi-purpose cereal crop which provides food, fodder and fuel on more than 31 million hectares worldwide (ICRISAT, 2016). It has a 2530 Mb genome size and a diploid chromosome number of 7, (2n = 14) (Bennett *et al.*, 2000). Pearl millet is the 6th most important cereal crop after wheat, rice, maize, sorghum and barley (Singh *et al.* 2003) and has evolved under the pressures of infertile soils, heat and drought, thus it has a natural ability to thrive in low moisture, nutrient deprived soils and at temperatures in excess of 40°C. It is cultivated throughout the arid and semi-arid regions of West and East Africa and in many parts of India (Oumar *et al.*, 2008). In these regions pearl millet constitutes up to 75% of the total cereal production and therefore represents an important part of local diets (Lestienne *et al.*, 2005). Pearl millet grains are naturally nutritious as compared to rice, maize and wheat. In light of this, nutrient-rich millets including pearl millet are being introduced to the mid-day meal schemes for many government aided schools in Karnataka and Telangana by the The Akshaya Patra Foundation on a pilot basis, with the aim of enhancing the nutritional intake of the students (Mahadevan *et al.*, 2013). This is because the millet based meals are considered more nutritious than wheat/rice based meals. Growing

pearl millet generally requires few chemical inputs; thus, investments in production tend to be low and more suitable for areas that have not benefited from dominant agricultural growth trajectories (Jalaja *et al.,* 2016). Pearl millet is a major source of energy, proteins, vitamins and micronutrients for millions of people living in poverty (Yadav & Rai, 2013). In addition, grains are reported to be rich in resistant starch, dietary fibre and antioxidants (Ragaee *et al.,* 2006). Pearl millet currently holds an orphan crop status, where it is acknowledged as regionally important, but scientific research in terms of nutritional enhancement is still somewhat limited. For example, pearl millet did not have a reference genome assembly until September 2017 (Varshney *et al.,* 2017).

Pearl millet is a monocot, short cycled and small seed size cereal crop (Ullah *et al.,* 2016). The growth and phenology of pearl millet is divided into three phases defined by growth stages GS1, GS2 and GS3 (Figure 1.2). GS1 includes seedling, root and leaf establishment, which takes 3-4 days at optimum conditions and with low rainfall. Tillering and panicle initiation takes place at around 20-25 days. Elongation of leaves, floral initiation in tillers and stem elongation take place at GS2, which usually occurs between 45-59 days. Pearl millet is a protogynous species, where the female reproductive organs mature before the male parts. Furthermore, the wind is a major mode of pollination, though several insects may also play a less significant role (Kaur & Soodnan, 2017). GS3 starts with fertilisation of florets, seed setting and grain filling and finally, maturity of the plant (Ullah *et al.,* 2016). Seed setting/grain formation occurs 60–65 days after germination and is usually completed within 10 days. The crop reaches physiological maturity 90–95 days after germination if optimal weather criteria are met.

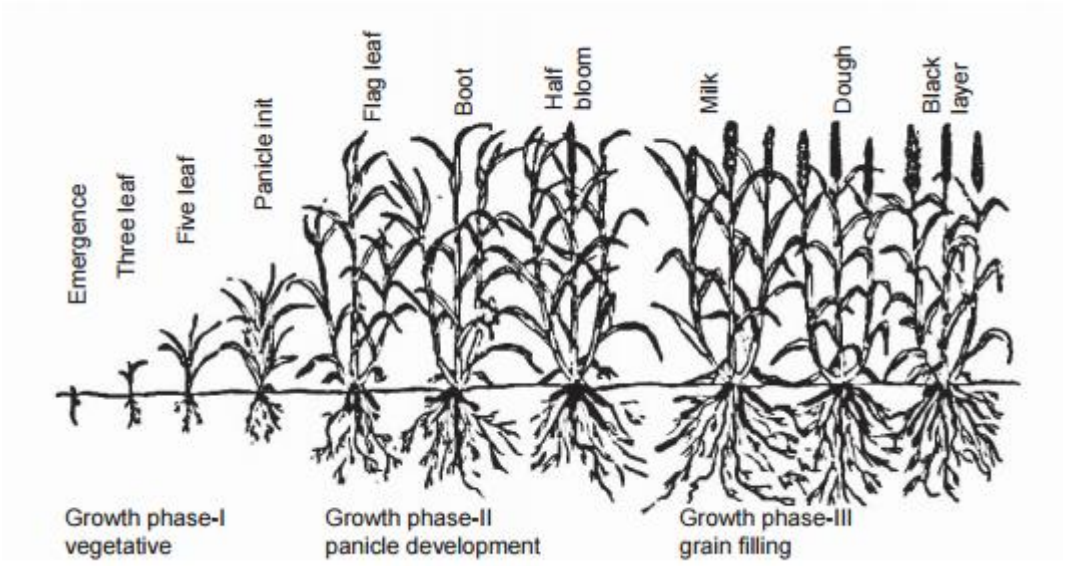


Figure 1.2: The growth stages of pearl millet (Khairwal *et al.,* 2007)

Crop production is particularly vulnerable to climate change because meteorological factors including UV radiation and rainfall determine resource availability and control the majority of processes concerned with plant growth and development (Meza and Silva, 2009). Increases in temperature and rainfall, associated with greenhouse gas emissions will cause significant changes in land suitability and crop yield (Schmidhuber & Tubiello, 2007). Climate change adversely affects food security on a global scale, with the biggest impact on developing countries where the majority of the population is fronting food insecurity and severe micronutrient malnutrition. Recently, Knox *et al.* (2012) noted that the average yield for all crops is predicted to reduce by 8% by 2050 with robust variations among type of crop and region; this highlights the importance of pearl millet, as it grows in areas where crops, considered to be of more economic importance, such as wheat and maize are unable to grow. Therefore, since pearl millet is adapted to harsher conditions, the crop and any associated research to do with harnessing and further improving its nutritional value becomes more valuable for populations afflicted with the consequences of food insecurity.

1.4 Biofortified Pearl Millet – Human Trials

The literature reports success from biofortified pearl millet. For example, Cercamondi *et al.*'s (2013) field trial was conducted on 20 African women from Benin to evaluate the potential of Fe-biofortified pearl millet as a source of

additional bioavailable Fe. Results showed that upon the consumption of two Fe biofortified pearl millet meals per day for 5 days, the total amount of Fe absorbed from Fe-biofortified pearl millet was up to three times higher than that from regular pearl millet. This suggests that biofortification of pearl millet is a valuable strategy in increasing bioavailable Fe to those living in millet consuming communities with limited access to conventional post-harvest fortified foods (Cercamondi *et al.,* 2013). These findings coincide with another recent study conducted in Karnataka, India. In this study by Kodkany *et al.* (2013), 40 Fe deficient children (aged 2 years) were fed pearl millet biofortified with both Zn and Fe. Findings revealed that the amount of Fe and Zn absorbed from the biofortified pearl millet meals was significantly greater than that from the non-biofortified pearl millet meals and the absorption of both Fe and Zn from the biofortified meals exceeded the minimum physiological requirement for children aged 2 years of 0.54 and 2.5 mg/d, respectively. These findings suggest that increased concentrations of Zn and Fe in pearl millet as a result of biofortification are more than sufficient in meeting the minimum physiological requirements of Zn and Fe in young children and biofortification is effective in eliminating MNDs without unpleasant side effects (Kodkany *et al.,* 2013).

1.5 Understanding Iron and Zinc Uptake- From Root to Seed

Any attempt to increase Fe and Zn concentrations in pearl millet grains using traditional breeding methods or genetic engineering must first consider how Fe and Zn are obtained from the environment, distributed and stored in edible parts (see Morrissey and Guerinot, 2009 for a comprehensive review). Even a small increase in bioavailable nutrient metals in pearl millet grains would have a significant impact on human health, particularly for those living in developing countries. It is also proven that biofortified crops can produce nutrient rich grains even when grown in relatively poor soils as long as target nutrients fall within the critical range. The critical soil levels of Fe and Zn are 2.6 to 4.5 mg/kg and 0.6 to 1.0 mg/kg, respectively (Kanatti *et al.,* 2014). Pearl millet can be grown successfully on clay, clay loam, or sandy loam soils. However, it does not grow well under calcareous soils (soils with moderate to excess lime).

Initial high levels of nutrient metals from the soil can be toxic. For example, unregulated high affinity binding of Zn to S-, N-, and O-containing functional

groups in certain biological molecules and uncontrolled displacement of essential metal cations, for example Mn^{2+} and $Fe^{2+,}$ can cause significant damage to the plant (Palmgren *et al.,* 2008). Because of this, the activity of metal ion transporters is selective and highly regulated (Philpott, 2014). This is in part achieved through membrane transporters such as metal tolerance proteins (MTPs) (Ricachenevsky *et al.,* 2013). Elucidating the mechanisms behind cation selectivity and regulation is necessary in understanding plant metabolism and development. If these pathways can be fully understood, they can be improved through biotechnological manipulation. A variety of platforms including, but not limited to, phylogenetic analysis, transcriptomics, gene expression analysis and sequence manipulation are available to help elucidate these mechanisms (see Ricachenevsky *et al.,* 2013 for a comprehensive review).

Fe availability in plants is dictated by a wide variety of factors including soil redox potential and pH. In soils that are at high pH, Fe is readily oxidised and presents itself as insoluble ferric oxides, however, at lower pH, ferric Fe is released from the oxide, making it available for root uptake via the activity of a ferric chelate reductase, FRO2 (Marschner and Rimmington, 1988). Fe is transported into the root epidermal cells by IRT1, a divalent metal transporter which is a member of the Zrt- and Irt-like protein (ZIP) family of transporters. AtIRT1 also transports Zn, Mn, Cd, Co and Ni (Morrissey and Guerinot, 2009). Unknown phenolics then control the extraction of Fe from the negatively charged cell walls, which allows transport into the root symplast (Morrissey and Guerinot, 2009). Fe is then bound by unknown chelators or chaperones, and moves symplastically through the connected cytoplasm of the root (Marschner & Rimmington, 1988). At the pericycle, Fe is effluxed into the xylem and moves towards the shoot via transpiration. When Fe enters the xylem, it complexes with citrate, without which, Fe will not move efficiently through the xylem and won't be utilised by the shoot (Curie *et al.,* 2009). Fe is then transported into the phloem by nicotianamine (NA) and Yellow Stripe Like (YSL) transporters (Morrissey and Guerinot, 2009). YSLs play an important role in the long distance transport of Fe complexes and in seed delivery (Inoue *et al.,* 2009). NA serves as a transporter that facilitates the movement of Fe in and out of the phloem via YSLs (Morrissey and Guerinot, 2009). It also complexes with Fe^{2+} and Fe^{3+} and binds readily to Cu^{2+}, Ni^{2+}, Co^{2+}, Zn^{2+}, and Mn^{2+} (Curie *et al.,* 2009). NA also plays an

important role in metal homeostasis (Takahashi *et al.*, 2003). Movement within the phloem occurs via Iron Transport Proteins (ITPs) (Morrissey and Guerinot, 2009). Fe moves into the seed via the phloem, and seeds in the early stages of development receive Fe from roots and senescent leaves. The loading of Fe into the seed occurs by NA and YSLs and is stored in the endosperm (Morrissey and Guerinot, 2009). A large abundance of Fe may be toxic to the seed embryo, therefore plants possess two damage preventing mechanisms; Fe can either be stored in large plastids with ferritins, which are able to store up to 4500 Fe atoms (Grillet *et al.*, 2014) or can be stored as phytate complexes.

Despite the importance of Zn as an essential micronutrient, there is a significant lack of literature detailing the mechanisms of Zn uptake compared to that of Fe. In the soil, Zn is taken up into the root epidermal cells in its water soluble $^{+2}$ oxidation state and unlike Fe^{2+}, it is redox stable (Broadley *et al.*, 2007). Several metal transporters of the ZIP family are considered to be the primary uptake systems for Zn (Guerinot, 2000). After uptake, Zn is present in living cells with a neutral pH; therefore it is prone to binding to a wide range of organic molecules. This restricts its movement and limits travel between living cells. Zn and Fe are thought to compete for the same uptake systems; therefore NA is also thought to be utilised as a transporter (Olsen and Palmgren, 2014). Transport from epidermal cells into the root xylem occurs via a symplastic pathway through a cytoplasmic continuum of cells, which are linked by plasmodesmata. The movement of Zn is then facilitated into the stellar apoplast (Lasat and Kochian, 2000). As discussed previously in the case of Fe, the chelator NA also contributes to long distance transport of Zn from the roots into shoots and seed. NA also modulates the vacuolar sequestration capacity, an essential mechanism controlling the way plant vacuoles provide temporary storage for micronutrients (Sperotto *et al.*, 2014). Zn uses the transporter Heavy Metal ATPASE 4 (HMA4) in shoot loading and once in the xylem, it is transported in aqueous form (Olsen and Palmgren, 2014). How Zn enters the phloem is not yet known, however, YSL proteins are likely to play a role in the process. In cereal grains, high concentrations of Zn are found in the embryo and the aleurone layer of the endosperm. A summary of the above can be seen in Figure 1.3.

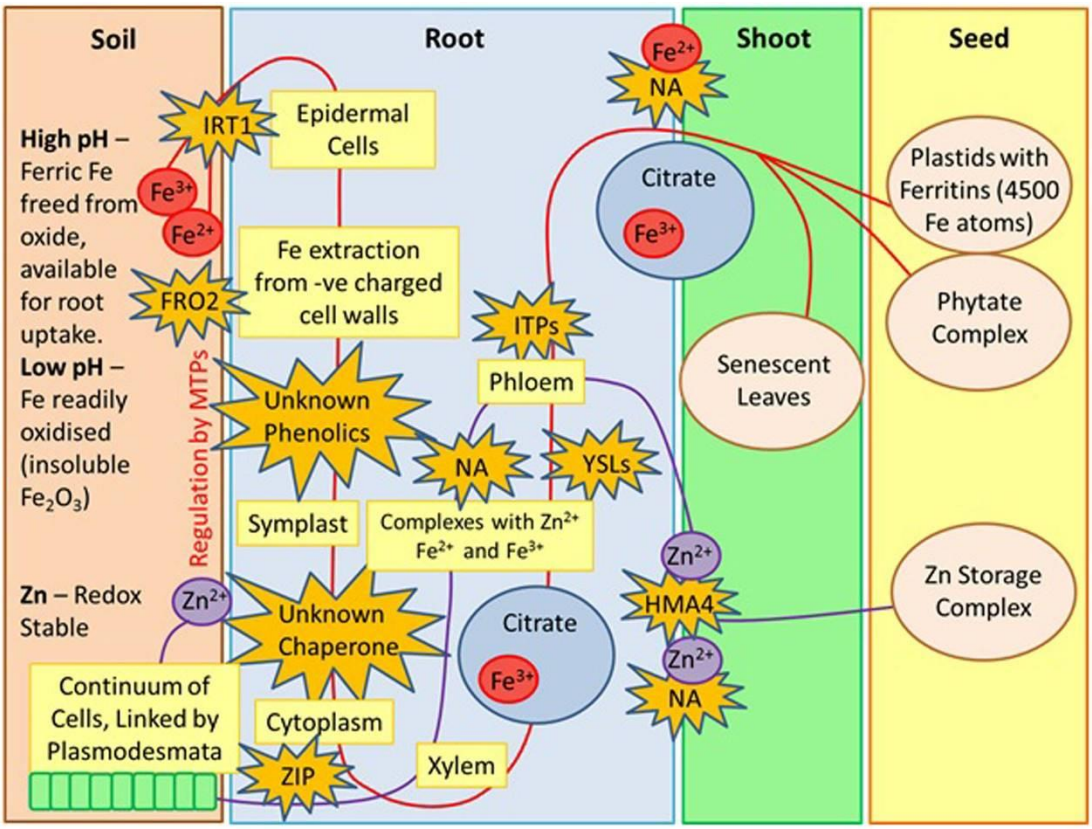


Figure 1.3, Zinc and iron uptake from soil to seed (Manwaring *et al.*, 2016). MTPs, metal tolerance proteins; FRO2, ferric chelate reductase; ZIP, zinc regulator transporter proteins; IRT1, divalent metal transporter; NA, nicotianamine; YSLs, yellow stripe like transporters; ITPs, iron transport proteins; HMA4, heavy metal ATPASE 4.

There are a variety of factors that affect the efficiency of Fe and Zn uptake including genotype, nutrition, soil type and climatic conditions. For example, plant-associated microorganisms present in the soil can stimulate growth and influence yield and quality of edible parts by affecting nutrient mobilisation and transport (Pii *et al.*, 2016). Because of this, the root rhizosphere microbiome, which consists of the area of soil surrounding the root where complex ecological and biological processes occur (Bais *et al.*, 2006), is considered to be one of the key determinants of productivity and plant health. Pii *et al.* (2016) showed using non-metric multidimensional scaling analysis, the presence of the different plant species coupled with nutritional status could promote a differentiation of the rhizosphere microbiome, which in turn has a significant effect on micronutrient uptake. More specifically, a study conducted by Borde *et al.*, (2010) showed that arbuscular mycorrhizal fungi (*Glomus fasciculatum*) inoculated pearl millet plants resulted in increased growth and nutrient uptake capacity. Root exudates are also thought to play a significant role in efficient micronutrient uptake, including that of Fe and Zn.

For example, a variety of low and high molecular weight organic compounds are triggered if plants are exposed to certain abiotic stressors (Dakora and Phillips, 2002) and due to their solubilising, chelating, redox capacities, they play a fundamental role in enhancing nutrient bioavailability, uptake, translocation and allocation. In recent decades it has also been proven that since Fe bioavailability is reduced in aerated and calcareous soils (Kobayashi and Nishizawa, 2012), plants have developed different mechanisms to compensate for Fe shortage including; (i) retrieving soluble Fe from the soil via the acidification of the rhizosphere through the release of protons, which causes the reduction of Fe^{3+} to Fe^{2+} by FRO2 and (ii) exudation of non-proteinogenic amino acids and uptake with the aid of YSL transporters (Kobayashi and Nishizawa, 2012).

1.6 Biofortification of Pearl Millet

Biofortification is the process of increasing the content and bioavailability of essential vitamins/minerals in staple crops, through traditional plant breeding or genetic engineering, with the aim of improving nutritional status (Bouis *et al.*, 2011). Modern agriculture has been largely successful in reducing poverty and meeting the energy needs of poor populations within developing countries. However, due to the global expanding population, plant breeders of staple crops have previously focused more on delivering calories, leaving nutritionists to cope with the need for dietary diversity to achieve well balanced nutrition. Recently, exploring the potential of improving the micronutrient quality of staple food crops, without compromising yield has become a key focus of scientific research. In order to achieve this, 5 key criteria need to be met including; (i) initial useful genetic variation for exploitation, (ii) traits need to be manageable within breeding programmes (simple screening and high heritability), (iii) phenotypes must be stable across a wide range of environments (for increased impact), (iv) nutrition related traits need to be combinable with traits for high yield to ensure farmers chose the elite lines and above all, (v) elite lines must significantly improve the health of those afflicted by MND's and the extra nutrients need to be bioavailable to the gut. The first 4 criteria have been proved extensively in the literature and it is now widely accepted that biofortification for nutritional quality is a practicable and cost-effective strategy. However, there are still gaps in the literature regarding to what extent crops can

contain increased amounts of bioavailable micronutrients and how effective these lines are in alleviating micronutrient malnutrition (Graham *et al.,* 2001).

There has been a remarkable amount of success relating to Fe biofortification in pearl millet in recent decades. Most notably, scientists at International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patencheru, India initiated a breeding programme in 2003 to develop high Fe hybrids with stable yields and good Fe performance for different agroecological zones in India. Scientists screened pearl millet germplasm accessions for Fe content and a range of 30 – 76ppm was revealed. High Fe genotypes were then selected to initiate crosses and as a result, breeding lines with more than 90ppm were identified and validated. As a result of this research an Fe open pollinated variety, ICTP 8203 was released in Maharashtra in 2013 and went on to become available to all pearl millet growing states in India by February 2014 (Bouis, 2014).

Biofortification is largely facilitated by drilling down into the genomes of crops to seek genes of interest and breeding these genes into new, improved varieties. Despite the lack of a reference genome until recently, genetic improvement of pearl millet previously resulted from the use of genetic resources including maps and sets of markers, which facilitated the selection and breeding of elite cultivars with high nutritional value. Genetic and genomic technologies drive the way forward for the discovery and transfer of genes and quantitative trait loci (QTL) associated with an improved nutritional profile from the diverse genetic resources of millets (Muthamilarasan *et al.,* 2016). A pearl millet sequencing consortium have since developed a ~1.79 Gb draft whole genome sequence of reference genotype Tift 23D2B1-P1-P5, which contains an estimated 38,579 genes and over 29 million SNPs. Findings became publically available as of September 2017 (http://ceg.icrisat.org/ipmgsc/).

1.7 Germplasms, Core and Mini Core Collections - A Good Place to Start

Continuous crop improvement largely depends on the discovery of new sources of genetic variation, accurate identification of lines with beneficial traits and their judicious use. Developing core collections to be made available to researchers globally has been recognised as “International Public Good” (Upadhyaya *et al.,* 2009). Managed core germplasm collections are available for pearl millet, and as is

the case within any plant breeding program, characterisation of genetic diversity within these collections is a necessary prelude to their efficient use (Varshney *et al.,* 2009). A wide variety of pearl millet germplasm collections exist, including the Pearl Millet inbred Germplasm Association Panel (PMiGAP) developed at ICRISAT. The PMiGAP has been drawn from a global core collection of pearl millet accessions, landraces and cultivars grown across three continents (Sehgal *et al.,* 2015). ICRISAT has also developed other regional collections including the Iniari germplasm, which uses landraces from West Africa. These have superior grain filling abilities under terminal drought stress, larger seeds, thicker panicles, and broader leaves (Ito *et al.,* 1999). Other collections include The USDA National Plant Germplasm System Pearl Millet Collection, which is maintained at the Plant Genetic Resources Conservation Unit located in Griffin, Ga, USA. It contains 1297 unique genotypes from 31 countries (https://scisoc.confex.com/scisoc/2015am/webprogram/Paper95640.html).

Germplasm collections contain a wide genetic base and promote plant diversity. However, it is often the case that these collections are impractically large in size (at ICRISAT there are currently 21,594 accessions of pearl millet from 50 countries), therefore there is a need to capture genetic variation from existing collections at a more manageable level. A core collection contains a subset of accessions from the entire collection that captures the majority of the available diversity within the species. This is usually about 10% of the entire collection and this promotes enhanced utilisation of germplasms in crop breeding. However, even these can be too large, thus not economically viable to analyse. A solution is the creation of mini core collections, which are generally about 10% of the core collection or 1% of the entire collection. These represent >80% variability from the core collection. Patterns of diversity are usually evalutated using descriptors developed especially for pearl millet. For example, in the last decade diversity in 21,954 accessions from <50 countries was assessed for 23 morphoagronomic traits of interest that revealed large genetic variability for flowering time, plant height, tiller number, panicle length and 1000-seed-weight (as a measure of yield). The core collection does not require equal cluster representation of the entire collection or for it to contain the largest possible diversity (if this were the case, the core would be biased towards vast numbers of distant wild relatives). Instead, the diversity should be as high as what would be consistent with the core being a representative genetic resource of practical use to

scientific researchers (Upadhyaya *et al.*, 2009). When developing a core/mini core collection a wide range of criteria must be met (Figure 1.4).

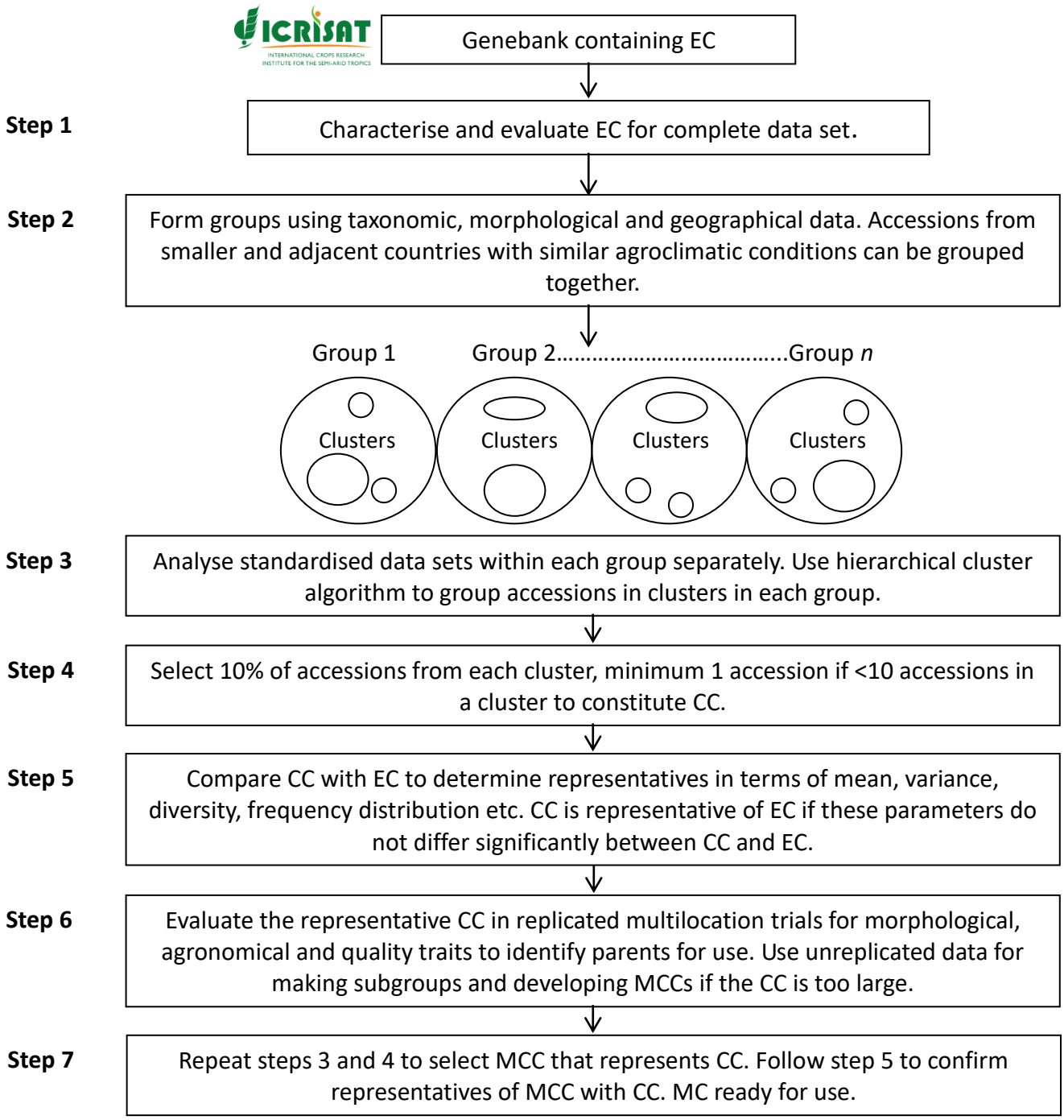


Figure 1.4: Flow diagram to establish core and mini core collections in a crop species (adapted from Upadhyaya *et al.*, 2009). EC, entire collection; CC, core collection; MCC, mini core collection.

In pearl millet, the mini core collection (PMiGAP) consists of 238 accessions selected from 2094 entries present in the core collection. These represent 46 countries. The composition of the mini core collection predominantly represents accessions from India and North West Africa, both representing dry, semi-arid, tropical ecology. It also captures 90% range variation of the core collection.

1.8 Traditional Breeding Methods

Traditional breeding methods would involve the selection of two parental lines with high Fe and Zn content and crossing them to create a hybrid that expresses the traits of interest. Successful crop improvement via plant breeding largely depends on the existence of genetic variation for the target traits within the gene pool. When breeding for elevated levels of Fe and Zn in edible parts, the task is further complicated by the fact that the grain micronutrient concentration is highly dependant on environmental conditions, including soil mineral composition (Feil *et al.,* 2005). Genotype Environment Interactions (GEI) are therefore a major factor in the development of stable and high-nutrition cultivars of pearl millet (Moghaddam and Pourdad, 2009). In light of this, multi-environmental trials are required to confirm the stability of phenotype data. The presence of GEI may reduce the validity of any downstream analysis, restrict the significance of findings, and limit the efficiency of selecting elite lines (Gurmu *et al.,* 2009). However, many studies have also reported in relation to Fe and Zn traits in pearl millet that although the environmental effect is strong, the genotype effect is fairly consistent across environments, implying the GEI interaction is not as serious as previously believed (Graham *et al.*, 2001).

1.9 A Stable Phenotype and Maintaining Yield

Many studies have identified potential high Fe and Zn lines with stable phenotypes in pearl millet. Velu *et al.* (2007) analysed a diverse range of genetic materials developed at ICRISAT for grain Fe and Zn content. Based on the average performance in two growing seasons, large genetic variability among the entries was found for both Fe and Zn. Most notably, well-adapted, highly utilised genotypes and their progenies from the Iniari germplasm contained high levels of grain Fe and Zn and the correlation between Fe and Zn content was found to be positive and highly significant. This suggests that the simultaneous selection for elevated levels of both micronutrients is possible, and selection within the Iniari germplasm is likely to provide excellent candidates for the development of elite varieties with increased grain Fe and Zn content (Velu *et al.,* 2007).

When considering traditional methods of biofortification for reducing MNDs, it is important to consider sustainability coupled with socioeconomic factors for

smallholder farmers – the people that will directly benefit from this research. For example, genes associated with micronutrient uptake and their relationship with grain yield has a direct bearing when formulating effective strategies for breeding elite lines (Kanatti *et al.,* 2014). In terms of increasing micronutrient content whilst keeping high yield, the study by Velu *et al*. (2007) also proved a highly significant positive correlation of 1000-grain weight with Fe and Zn content in pearl millet grain, indicating that breeding for elevated levels of these micronutrients is possible without compromising yield.

Kanatti *et al.* (2014) demonstrated, using 196 hybrids and their 28 parental lines of pearl millet, that in order to breed successful hybrids that express elevated grain Fe and Zn levels, the same genes for Fe and Zn content should be incorporated into both parents (Velu *et al.,* 2011). Hybrids were found to express no better-parent heterosis as barely any hybrid was found transgressing the parental lines for increased grain Fe content. The study showed that the underlying physiological processes that determine grain Fe and Zn content are primarily under additive genetic control (Govindaraj *et al.,* 2013; Kanatti *et al.,* 2014) and a large amount of the mid-parent heterosis values were in the negative direction. This indicates that the involvement of genes, with the exception of those with additive gene action, where alleles determining lower Fe and Zn densities, are partially dominant. However, when considering additive gene action, if the same source is used to transfer the genes associated with Fe and Zn content in both parental lines, this would cause the amount of genetic diversity between lines for other traits to be reduced. This may lead to a reduction in heterosis for yield, which is controlled by non-additive gene effects (Kanatti *et al.,* 2014). It was also found that Fe and Zn content from inbred lines and their general combining ability were positively correlated and highly significant. This suggests that recurrent selection could be used to significantly improve breeding populations for grain Fe and Zn content and breeding lines selected for high Fe and Zn levels are more likely to include those with high general combining ability for these micronutrients (Govindaraj *et al.,* 2013).

1.10 End Use Quality

The acceptability of a cultivar by farmers and consumers is highly based on how the grain is processed and the end-use quality (Ortiz-Monasterio *et al.,* 2007).

Additionally, the concentration and bioavailability of micronutrients in pearl millet, as in other cereal crops, may be enhanced or reduced by various methods of processing; this is achieved by fortification with certain ingredients as well as meal preparation techniques (Welch & Graham, 2004). Therefore, end-use quality traits including protein content, grain hardness and baking properties must be considered when creating elite lines in pearl millet. For example, in maize grain hardness and factors affecting the gelatinisation and pasting properties of starch are considered when determining end-use quality (Ramirez-Wong *et al.,* 1994). Micronutrient-enhanced lines should also be screened for desirable end-use quality traits. For example, various processing treatments of pearl millet including germination, autoclaving, and debranning are known to be effective in reducing levels of phytate (Sharma & Kapoor, 1996). However, some studies suggest that certain methods of processing raw pearl millet grain may result in decreased levels of Fe. For example, soaking grains may result in up to 25% loss of Fe (Eyzaguirre *et al.,* 2006). In light of this, methods of processing should be considered that are not detrimental to the levels of these micronutrients.

1.11 Deleterious Effects – Toxicity and Antinutrients

Any potential products of biofortification should be carefully evaluated under real conditions. This can be achieved via the assessment of trace element bioavailability to humans and investigation into any drawbacks such as enhanced uptake of toxic metals from the soil (e.g., Cd, which is deleterious to all organisms). Toxic metals such as Cd share the same transporters as some micronutrient metals (e.g., Fe and Zn) (Zhao & McGrath, 2009). Several studies have investigated the possible unwanted side effects of biofortification, for example the enhancement of ZIPs. In addition to aiding Zn uptake by roots, ZIPs can also aid the uptake for other, non-specific cations such as Cd. Two well-characterised ZIP family proteins, IRT1 and IRT2, represent the main Fe^{2+} uptake systems in *Arabidopsis thaliana* root cells. IRT1 also facilitates the uptake of Zn and Cd (Guerinot, 2000). If this pathway were enhanced by biofortification, this could potentially cause the plant to become toxic and dangerous for consumption.

The improvement of Fe and Zn levels would also need to include the reduction of antinutrient compounds such as phytate. Due to high phytate content within the

endosperm of pearl millet seed, this often limits the bioavailability of many important micronutrients such as Fe and Zn upon human consumption (Shanmuganathan *et al.,* 2006). The deleterious effect of phytate on mineral bioavailability has been confirmed by Egli *et al*. (2004) whose *in vivo* radioactive and stable isotope studies demonstrate that, absorption of Fe, Zn, and Ca are significantly lower from diets with a high content of phytate than from diets that contain low levels of phytate.

Phytate exists as a phosphorylated myo-inositol ring which strongly chelates metal cations, including Fe^{2+} and Zn^{2+} (Urbano *et al.,* 2000). When Fe and Zn bind to phytate, an insoluble precipitate is formed. This is not efficiently absorbed by the intestines due to the absence of intestinal phytase enzymes (Nielsen *et al.,* 2013). This poor absorption can therefore exacerbate Fe and Zn deficiencies (Hurrell, 2003). The adverse effect of phytate on Fe and Zn absorption is dose-dependent (Gibson *et al.,* 2010). For example, studies have proved a negative relationship between Zn absorption and phytate over a wide range of phytate:Zn molar ratios (Hambidge *et al.,* 2004). With respect to Fe, the inhibitory effect of phytic acid is still strong at very low phytate levels, when ratios are as low as 0.2 Phytate:1.0 Fe (Hallberg *et al.,* 1989).

Although indigestible to humans, phytic acid plays several key roles in the development of seedlings. For example, phytic acid acts as the principal storage form of P and also acts as a source myoinositol, which is required for cell wall development (Reddy *et al.,* 1982). Because of this, phytate can never be completely eliminated from the crop. However, conscious efforts should be made to significantly reduce phytate so that it does not become a limiting factor in micronutrient absorption. A possible solution is the development of low phytate varieties of pearl millet. For example, a study by Shanmuganathan *et al.* (2006) showed that many parents and crosses can be successfully exploited for the development of pearl millet genotypes with low phytate content. The study involved crossing 11 pearl millet genotypes to create 55 hybrids and measuring the hybrids for phytate content in order to understand the nature of gene action and to evaluate the parents and hybrids for combining ability with respect to low phytate content. Several crosses were identified as having low phytate content by measuring negative specific combining ability.

Multiple pearl millet lines have been associated with low phytate content, for example in a study by Simwemba *et al.* (1984), two lines (TIFT 2 23 DAE X 656 + 653) were identified to have lower phytate content as compared to certain varieties of wheat. It was also found that the environment (as well as genetics) also plays an important role in determining phytate content (Simwemba *et al.,* 1984). There also appears to be a significant gap in the literature as no genes/markers have been associated with low phytate content in pearl millet. It is hoped that using synteny studies with other crops may facilitate the discovery of markers/QTLs/genes associated with low phytate content (Rawat *et al.,* 2013). For example, QTL mapping for phytate content has been extensively accomplished in rice (Stangoulis *et al.,* 2007), soy bean (Walker *et al.,* 2006), and bean (Cichy *et al.,* 2009). It has also been proved that loci affecting phytate content are different to loci affecting grain micronutrient content, this suggests that the simultaneous increase of grain micronutrient content and a decrease in phytate content is possible (White & Broadley, 2011). For example, in the case of Stangoulis *et al.'s* (2007) rice study, QTLs for grain phytate, Zn and Fe concentration were identified in several rice lines from an IR64 × Azucena doubled haploid population. Findings showed that there were significant positive correlations between phytate levels and Fe/Zn concentrations. Since the QTLs of phytate are located on different chromosomal regions as compared to those found for Fe and Zn, this suggests that they are genetically distinct and it should be possible to use molecular markers for breeding and selection purposes to modify the phytate concentration without affecting grain micronutrient content (Stangoulis *et al.,* 2007). Following these studies, further population improvement could be implemented using recurrent selection to breed for low phytate content, whilst keeping yield and micronutrient uptake high.

A wide variety of polyphenolic compounds including flavonoids are present in pearl millet, and like phytate, many of these are found to have adverse effects on mineral bioavailability via their metal chelating properties (Al-Sa'aidi, 2003, Cook & Samman, 1996, Tako *et al.,* 2015). For example some, including apigenin and luteolin interfere with Fe absorption by forming insoluble complexes in the gastrointestinal lumen thus reducing Fe/Zn bioavailability (Brune *et al.,* 1989). The relationship between polyphenolic compounds and bioavailability of micronutrients has been reported extensively in the literature. For example, using the Caco-2 Fe

uptake assay, Miret *et al.,* studied different food matrices containing a variety of Fe forms. Results suggested that Fe uptake was significantly reduced by the food matrices which contained a rich source of polyphenols. A possible explanation why Fe bioavailability was reduced could be attributed to the polyphenols causing an increase in pepsin activity, which may influence the digestion of haemoglobin and the solubility of the released heme. For example, extensive digestion of the peptides could decrease heme solubility and consequently, heme-iron bioavailability. Polyphenols have also been known to inhibit proteolytic enzymes, whose function include catalysing hydrolysis of proteins, and amylolytic enzymes, which are involved in starch degradation. This significantly reduces the digestibility of proteins and starch (Knuckles, 1985, Sharma *et al.,* 1978).

1.12 Tools that Harness the Potential of Pearl Millet in the Fields of Genetics and Genomics

1.12.1 Genetic Maps

In recent decades, the potential of pearl millet and a vast number of genetic variations associated with useful traits has been extensively documented. However, the utilisation of molecular breeding technologies for the genetic improvement of pearl millet is still limited and progress has been slow due to insufficient numbers of PCR compatible co-dominant markers (Senthilvel *et al.,* 2008). In genetic mapping, the location and distance between genetic markers on chromosomes are determined. Recent advances in biotechnology have facilitated the creation of high-density maps that consist of thousands of molecular markers (Hyten & Lee, 2016). Recently, the use of an F2 population of 93 progenies and 9 cultivated pearl millet crosses has facilitated the production of a genetic map with higher density and better uniformity of markers than previously published maps. This was achieved using a modified Genotyping by Sequencing (GBS) platform, which involved the use of two restriction enzymes (Pst1–Msp1) and PCR amplification with primers including three selective bases. These efforts resulted in 3,321 SNPs generated for public use (Moumouni *et al.,* 2015). The availability of large numbers of SNP markers and high-density genetic maps will enhance the progress of gene and QTL mapping in bi-parental populations significantly and also facilitate association analyses on panels of diverse and unrelated lines. The progress of this entails different techniques

such as QTL mapping and Genome Wide Association Studies (GWAS) as discussed below and observed in Figure 1.5.

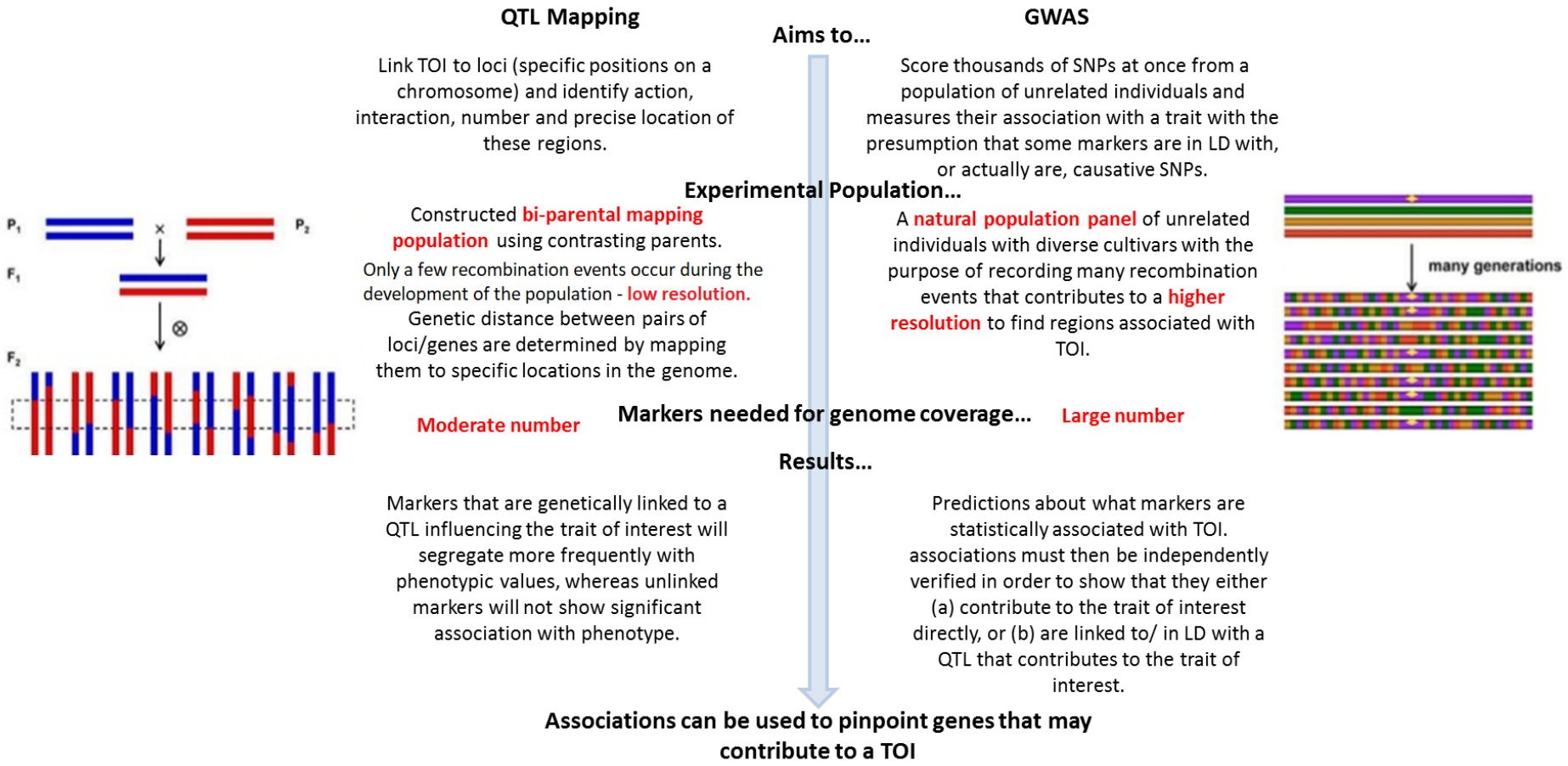


Figure 1.5: QTL mapping vs GWAS – key differences. LD: linkage disequilibrium; TOI: trait of interest (derived by Author).

1.12.2 Genome Wide Association Studies and Association Mapping

In GWAS, phenotypic data is combined with marker data in order to identify genomic regions controlling traits of interest. GWAS have expanded into a powerful tool for investigating the genetic architecture of many staple crops. GWAS exploits the natural diversity generated by multi-generational recombination events that occur in a population or in germplasm panels (Deschamps *et al.*, 2012). This approach results in increased mapping resolution as compared to linkage mapping populations. Genetic sources of phenotypic variation are an essential component of plant genetics. Taking on the lessons learned from model species, such as rice and maize, future developments are being applied to staple crops as well as orphan crops. Comprehensive maps of genome variations will facilitate GWAS of complex agriculture benefitting traits. The result of which will greatly accelerate crop improvement via genomics-assisted breeding (Huang & Han, 2014). The recent pearl millet genome sequence together with the resequencing of the entries of the various germplasm populations will certainly assist in such endeavours. Findings from GWAS will be the catalyst in the mining of candidate genes. These candidate genes can be verified through Transfer (T)-DNA mutants or genetic transformation, which will then facilitate the genetic modification or marker assisted selection (MAS) for validated genes. These steps will then lead to more nutrient rich, improved varieties (Huang & Han, 2014).

Even though research has been previously limited, the first attempt to mine favourable alleles for grain Fe and Zn content through association mapping in pearl millet was conducted by Anuradha *et al.*, (2017), using a total of 267 polymorphic markers, of which 250 were SSR markers and 17 were genic markers. Using an association mapping panel of 130 diverse pearl millet lines that represent three agroclimatic zones in India, favourable alleles and promising lines were identified across multiple and specific environments. Alleles having positive effect were considered as favourable for both Fe and Zn content. A total of six alleles were detected for grain Fe content, and five alleles were detected for grain Zn content (Anuradha *et al.*, 2017).

1.12.3 Synteny Studies and Resources from Major Crop Species

Synteny studies among the cereal family set the stage for a comparative study of millets and non-millet cereals to trace common genes associated with nutrition biosynthesis pathways (Muthamilarasan *et al.,* 2016, Gale and Devos, 1998). When these genes, alleles and QTLs are discovered, they can then be incorporated into elite lines through the use of molecular marker assisted breeding or transgene based methods. Synteny studies may also facilitate the introgression of these traits of interest from the major cereals into millets. This will largely be achieved with the role of genomics, bioinformatics, transcriptomics, proteomics, metabolomics, and ionomics (Muthamilarasan *et al.,* 2016). The genomic resources that have characterised most major cereal crops is going to benefit pearl millet directly. The high throughput and low cost of Next Generation Sequencing (NGS) technologies has made it possible to sequence orphan crops for the development of elite cultivars with desirable traits. Because significant genomic collinearity has been reported in many cereal crops (Devos, 2005), comparative genomics methods that are facilitated by the use of genomic resources and bioinformatics tools will allow the transfer of genes from model/major crops to minor crops (Varshney *et al.,* 2006). The benefits include: (i) Improved analysis of cereal biodiversity and the identification of useful variants; (ii) MAS of alleles and allele combinations of interest; and (iii) Cloning and efficient transfer of useful alleles among members of the cereal family (Nelson *et al.,* 2004). Resources from well sequenced species will enable functional definition of many key genes and pathways (Chen *et al.,* 2016). For example, Chen *et al.* (2016) studied the metabolic and phenotypic GWAS (mGWAS and pGWAS) in rice grain and maize kernels and identified new candidate genes that could be the cause of variation in traits such as grain colour and size. It was found that distinct and overlapping aspects of genetic control of metabolism exist within and between species and the mGWAS analysis indicated that rice and maize are likely to share common genetic control strategies for a variety of metabolites. A search for homologous loci mapped by the same metabolites (or metabolites with similar structures) identified 42 loci underlying the 23 co-detected metabolic features between maize and rice. This data suggests that there is potential for the identification of genes associated with traits of interest between other cereals using

mGWAS and pGWAS and genetic analysis of the metabolome could improve what is currently known about these complex traits.

1.12.4 QTL Fine Mapping

QTL associated with increased Zn and Fe accumulation in pearl millet are an important asset for targeting candidate genes associated with these traits and have driven biofortification research for many years. Focus should be made on what resources could provide potential candidates for the identification of QTLs. As discussed, there are several germplasm panels that hold lines associated with high micronutrient accumulation that cover global diversity. Potential candidates for QTL fine mapping exist within these germplasm banks. For example, in the Iniadi and PMiGAP germplasms, certain lines have been found to be particularly high in Fe and Zn, with a highly significant and high positive correlation between these two micronutrients, as previously discussed. One of these lines, ICTP 8203 has been released in India and was cultivated on more than 0.8 million hectares (Rai *et al.,* 2013) and as of today, has been marketed to more than 70,000 farmers, predominantly within Maharashtra, India (Andersson *et al.,* 2017). Simultaneous accumulation has been reported in a wide variety of crops, including pearl millet and these positive correlations could be due to common and overlapping QTLs for grain Fe and Zn densities (Kumar, 2011).

Research into identifying QTLs and candidate genes for elevated levels of Fe and Zn in pearl millet is limited at present, perhaps due to resource constraints such as lack of a reference genome until recently. The fact that QTLs associated with increased Fe and Zn content have been identified in other crops will benefit QTL fine mapping in pearl millet through synteny studies. For example, QTLs for grain Fe and Zn densities are reported in wheat (Peleg *et al.,* 2009), rice (Stangoulis *et al.*, 2007), and bean (Cichy *et al.*, 2009). In a tetraploid wheat population of 152 RILs, 82 QTLs were mapped for 10 minerals, including Fe and Zn with LOD score range of 3.2–16.7. These were located in 32 non-overlapping genomic regions. A strong association was found between QTLs conferring Zn and QTLs for Fe, which is indicative of a strong genetic association between mechanisms affecting grain Zn and Fe levels (Peleg *et al.,* 2009).

1.12.5 The Pearl Millet Genome Sequence

Thanks to Varshney *et al.*, (2017), the recent ~1.79Gb whole genome sequence of pearl millet, which contains 38,579 genes can be used as an important resource to greatly accelerate the improvement of this far-reaching cereal crop on a larger scale than what was previously thought possible. It is hoped that the current status of 'orphan crop' will soon be shifted to 'staple crop' as there is now increased awareness of how vital this crop is, not only for those already dependant on it but on a global scale owing to its growth, concurrent with climate change. This is paramount because global temperatures are expected to increase from 1-6°C by 2100 (NRC, 2011). Prior to its release in 2017, the foxtail millet genome (Zhang *et al.*, 2012) was considered the genome of choice for pearl millet genetic analysis requiring a reference genome, due to its close relationship with pearl millet. A comparison of how the two were achieved can be seen in Figure1.6.

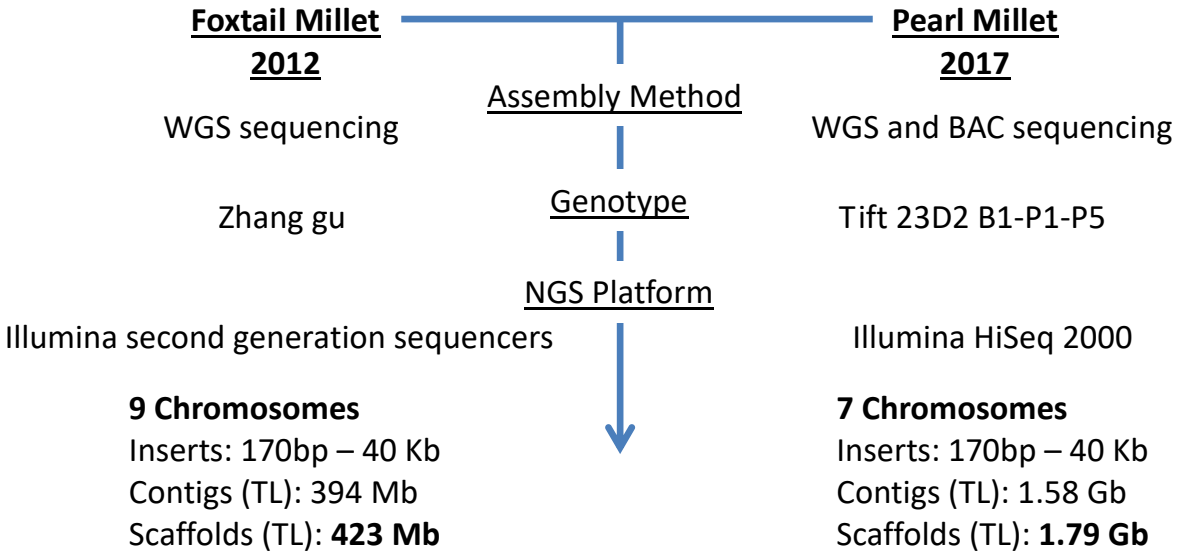


Figure 1.6 Details of the foxtail millet and pearl millet genome sequences (WGS = whole genome shotgun, BAC = bacterial artificial chromosome, NGS = next generation sequencing), derived by Author.

The whole pearl millet genome was sequenced using whole genome shotgun (WGS) and bacterial artificial chromosome (BAC) sequencing from reference genotype Tift 23D2B1-P1-P5 (Varshney *et al.*, 2017) and 1.56 Gb was assembled into 7 pseudomolecules. Additionally, 994 pearl millet genotypes were also resequenced, including 963 inbred lines and 31 wild accessions to gain insight into the population structure, genetic diversity and domestication. 88,256 simple sequence repeat (SSR)

motifs from the pearl millet genome sequence and 32,901,665 SNPs from PMiGAP lines were identified.

In terms of population structure, using 450,000 filtered, high quality SNPs, principal component analysis (PCA) and phylogenetic trees revealed 4 main clusters; 3 contained wild accessions and separated into East, Central and West African clusters, and 1 represented the PMiGAP ~ cultivated lines. Strong population structure could be observed in wild accessions and weak population structure for PMiGAP lines. The weak population structure is explained by the lack of major genetic bottlenecks during the rapid spread of pearl millet agriculture within India and Africa (Varshney *et al.,* 2017).

The genetic basis of the exceptional heat and drought tolerance characteristics of pearl millet were also investigated via the identification of gene families containing cutin, suberin and wax biosynthesis genes. Increased cuticular wax synthesis has long been proven to improve drought tolerance in *Arabidopsis* species (Seo *et al.,* 2011). Genome wide SNP data was used to carry out GWAS across 258 PMiGAP lines for 15 traits. From the subsequent analysis, 1054 highly significant marker trait associations (MTAs) were identified for useful agronomic traits including grain number per panicle, grains per square meter, dry stover yield, fresh stover yield, tillers per plants, panicle diameter, panicle harvest index, panicle length, panicle yield, panicle number, plant population, grain yield, grain harvest index, plant height and 1000-grain-weight. These MTAs explain 9-27% of the phenotypic variation (Varshney *et al.,* 2017). This break-through research will enable a better understanding of trait variation and greatly accelerate the genetic improvement of pearl millet for a more food secure future.

1.12.6 Recombinant DNA Technology and Genetic Modification

Even though recombinant DNA technology and Genetic Modification (GM) is not within the scope of this thesis, they are still acknowledged as powerful tools that would facilitate the improvement of the pearl millet gene pool (O'Kennedy *et al.,* 2006). Research into DNA technology has been developed extensively in major cereal crops, more so than for pearl millet. Although recent advances for the improvement of pearl millet have been well established via traditional breeding methods and MAS, genetic engineering and *in-vitro* culture allows the gene pool to

be expanded further than previously thought possible. This may be facilitated by the transfer of genes which control well-defined traits between species (O'Kennedy *et al.,* 2006).

The first cereal embryogenic *in-vitro* culture systems were established for pearl millet in the 1980's (Vasil and Vasil, 1981) and since then, efficient regeneration systems for pearl millet breeding lines have facilitated the development of reliable transformation systems (O'Kennedy *et al.,* 2006). For the improvement of nutritional quality, genetic engineering methods can be used to elevate mineral and vitamin content in the starchy endosperm of cereal seeds. This has been accomplished in rice, where expression of soybean ferritin (a Fe binding protein) in developing rice seeds has resulted in a threefold increase in endosperm Fe content compared to the non-transformant (Qu *et al.,* 2005). This work was achieved using two types of ferritin hyper-expressing rice lines, which were synthesised via the introduction of a soybean ferritin SoyferH-1 gene under the control of the rice seed storage glutelin gene promoter, GluB-1 and the rice seed storage globulin gene promoter, Glb-1, (Double transformation line GluB-1/SoyferH-1 and Glb-1/SoyferH-1) and by introducing the SoyferH-1 gene under the control of Glb-1 promoter alone (Single transformation line with Glb-1/SoyferH-1). Similar findings were also reported by Goto *et al.* (1999), where the coding sequence of the soybean ferritin gene was transferred into Asian rice by Agrobacterium-mediated transformation. GluB-1 was used to facilitate the expression of the soybean gene in developing, self-pollinated seeds of transgenic plants. Findings showed that the Fe content of seeds from the transgenic plants was up to three times greater than that of their untransformed counterparts (Goto *et al.,* 1999). The same techniques could be applied to pearl millet for the increased Fe accumulation, when more is known about the Fe pathways.

Harnessing tools that facilitate genetic engineering have also been established for anti-nutrient compounds such as phytate. Techniques involve the development of new varieties by generating and utilising genetic variation, via chemical/physical mutagenesis (Oladosu *et al.,* 2016). The pathway of phytate from myoinositol is considered to be well understood and the screening of mutant populations for reduced phytate accumulation is now possible. For example, the identification of low phytate mutants in maize, barley, wheat, soybean, and rice will assist in the selection of similar mutations in millets and these can be incorporated into breeding programs.

Rasmussen and Hatzack (1998) mutagenised barley grains with sodium azide and screened for high levels of free phosphate for the identification of low-phytate mutants. Results showed that 9 out of 27 mutants had increased free phosphate content in the grain and this was correlated with a significant decrease in levels of phytate. Allelic testing of 4 out of the 9 mutants showed that at least two distinct loci control the biosynthesis of grain phytin (a Ca-Mg salt of phytic acid). It is therefore possible to screen for and isolate low phytate mutants through the identification of genes involved in the biosynthetic pathway of phytin. This contributes to the development of low-phytin crops with higher nutritional value (Rasmussen & Hatzack, 1998).

As previously discussed, NA plays a key role in metal assimilation and homeostasis (Morrissey and Guerinot, 2009). Therefore, manipulation of cellular NA concentrations should be considered for the improvement of Fe and Zn content in pearl millet. This has previously been achieved through the use of activation and knockout mutants in rice and tobacco. A study by Inoue *et al.* (2003) demonstrates that among the three *NAS* genes present in rice, *OsNAS1* and *OsNAS2* transcripts are elevated in roots and leaves in response to reduced Fe levels, whereas *OsNAS3* expression is induced in roots but suppressed in leaves when Fe is insufficient. Activation and knockout mutants were used to examine the functioning of *OsNAS3* in metal homeostasis in rice plants and it was found that there was an increase in NA by activation of *OsNAS3*, causing increased levels of Fe and Zn in both leaves and seeds (Lee *et al.,* 2009). *NAS* genes could therefore be potential candidates for the improvement of Fe and Zn in rice. Constitutive overexpression of *NAS* genes also resulted in elevated levels of Fe and Zn in transgenic tobacco plants (Douchkov *et al.,* 2005); this suggests findings may also be relevant to other crops such as pearl millet.

Due to the orphan status of pearl millet, little work has been performed so far on the nutritional enhancement of grains via genetic engineering – thus presenting a significant gap in the literature. However, work on major cereals via reliable techniques and protocols have demonstrated that genetic improvements are possible using genetic engineering approaches (O'Kennedy *et al.,* 2006). In order to employ recombinant DNA technology and GM methods, there needs to be an increase in knowledge about the Zn and Fe pathways in pearl millet and it is also important to

consider factors such as cost, consumer acceptability and socio-economic barriers to smallholder farmers. In many countries, farmers are free to choose which technologies they wish to employ. For example, in the USA the use of GM seeds with respect to maize, cotton and soybean exceeds 90% (James, 2015), the majority of which is used as feed for food-producing animals (Van and Young, 2014). The main reasons for this choice are better performance and increased profits. However, in particularly vulnerable and developing regions, such as where the majority of pearl millet is grown for human consumption there may be cost constraits to smallholder farmers due to expensive patented seed (Lucht, 2015) alongsidescepticism, fears of eroding indigenous crop diversity and increased pesticide costs (Makanya, 2004). Therefore, many Indian and African small-scale farmers may reject GM crops and instead opt for traditional plant breeding systems that take into account local pests, soils and weather patterns.

1.13 Aims of Thesis and Chapter Descriptions

Micronutrient malnutrition of Fe and Zn is a persistent challenge for global development and predominantly affects low and middle income populations living in India and Africa. Therefore, enhancing levels of these micronutrients by crop biofortification is increasingly being recognised as a cost-effective and sustainable solution. A suitable crop to focus these efforts on is pearl millet, due to its concurrent growth with climate change and the fact that it is highly accessible to those who suffer from MNDs. The identification of genomic regions associated with elevated levels of grain Fe/Zn will aid the breeding of elite lines, with enhanced micronutrient content, leading to improved health and well-being, especially for women and children. Furthermore, the improvement of pearl millet will be greatly accelerated, due to the recent pearl millet genome assembly, the sequencing of which proves the importance of this crop is being recognised.

The aim of this project was to identify genes associated with Fe/Zn uptake in the PMiGAP. The first objective was to phenotype levels of grain Fe and Zn by Inductively-coupled Plasma Atomic-Emission Spectroscopy (ICPAES) in glasshouse and field populations and to analyse the influence of environment on micronutrient uptake (Chapter 3). Other phenotyping objectives included assessing factors that may affect the bioavailability of Fe and Zn, including of levels of phytate (Chapter

4) and metal-chelating phenolic compounds, identified using High Performance Liquid Chromatography-Mass Spectrometry (HPLC-MS) (Chapter 5). In Chapter 6, GBS libraries were created from PMiGAP DNA. The resulting GBS sequence reads were then mapped to a foxtail millet *Setaria italica* reference genome (pearl millet did not have a reference genome at the time of this research). As a result, 663 polymorphic SNP markers were generated for GWAS, to identify markers associated with mineral uptake. Prior to GWAS, the population structure in the PMiGAP was analysed to minimise the rate of false positives between MTAs. The extent of Linkage Disequilibrium (LD) was also assessed among all pairs of loci, using the most commonly used LD measure, r^2. After GWAS, a 4kb region surrounding each significant SNP was BLASTed against the most annotated *Setaria italica* reference genome, for the identification of candidate genes. In September 2017, a pearl millet genome assembly and 32,901,665 SNPs became available for public use (Varshney *et al.*, 2017). The SNPs were downloaded and filtered into two data sets, which contained >37,000 SNPs (as a result of stringent filtering) and >3,000,000 SNPs (as a result of less stringent filtering), respectively. These were used to conduct two additional GWAS', in hope of yielding more significant findings. In the final chapter (Chapter 7), haplotypes covering the YUCCA11 gene were identified and their association with Fe/Zn uptake was assessed. Figure 1.7 demonstrates how the project proceeded to address aims and objectives.

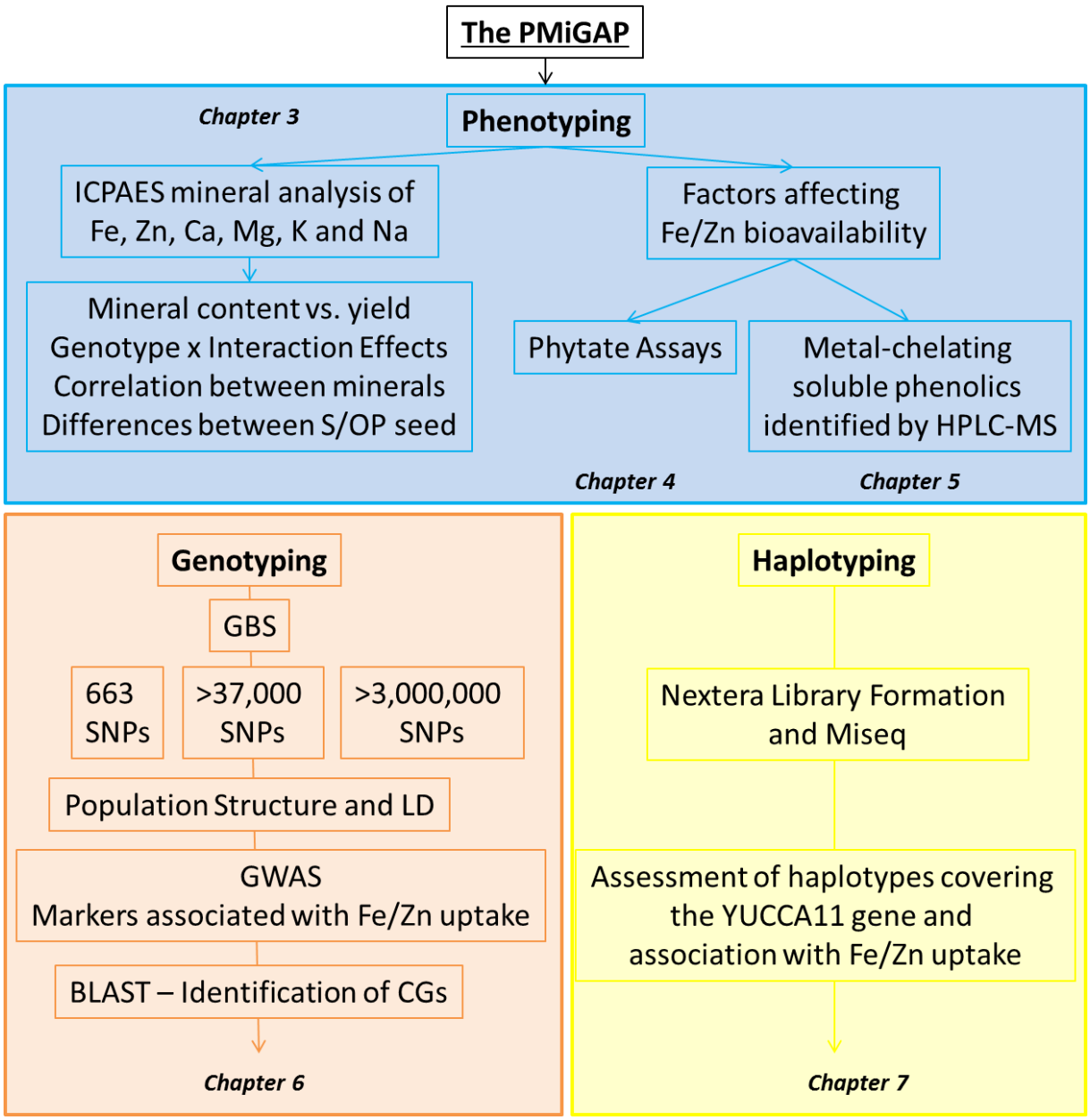


Figure 1.7: The project process. Blue shows aspects covered in chapter 3, 4 and 5; orange shows processes covered in chapter 6 and yellow shows areas of the project covered in chapter 7. S = Selfed; OP = Open-Pollinated, CG = Candidate Gene.

Chapter 2: Plant Material for Phenotyping/Genotyping

2.1 Summary

230 selfed and 235 Open-Pollinated (OP) PMiGAP lines were obtained from ICRISAT, Patancheru, India according to the ICRISAT guidelines for germplasm distribution. Guidelines state that researchers may request up to 5g per line. In order to bulk enough seed for phenotyping/genotyping, three in-house seed multiplication trials were conducted under glasshouse conditions using selfed imported seed. Mature seed was harvested for phenotyping purposes, as well as young leaf tissue (harvested at four weeks), which was subsequently used for DNA extractions.

2.2 Introduction

The findings drawn from this research are based on phenotypic/genotypic analysis of the PMiGAP. As previously discussed, the PMiGAP has been drawn from a core collection, consisting of 2094 pearl millet accessions, landraces and cultivars grown across three continents. The resulting lines represent the entire cultivated global diversity of pearl millet and cover a wide range of genetic variations present in the pearl millet germplasm (Bhattacharjee *et al.*, 2007). The original 2094 accessions were analysed at ICRISAT, India, using genome wide SSR markers, and grouped into clusters based on genetic similarity. 24 core clusters were identified and a number of lines were selected from each cluster (ranging from 3-14). As a result, the PMiGAP is classed as a mini core collection, consisting of 250 lines in total (Yadav *et al.*, 2010).

PMiGAP seed was obtained from the ICRISAT genebank, which serves as a world repository for the collection of pearl millet germplasm, as well as for sorghum, chickpea, pigeon-pea, groundnut, finger millet; and five small millets ~ foxtail millet, little millet, kodo millet, proso millet and barnyard millet. In-house seed multiplication under venlo glasshouse conditions was necessary to ensure sufficient amounts of plant material for phenotyping/genotyping.

2.3 Materials and Methods

All imported PMiGAP lines were grown by ICRISAT staff in the fields at ICRISAT, India. These were imported according to the ICRISAT guidelines for germplasm distribution (http://genebank.icrisat.org/PDF/Section/Section8.pdf).

2.3.1 PMiGAP Selfed Seed Multiplication, ICRISAT, 2010

230 PMiGAP inbred lines were sown at ICRISAT, Patancheru (17°31'4"N 78°16'43"E) India, in shallow sandy loam (alfisol) soil in the late rainy season on 23rd August 2010. These were hand sown at 2-3 cm depth in plots of 4 m length with 1 m paths in between. The plants were thinned 3 weeks after sowing to maintain 10-15 cm distance between them. The distance between the ridges was 75 cm. Crop nutritional requirements were met by applying 40 kg of Nitrogen (N) and 18 kg of Phosphorus (P) ha^{-1} into the ridges before sowing and side dressing an additional 45 kg of N ha^{-1} after thinning. Irrigation was provided at periodical intervals until harvest, which took place on 15th December 2010. One selfed panicle was harvested from each accession and the seed was threshed and retained (Rao and Bramel, 2000).

Seed Viability

Seed viability was tested at IBERS, Aberystwyth University in 2014 by observing the germination rate of 5 seeds per line. Seeds were placed in petri dishes lined with filter paper (Camlab, UK) dampened with water. They were subsequently placed in a heated oven at 20°C and observed daily for signs of germination.

2.3.2 Multiplication of PMiGAP Open Pollinated Lines, ICRISAT, 2013

235 PMiGAP lines were sown at ICRISAT, Patancheru (17°31'4"N 78°16'43"E) India, in shallow sandy loam (alfisol) soil in the late rainy season in August 2013. These were hand sown at 2-3 cm depth in plots of 4 m length with 1 m paths in between. The plants were thinned 3 weeks after sowing to maintain 10-15 cm distance between them. The distance between the ridges was 75 cm. Crop nutritional requirements were met by applying 40 kg of Nitrogen (N) and 18 kg of Phosphorus (P) ha^{-1} into the ridges before sowing and side dressing an additional 45 kg of N ha^{-1} after thinning. Irrigation was provided at periodical intervals until harvest, which

took place in December 2013. One OP panicle was harvested from each accession and the seed was threshed and retained (Rao and Bramel, 2000).

2.3.3 PMiGAP Seed Multiplication, Glasshouse Conditions, Aberystwyth University

Three independent seed multiplication trials took place under venlo glasshouse conditions at IBERS, Aberystwyth University (Trial 1: October 2014 – February 2015, Trial 2: February – June 2015, Trial 3: August 2015- January 2016).

Trial 1: In October 2014, 230 27.5 × 27 × 24 cm pots were filled with John Innes potting compost No.3 and placed in rows of four in a temperature controlled venlo glasshouse system (Figure 2.1).

Figure 2.1: Pearl millet growing under glasshouse conditions, IBERS, Aberystwyth University (photo taken by Author).

The concrete floor area was lined with a waterproof sheet plus an absorbent felt material. The temperature was maintained at 28°C and 10 hours daylight was artificially provided by Philips Son T Plus High Pressure Sodium (HPS) Bulbs (400 watts). HPS lamps produce ‘redder’ light and are widely used to extend the natural ‘day length’ that a plant is subjected to, thereby imitating summer conditions.

In each pot, 10 seeds per genotype were sown at a depth of 2-3cm from the top of the soil (Figure 2.2) and watered daily using a hose with a spray nozzle attachment.

Figure 2.2: Sprouting pearl millet seedlings (photo taken by Author).

After one month the artificial daylight hours were reduced to 8 hours and watering took place on alternate days. The first plant to reach the panicle initiation stage per pot was covered with a 30cm × 10cm agricultural parchment paper bag and fastened with a staple, for the production of selfed seed. The other plants were able to exchange pollen, to produce Open Pollinated Varieties (OPV's). After two months, the majority of genotypes started to produce pollen and a fan was installed to aid pollen circulation. After three months, most plants reached physiological maturity as indicated by yellowing stems, a dried up appearance and a firm panicle with seeds that had a "black layer" at the base. The panicles were harvested, placed in cotton storage bags and dried in an on-site desiccator oven at 28°C for 5 days. Panicles were subsequently threshed and separated from the chaff by hand. The collected seeds were then weighed and stored in paper envelopes at 4°C, in the presence of silica gel in a Nalgene desiccator cabinet (Cat. No 3517). This process was repeated in February 2015 (trial 2) and August 2015 (trial 3).

Observations

Following observations from trial 1, several changes took place in the second and third trials with the aim of improving yield:

During the first trial, it was found that the pots were placed too close together; this meant that genotypes which took longer to germinate were generally unsuccessful due to the surrounding larger plants shading the seedlings during the initial growth stages. This suggests that the competition for light had an effect on successful growth. In light of this, 230 pots were spread across two venlo compartments in rows

of three during the second and third trial to allow more space between the pots (7.5 inches between each pot). For the production of OP seed, pollen could also be exchanged across the venlo compartments via the removal of 3 large glass panels separating the two compartments.

The natural day length was longer throughout February's seed multiplication trial due to daylight saving time. This was a factor beyond experimental control and tests have been conducted determine whether this had an effect on mineral uptake (Chapter 3). As a result of the longer daylight hours, plants grew much faster and exhibited more vegetative growth as compared to trial 1.

Aphid infestations - During trial 2, there were two aphid infestations during the months of May and June, lasting around two weeks each. Additionally, during trial 3, there was one infestation in November, lasting 1 week. The treatment was a chemical spray (APHOX, applied at 1.4 gms per litre of water), applied by hand.

Due to limited amounts of selfed PMiGAP seed stock, the number of plants covered for the production of selfed seed increased from 1 to 4 during trials 2 and 3.

2.3.4 Young Leaf Tissue for DNA Extractions

PMiGAP young leaf tissue (1-2 leaves per plant) was harvested from seed multiplication trial 2, when plants had been growing for 1 month. Upon harvest, the leaves were immediately placed in liquid nitrogen, and then stored at -80°C in 5 × 10 cm plastic bags.

2.3.5 Pearl Millet *HHB67 Improved*

HHB67 Improved was imported from ICRISAT according to the ICRISAT guidelines for seed distribution.

Prior to the development of *HHB67 Improved*, the hybrid, *HHB67* was released in 1990 by CCS Haryana Agricultural University, India, due to its extra-early maturity (65 days from sowing to grain maturity). In the late 1990's it was grown on over 100,000 ha in Haryana and Rajasthan states in India. Unfortunately, the hybrid was found to be susceptible to Downy Mildew (DM) disease, which developed in up to 30% of crops planted. In light of this, marker-assisted backcrossing with the elite donor parent ICMP 451 took place to add DM resistance to the male parent H

77/833-2 and additional DM resistance genes were backcrossed into female parent 843A/B from donor ICML 22 using traditional progeny-based greenhouse screening of pot-grown seedlings. Greenhouse disease screening and subsequent field testing across six environments confirmed DM resistance improvement in the new parental lines and their hybrids. As a result, farmers have since expressed a preference for *HHB67 improved*, which was also found to have significantly higher grain and stover yields (5–10%) than that of the original *HHB67* (Khairwal & Hash, 2007).

2.3.6 Other Cereals

9 cereal samples were donated by Aberystwyth University (Wheat samples), Bangor University (Barley) and Aberdeen University (Rice) (Table 2.1).

Table 2.1, Cereals used for comparison against *HHB67 Improved* (DM = downy mildew, AU[1] = Aberystwyth University, AU[2] = Aberdeen University, BU = Bangor University, BBL = Bangor breeding line).

Analysis	Crop	Sample	Sample Type	Obtained from	No. of Lines	Information
Mineral Quantification (Fe, Zn, Na, K, Ca, Mg) ICPAES	Pearl Millet	*HHB67 Improved*	Mature Seed	ICRISAT	1	Control.
	Rice	Kalinga III	Mature Seed	AU[2]	1	A popular variety grown in the upland areas of Eastern India. Tall with long slender grains, resistant to brown spot, cold tolerant and produces average yield (Steele *et al.*, 2006).
	Wheat	CV Viscount	Mature Seed	AU[1]	7	Soft, low-nitrogen wheat. Popular in Scotland. Moderate winter dormancy. Good tillering ability and retention, backed by reasonable disease resistance (Green *et al.*, 2015).
		CV Revelation	Mature Seed	AU[1]		Produced in Europe. Does not appear to have high yield potential. Known for reliability in terms of an excellent all-round disease resistance profile. (Limagrain Europe, 2015).
		CV Riband	Mature Seed	AU[1]		A soft (low gluten content) variety. Low in protein (11.0 – 11.5%)
		CV Batalion	Mature Seed	AU[1]		Common in Europe and has temperature-sensitive resistance to Brown Rust which is ineffective at 25°C but is effective at 10°C (Jones, 1997).
		CV Hereward	Mature Seed	AU[1]		Known for good bread making qualities due to increased protein content, as compared to CV Riband (Khatkar *et al.*, 1995).
		CV Cordiale	Mature Seed	AU[1]		Known for good bread making qualities.
		CV Hereward II	Mature Seed	AU[1]		Known for good bread making qualities.
	Barley	Naked Barley (BBL)	Mature Seed	BU	1	Favourable taste and has high levels of beta-glucan soluble fibre. It has potential for delivering public health benefits in reducing rates of obesity and Type-2 diabetes. Susceptible to lodging and foliar diseases when grown in the UK (Dicken *et al.*, 2011).

2.4 Results and Discussion

2.4.1 Seed Viability

The data in Table 2.2 suggests that the selfed PMiGAP seed population imported from ICRISAT in 2010 is a reliable population for in-house multiplication, since 90% of the lines tested achieved a 100% germination rate when grown on dampened filter paper, in the laboratory. This high reliability rate may be due to proper storage at 4°C in the presence of silica gel beads.

Table 2.2, % Seed viability for the selfed PMiGAP population imported from ICRISAT, 2010**.**

% Seed viability*	No. of lines
100%	206
40%	6
20%	6
0%	12
	Total: 230 PMiGAP lines

* Seed viability measured as a percentage of how many, out of a total of 10 seeds germinated when grown on damp filter paper in a 10 day period.

2.4.2 Yield

The data in Figure 2.3 suggests that across the seed multiplication trials, enough seed was produced to meet the requirements for the experiments detailed in Table 2.3. Due to increased distance between pots for trials 2 and 3, as compared to trial 1, yield increased significantly, suggesting that adequate daylight is an extremely important factor during the initial seedling growth stages, due to the growth rate variation between PMiGAP accessions.

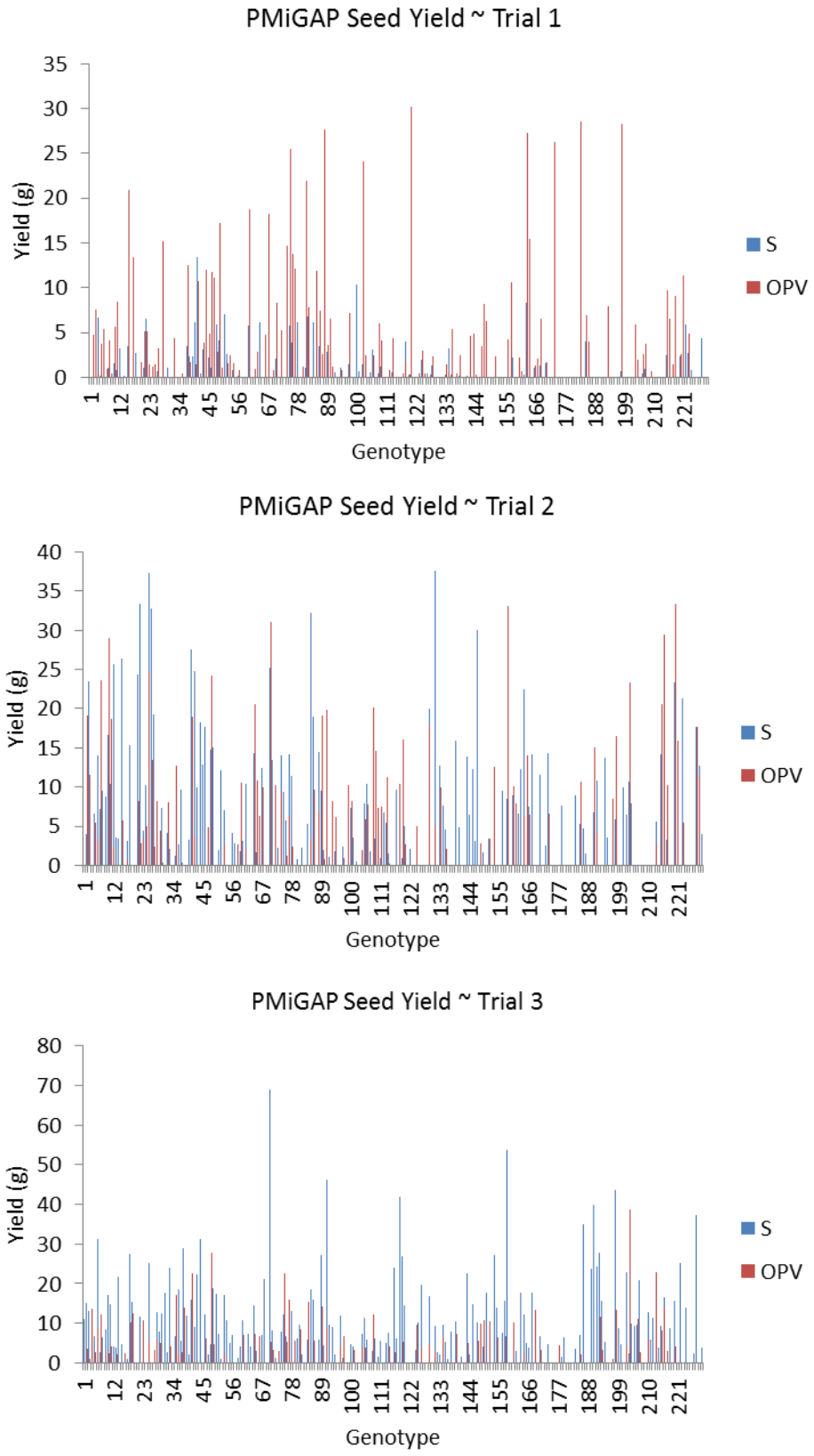


Figure 2.3, Seed yield per trial. S = selfed seed, OPV = open pollinated varieties

Table 2.3, Summary of the plant material used within this thesis, per chapter and experiment within. (* = See Table 2.1, S = self, OPV = open pollinated variety, SMT = seed multiplication trial, I = ICRISAT, P = phytate assays, UA = untargeted analysis, TA = targeted analysis, MD = method development, MS = mature seed, YL = young leaf, Exp. = experiment, HTA = haplotype trait associations).

					No. of Genotypes Obtained from…							
Chapter, Analysis	**Exp. No.**	**Plant Material**	**No. of Genotypes**	**Weight Required/ Genotype (g)**	**I S**	**I OPV**	**SMT1**	**SMT 2**	**SMT3**	***HHB67 Improved***	***Other Cereals**	**I S (2012)**
3 ICPAES	1	Mature Seed	10	1	-	-	-	-	-	1	9	-
	2	Mature Seed	229	1	229	-	-	-	-	-	-	-
	3	Mature Seed	40	1	-	-	40	40	-	-	-	-
	4 (Pilot)	Mature Seed	20	1	20	-	20	20	-	-	-	-
	4	Mature Seed	99	1	99	-	-	-	99	-	-	-
	5	Mature Seed	131	1	131	131	-	-	-	-	-	-
4 Phytate assays + ICPAES	1	Mature Seed	84	0.5	-	-	42 (P) 42 (ICPAES)	-	42 (P) 42 (ICPAES)	-	-	-
	2	Mature Seed	235	0.5	-	235 (P) 235 (ICPAES)	-	-	-	-	-	-
5 UV-HPLC + HPLC-MS	MD	Mature Seed	4	12	-	-	4	-	-	-	-	-
	1	Mature Seed	34	0.5	-	-	34	-	-	-	-	-
	2	Mature Seed	57	1	-	-	-	-	57 (UA) + 57 (TA)	-	-	-
	3	Mature Seed	1	1	-	-	-	-	-	1	-	-
	4	Mature Seed	185	0.1	-	185	-	-	-	-	-	-
6 ICPAES	GWAS 1	Mature Seed + YL	223	1	223 (MS ICPAES)	-	-	-	-	-	-	223 (YL DNA)
	GWAS 2	Mature Seed	223	1	223 (MS ICPAES)	-	-	-	-	-	-	-

	GWAS 3	Mature Seed	223	1	223 (MS ICPAES)	-	-	-	-	-	-	-
7 DNA Extractions +ICPAES	DNA Extraction	YL	48	1	-	-	-	48	-	-	-	-
	HTA	Mature Seed	42	1	42 (ICPAES)	-	-	-	-	-	-	-

Where sub-populations of PMiGAP lines have been used (for example in Chapter 3, experiment 4, where 20 lines were analysed for GEI effects), these samples were selected based on yield. Lines that produced a total yield of >5g were selected for analysis. Therefore, it was considered that this would prevent the selection of a truly random sample. It was considered that the lines that produced >5g of seed for this study may have been hardier genotypes, which may have caused some degree of bias. In order to test this, the selected lines were plotted on a Zn/Fe scatter plot, as measured by ICPAES and compared to that of the entire PMiGAP population. From the results in Figure 2.4, no obvious clustering can be seen, therefore there is some degree of random selection. Details of how other PMiGAP sub-populations were selected can be seen in Table 2.4.

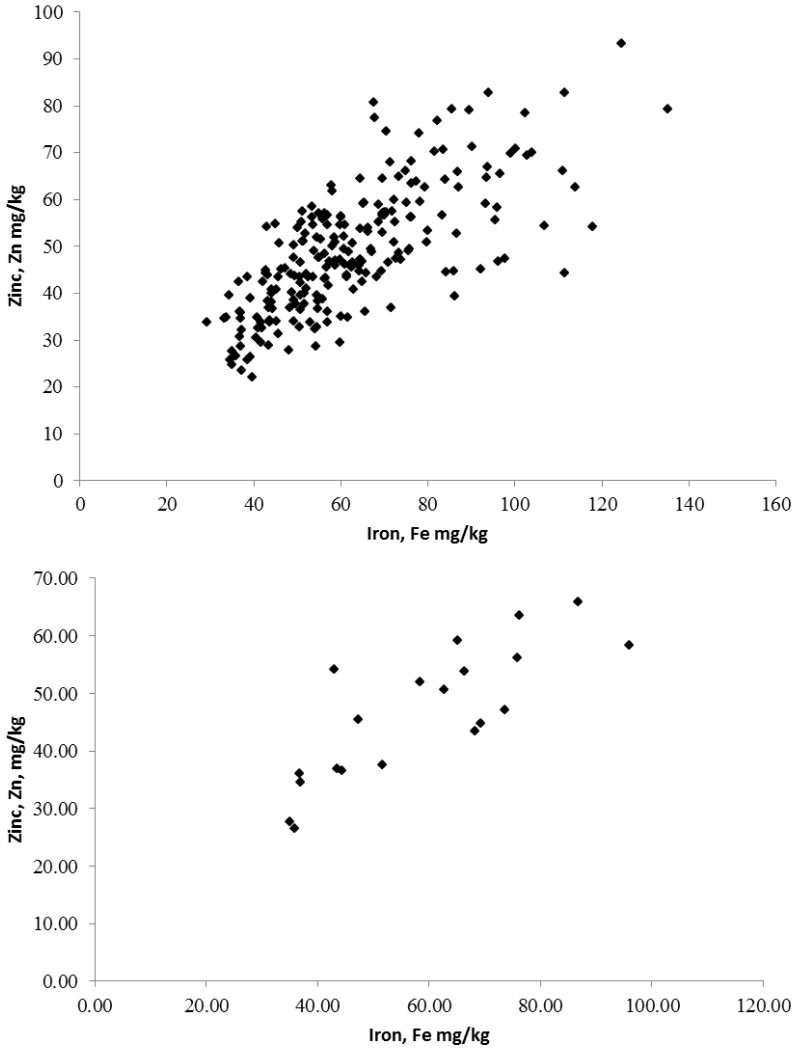


Figure 2.4, Top: Zn and Fe content in 229 PMiGAP lines. Bottom: Zn and Fe content in 20 PMiGAP lines selected for a GEI Experiment, which produced >5g total yield.

Table 2.4, Selection of PMiGAP sub-populations (SMT = seed multiplication trial, MD = method development, S = selfed, OPV = open pollinated variety, HTA = haplotype trait association).

Chapter, Analysis	**Exp No.**	**No. of Genotypes**	**Selection Based On…**
3 ICPAES	1	10	N/A
	2	229	Imported Seed (ICRISAT) S
	3	40	Yield (SMT 1 + 2)
	4 (Pilot)	20	Yield (SMT 1 + 2)
	4	99	Yield (SMT 3)
	5	131	Imported Seed (ICRISAT) OPV
4 Phytate assays + ICPAES	1	84	Yield (SMT 1 + 2)
	2	235	Imported Seed (ICRISAT) OPV
5 UV-HPLC + HPLC-MS	MD	4	Yield (SMT 1)
	1	34	Yield (SMT 1)
	2	57	Yield (SMT 3)
	3	1	Imported Seed (ICRISAT) *HHB67 Improved*
	4	185	Imported Seed (ICRISAT) OPV
6 ICPAES	GWAS 1	223	Trait data - Imported Seed (ICRISAT) S Imported DNA (ICRISAT) S
	GWAS 2	223	Trait data - Imported Seed (ICRISAT) S SNP Markers (Varshney *et al.*, 2017)
	GWAS 3	223	Trait data - Imported Seed (ICRISAT) S SNP Markers (Varshney *et al.*, 2017)
7 DNA Extractions +ICPAES	1 DNA Extraction	48	High, Med Low Combined Fe/Zn content
	2 HTA	42	High, Med Low Combined Fe/Zn content

2.4.3 The Limitations of Glasshouse Grown Seed

The use of an in-house glasshouse system facilitated the multiplication of pearl millet seed in the UK, without which growth would have been impossible due to the cooler climate as compared to India and Africa, where pearl millet is grown naturally. Whilst there were many advantages to the use of such a system, including minimal external threats of adverse weather conditions and grazers, as well as 24-hour controlled temperature and light, there were also many limitations including high operating expenses and increased risk of insect pests. As previously discussed, a total of three aphid infestations occurred during the seed multiplication trials. This required the use of a chemical treatment spray (APHOX). Although the chemical treatment spray was not thought to have any significant impact on seed quality, the stress of the infestation on the plants may have had a negative impact since aphids in large numbers are known to cause damage, which in-turn may reduce yield and seed size (Kolbe, 1970).

2.5 Conclusions

The in-house seed multiplication trials are believed to be the first conducted at IBERS, Aberystwyth University. Therefore, the detailed observations from trial 1 were crucial for the improvement of yield in trials 2 and 3. Whether the challenges faced had a significant impact on micronutrient uptake is determined in Chapter 3, where GEI effects were analysed between lines grown under glasshouse and field conditions. Although the imported samples and seed multiplication trials gave an adequate amount of plant material for the unreplicated experiments listed in Table 2.3, the seed resulting from glasshouse trials was shared between other members of the research group, for other projects not relating to this one. Therefore, there are regrettably no experimental replicates of PMiGAP lines obtained under field or glasshouse conditions, use in the experiments listed in Table 2.3. In light of this, it is recommended that seed is bulked on much a larger scale for future work, with the aim of producing enough for at least three replicates per experiment. This will greatly improve the reliability and validity of findings in the downstream analysis.

Chapter 3: Quantification of Grain Trace Elements by ICPAES

3.1 Summary

Comparing the levels of 6 grain micronutrients (Fe, Zn, K, Mg, Na and Ca) present in the pearl millet hybrid line *HHB67 Improved* against that of 9 major staples (including wheat, rice and barley) revealed higher levels of Fe, Zn, Mg and Na in pearl millet. Therefore, pearl millet is deemed an excellent candidate for the promotion of good health and nutrition, even though its nutrition credentials are currently undersold.

To generate trait data for GWAS, inductively coupled plasma atomic emission spectroscopy (ICPAES) mineral analysis was conducted on 229 selfed PMiGAP lines. Results showed that grain Fe and Zn content was highly variable and positively correlated ($R=0.72$; $P\leq0.05$), this suggests the possibility of simultaneous selection for both minerals. To test suitability for agriculture, ten lines from the 90^{th} percentile for grain Fe and Zn levels were compared against the total yield of three in-house seed multiplication trials, per genotype. Findings revealed 5 high Fe/Zn lines with above-average yield. For example, the line IP19405 contained 102 mg/kg Fe, 78 mg/kg Zn and produced a total of 71g seed, where the average total yield was 26g.

The size of genotype × environmental interaction (GEI) effects on mineral uptake was examined in multi-environmental trials to test the reliability of trait data. Between 2 trials conducted during different seasons, under glasshouse conditions, levels of Fe, Zn and Mg did not differ significantly, but did for Na, Ca and K. Furthermore, mineral levels between glasshouse and field populations revealed that Fe, Na and Zn levels differed significantly between the two populations, but Ca, K and Mg levels did not. Therefore, the differences indicate that the environment has a stronger effect on mineral uptake than the genetics, indicating that the data should be approached with caution. Additionally, the mineral content in OP and selfed seed harvested from field populations was also compared. It was found that mineral levels were significantly higher in selfed populations, which may be indicative of seed concentration effects.

3.2 Introduction

3.2.1 ICPAES

During the last century, a major objective of conventional plant breeding was to breed high yielding cereal crops with increased total calorific value, to meet an increased demand for food, owing to an increasing global human population. As a result, the fibre, protein and micronutrient content of grains were largely ignored, leading to an unintentional increase in micronutrient malnutrition, i.e. hidden hunger (Gupta *et al.,* 2009). The resulting increasing prevalence of Fe and Zn deficiencies in human populations on a global scale has stressed the need for more research focusing on the distribution and chemical speciation of these elements in cereal products (Persson *et al.,* 2009). This is a prerequisite for efficient breeding/genetic engineering of elite, micronutrient rich cultivars, for a more nutrition secure future. The levels of 6 mineral elements (Fe Mg, Na, K, Ca and Zn), present in PMiGAP lines were determined by ICPAES. This method is proven to be highly sensitive, accurate, and can be used to determine many elements at the same time (Du & Du, 2009). The results will provide reliable trait data for GWAS, to ultimately determine significant MTAs and in turn candidate genes. The principles of ICPAES can be seen in Figure 3.1.

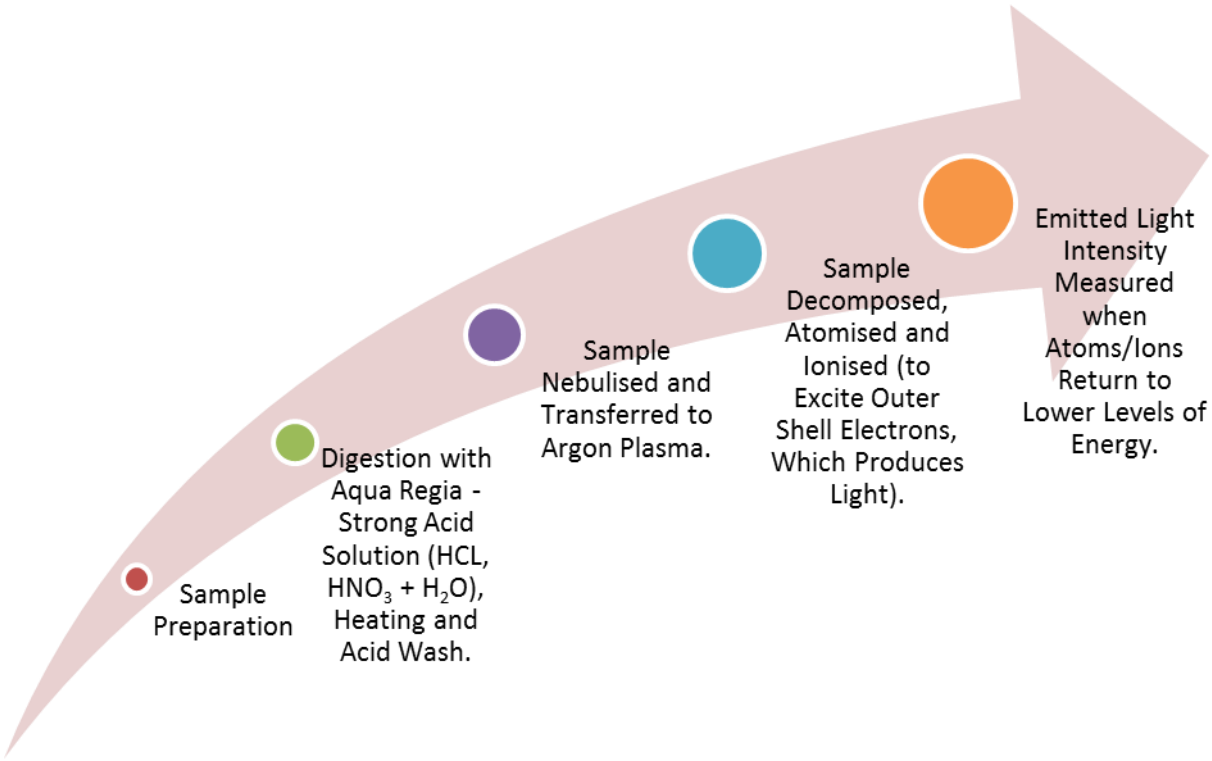


Figure 3.1: The principles of ICPAES, derived by author.

3.2.2 The Mineral Content of Pearl Millet

As previously discussed, plant based food products are a staple food for people in many parts of the world either by choice, circumstance or religious constraint. They constitute an important source of carbohydrates, protein, dietary fibre, vitamins and minerals (Berwal *et al.*, 2017). Pearl millet grains are one of the cheapest sources of Fe and Zn as compared to other cereals and vegetables (Rao *et al.*, 2006) and many health benefits associated with their consumption have been well documented, this includes many accounts that the mineral profile of pearl millet is better than that of other cereals, including wheat and maize (Manwaring *et al.*, 2016). However, the bioavailability of bivalent minerals including Fe, Mg etc. is lower in pearl millet due to the presence of anti-nutrient factors, such as phytate.

Successful crop improvement, facilitated by traditional plant breeding is dependent on the presence of natural genetic variation for the target traits in the gene pool. The amount of influence that the environment has on the phenotype is a hotly debated topic known as GEI, which is most commonly observed as a change in genotype performance across different environments (Cooper and Delacy, 1994). Significant

GEI may seriously limit the feasibility of selecting superior genotypes for the breeding of elite lines (Gurmu *et al.*, 2009, Živanović *et al.*, 2012). When genotypes are selected for wide adaptation, plant breeders prefer minimal GEI (Matus-Cadiz *et al.*, 2003) because when this is the case the target trait is likely to be expressed across multiple environments (Basford & Cooper, 1998). Breeding for elevated grain Fe/Zn levels is complicated because many studies document that micronutrient concentration is highly dependent on the environment, particularly soil mineral composition (Feil *et al.*, 2005).

For this study, analysing the mineral content of PMiGAP lines belonging to just one growing season is not a sufficient method to determine a true estimate of mineral content. Many other factors must be considered (Figure 3.2), including: The differences that occur between environments and if these will have an impact on mineral uptake (GEI effects), any correlation between the uptake of certain minerals - this is important when elucidating the uptake pathways and understanding how they can be improved, whether lines associated with high mineral content also produce acceptable yield and lastly, whether any seed-set differences between OP and selfed varieties will cause a significant difference to overall mineral concentrations. A robust analysis during these early stages will allow any conclusions from the GWAS to be met with increased confidence.

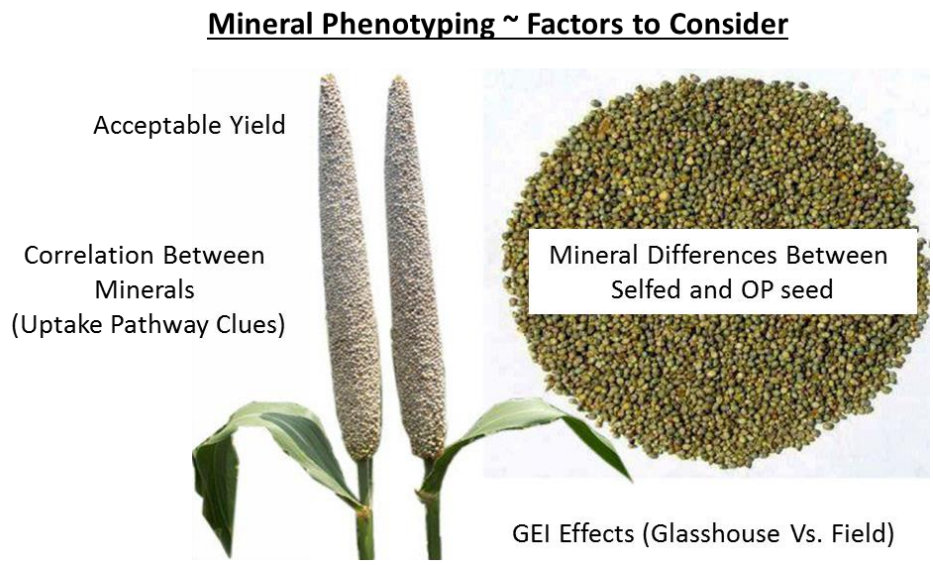


Figure 3.2, Mineral phenotyping :Factors to consider (derived by Author).

3.2.3 Experimental Aims

ICPAES analysis for the quantification of 6 minerals has been conducted on 9 different staples, including wheat, barley and rice. The data was compared to the nutrition credentials of a popular variety of pearl millet ~ *HHB67 Improved*, which is grown on over 500,000ha in Haryana and Rajasthan, India (Hash *et al.,* 2006). Mineral levels were also quantified in 229 selfed PMiGAP lines, grown under field conditions to determine if any correlations between the minerals exist and to identify high and low Fe/Zn lines for the future GWAS.

A negative correlaton exists between elevated grain Fe/Zn content and grain yield in some crops, such as maize (Banziger & Long, 2000) and sorghum (Reddy *et al.*, 2005). This indicates some difficulty in breeding simultaneously high levels of micronutrients and acceptable grain yield. In light of this, once high Fe/Zn lines were identified, the lines that fell within the 90th percentile were compared with the combined yields of three in-house seed multiplication trials, per genotype, for the selection of superior lines with increased grain Fe/Zn levels coupled with high yield.

To gain insight into GEI effects on mineral uptake, differences in mineral concentration between lines grown in multi-environments, including glasshouse and field conditions were assessed. In the pilot phase, GEI effects were determined by comparing the mineral levels of 10 PMiGAP lines grown under field conditions, to

those grown in glasshouse conditions. Analysis was then conducted on a larger scale, using 99 PMiGAP lines and the GEI effects between the two unrelated environments were evaluated.

Both selfed and OP PMiGAP lines have been used in this study for mineral analysis. Since reliable estimates of mineral concentrations have a direct bearing on breeding efficiency (Rai *et al.,* 2015), selfed and OP seed from 131 lines grown under field conditions were examined for grain micronutrient density and potential factors that may cause differences between the two populations are discussed.

3.3 Materials and Methods

3.3.1 Plant Material

See Chapter 2, Table 2.1 and Table 2.3.

3.3.2 ICPAES Analysis

ICPAES analysis was conducted by Dr Sue Lister and Mrs Delma Jones using an accredited service at the Institute of Biological, Environmental and Rural Sciences, Aberystwyth University, UK.

Sample digestion

Two analytical quality control (AQC) samples and one blank were run with every batch of 40. Using an electric powered grinding mill (Retsch Mortar Grinder Mill, 110V/60Hz OY-04181-10) 1g of whole pearl millet seed from each genotype was ground to obtain a fine powder, which was able to pass through a 1mm sieve.1 g of sample was placed into a 100mL Kjeldahl and 15mL digestion acid (Aqua regia, 780mL of concentrated hydrocloric acid + 500mL nitric acid + 7.20mL deionised water) was added and left to soak in a fume cupboard overnight. Samples were digested on a Kjeldahl heating block at 120°C for 3 hours and cooled for a further 1 hour. The solution was washed into a 50mL volumetric flask using diluent acid (250mL nitric acid + 2L deionised water) and shaken. The contents of each volumetric flask was filtered through Whatman No. 1 filter paper (25mm) into a 100mL conical flask, using the initial filtered extract to rinse the flask.

Examination of extracts

Standards were prepared by pipetting the relevant volume of single element stock solutions into the flasks (as below) and made up to 1L with diluent acid. Standards were poured into 25mL test tubes and the digested samples into 8mL test tubes, and then measured by ICP-AES.

Preparation details for Standard Solutions (Ppm = Parts per million):

Mineral	Std	Concentration of single element stock (ppm)	ML of stock to add to 1L	Ppm	Ppm in plant material
Zn	1	1000	0.1	0.1	5
	2		0.5	0.5	25
	3		1.0	1.0	50
Fe	1	1000	1.0	1.0	50
	2		2.0	2.0	100
	3		4.0	4.0	200

Operating conditions of the ICP-AES instrumentation:

Template: Fe, Zn

Element	Wavelength (nm)	Rep	Curve
Fe	259.940	3	Linear
Zn	213.856	3	Linear

Power:	1.2 kW
Plasma gas flow:	15.0 L/min
Auxiliary gas flow:	1.5 L/min
Spray chamber type:	Glass cyclonic
Torch type:	Axial Quartz with 90° bend
Nebuliser type:	Conikal glass concentric/OneNeb
Nebuliser flow:	0.75 L/min
Pump tube:	Orange-orange (inlet)
	Blue-blue (outlet)
Pump rate:	15 rpm
Replicate read time:	5 s
Sample delay time:	25 s
Rinse time:	10 s
Stabilisation time:	15 s
Background correction:	None
PMT voltage:	650 V

Filter position: Default

Template: Ca, K, Mg, Na

Element	Wavelength (nm)	Rep	Curve
Ca	317.933	3	Linear
K	769.897	3	Quadratic
Mg	285.213	3	Linear
Na	589.592	3	Linear

Power:	1.2 kW
Plasma gas flow:	15.0 L/min
Auxiliary gas flow:	1.5 L/min
Spray chamber type:	Glass cyclonic
Torch type:	Axial Quartz with 90° bend
Nebuliser type:	Conikal glass concentric/OneNeb
Nebuliser flow:	0.75 L/min
Pump tube:	Orange-orange (inlet)
	Blue-blue (outlet)
Pump rate:	15 rpm
Replicate read time:	5 s
Sample delay time:	30 s
Rinse time:	10 s
Stabilisation time:	15 s
Background correction:	None
PMT voltage:	650 V
Filter position:	Default

3.3.3 Statistical Analysis

Data were analysed using the software GenStat (16th edition, VSN International Ltd, Hemel Hempstead, UK) using the 2 sample *t*-test, one-way analysis of variance (ANOVA) with location as the fixed effect, Spearman's Rank correlation, and Pearson's product moment.

3.4 Results and Discussion

3.4.1 Comparing Micronutrient Levels in Pearl Millet Wheat, Rice and Barley

Table 3.1: Comparison of pearl millet to wheat, rice and barley (*[1Lowest Ranking Value-10Highest ranking Value], TMC = total mineral content, BBL = Bangor breeding line).

Analyte	**Ca**	**K**	**Mg**	**Na**	**Fe**	**Zn**	**TMC**
Units 1	mg/100g*	mg/100g	mg/100g	mg/100g	mg/kg	mg/kg	mg/kg
HHB67 Improved (Pearl Millet)	51.05[3]	369.4[6]	**135.8[10]**	37.05[9]	46.19[9]	32.16[8]	6011.35[8]
KALINGA III (Rice)	**92.31[10]**	347.1[3]	115.1[7]	33.33[5]	21.75[1]	**37.89[10]**	5938.04[7]
CV Viscount (Wheat)	47.63[2]	403.2[8]	90[3]	22.97[1]	24.89[2]	22.12[4]	5685.01[3]
CV Revelation (Wheat)	54.75[4]	386.8[7]	89.9[2]	33.96[6]	39.46[6]	18.02[2]	5711.58[4]
CV Riband (Wheat)	69.04[8]	**619.1[10]**	99[4]	34.41[7]	**50.14[10]**	24.2[6]	**8289.84[10]**
CV Battalion (Wheat)	69.11[9]	367.3[5]	111.8[6]	27.35[2]	38.87[5]	20.42[3]	5814.89[6]
CV Hereward (Wheat)	66.83[7]	347.7[4]	105.3[5]	36.93[8]	38.68[4]	23.21[5]	5629.49[2]
CV Cordiale (Wheat)	47.59[1]	323.7[1]	85.6[1]	27.43[3]	25.52[3]	17.01[1]	4885.73[1]
CV Hereward II (Wheat)	65.36[6]	415.6[9]	120.8[8]	**42.66[10]**	41.07[7]	26.92[7]	6512.19[9]
Naked Barley (BBL)	60.74[5]	345.7[2]	130.1[9]	32.11[4]	45.96[8]	34.29[9]	5766.75[5]

With the exception of Ca, pearl millet line *HHB67 Improved* performed above average (defined as a ranking of 5 or higher) for all other minerals and had the third highest total mineral content of all lines. The data in Table 3.1 suggest that pearl millet is an excellent candidate for good health and nutrition in comparison to common varieties of rice, wheat and barley, For example, in terms of Fe and Zn, *HHB67 Improved* ranks 9/10 and 8/10, respectively. Additionally, the concentration of Mg is the highest in pearl millet, being 37% higher than that of CV Cordiale (wheat). It also exceeds that of Naked Barley. Therefore pearl millet can be seen as a good alternative Mg source to those who are gluten intolerant. It must be noted that samples within this comparison are predominantly wheat (7/10 samples). Any future comparison should include a variety of other crops including maize, sorghum and oat for a more holistic comparison.

3.4.2 Micronutrient Analysis of 229 PMiGAP lines

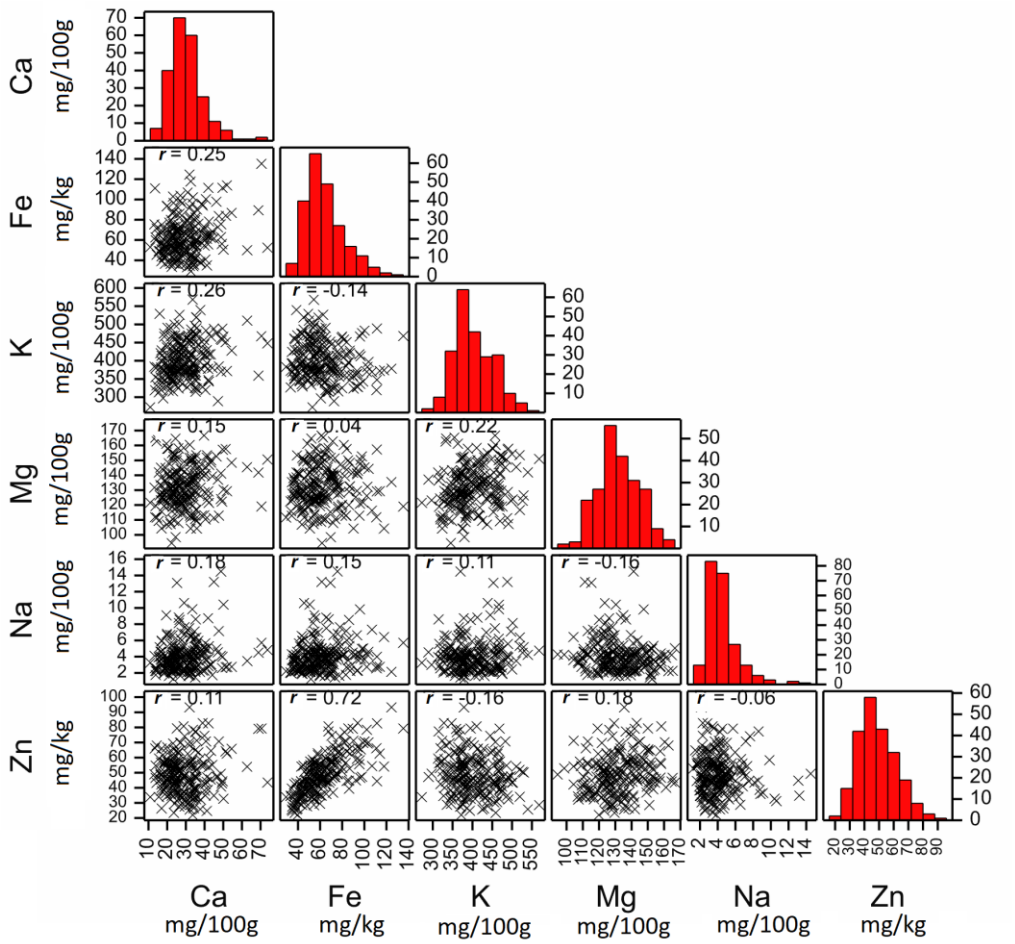


Figure 3.3: Pairwise association and correlation (*r*; Pearson's product moment) between variables from 229 PMiGAP lines grown at ICRISAT, Patencheru, under field conditions. *r* (ranging from -1 and +1) represents the correlation coefficient, which measures the strength and direction of a linear relationship between two minerals. The histograms represent the sample distribution, where each bar represents the number of samples that fall within a set range. The distribution of samples is split into 10 equal bars that represent the range.

The data in Figure 3.3 shows the correlation between different minerals in 229 selfed PMiGAP lines grown under field conditions. The correlation between Fe and Zn was positive and highly significant (R=0.72; P≤0.01). However, there is no significant relationship between any other minerals. These results indicate good prospects for the simultaneous selection for both Fe and Zn, and selection within the PMiGAP, is likely to provide good opportunities for the development of elite lines with elevated grain Fe and Zn levels. Findings also coincide with other reports in the literature, including that of Velu *et al.*, (2007), who investigated the Fe and Zn content within the Iniadi germplasm of pearl millet, also developed at ICRISAT, which is

considered to be a valuable genetic resource for high grain Fe and Zn densities. The positive association between Fe and Zn has also been proven in a variety of other crops and is thought to be due to common molecular mechanisms that control the uptake and metabolism of these minerals or common transporters controlling the movement of these minerals within plants (Vreugdenhil *et al.*, 2004; Ghandilyan *et al.*, 2006).

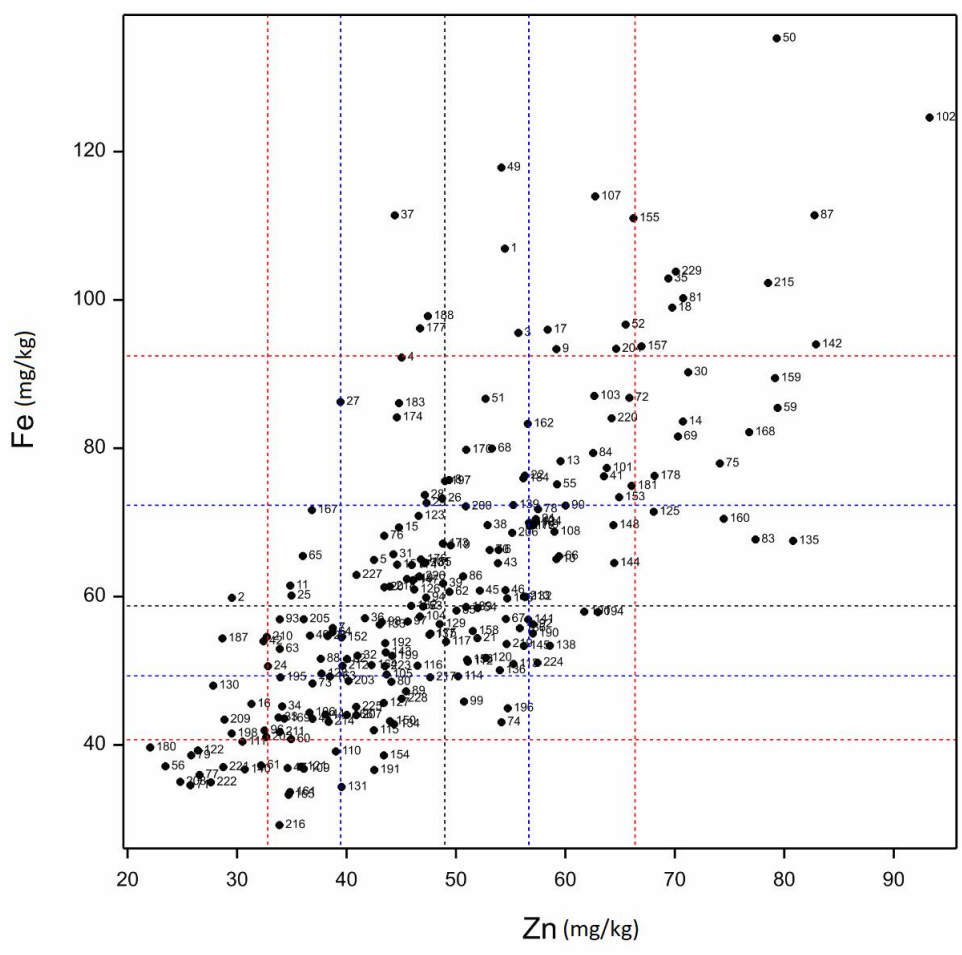


Figure 3.4: Median (grey dotted lines), quartiles (blue dotted lines) and 10th and 90th percentiles (red dotted lines) for grain Zn and Fe content in 229 PMiGAP lines grown at ICRISAT, Patencheru, under field conditions.

The data in Figure 3.4 suggests large within population genetic variability for both Fe and Zn across 229 PMiGAP lines. Fe concentration ranges from 29.18 – 135.27 mg/kg and Zn concentration ranges from 22.07 – 93.28 mg/kg. From the data, it is possible to identify individual high and low Fe and Zn lines. High Fe/Zn lines are identified as those that fall within the top 90th percentile and low Fe/Zn lines are identified as those that fall within the bottom 10th percentile.

Table 3.2: 90th percentiles for grain Zn and Fe content in 229 PMiGAP lines, in relation to Figure 3.4 (OPV = open pollinated variety. DM = downy mildew).

Sample No.	IP No.	Origin	Known Traits	Citation
50	ICTP 8203	ICRISAT, Patencheru	High yielding and high Fe, content OPV, DM resistant	Saltzman *et al.,* 2013
102	IP 8955	Togo	Drought Tolerant	Sehgal *et al.*, 2015
87	IP 15344	India	High yielding, productive tillers	Sehgal *et al.*, 2015
215	IP 19405	Chad	Tall, increased vegetative growth	Khairwal *et al.,* 2007
142	IP 9971	Zambia	-	Icrisat.org
157	IP 7886	India	-	Icrisat.org
18	IP 12925	Ghana	High yielding	Khairwal *et al.,* 2007
81	IP 6460	Mali	-	-
35	GB 8735	ICRISAT, Patencheru	High Fe line	Andrews & Kumar, 1996
229	IP 4952	Uganda	-	Icrisat.org

Table 3.2 shows the lines within the 90th percentile for grain Zn and Fe content in 229 PMiGAP lines. The data in Figure 3.4 suggests that ICTP 8203 (50) had the highest levels of both Fe and Zn. It also contains other useful traits such as high yield and resistance to DM. Development of ICTP 8203 was achieved by randomly mating five S2 progenies selected at ICRISAT and is related to an Iniadi germplasm line, known for high Fe/Zn levels, as previously discussed (Rai *et al.* 1990).

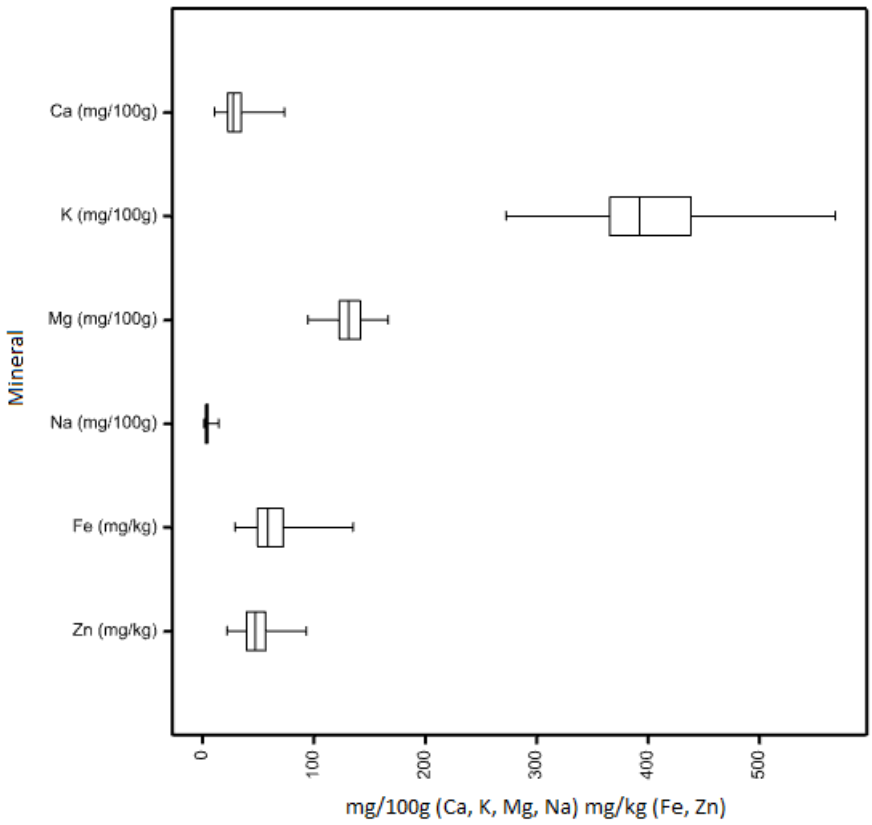


Figure 3.5: Boxplots displaying the distribution of minerals Ca, Fe, K, Mg, Na and Zn data based on the minimum, first quartile, median, third quartile and the maximum in 229 selfed PMiGAP lines grown at ICRISAT, Patencheru, under field conditions.

In addition to measuring Zn and Fe, other elements including Ca, K, Mg and Na were also quantified in PMiGAP lines. Findings suggest that the predominant mineral (mineral at highest levels) was K, which varied between 272.5 mg/100g and 568.1 mg/100g. This was also demonstrated by Oshodi *et al.*, 1999. For all other lines, the Mg content ranged from 98.5 mg/100g to 166.7 mg/100g, Na content ranged from 1.2 mg/100g to 14.5 mg/100g, Ca content ranged from 10.9 mg/100g to 73.7 mg/100g, Fe content ranged from 29.2 mg/kg to 93.3 mg/kg and Zn content ranged from 22.1 mg/kg to 92.3 mg/kg. The seeds were richer in K, Mg and Ca while Na, Fe and Zn were evenly distributed; this was also reported by Oshodi *et al.*, 1999.

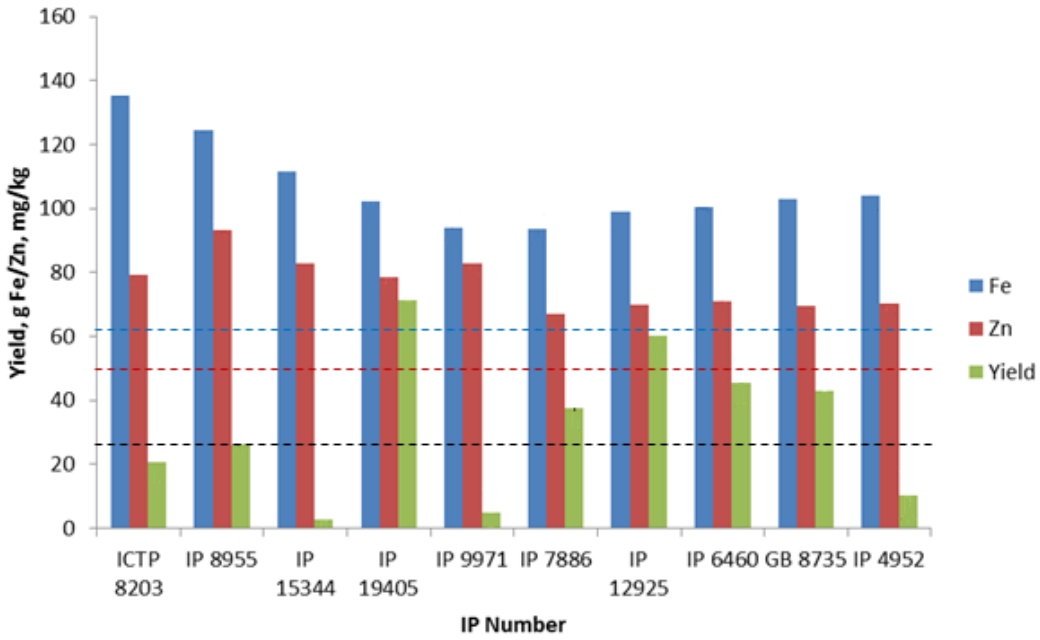


Figure 3.6: A comparison of grain Zn and Fe content to the combined yield of 3 seed multiplication trials. Black dotted line denotes average PMiGAP yield, red dotted line denotes average PMiGAP Zn content and blue dotted line denotes average PMiGAP Fe content.

The data in Figure 3.6 shows a comparison of the 90th percentile Fe/Zn lines with the combined yields of three in-house seed multiplication trials, conducted under glasshouse conditions. Average yield for the entire PMiGAP population was 26.3g per line, therefore 5 lines: IP19405, IP7885, IP12925, IP6460 and GB8735 performed above average. The best line for elevated Fe/Zn content coupled with high yield was IP19405 and may be used as a parent for the development of elite lines. IP15344 has been cited in the literature to be high yielding (Sehgal *et al.*, 2015); however this is not reflected during the three seed multiplication trials. This may be due to GEI factors, where the environmental effects are stronger than the genetics. Although the line ICTP 8203 described previously has the highest readings for Fe and Zn content, this is not reflected in the yield as it performed below average.

3.4.3 A Comparison between 2 Glasshouse Seed Multiplication Trials

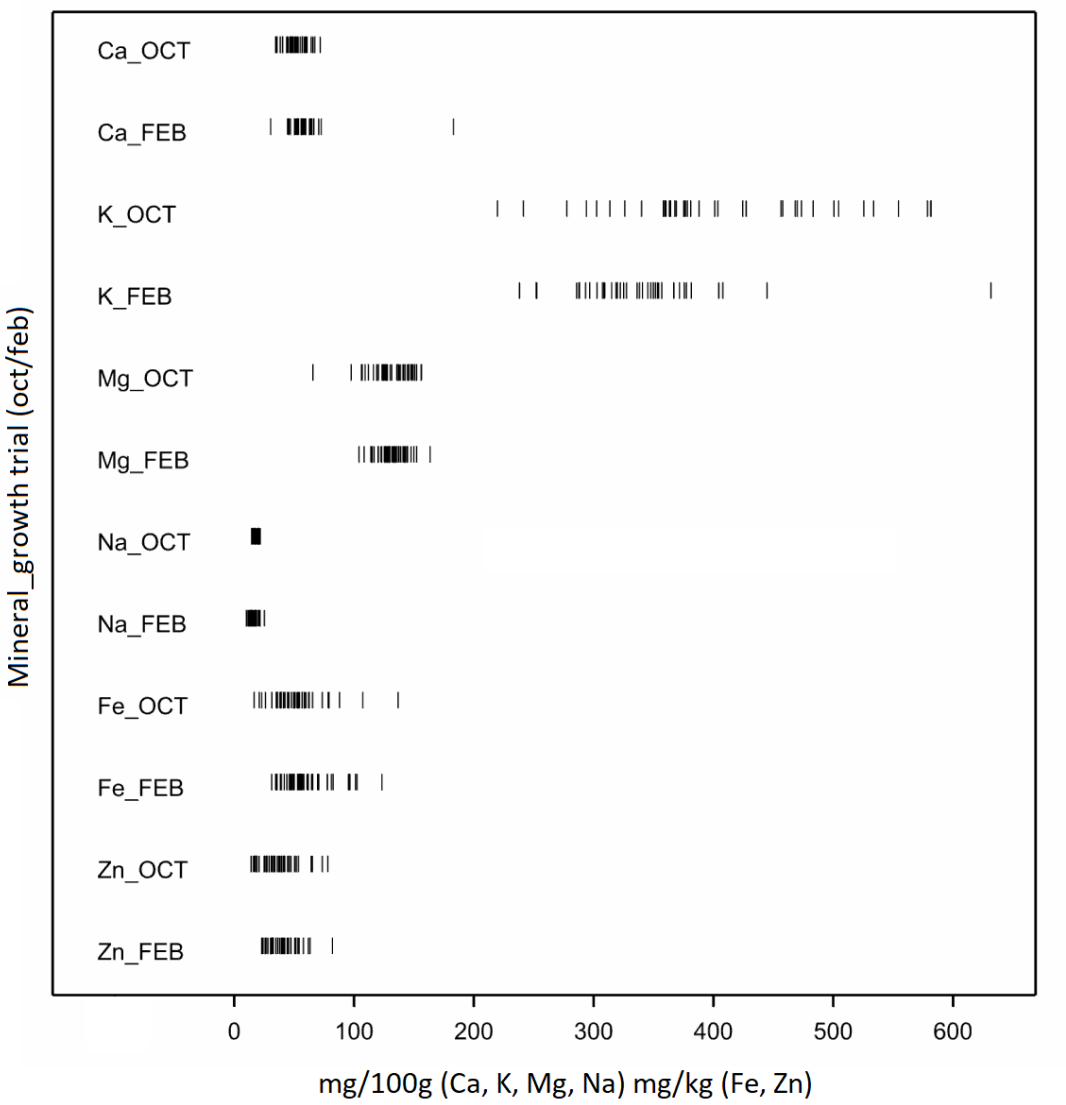


Figure 3.7: Rug plot to show the distribution of minerals Ca, Fe, K, Mg, Na and Zn between 2 glasshouse grown PMiGAP populations (n = 40) conducted in October (OCT) 2014 and February (FEB) 2015 at Aberystwyth University. Ca, K, Mg, Na are measured in mg/100g. Fe and Zn are measured in mg/kg. The rug plot depicts each observation per mineral, per population, respectively. Ca data and K data for February 2015 both included an outlier (2023 S and 2035 S, respectively). These data were excluded from further analyses. All data were approximately normally distributed.

The findings from Figure 3.7 suggest that the predominant mineral in the two glasshouse populations was K; this is consistent with the results from Figure 3.5. Values varied between 220.5 mg/100g and 583.3 mg/100g for October's population and between 237.5 mg/100g and 445.3 mg/100g for February's population. The Mg content ranged from 65.4 mg/100g to 156.7 mg/100g for October and from 104.3 mg/100g to 163.5 mg/100g for February, Na content ranged from 14.41 mg/100g to

21.69 mg/100g for October and from 10.02 mg/100g to 25.2 mg/100g for February, Ca content ranged from 34.64 mg/100g to 71.65 mg/100g for October and from 30.38 mg/100g to 71.65 mg/100g for February, Fe content ranged from 17.04 mg/kg to 137.16 mg/kg for October and from 30.88 mg/kg to 123.11 mg/kg for February and Zn content ranged from 14.36 mg/kg to 78.12 mg/kg for October and from 22.81 mg/kg to 81.98 mg/kg for February.

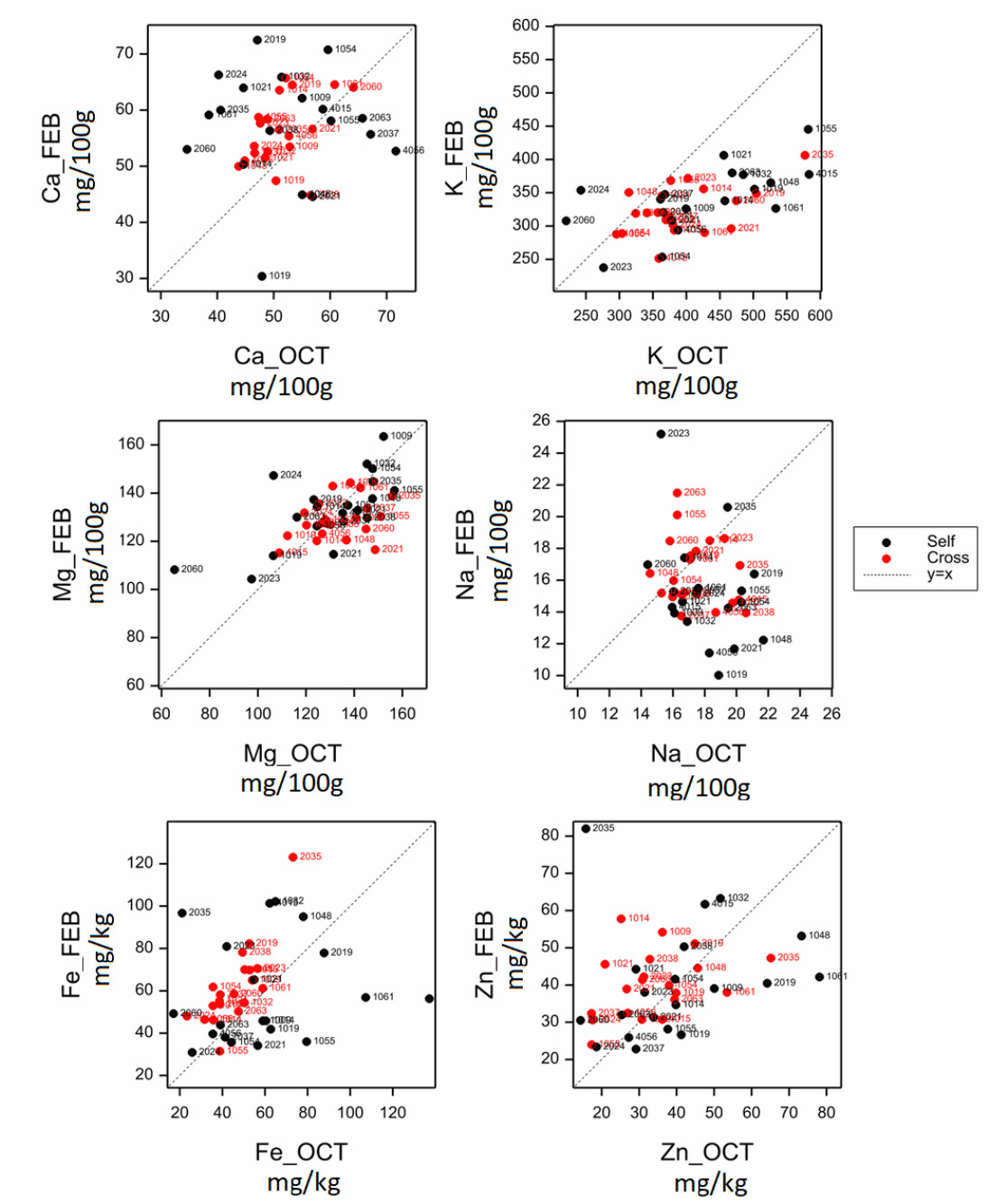


Figure 3.8: Association between mineral contents of 40 PMiGAP lines from February and October glasshouse trials, grown at Aberystywth University. Broken line denotes y = x, red data points = OP lines, black data points = selfed lines.

The data in Figure 3.8 determines the relationship between the mineral content of lines from the October 2014 and February 2015 populations. The two-sample *t*-Test was used to determine whether mean mineral concentrations differed between the populations (Table 3.3).

Table 3.3: Comparison of the mean mineral content of 40 PMiGAP lines from the October 2014 and February 2015 glasshouse trials, conducted at Aberystwyth University.

Mineral	**October**			**February**			*t* prob[1]
	N	Mean	s.e.m.	N	Mean	s.e.m.	
Ca (mg/100 g)	40	51.2	1.31	39	56.6	1.30	**0.005**
K (mg/100 g)	40	408.3	14.46	39	331.7	6.97	**<0.001**
Mg (mg/100 g)	40	131.1	2.86	40	131.1	1.93	0.998
Na (mg/100 g)	40	17.7	0.30	40	15.9	0.45	**0.001**
Fe (mg/kg)	40	52.1	3.42	40	60.3	3.62	0.107
Zn (mg/kg)	40	36.7	2.42	40	40.4	1.94	0.237

s.e.m; Standard error of the mean
[1]; H_0: Mean February = Mean October; H_1: Mean February ≠ Mean October.

The two-sample *t*-Test requires the establishment of H_0 and H_1 (Table 3.3). Using the mean of both populations, the s.e.m. and the number of observations in the two populations, F prob and T prob values were generated. Having tested the t-statistic and compared the t-value with a standard table of t-values, it was possible to determine whether the T prob value exceeded the threshold of statistical significance. From the results it can be seen that there is a difference between the mean values of the two populations for Ca, K and Na, therefore H_0 is rejected and H_1 is accepted, that there are differences between the mean values between the populations. However, for Mg, Fe and Zn, the results are not statistically significant, therefore H_0 is accepted and it is determined that the mean of the minerals between the populations is equal and does not reach the threshold for significance. In terms of phenotypic stability, the data suggests that since the threshold for significance has not been met for Mg, Fe and Zn according to T prob, the phenotype is stable across both environments, but this is not the case for Ca, K and Na.

To assess the stability of the phenotype, Spearman's rank correlation coefficients between the two populations were compared.

Table 3.4: Spearman's rank and Bootstrapped Pearson's correlation values per mineral between October and February populations grown under glasshouse conditions at Aberystwyth University.

Mineral	Spearman's Rank	Prob[1]	Bootstrapped Pearson's Correlation	Prob[1]
Ca (mg/100 g)	0.093	0.144	0.053	>0.05
K (mg/100 g)	**0.564**	**<0.001**	**0.596**	**<0.001**
Mg (mg/100 g)	**0.513**	**<0.001**	**0.565**	**<0.001**
Na (mg/100 g)	-0.284	0.019	-0.299	>0.05
Fe (mg/kg)	0.318	0.011	0.222	>0.05
Zn (mg/kg)	0.388	0.003	0.255	>0.05

[1]; H_0: $r = 0$; H_1: $r \neq 0$.

According to the data in Table 3.4, the Spearman's rank probability values point to weak positive relationships between K and Mg values from the two populations ($P \leq 0.001$). The weak positive relationships suggest some degree of phenotypic stability. However, since correlations are not ≥ 0.70, these lower positive correlations are met with caution. Between seasons, the data overall is not consistent, which suggests that the environment has a stronger effect than the genetics.

Associations between the different minerals were also examined within each population (Figure 3.9 and 3.10).

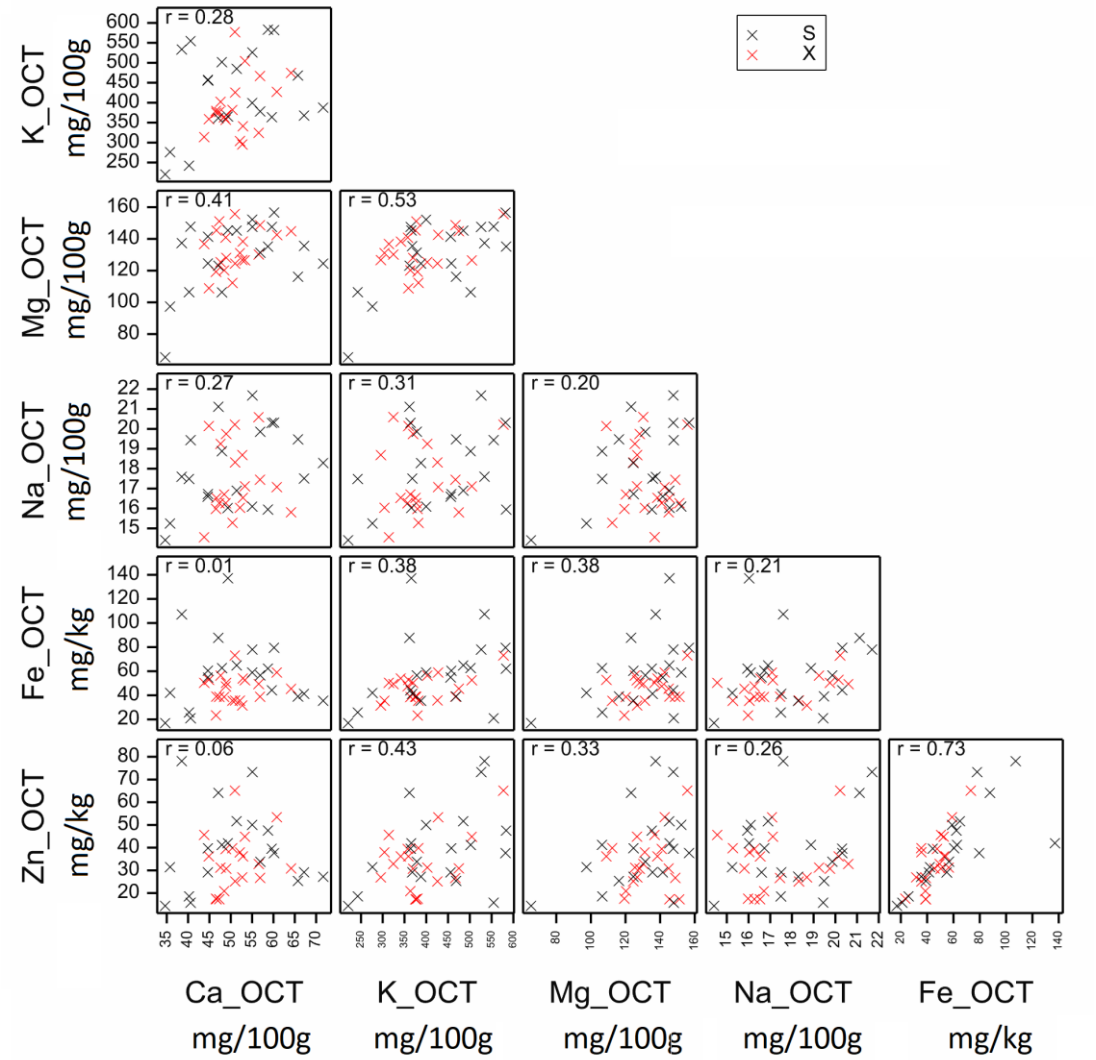


Figure 3.9: Pairwise association and correlation (*r*; Pearson's product moment) between variables from 40 PMiGAP lines from the October 2014 glasshouse trial, Aberystwyth University. *r* (ranging from -1 and +1) respresents the correlation coefficient (S – Selfed samples, X – OP samples).

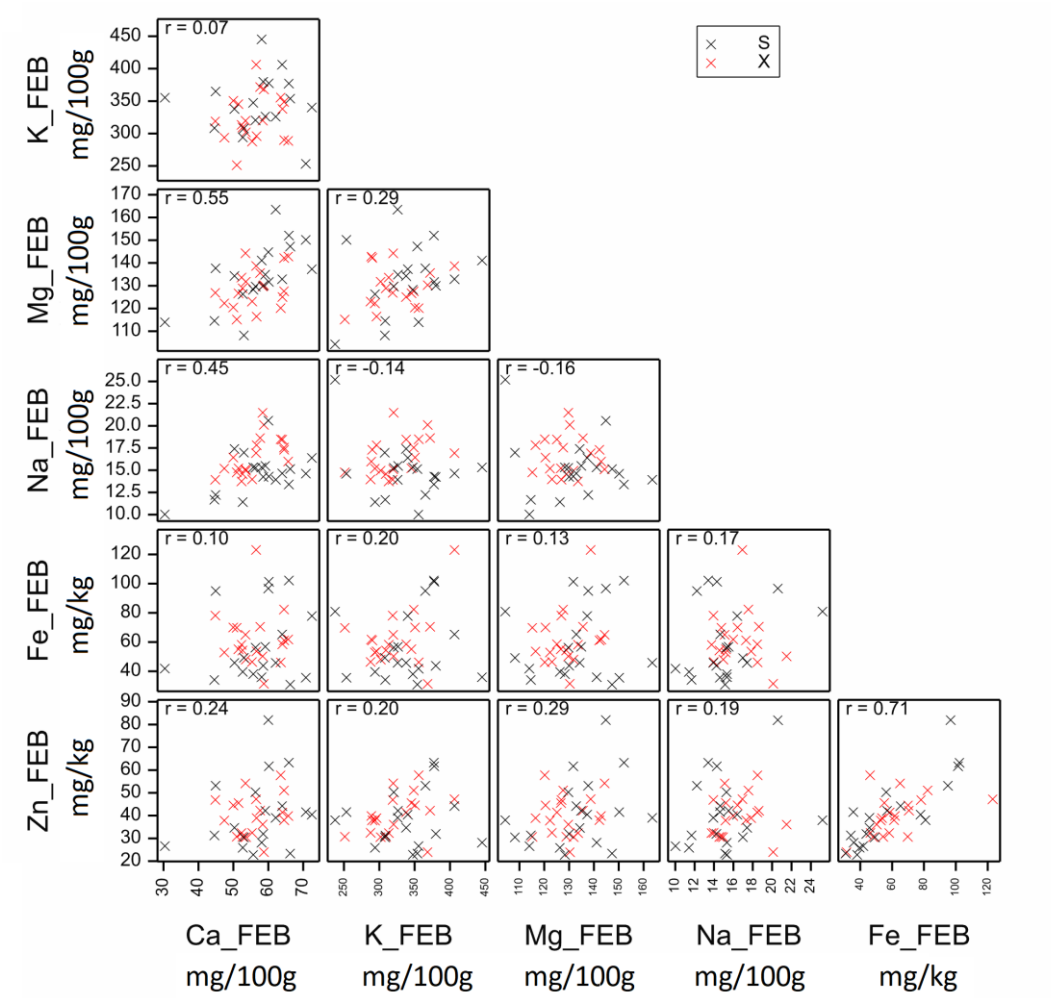


Figure 3.10: Pairwise association and correlation (*r*; Pearson's product moment) between variables from 40 PMiGAP lines from the February 2015 glasshouse trial, Aberystwyth University. *r* (ranging from -1 and +1) respresents the correlation coefficient (S – Selfed samples, X – OP samples).

The results in Figure 3.9 and 3.10 suggest that the only strong association is the relationship between Fe and Zn, this is consistent for both populations and coincides with the data from Figure 3.3. It can therefore be concluded that Fe and Zn are likely to be positively correlated together, even in the case of multi-environments.

3.4.4 Multi-Environmental Trials (Field and Glasshouse)

Pilot Phase

During the pilot phase, a total of 20 selfed PMiGAP entries grown in different environments were analysed by *t*-test and Spearman's Rank statistical tests. A larger sample size would have been favourable but was not possible to obtain due to low seed availability at the time of analysis.

Table 3.5: Mean grain Fe and Zn content of 20 genotypes across 3 sites by one-way ANOVA. (ICRISAT, field trial, Patencheru; Aberystwyth University, glasshouse trial, conducted in October 2014 and Aberystywth University, glasshouse trial conducted in February 2015).

	October '14		**February '15**		**ICRISAT**		Combined	Prob
	Mean	s.e.m.	Mean	s.e.m.	Mean	s.e.m.	s.e.m.	
Fe (mg/kg)	58.8	6.53	58.7	5.51	58.6	4.05	5.46	1.000
Zn (mg/kg)	39.5	3.93	40.6	3.41	46.5	2.57	3.35	0.286

s.e.m,; Standard error of the mean

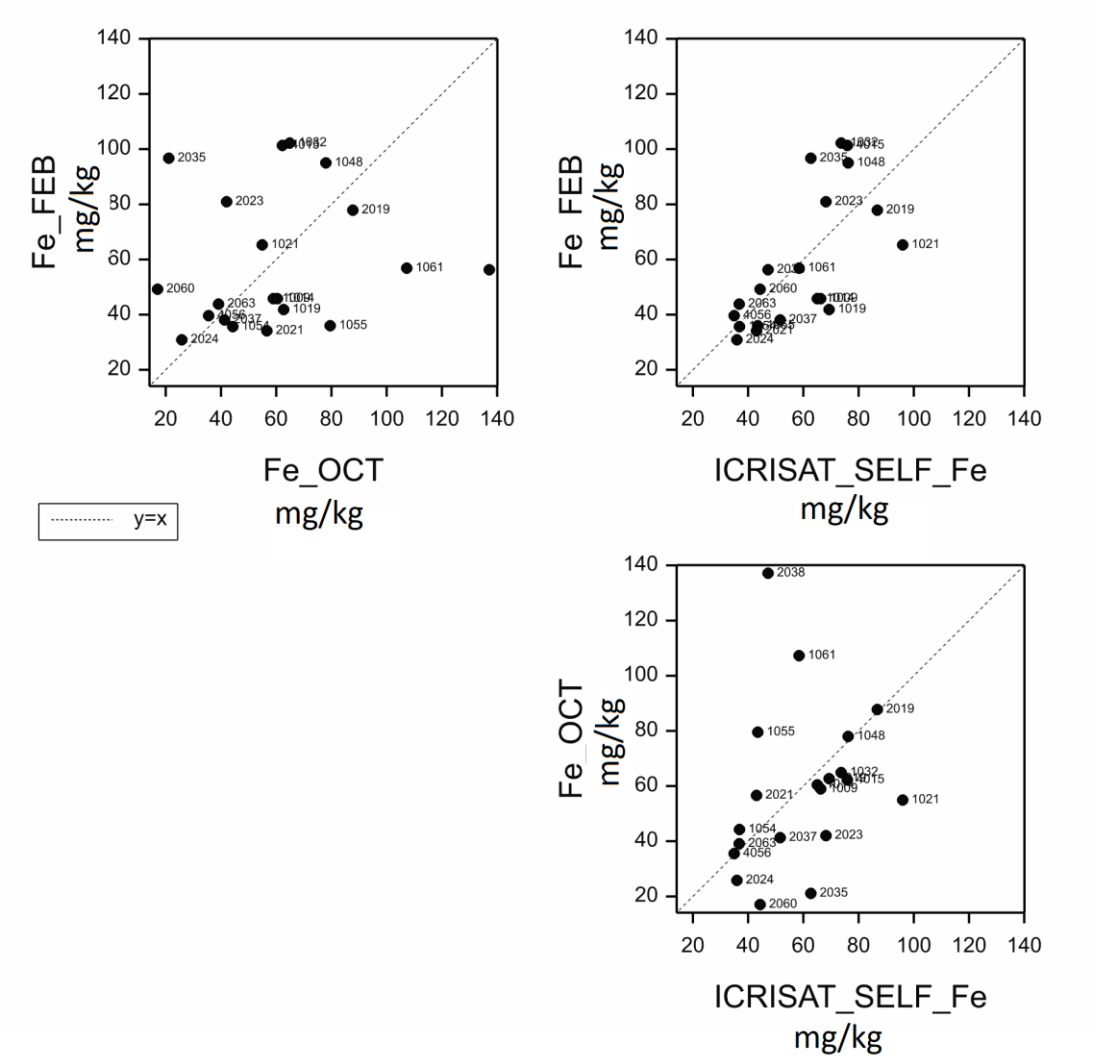


Figure 3.11: Comparing grain Fe levels between three different environments (ICRISAT, field trial, Patencheru; Aberystwyth University, glasshouse trial, conducted in October 2014 and Aberystywth University, glasshouse trial conducted in February 2015).

Table 3.6: Pairwise Spearman's rank correlation coefficients for grain Fe content across 3 environments (ICRISAT, field trial, Patencheru; Aberystwyth University, glasshouse trial, conducted in October 2014 and Aberystywth University, glasshouse trial conducted in February 2015).

Fe_FEB	0.273	**0.751**
	(P=0.061)	(P<0.001)
Fe_OCT	-	0.456
		(P=0.011)
	Fe_OCT	ICRISAT_SELF_Fe

With respect to the data in Table 3.6, although there is a strong association between the February glasshouse and the ICRISAT field populations (R=0.751;P<0.0011), there is no relationship between the February and October glasshouse populations. This is also the case for the October glasshouse and ICRISAT field population. Therefore, the overall data points to the environmental effects overriding the genetic effects of Fe uptake.

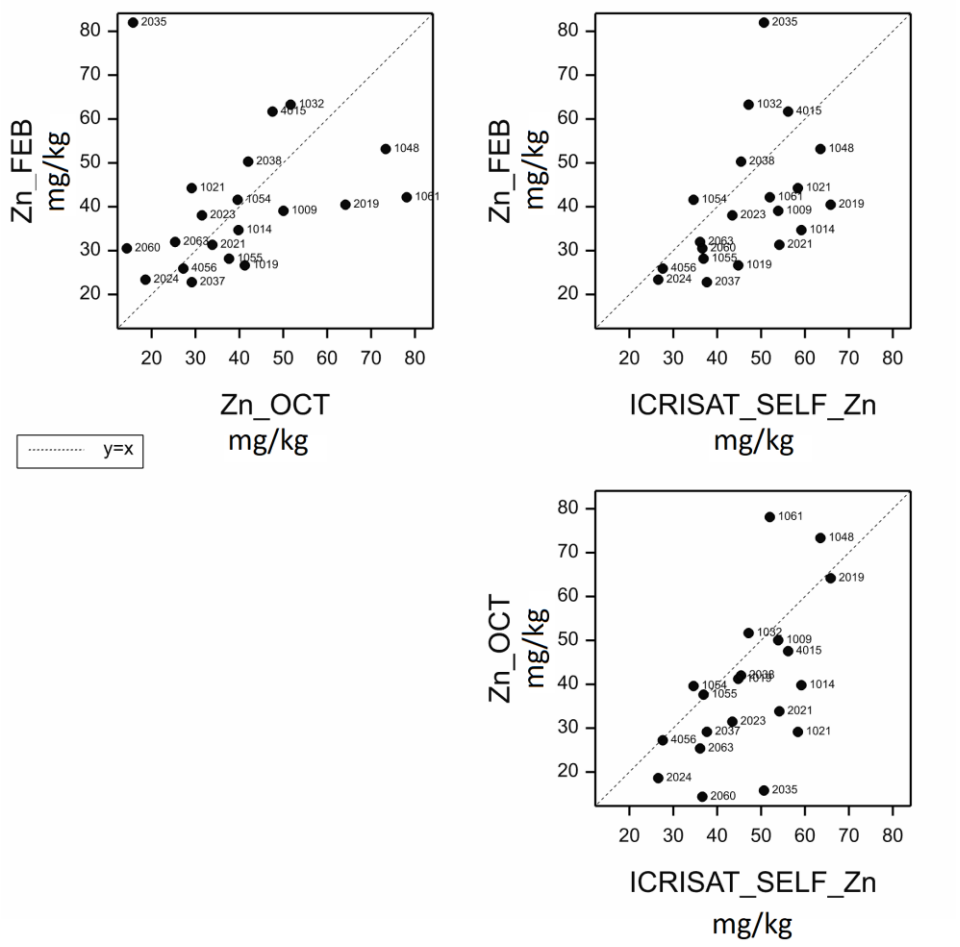


Figure 3.12: Comparing grain Zn levels between three different environments (ICRISAT, field trial, Patencheru; Aberystwyth University, glasshouse trial,

conducted in October 2014 and Aberystywth University, glasshouse trial conducted in February 2015).

Table 3.7, Pairwise Spearman's rank correlation coefficients for grain Zn content across 3 environments (ICRISAT, field trial, Patencheru; Aberystwyth University, glasshouse trial, conducted in October 2014 and Aberystywth University, glasshouse trial conducted in February 2015).

Zn_FEB	0.450	0.564
	(P=0.012)	(P=0.003)
Zn_OCT	-	0.603
		(P=0.001)
	Zn_OCT	ICRISAT_SELF_Zn

According to Table 3.7, with respect to Zn, the association between genotypes grown in different environments does not exceed the Spearman's rank threshold of 0.7, for a positive relationship. Therefore the data points to the environmental effects being stronger the genetic effects of Zn uptake.

Post-Pilot Phase (n=99)

A larger data set, which consisted of 99 selfed PMiGAP lines, became available due to an additional growing season, under glasshouse conditions in August 2015. The GEI effects between the two unrelated environments were determined by comparing mineral levels from the ICRISAT field population to the in-house glasshouse population, conducted in August.

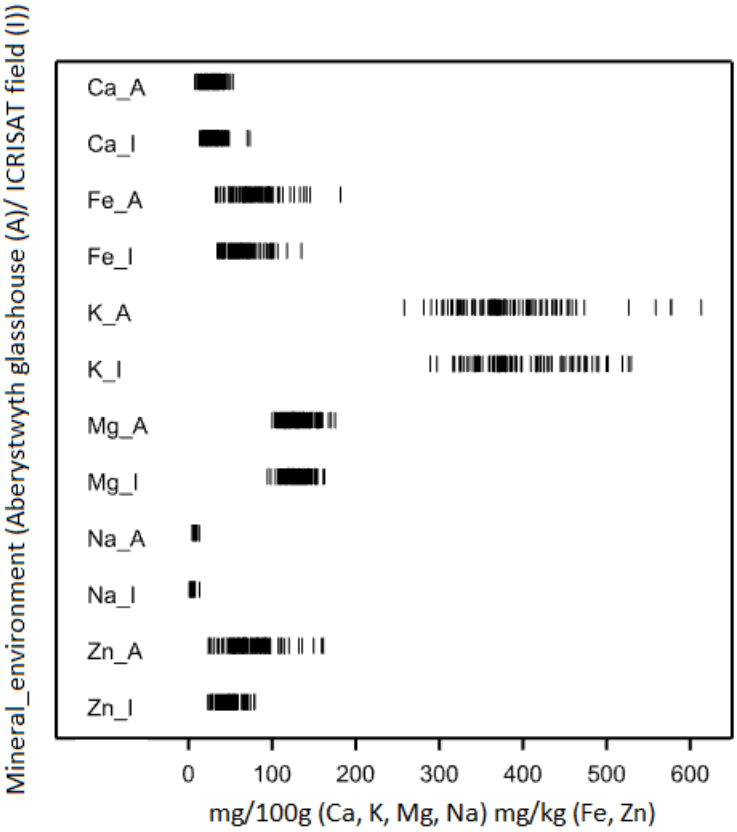


Figure 3.13: Rug plot to show the distribution ofminerals Ca, Fe, K, Mg, Na and Zn between populations grown under glasshouse conditions in Aberystwyth (August 2015) and under field conditions at ICRISAT, patencheru. A = August '15 trial, glasshouse and I = ICRISAT, field trial. Ca, K, Mg, Na are measured in mg/100g. Fe and Zn are measured in mg/kg. The rug plot depicts each observation per mineral, per population, respectively. All data were approximately normally distributed.

Values for K ranged between 289.2 mg/100g and 529.3 mg/100g for ICRISAT field grown population and between 257.92 mg/100g and 613.23 mg/100g for glasshouse population. The Mg content ranged from 94.5 mg/100g to 162.8 mg/100g for ICRISAT field population and from 100.45 mg/100g to 175.43 mg/100g for the glasshouse population, Na content ranged from 1.26 mg/100g to 13.21 mg/100g for ICRISAT field population and from 4.51 mg/100g to 13.11 mg/100g for the glasshouse population, Ca content ranged from 13.2 mg/100g to 73.71 mg/100g for ICRISAT field population and from 7.78 mg/100g to 53.08 mg/100g for the glasshouse population, Fe content ranged from 34.33 mg/kg to 135.27 mg/kg for ICRISAT field population and from 32.55 mg/kg to 182 mg/kg for the glasshouse population and Zn content ranged from 23.46 mg/kg to 79.31 mg/kg for ICRISAT field population and from 24.05 mg/kg to 161.06 mg/kg for the glasshouse population. The Zn levels in lines grown under glasshouse conditions show a far

greater distribution of values, whereas the distribution in the field population is tighter and lower. Therefore, there may be environmental factors affecting the gene expression, perhaps within the soil, thus limiting the Zn. This is also the case for Fe, although to a lesser extent.

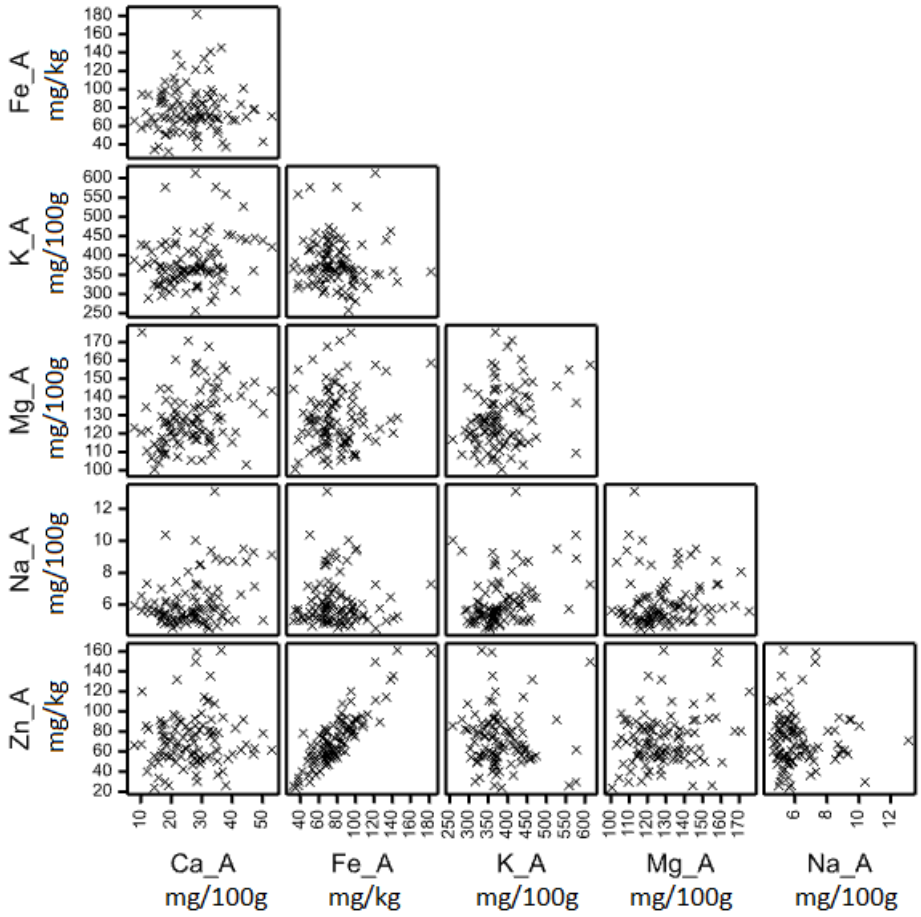


Figure 3.14: Scatter plot matrix to show the relationship between minerals in 99 PMiGAP lines grown under glasshouse conditions at Aberystwyth University in August 2015 , A..

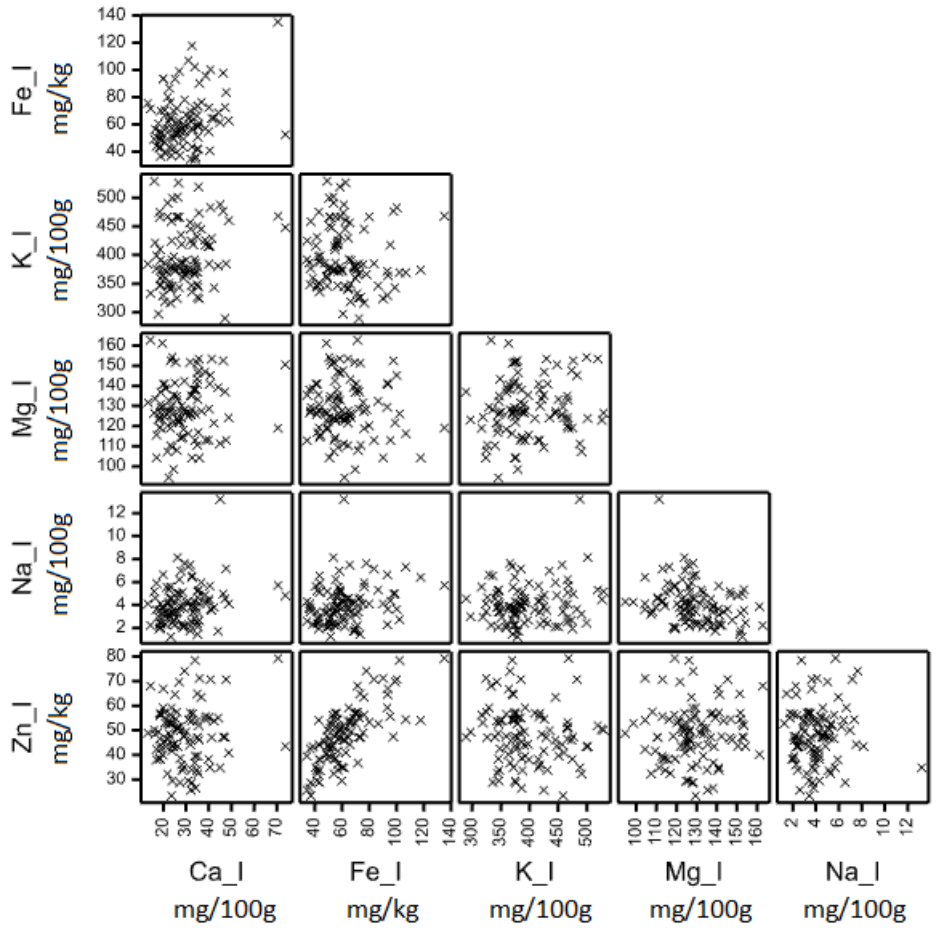


Figure 3.15: Scatter plot matrix to show the relationship between minerals in 99 PMiGAP lines grown under field conditions at ICRISAT, Patencheru in 2010, I.

From the data in Figure 3.14 and 3.15, the strongest relationship is between Fe and Zn for both environments, and is further demonstrated by the Spearman's Rank values in Table 3.8. This indicates a genetic link, regardless of the environment.

Table 3.8: Spearman's Rank correlations between 99 PMiGAP lines grown under glasshouse (G) conditions at Aberystwyth University, in 2015 and field (F) conditions at ICRISAT, Patencheru in 2010.

Mineral	**Aberystwyth G Aug'15**			**ICRISAT F 2010**		
	n	Spearman's Rank	Prob[1]	n	Spearman's Rank	Prob[1]
Fe/Ca	99	0.027	0.790	99	0.249	0.013
K/Ca	99	0.218	0.030	99	0.191	0.058
Mg/Ca	99	0.329	<0.001	99	0.104	0.304
Na/Ca	99	0.214	0.034	99	0.249	0.013
Zn/Ca	99	0.078	0.444	99	0.013	0.897
K/Fe	99	-0.137	0.178	99	-0.186	0.066
Mg/Fe	99	0.109	0.281	99	-0.064	0.531
Na/Fe	99	-0.019	0.855	99	0.253	0.011
Zn/Fe	99	**0.807**	**<0.001**	99	**0.721**	**<0.001**
Mg/K	99	0.252	0.012	99	0.105	0.301
Na/K	99	0.343	<0.001	99	0.085	0.402
Zn/K	99	-0.137	0.176	99	-0.189	0.061
Na/Mg	99	0.175	0.084	99	-0.335	<0.001
Zn/Mg	99	0.146	0.150	99	0.088	0.387
Na/Zn	99	-0.032	0.751	99	0.028	0.782

[1]; H_0: $r = 0$; H_1: $r \neq 0$.

From the data in Table 3.8, it can be seen that the only strong positive relationship that occurs for both environments is that of Fe and Zn (P=<0.001). Therefore, H_0 is rejected for these combinations and it is concluded that there is a relationship between these minerals. The data suggests that due to the occurrence of this relationship in both environments, this phenotype is deemed stable.

The two sample *t*-Test was used to determine whether mean mineral concentrations differed between glasshouse and field populations (Table 3.9).

Table 3.9: Comparison of the mean grain mineral content of 99 PMiGAP lines grown under glasshouse (G) conditions at Aberystwyth University, in 2015 and field (F) conditions at ICRISAT, Patencheru in 2010.

Mineral	**Aberystwyth (G) Aug'15**			**ICRISAT (F) 2010**			*t* prob[1]
	N	Mean	s.e.m.	N	Mean	s.e.m.	
Ca (mg/100 g)	99	26.8	0.95	99	29.4	1.04	0.062
Fe (mg/kg)	99	78.6	2.58	99	62.7	1.92	**<0.001**
K (mg/100 g)	99	382.9	6.34	99	397.5	5.44	0.081
Mg (mg/100 g)	99	128.6	1.61	99	129.0	1.421	0.885
Na (mg/kg)	99	6.2	0.15	99	4.1	0.18	**<0.001**
Zn (mg/kg)	99	72.3	2.71	99	48.5	1.18	**<0.001**

s.e.m; Standard error of the mean

[1]; H_0: Mean February = Mean October; H_1: Mean February ≠ Mean October.

From the results it can be seen that there are differences between the mean mineral content values of the two populations for Fe, Na and Zn, therefore H_0 is rejected and H_1 is accepted, that the mean mineral content from the glasshouse population is not equal to mean of the field population. This is indicative of phenotypic instability between the two environments. However, for Ca, K and Mg, the results are not statistically significant and H_0 is accepted, that the mean mineral content from the field population is equal to mean of the glasshouse population, this is indicative of some degree of phenotypic stability for these three elements.

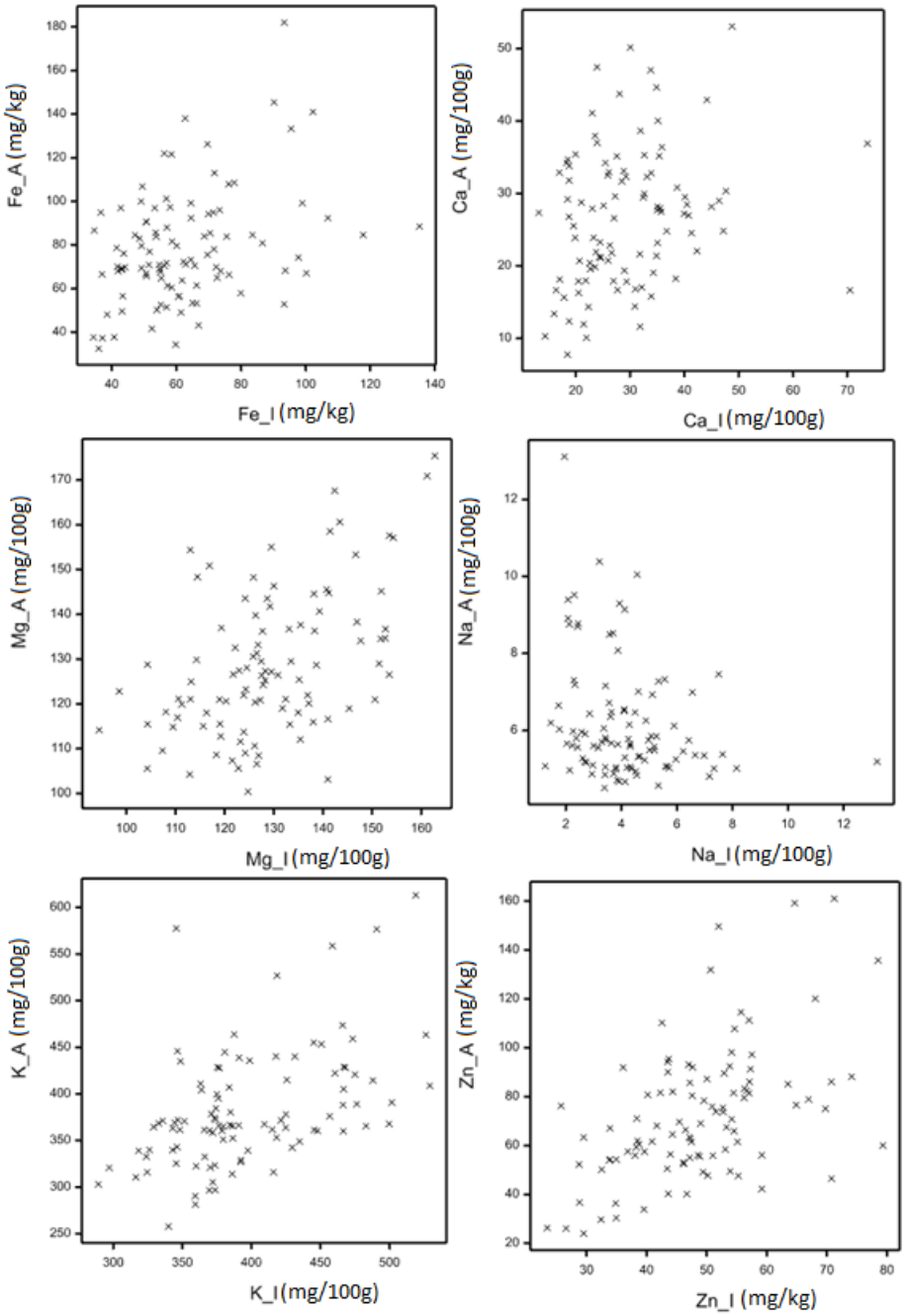


Figure 3.16: Pairwise association between grain mineral content from 99 PMiGAP lines grown under glasshouse conditions in 2015, at Aberystywth University (A) and field conditions in 2010, at ICRISAT, Patencheru (I).

Table 3.10: Spearman's Rank correlations between grain mineral content from 99 PMiGAP lines grown under glasshouse at Aberystwyth University in 2015 and field conditions at ICRISAT, Patencheru in 2010.

Mineral	Spearman's Rank	Prob[1]
Ca (mg/100 g)	0.261	0.009
Fe (mg/kg)	0.291	0.003
K (mg/100 g)	**0.496**	**<0.001**
Mg (mg/100 g)	**0.468**	**<0.001**
Na (mg/kg)	-0.262	0.009
Zn (mg/kg)	**0.483**	**<0.001**

[1]; H_0: r = 0; H_1: r ≠ 0.

From the results in Table 3.10, it can be seen that there is a weak positive relationship (above R=0.4) for K, Mg and Zn between the two populations (P<0.001). However, the extent of this is less for Ca and Fe and a negative correlation of -0.262 can be seen for Na. H_0 is rejected due to P=<0.001 for K, Mg and Zn and accepted for Ca, Fe and Na. Although the positive association is statistically significant according to the data, it must be noted that since correlations are not ≥0.70, lower positive correlations should be met with caution.

3.4.5 Selfed vs. OP PMiGAP Lines

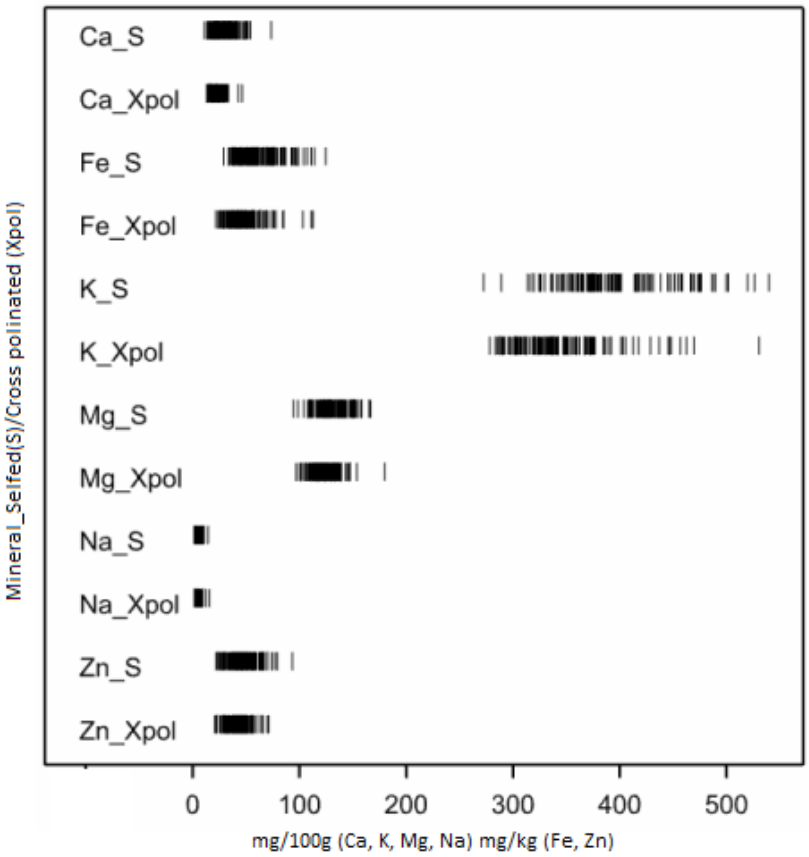


Figure 3.17: Rug plot to show the distribution of minerals Ca, Fe, K, Mg, Na and Zn between 131 selfed and OP PMiGAP entries grown under field conditions at ICRISAT, Patencheru. (S = Selfed samples, Xpol = open pollinated varieties).. The rug plot depicts each observation per mineral.. All data were approximately normally distributed.

When selfed and OP seed from the same environment were compared, the predominant mineral for both populations was K (Figure 3.17). Values for K varied between 272.5 mg/100g and 539.7 mg/100g for selfed samples and 278.1 mg/100g and 530.6 mg/100g for OP samples. Values for Ca content varied between 10.94 mg/100g and 73.71 mg/100g for selfed entries and 13.23 mg/100g and 46.36 mg/100g for OP entries, Fe content varied between 29.18 mg/kg and 124.6 mg/kg for selfed entries and 21.9 mg/kg and 112.8 mg/kg for OP entries. Na content varied from 1.46 mg/100g and 14.5 mg/100g for selfed entries and 1.5 mg/100g and 15.51 mg/100g for OP entries and Zn content varied from 22.07 mg/kg and 71.11 mg/kg for OP entries and 22.07 mg/kg and 93.28 mg/kg. Mg content varied from 96.9

mg/100g and 179.6 mg/100g for OP entries and 94.5 mg/100g and 166.7 mg/100g for selfed entries.

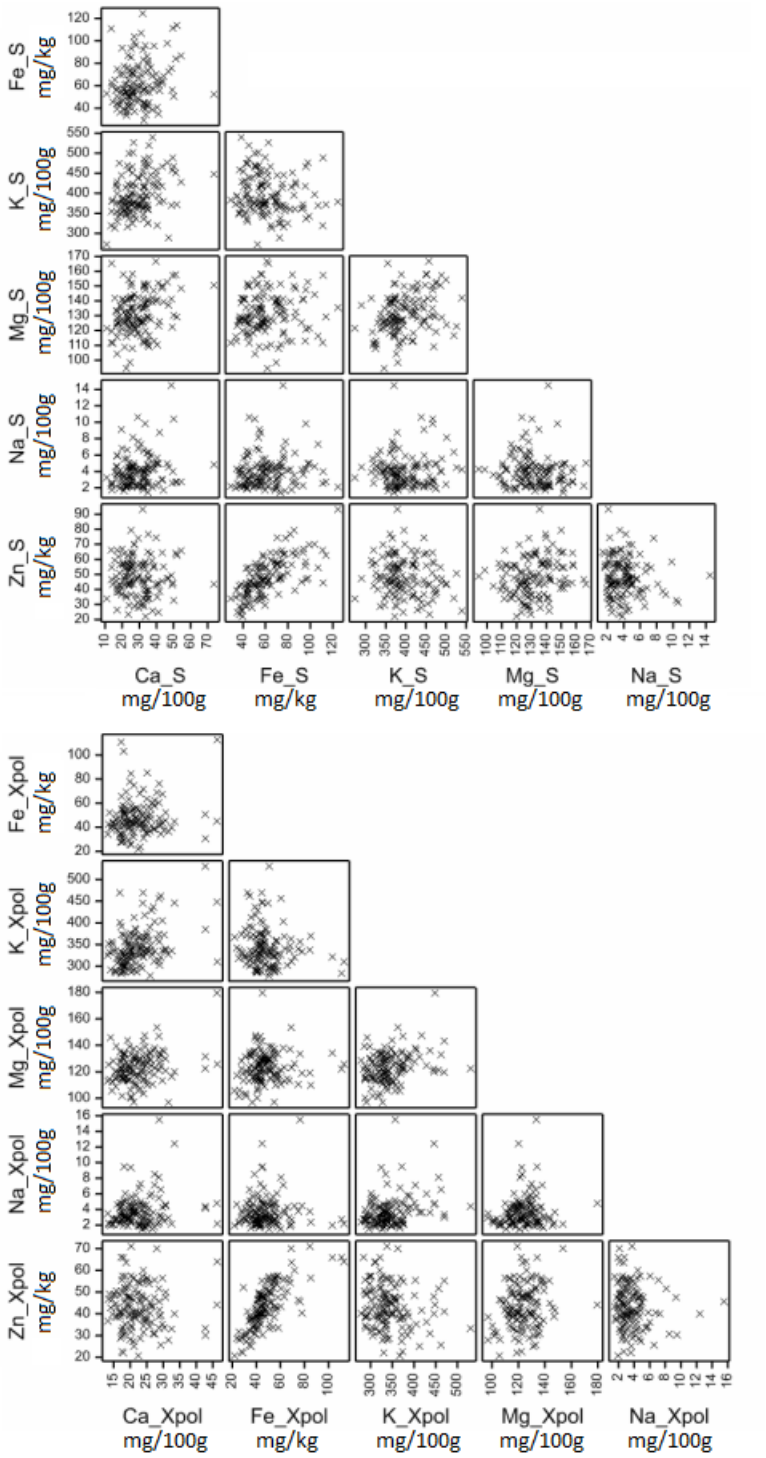


Figure 3.18: Scatter plot matrix to show the relationship between minerals in 131 selfed (top) and OP (bottom) PMiGAP lines grown under field conditions at ICRISAT, Patencheru.

Table 3.11: Spearman's Rank Correlations between grain mineral content from 131 selfed/OP PMiGAP lines grown under field conditions, at ICRISAT, Patencheru.

Mineral	Selfed			OPV's		
	n	Spearman's Rank	Prob[1]	N	Spearman's Rank	Prob[1]
Fe/Ca	131	0.203	0.020	131	0.107	0.226
K/Ca	131	0.343	<0.001	131	0.367	<0.001
Mg/Ca	131	0.253	0.004	131	0.206	0.018
Na/Ca	131	0.158	0.072	131	0.142	0.106
Zn/Ca	131	-0.059	0.505	131	-0.064	0.468
K/Fe	131	-0.183	0.036	131	-0.113	0.199
Mg/Fe	131	0.126	0.153	131	0.106	0.229
Na/Fe	131	0.127	0.149	131	0.021	0.812
Zn/Fe	131	**0.717**	**<0.001**	131	**0.737**	**<0.001**
Mg/K	131	0.330	<0.001	131	0.257	0.003
Na/K	131	-0.043	0.627	131	0.285	0.001
Zn/K	131	-0.147	0.095	131	-0.185	0.034
Na/Mg	131	-0.124	0.159	131	0.107	0.223
Zn/Mg	131	0.218	0.012	131	0.174	0.047
Na/Zn	131	-0.077	0.381	131	-0.131	0.136

[1]; H_0: $r = 0$; H_1: $r \neq 0$.
OPV's = Open Pollinated Varieties

From the scatter plot matrices in Figure 3.18, the strongest observable relationship exists only between Fe and Zn for both selfed and OP populations. This is further demonstrated by the Spearman's Rank values in Table 3.11 ($P<0.001$).

Table 3.12: Comparison of the mean mineral content between 131 selfed and OP PMiGAP lines grown under field conditions at ICRISAT, Patencheru.

Mineral	Selfed			OPV's			*t* prob[1]
	N	Mean	s.e.m.	N	Mean	s.e.m.	
Ca (mg/100 g)	131	29.50	0.8782	131	22.46	0.5235	<**0.001**
Fe (mg/kg)	131	62.16	1.696	131	48.13	1.323	<**0.001**
K (mg/100 g)	131	397.3	4.462	131	344.8	3.957	<**0.001**
Mg (mg/100 g)	131	131.5	1.203	131	123.7	1.030	<**0.001**
Na (mg/kg)	131	4.064	0.1774	131	3.759	0.1759	0.223
Zn (mg/kg)	131	48.40	1.118	131	43.43	0.888	<**0.001**

s.e.m; Standard error of the mean OPV's – Open Pollinated Varieties
[1]; H_0: Mean February = Mean October; H_1: Mean February ≠ Mean October.

A 2 sample *t*-test was used to determine whether mean mineral contents varied between selfed and OP populations. The general trend in this data is that the mean mineral content is higher in selfed populations as compared to OP populations. From the results, it can be seen that there are differences between mean mineral content for Fe, K, Mg, Zn and Ca, therefore H_0 is rejected and H_1 is accepted, that the mean of these minerals from the selfed population is not equal to the mean of the OP population. This is indicative of phenotypic instability, perhaps by GEI effects (as they were not grown at the same time) or by differences in OP/selfed populations owing to factors such as seed set or dilution/concentration effects. However, for Na, the results do not reach the threshold for statistical significance, therefore H_0 is accepted, that the mean mineral content from the selfed population is equal to the mean of the OP population.

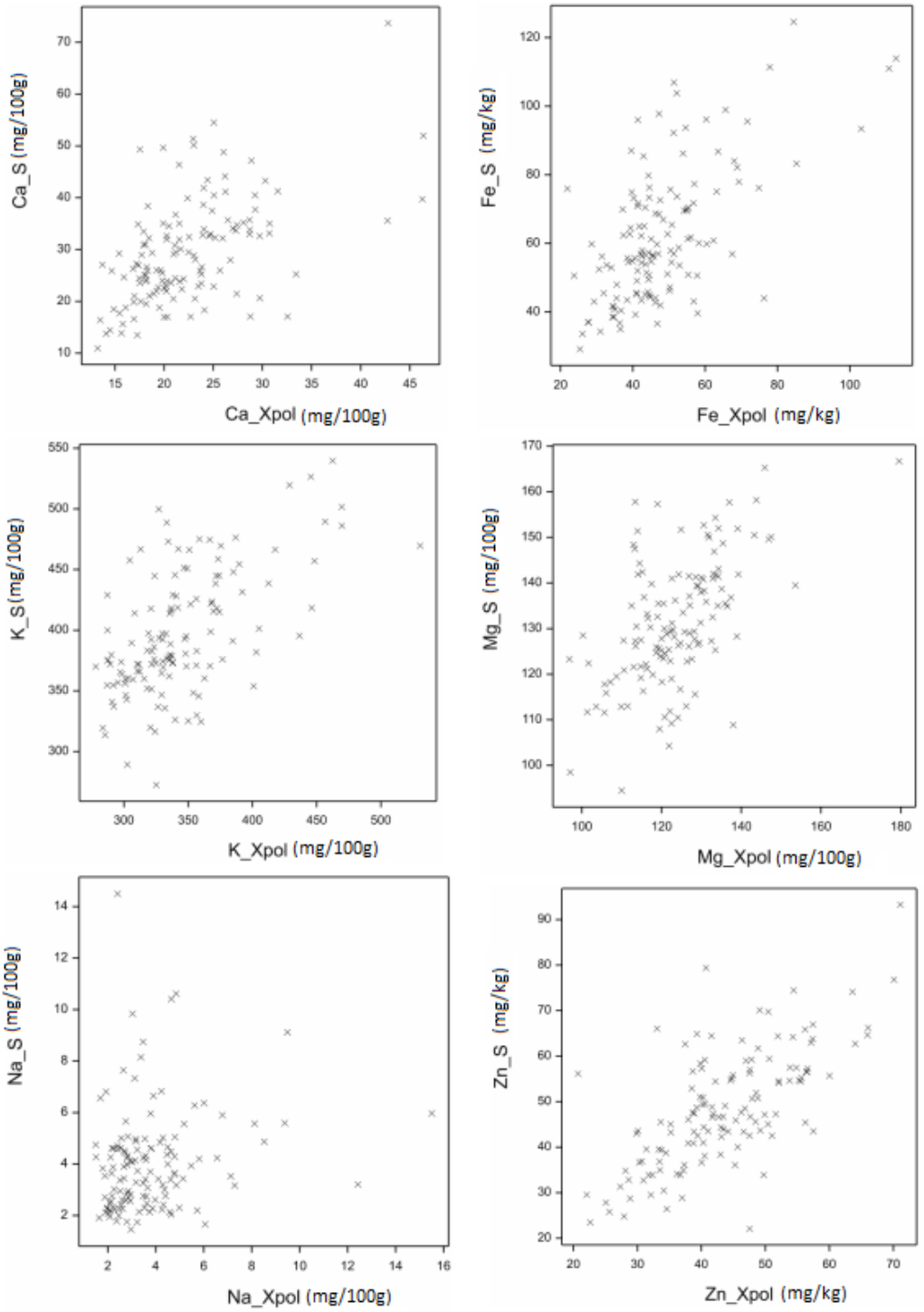


Figure 3.19: Pairwise association between grain mineral content from 131 selfed (S) and open pollinated (Xpol) PMiGAP lines grown under field conditions at ICRISAT, Patencheru.

Table 3.13: Spearman's Rank correlations for grain mineral content from 131 selfed and OP PMiGAP lines grown under field conditions at ICRISAT, Patencheru.

Mineral	Spearman's Rank	Prob[1]
Ca (mg/100 g)	**0.522**	**<0.001**
Fe (mg/kg)	**0.542**	**<0.001**
K (mg/100 g)	**0.548**	**<0.001**
Mg (mg/100 g)	**0.520**	**<0.001**
Na (mg/kg)	0.221	0.011
Zn (mg/kg)	**0.660**	**<0.001**

[1]; H_0: r = 0; H_1: r ≠ 0.

From the results in Table 3.13, it can be seen that there are positive correlations above R=0.5 for Ca, Fe, K, Mg and Zn values between the selfed and OP populations (P= <0.001), the strongest of which is for Zn , which is indicative of some degree of phenotypic stability between the two populations. However, for a much weaker correlation of R=0.221 can be seen for Na, therefore H_0 is rejected for Ca, Fe, K, Mg and Zn and accepted for Na.

3.5 Conclusions

GEI effects are determined based on various statistical analyses that examine the correlation of minerals between lines grown in multi-environments (Spearman's rank and Pearson's product moment) in addition to a comparison of mean mineral levels between environments (*t*-tests). From the results, the most abundant element in pearl millet is K in all environments, also demonstrated by Oshodi *et al.*, (1999). Although the *t*-test results in Table 3.3 indicate that with respect to the October and February glasshouse populations, the means significantly differ for Ca K and Na, but they do not for Mg Fe and Zn. These results differ to those in Table 3.9, which compares the mineral concentrations of field and glasshouse populations. These results indicate that the means significantly differ for Fe, Na and Zn, but do not for Ca K and Mg. Although this was not the case for Mg, the results show that there are significant GEI effects at work for all other elements between growth trials meaning that the environmental effect is often stronger than the genetic effect, overall. Plant breeders need to take this into account when selecting elite lines and any results from subsequent genetic analysis must be approached with caution. When a smaller PMiGAP data set was used for the comparison of Fe and Zn concentrations in two glasshouse populations and one field population (pilot phase, Table 3.5), it was

found that the means did not differ significantly, according to a 1 way ANOVA statistical analysis. However, this could be due to the small sample size (n=20), which is not a true representation of the entire PMiGAP.

According to the Spearman's Rank values, a strong positive association is observed between Fe and Zn in all populations ($R \geq 0.7$; $P \leq 0.001$), this coincides with many accounts in the literature and could be owing to common molecular mecahnisms that control the uptake/metabolism of these minerals or common transporters that control the movement of these minerals within plants (Vreugdenhil *et al.,* 2004; Ghandilyan *et al.,* 2006). The Spearman's rank test was also used to determine the relationship between minerals measured in multi-environments and it was found that a positive relationship exists between K and Mg for the October and February glasshouse trials ($P<0.001$) and between the field and glasshouse trials ($P<0.001$), this suggests some degree of phenotypic stability when pearl millet is grown in multi-environments. Regretably, this was not the case for other elements, as the threshold for significance was not met.

It has been well documented that a reduction in seed-set, as a result of selfing is a genotype-dependent and common trait that exists within pearl millet (Rai *et al.,* 2015). This can lead to an over-estimation of mineral density, indicating that data from selfed populations must be approached with caution. Inbreeding depression is also associated with selfed populations, due to increased rates of homozygosity (Kaur & Soodan, 2017). Contrastingly, pearl millet OP lines are documented to have >90% seed set and produce higher average seed weights, more stable yields and fewer seedling losses during plant establishment (Hanna *et al.,* 1986, Gray & Steckel, 1986, Kaur & Soodan, 2017). Because of this, OP seed production is considered to be more cost-effective and data from these populations can be used for reliable estimation of mineral density (Rai *et al.,* 2015). A study by Kumar *et al.,* (2016), compared mineral levels from selfed and OP populations of pearl millet. It was found that the average selfed grain mineral content was significantly higher than that of OP populations, which coincides with the findings from this study (Table 3.12). It was also reported that the average selfed grain Fe content was higher than the selfed grain Zn content, which also coincides with this study (Table 3.12) and another study by Voorrips *et al.,* (2002). Furthermore, Kumar *et al.'s* (2016) study also suggests that the effect of seasonal variation is greater on selfed seed. For

example, in the summer season, the mineral contents in selfed seeds were found to be significantly higher than that of the rainy season. This indicates that environmental factors are stronger than the genetic effects on selfed seed set, mineral distribution and uptake. Findings from other studies reveal contrasting results that the rainy season is more favourable for mineral content in seeds (Velu *et al.*, 2007). In spite of growing season, it has been widely documented in a variety of different species that alterations in the environment or plant physiology can have vast effects on the accumulation of multiple elements, simultaneously (Vreugdenhil *et al.*, 2004). For example, growing the same genotype of *Silene vulgaris* (Bladder Campion) on different soil types is proven to result in variation of micronutrients by a factor of 2.4, 3.6, 4.4 and 4.2 for Fe, Cu, Mn and Zn, respectively (Ernst *et al.*, 2000).

For selfed seed, it is also important to consider the possibility that an increase in grain mineral content may be due to "concentration effects" (Gomez-Becerra *et al.*, 2010) as a result of small seed, reduced seed set or low yield capacity. For example, in wild wheat, researchers have explored the possibility of increased mineral content as a result of higher concentrations of minerals in small seeds and/or yield capacity because the existence of this inverse relationship between grain yield and mineral concentration has been well documented in the past. However, it was found that increased grain concentration of Fe and Zn in wild wheat is under genetic control and not entirely attributable to differences in grain yield/size (Gomez-Becerra *et al.*, 2010).

As previously discussed, pearl millet is a protogynous species, in which stigmas are usually fertilised by wind bourne pollen from other sources prior to pollen shed from flowers from the same panicle, thus open pollination is favoured. Due to a reduction in seed size, seed set and yield in selfed panicles as compared to OPV's, the differences in mineral content as seen in Table 3.12 may be attributable to concentration effects due to Fe and Zn being predominantly present in the outer seed layers and a reduction in bulk endosperm fraction (Garvin *et al.*, 2006). Contrastingly, it would be expected that increased weight due to increased endosperm and yield in OPVs results in a dilution effect.. This suggests the importance of acknowledging different individual grain weights in PMiGAP entries and that if individual grain weight differs among varieties, this should be accounted

for in results, possibly by measuring mineral contents per 1000 seeds instead of a set weight per gram only.

Another contributing factor to differences in mineral content between selfed and OP seed could also be due to the pollen source, i.e. the xenia effect. The term xenia describes the effect of foreign pollen on the development and characteristics of the seed (Pahlavani & Abolhasani, 2006). This is because the genotype of the seed embryo is decided after pollination, therefore, OP seeds could not only be more numerous per panicle as compared to selfed seeds, but grain mineral content could also be affected (Rai *et al.,* 2015). Upon further investigation into this theory, Rai *et al.'s* (2015) research suggests that OPV's of pearl millet may be the best option for cost effective and reliable estimation of Fe and Zn density and are not affected by the pollen source (xenia effect). Furthermore, results from this study demonstrate no xenia effect on seed Fe and Zn density in OP lines. Other studies report contrasting results on the xenia effect on seed size in pearl millet (Burton, 1952) and the only other study on the xenia effect for mineral content is in maize by Pletsch-Rivera & Kaeppler, (2007), which looked the xenia effect on P content in cross pollinated maize. Results indicated that no xenia effect was observed (Pletsch-Rivera & Kaeppler, 2007).

Findings from the experiments conducted in this chapter will contribute to the decision of which population from which seed multiplication trial (field/glasshouse) to use as trait data in the downstream genetic analysis, i.e. the GWAS. Whilst there were many interesting conclusions drawn, quite a few limitations came to light. For example, the lack of replicates owing to limited seed stocks and whether the mineral concentrations can be trusted owing to the GEI or mineral concentration effects, as a result of selfing. In order to investigate the issue of mineral concentration effects further, it is recommended that instead of submitting 1g of seed for mineral analysis, 1000-seed-weight should be submitted so that the mineral concentrations are relative to the size of the seed/the seed-set.

According to Viana *et al.,* (2016), GWAS is considered ineffective for smaller population sizes, consisting of less than 200 individuals. Therefore the population used for the GWAS conducted in this study must include as many PMiGAP lines as possible, exceeding 200 lines. The population with the largest number of lines

phenotyped was the ICRISAT field population, which consisted of 229 selfed genotypes. Therefore, this population is deemed the most suitable for GWAS. This population was also grown in field conditions, more representative of the weather conditions likely to occur where this crop can be grown naturally. It is also recommended by Viana *et al.,* (2016) that high trait heritability is achieved at a rate of 70-80% by phenotyping replicated populations. This was not possible to achieve during this study, due to limited seed stocks, therefore no broad/narrow sense heritability studies could take place. Instead, *t*-tests were used to determine phenotypic stability across the populations grown in multi-environments. The results from these tests suggest that the resulting data needs to be treated with caution due to the environmental effect being stronger than the genetics in the majority of findings.

Chapter 4: The Effect of Phytate on Iron and Zinc Bioavailability

4.1 Summary

Phytate assays were performed to quantify levels of phytate in 42 PMiGAP lines, from two glasshouse populations. Findings revealed that levels of phytate differed significantly between the two populations (P=<0.05). Additionally, Pearson's product moment and Spearman's rank correlation coefficients showed no significant correlation between the two populations. This may be attributable to the effects of cultivars, environment and their interactions with phytate being highly significant. Furthermore, there was no correlation between Fe/Zn content and phytate, which is beneficial in terms of plant breeding because this indicates the possibility of breeding elite pearl millet lines with low phytate content and high grain Fe/Zn levels.

As discussed previously, phytate has a negative effect on the bioavailability of Fe and Zn due to its metal chelating properties. It is possible to predict the bioavailability of Fe and Zn using the molar ratios of phytate: minerals. The literature documents that if the phytate: Fe ratio exceeds a critical value of 1 and if the phytate: Zn ratio exceeds a critical value of 15; this may result in poor Fe/Zn bioavailability (Al Hasan *et al.,* 2016). Using phytate data from 235 PMiGAP lines grown at ICRISAT and their comparison to grain Fe/Zn levels, it was found that 100% of the population exceeded the critical value of 1 for Fe and 69% exceeded the critical value of 15 for Zn. This suggests that the vast majority of the PMiGAP for breeding is currently limited in terms of bioavailability of Fe and Zn.

4.2 Introduction

Phytates represent a complex class of naturally occurring phosphorus storage compounds that can significantly influence the functional and nutritional properties of foods by chelating with metal cations, including Fe^{2+} and Zn^{2+}. This creates an insoluble complex (Urbano *et al.,* 2000). Additionally, phytate can be found complexed to proteins, in free form and during internal digestion and can also bind with micronutrients in other foods. The hindering effect of phytate on mineral bioavailability has been previously discussed in Chapter 1, where it is demonstrate that, absorption of Fe, Zn, and Ca are significantly lower from diets with high levels of phytate than from diets that contain low levels of phytate (Egli *et al.,* 2004).

Although indigestible to humans, phytate plays several key roles in the seed development of cereals and legumes and reaches its highest level at seed maturity. As previously discussed, due to its role in storage and as a source of myoinositol, it can never be completely eliminated from the crop (Reddy *et al.*, 1989). In spite of this, conscious efforts should be made to significantly reduce phytate content so that it does not become a limiting factor in micronutrient absorption. The phytate concentration in morphological components of cereal grains is presented in Table 4.1. In pearl millet, the majority of the phytate appears to be present in germ and bran fractions. Other studies indicate it is also present in the aleurone layer (Kulp, 2000).

Table 4.1: Phytate concentration and distribution in morphological components of cereals (Reddy *et al.*, 1989)

Cereal	**Morphological component**	**Phytate phosphorus (%)**	**Phytate (%)[a]**	**Distribution (%)[b]**
Corn	Commercial hybrid	0.25	0.89	-
	Endosperm	0.01	0.04	3.20
	Germ	1.80	6.39	88.00
	Hull	0.02	0.07	0.04
Corn	High lysine	0.27	0.96	-
	Endosperm	0.01	0.04	3.00
	Germ	1.61	5.72	88.90
	Hull	0.07	0.25	1.50
Wheat	Soft	0.32	1.14	-
	Endosperm	0.001	0.004	2.20
	Germ	1.10	3.91	12.90
	Aleurone layer	1.16	4.12	87.10
Rice	Brown	0.25	0.89	-
	Endosperm	0.004	0.01	1.20
	Germ	0.98	3.48	7.60
	Pericarp	0.95	3.37	80.00
Pearl Millet	Whole	0.25	0.89	-
	Endosperm	0.09	0.32	-
	Germ	0.75	2.66	-
	Bran	0.28	0.99	-

[a] Phytate content is calculated by assuming that it contains 28.20% phosphorus.
[b] Percentage of phytate in the component part.

Decreasing levels of phytate is advantageous to the consumer, due to its influence on nutritional end use quality (Samia *et al.*, 2005). Soaking in water, dehulling, germination and fermentation are traditional methods of reducing levels of phytate in pearl millet. Among these methods, fermentation is proven to further decrease levels of antinutrients in grains, whilst increasing mineral extractability (Badau *et al.*,

2005). In a study by Eltayeb *et al.,* (2016), two common pearl millet cultivars: *Gazira* and *Gadarif*, were obtained from Khartoum North local market, Sudan. It was found that soaking of grains in water reduced phytate content by 39% for the *Gazira* cultivar and 24% for the *Gadarif* cultivar. Germination proved to be the most effective method in reducing phytate ($P<0.05$) as compared to all other processing methods and resulted in a reduction in phytate content by 67% for the *Gazira* cultivar and 65% for the *Gadarif* cultivar. These findings also coincided with research conducted by Badau *et al.,* (2005). Fermentation for 12 hours of all treated grain was also found to further significantly decrease phytate content in both cultivars, owing to the action of the enzyme phytase, which is released by microorganisms during the fermentation process (Eltayeb *et al.,* 2016).

Fe and Zn biofortification is applicable to a wide variety of cereals including wheat, rice, and pearl millet (Prentice *et al.,* 2017). However human absorption of non-heme Fe and Zn is subject to strong interference from phytate. The high affinity of phytate to bind to Fe and Zn should limit the use of Fe by microorganisms in the gut and intestines, however research suggests that phytases derived from various sources (plants or other dietary components), synthesised by gut bacteria, or produced by the intestinal mucosal cells can degrade phytate in the gastrointestinal tract and therefore free the chelated Fe and Zn to some extent (Sandberg & Andlid, 2002). However, the extent of phytase activity in the human small intestine is naturally extremely limited. For example, in a study by Iqbal *et al.,* (1994), small intestinal phytase activity was measured *in-vitro* in mucosal homogenates from two human small intestinal samples obtained from transplant donor patients. Rat intestine was also studied as a comparison. Phytase activity was found in human small intestine samples at low values, up to 30 × less than that of the rat tissue (Iqbal *et al.,* 1994). In order to enhance the phytase activity in the human body, studies have tested the efficiency of phytase supplements, which are added to food directly before consumption. For example, in a study by Brnić *et al.,* (2016) fractional absorption of Zn was assessed in 35 children using the double-isotopic tracer ratio approach with ^{67}Zn as oral tracer and ^{70}Zn as intravenous tracer, in test meals which consisted of millet based porridge containing 1.4mg total Zn with the addition of 20.5 phytase units. The results indicated that mean fractional absorption of Zn increased from 9.5±3.4 to 16.0±5.1% ($P<0.0001$), when phytase was added to the test meal. This suggests that adding

phytase prior to the consumption of a Zn-fortified cereal-based complementary food can significantly improve Zn absorption in young children (Brnić *et al.*, 2016).

4.2.1 Experimental Aims

The staple diet of most of the global population is provided by cereal crops, which present levels of Fe and Zn below the nutritional demand of human beings. Factors that may interfere with the bioavailability of Fe and Zn include antinutritional compounds present in the edible parts of plants, such as phytate. Thus, it is necessary to evaluate the variation of Fe/Zn/phytate levels in PMiGAP lines for the identification of cultivars with potential for biofortification. For the development of low-phytate lines, it is also necessary to determine the consistency of phytate levels between multi-environmental growth trials, which may be affected by GEI factors, since different species of plants as well as cultivars of the same species, have differing abilities to absorb, translocate and accumulate nutrients and antinutrients and often exhibit a wide range of genetic variability for these traits (White & Broadley, 2009). 42 selfed PMiGAP lines were selected from two glasshouse populations and phytate levels were compared to determine if certain lines were characterised by high/low phytate and to determine if GEI factors were at work. The effect of phytate on Fe/Zn and bioavailability depends on the ratio of phytate: minerals in the diet. Mineral and phytate levels were quantified in 235 OP PMiGAP lines and the relative bioavailability of minerals was predicted from the molar ratio of phytate: minerals (Ma *et al.*, 2005).

4.3 Materials and Methods

4.3.1 Plant Material

See Chapter 2, Table 2.3.

4.3.2 Phytate Assays

Phytate content was determined using the Phytic Acid (Total Phosphorus) Assay Kit (K-PHYT), Megazyme International Ireland Limited, Bray, Ireland.

Sample extraction

Using an electric powered grinding mill (Retsch Mortar Grinder Mill, 110V/60Hz OY-04181-10) 0.5g of whole pearl millet seed from each genotype was ground to obtain a fine powder, which was able to pass through a 1mm sieve. This was transferred into a 75mL glass beaker and 20mL 0.66M hydrochloric acid was added. The sample was left to stir overnight. 1mL of the extract was centrifuged at 13,000 rpm for 10 mins and 0.5mL of the supernatant was transferred to a fresh 1.5mL microfuge tube. The sample was then neutralised by the addition of 0.5mL sodium hydroxide solution (0.75M).

Enzymatic Dephosphorylation Reaction

Table 4.2: Enzymatic dephosphorylation reaction protocol.

Amount pipetted into 1.5mL microfuge tube	Free Phosphorus	Total Phorphorus
Distilled Water	0.62mL	0.60mL
Buffer plus Sodium azide, pH 5.5	0.20mL	0.20mL
Sample extract	0.05mL	0.05mL
Phytase suspension	-	0.02mL
Samples mixed by vortex and incubated in a water bath set to 40ºC for 10min.		
After 10min next reaction started by addition of:		
Distilled Water	0.02mL	-
Buffer + $MgCl_2$, $ZnSo_4$ and sodium azide	0.20mL	0.20mL
Alkaline phosphatase suspension	-	0.02mL
Samples mixed by vortex and incubated in a water bath set at 40ºC for 15min.		
After 15 min the reaction was stopped by addition of:		
Trichloroacetic acid (50% w/v)	0.30mL	0.30mL
Terminated reaction centrifuged at 13,000 rpm for 10min.		

The supernatant was extracted and used for the colourimetric determination of phosphorus.

Colourimetric Determination of Phosphorus

For every batch of samples that was applied to the colourimetric determination of phosphorus, a phosphorus calibration curve was performed concurrently using the same batch of colour reagent (ammonium molybdate , 5% w/v added to ascorbic acid, 10% w/v / 1M sulphuric acid). Samples were read against air, without a cuvette in the light path.

Table 4.3: Ratio of sample/standard : colour reagent.

Pipetted into a 1.5mL microfuge tube	Sample
Sample/ Phosphorus standard	1.00mL
Colour reagent	0.50mL
Mixed by vortex and incubated in a water bath set to 40°C for 1 hour. After 1 hour, samples were mixed by vortex, then 1mL was transferred to a semi-micro cuvette and the absorbance was read at 665nm (A_{665}), within 3 hours.	

Preparation of the Phosphorus Calibration Curve

Preparation of standard phosphorus solutions are described in Table 4.4 and treated as samples for the colourmetric determination of phosphorus.

Table 4.4: Preparation of standard phosphorus solutions

Pipetted into a 13mL polypropylene tube	STD 0 (0µg)	STD 1 (0.5µg)	STD 2 (2.5µg)	STD 3 (5 µg)	STD 4 (7.5 µg)
Distilled Water	5.00mL	4.95mL	4.75mL	4.50mL	4.25mL
Phosphorus standard solution	-	0.05mL	0.25mL	0.50mL	0.75mL
Total volume	5.00mL	5.00mL	5.00mL	5.00mL	5.00mL

Calculation

Results were entered into a downloadable spreadsheet (https://secure.megazyme.com/phytic-acid-total-phosphorus-assay-kit) to determine phytate content.

4.3.3 Statistical Analysis

Data were analysed using the software GenStat (16th edition, VSN International Ltd, Hemel Hempstead, UK) using the 2 sample *t*-test, Spearman's rank correlation, and Pearson's product moment. Phytate to mineral millimolar ratios were used to estimate the inhibitory effects of phytate on the bioavailability of minerals.

4.4 Results and Discussion

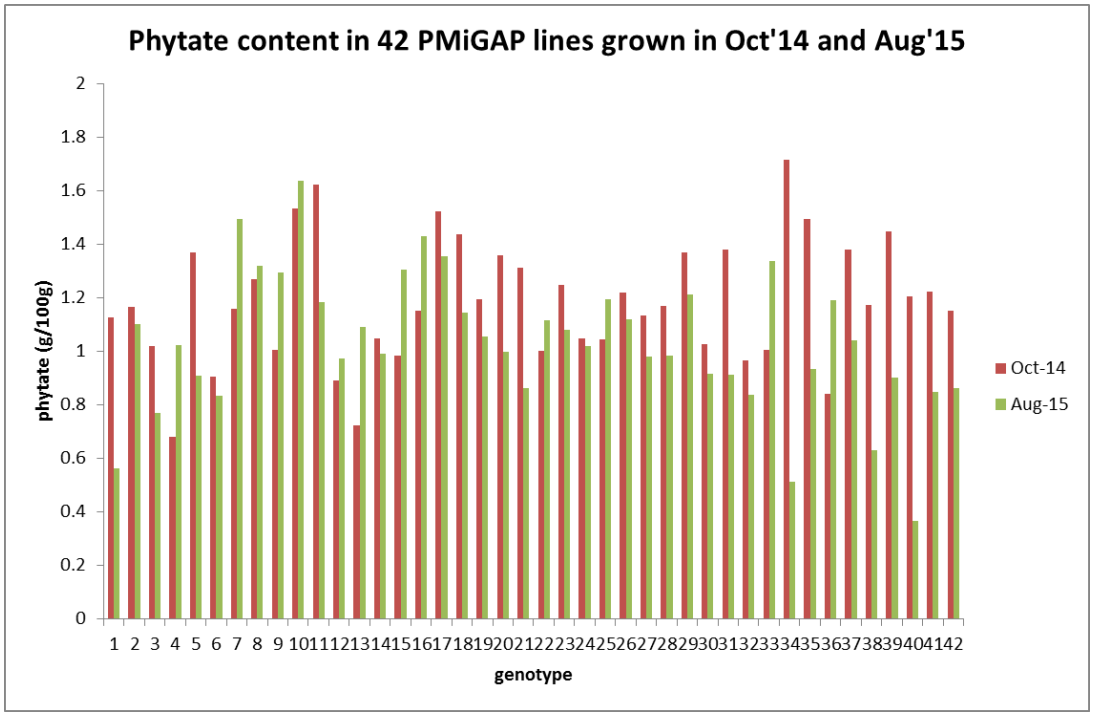


Figure 4.1: Average phytate content in grains of 42 PMiGAP lines grown under glasshouse conditions (Aberystwyth University, October 2014 and August 2015).

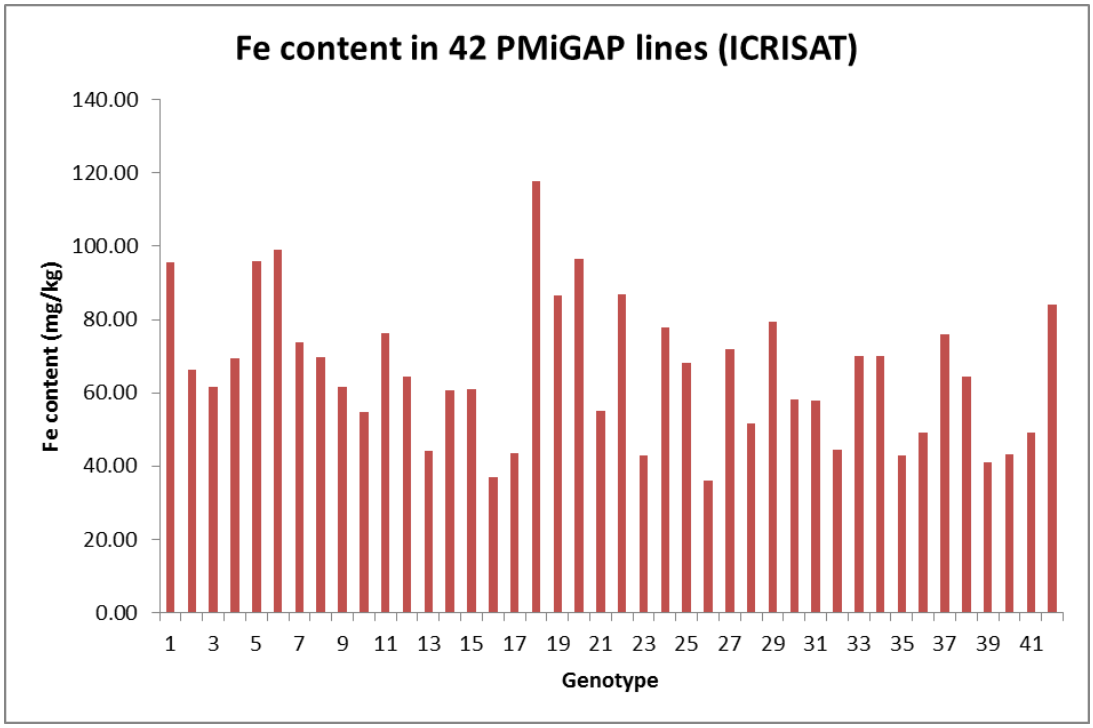


Figure 4.2: Average Fe content in grains of 42 PMiGAP lines grown under field conditions (ICRISAT, Patencheru, 2010).

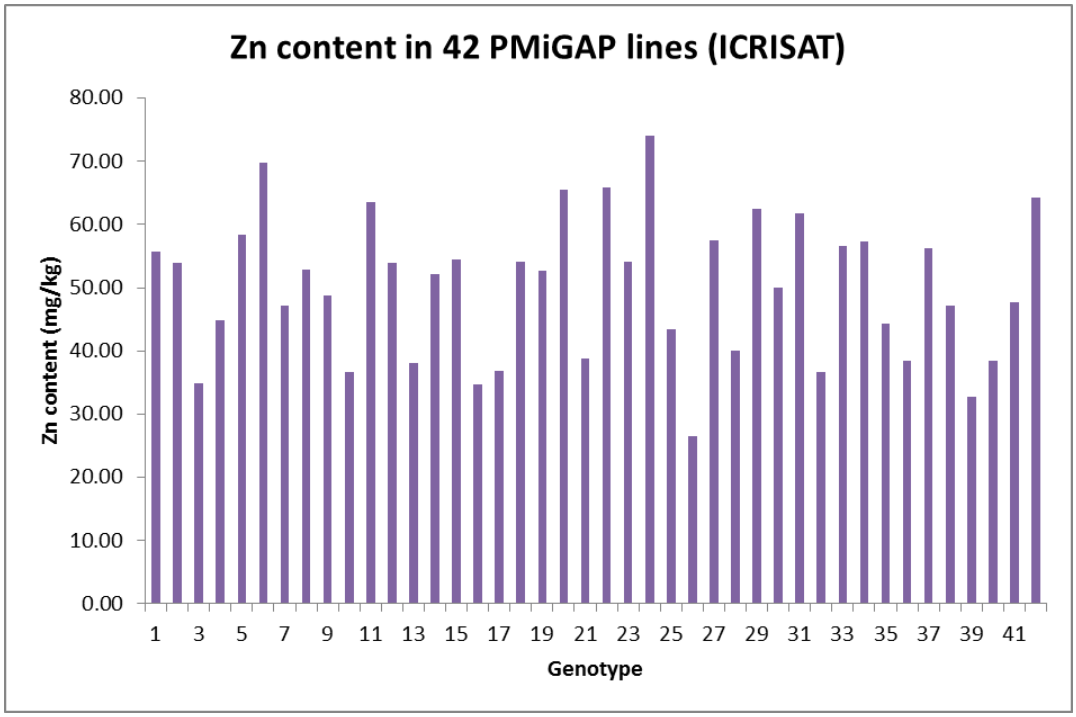


Figure 4.3: Average Zn content in grains of 42 PMiGAP lines grown under field conditions (ICRISAT, Patencheru, 2010)..

Levels of phytate, Zn and Fe varied considerably between genotypes, according to Figures 4.1, 4.2 and 4.3.

The average content of phytate found in grains of different PMiGAP cultivars grown in October 2014 was 1.18 g/100g, with values varying by approximately 60%, from 0.68 to 1.72 g/100g, while the average content of phytate for PMiGAP entries grown in August 2015 was 1.03 g/100g, ranging from 0.37 to 1.64 g/100g, that is, a difference of about 77% among cultivars (Figure 4.1). The average content of Fe found in grains of different PMiGAP cultivars was 65.61 mg/kg, with values varying by approximately 70%, from 35.94 to 117.84 mg/kg, while the average content of Zn was 50.08 mg/kg, ranging from 26.57 to 74.12 mg/kg, that is, a difference of about 64% among cultivars (Figures 4.2 and 4.3).

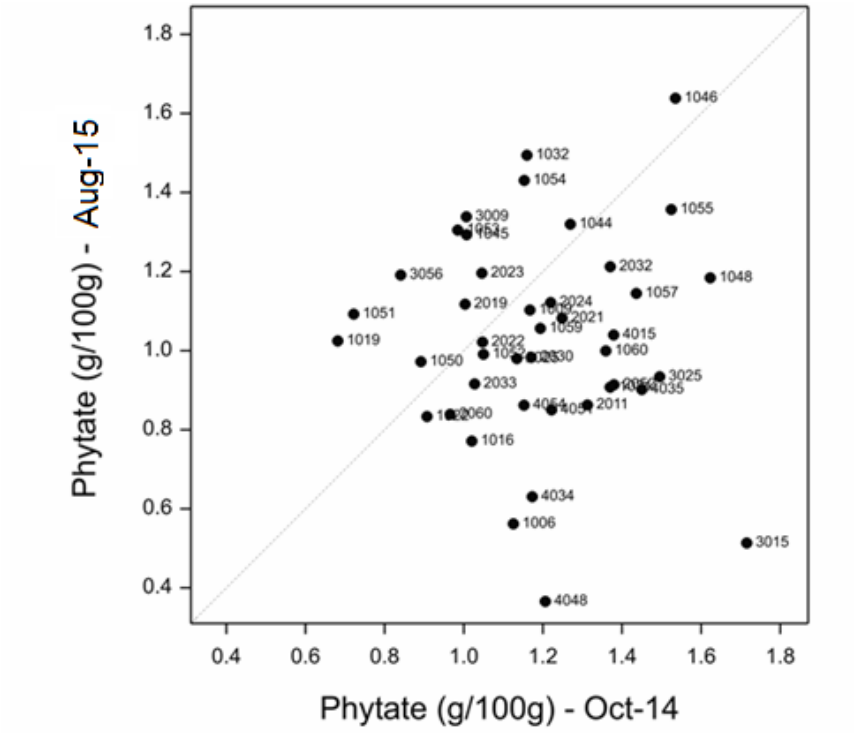


Figure 4.4: Scatter graph to show phytate content in 42 PMiGAP lines grown under glasshouse conditions (Aberystwyth University, October 2014 and August 2015).

Table 4.5: Comparison of the mean phytate content of 42 PMiGAP lines from the October and August glasshouse trials.

	August 2015			October 2014			F	*t*
	n	Mean	s.e.m	n	Mean	s.e.m	Prob[1]	Prob[2]
Phytate (g/100g)	42	1.032	0.3978	42	1.184	0.05361	0.49	**0.006**

s.e.m; Standard error of the mean

[1];H_0: Variance September 2015 = Variance October 2014; H_1: Variance September 2014 ≠ Variance October 2014

[2]; H_0: Mean September 2015 = Mean October 2014; H_1: Mean September 2014 ≠ Mean October 2014

From the results in Table 4.5, it can be seen that there are differences between the mean phytate content of the two populations at a P<0.006 significance threshold, upon the rejection of H_0.

Table 4.6: Spearman's Rank correlation coefficient between 42 PMiGAP lines grown under glasshouse conditions in October 2014 and August 2015.

Variates	**n**	**Spearman's Rank Correlation**	**Exact Probability**	**P-Value***
Aug'15/Oct'14	42	0.007	0.241	0.9494

*t approximation 0.05 on 40d.f.

From the results in Table 4.6, it can be seen that there is no significant correlation between the two populations.

The relationship between grain Fe/Zn levels and phytate content was determined by correlating Fe/Zn data to phytic acid assay data (Figure 4.5, Table 4.8 Table 4.9).

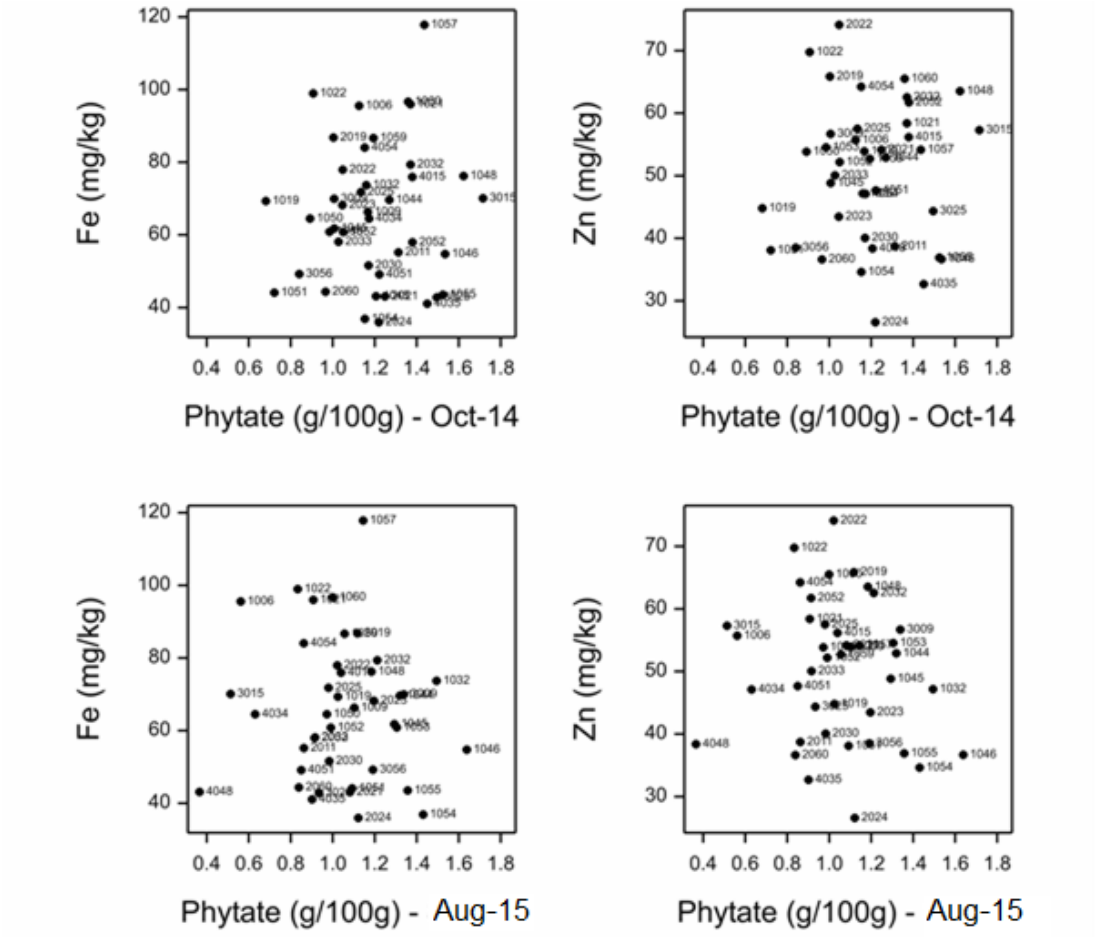


Figure 4.5: Scatter graphs to show Fe/Zn content Vs Phytate content from two glasshousegrowth trials conducted in Oct'14 and Aug '15 at Aberystwyth University.

Table 4.7: Spearman's Rank Correlation Coefficient between phytate levels and Fe/Zn content in 41 PMiGAP lines.

Varieties	N	Spearman's Rank Correlation	Exact Probability	*t*-Prob
Fe(mg/kg) ICRISAT/Phytate(g/100g) Oct'14	41	-0.041	0.200	0.799
Zn(mg/kg) ICRISAT/Phytate(g/100g) Oct'14	41	0.013	0.234	0.935
Fe(mg/kg) ICRISAT/Phytate(g/100g) Aug'15	41	-0.042	0.199	0.794
Zn(mg/kg) ICRISAT/Phytate(g/100g) Aug'15	41	-0.136	0.099	0.396

Although the literature reports a close positive association between phytate content and mineral elements, including Ca, Fe, Mg, and Zn, as indicated in studies on soybean, winter wheat, and maize (Feil & Fossati, 1997, Raboy *et al.*, 1984, Raboy *et al.*, 1989), this experiment shows contrasting results (Figure 4.5, Table 4.7,), that there is no correlation between phytate content and Fe/Zn content within the two populations. This suggests the strong possibility of breeding pearl millet cultivars with low phytate content and high Fe/Zn content as they seem to be associated with different pathways.

Using the phytate data from 235 PMiGAP lines grown at ICRISAT, it was possible to compare these to grain Fe/Zn levels (Figure 4.6).

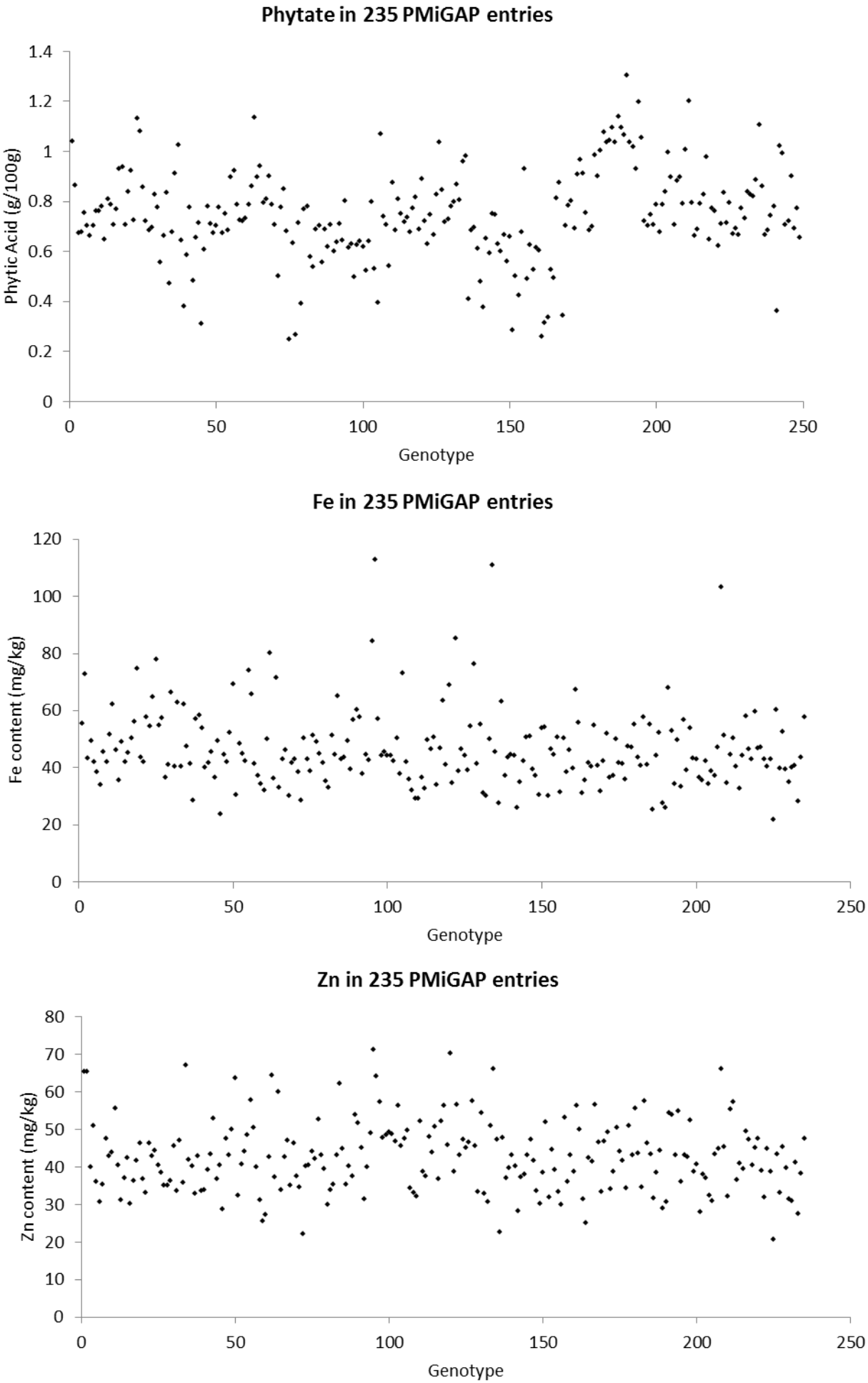


Figure 4.6: Phytate, Fe and Zn content in 235 OP PMiGAP entries grown under

field condition, ICRISAT, Patencheru, 2015.

The average content of phytate was 7473 mg/kg, with values varying by approximately 47%, from 6857 to 13032 mg/kg. The average content of Fe found in grains was 46.6 mg/kg, with values varying by approximately 74%, from 21.9 to 84.49 mg/kg, while the average content of Zn was 42.6 mg/kg, ranging from 20.72 to 71.11 mg/kg that is a difference of about 71% among PMiGAP entries.

The literature reports a variety of studies in humans that indicate that the absorption of Zn and Fe from a meal corresponds directly to its phytate levels (Ma *et al.*, 2005). Phytate/mineral molar ratios are typically used to predict the inhibitory effect of phytate on the bioavailability of minerals based on a 'critical value'. For example, if the phytate/Fe molar ratio is >1, this is indicative of poor Fe bioavailability (Hallberg *et al.*, 1989). Zn absorption is also greatly reduced when bound to phytate, although to a lesser extent and results in a negative Zn balance when the phytate/Zn molar ratio is >15 (Turnlund *et al.*, 1984). Phytate exerts its inhibitory effect on the bioavailability of Fe and Zn when the critical values are exceeded (Al Hassan *et al.*, 2016).

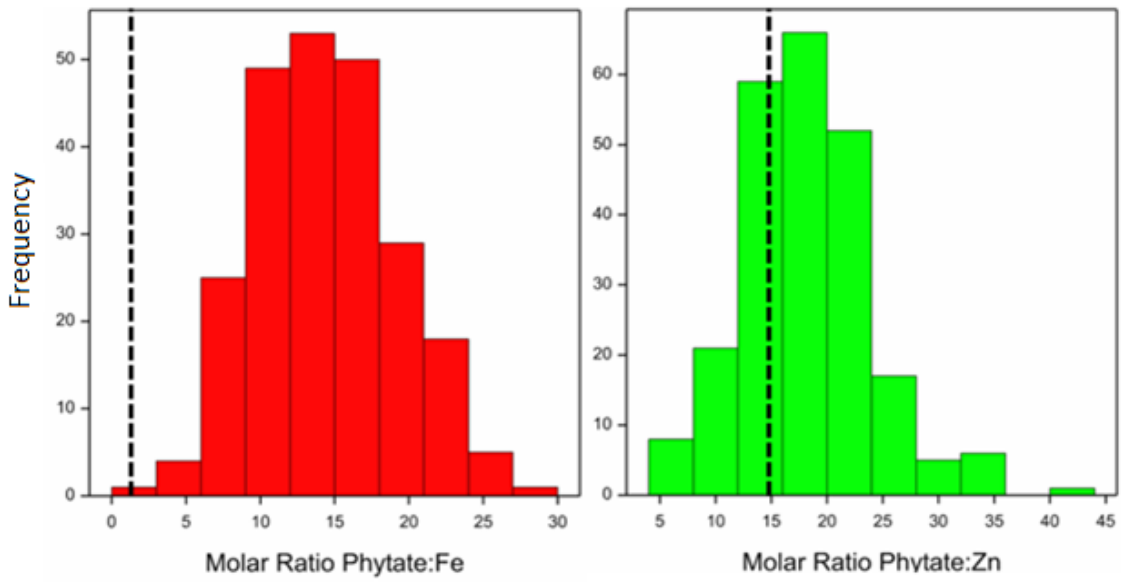


Figure 4.7: Histograms showing Phytate/Fe and Phytate/Zn molar ratios in 235 PMiGAP lines grown under field conditions at ICRISAT, Patencheru, dashed line denotes the critical values of 1 and 15, respectively.

According to the data in Figure 4.7, molar ratios exceeding the critical values indicate the proportion of the population that is likely to have reduced mineral

bioavailability. For Fe, all PMiGAP entries are likely to have reduced mineral bioavailability. However, there are 73 lines that do not meet the Zn critical value, which indicates that the mineral bioavailability would not be affected. These have potential use in biofortification studies for the development of elite lines. On the basis of phytate/Zn molar ratio, a set of 10 lines are suggested (Table 4.8), from the 73 which did not meet the Zn critical value of 15. These can be used in crossing programs for the creation of elite lines with low phytate/Zn molar ratio. Interestingly, 70% of lines are of African origin, which may suggest that lines from this region may naturally have a lower phytate/Zn molar ratio, as compared to Indian lines. Additionally, 70% of these lines achieved above-average combined Fe/Zn content.

Table 4.8: Suggested lines for the creation of elite lines selected based on phytate/Zn molar ratio. High Fe/Zn content was defined as above an average of 89.2mg/kg, for the total 235 lines analysed.

	Line IP. No	**Origin**	**Fe Critical Value**	**Zn Critical Value**	**Combined Fe/Zn (mg/kg)**
1	IP 9282	Togo	2.9	**5.7**	**176.9**
2	IP 13817	Burkina Faso	4.7	**5.8**	**91.25**
3	IP 18157	Mali	4.5	**6.6**	**90.56**
4	IP 8276	ICRISAT	6.3	**7.1**	**85.14**
5	IP 11353	Burkina Faso	5.5	**7.2**	72.97
6	IP 22419	ICRISAT	6.9	**7.6**	72.15
7	IP 8187	ICRISAT	6.3	**7.9**	**89.29**
8	IP 10379	Nigeria	7.5	**8.0**	**93.58**
9	IP 9406	Ghana	3.9	**8.1**	**176.84**
10	IP 10471	Zimbabwe	6.4	**8.5**	83.81

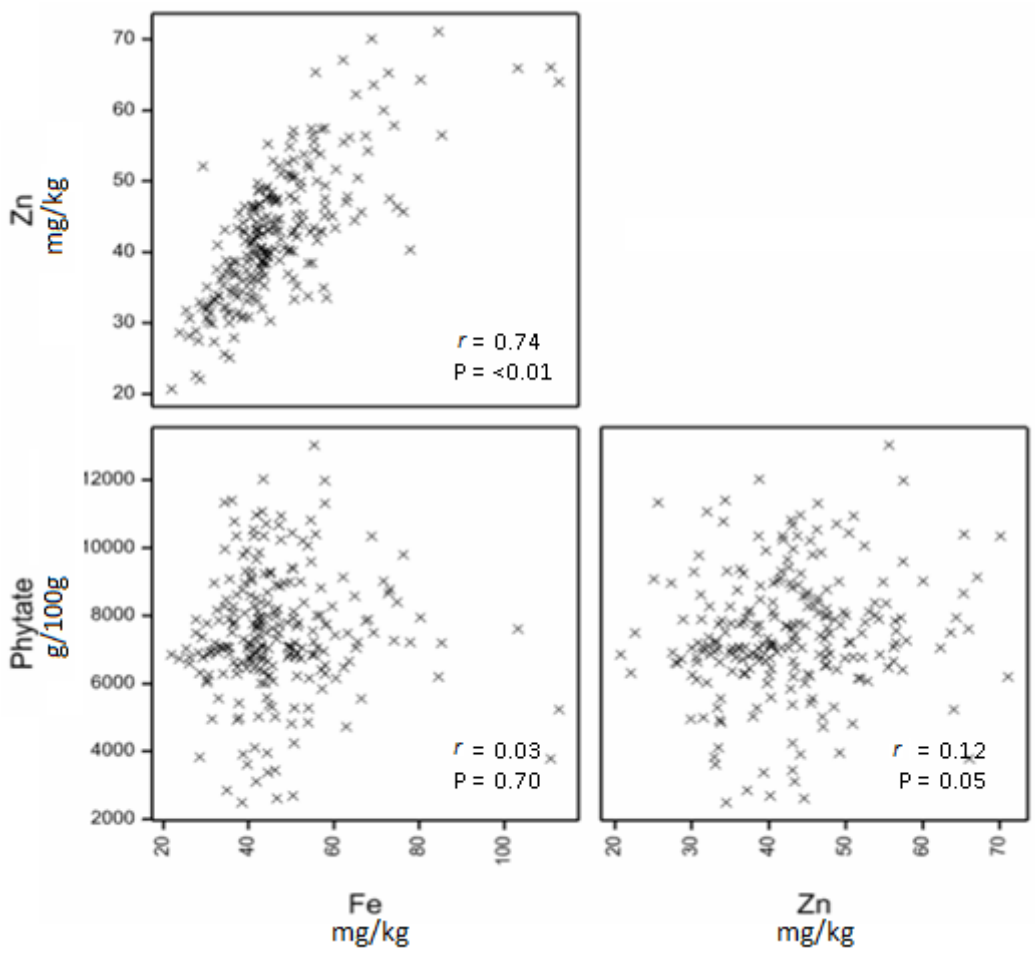


Figure 4.8: Scatter plot matrix for Fe, Zn and Phytate measured in mg/kg with Pearson's product moment correlations (*r*).

The results in Figure 4.8 indicate that although Fe and Zn levels are strongly correlated, levels of phytate and these minerals are not. This suggests that levels of phytate may be decreased without affecting levels of Fe and Zn.

4.5 Conclusions

This experiment primarily focuses on the inhibitory effect of phytate on Fe and Zn bioavailability in matured raw pearl millet grains. However, the influence of phytate on the bioavailability of essential micronutrients not only depends on the phytate content itself, but also phytate and mineral interactions. Phytate to mineral molar ratios have been extensively used to forecast the inhibitory effect of phytate on the bioavailability of minerals (Al Hasan *et al.*, 2016). The mean phytate to Fe millimolar ratio was found to be 14.5 and the mean phytate to Zn millimolar ratio was 18.1. Both average values exceeded the critical values of 1 and 15, respectively to significantly impair Fe and Zn absorption upon consumption. 100% of genotypes

are likely to reduce Fe bioavailability. However, a significantly lower percentage of the population, 69% are likely to reduce Zn bioavailability (Figure 4.7). This suggests that Zn is less affected by phytate than Fe, although the rate is still highly prevalent in the population.

Al Hasan's (2016) study states that among a variety of predictors of the inhibitory effects of phytate on micronutrient bioavailability, including; phytate intake, inadequate micronutrient bioavailability, inadequate micronutrient intake, age and total energy intake, phytate was the strongest inhibitory predictor of Fe and Zn and it was proven that phytate to Fe and Zn molar ratios would be expected to be 2.48 and 1.96 points higher for every 100mg increment in daily phytate intake, respectively. This demonstrates the potency of phytate as an inhibitor of Fe and Zn bioavailability from grain based diets.

Additionally, the dietary patterns of individuals living in pearl millet consuming communities must also be taken into consideration. These diets are often characterised by lack of diversity, with infrequent consumption of nutrient rich foods, meat and milk products (Torheim *et al.*, 2010). Studies also suggest that age and pregnancy status has a significant effect on the bioavailability of Fe and Zn. For example children and pregnant women are more likely to be effected by high levels of phytate to Fe/Zn ratios. This may be largely attributable to the increasing demands for these micronutrients during childhood and pregnancy (Butte *et al.*,2004).

The results from this experiment point to an urgent need to address the challenges associated with poor mineral bioavailability for pearl millet consuming communities and they also prove that bioavailability of micronutrients as a result of phytate must be considered a significant factor in plant breeding. The effect of phytate intake on the bioavailability of minerals means that their micronutrient consumption will likely fail to meet their mineral needs without intervention.

Processing raw grains may play a huge role in enhancing bioavailability of Fe/Zn since phytate cannot ever be completely eliminated. In this study, although readings were taken from raw milled grains, it has been proven that if grains were germinated (spouted in water for 24 hours) before consumption, this would significantly reduce the levels of phytate (Kumar & Chauhan, 1993). Other effective methods of processing to reduce phytate content include; soaking, autoclaving, fermentation and

debranning, as previously discussed. However, soaking of grains may result in a 25% loss of total Fe but *in-vitro* soluble Fe would remain relatively high. This is due to its distribution in the grain. Even though some Fe is lost, the 25% loss is more acceptable than if the levels of phytate were to remain high. Germination and fermentation of whole grains also maintains high levels of *in-vitro* soluble Fe. A molar ratio of <1:1 was achieved by this method in a study by Eyzaguirre *et al.,* (2006). Levels of Zn are generally less affected (Eyzaguirre *et al.,* 2006), which is due to Zn being more evenly distributed in the endosperm as opposed to Fe, which is mostly present in the testa. A ratio of <10:1 was achieved by soaking, germination and fermentation.

Chapter 5: Identifying and Quantifying Metabolites That Affect the Bioavailability of Absorbable Fe/Zn using HPLC

5.1 Summary

Even though pearl millet grains are rich in Fe and Zn, the bioavailability of these important micronutrients is limited due to the presence of metal chelating phenolic compounds, such as the flavonones; apigenin and luteolin (Tako *et al.*, 2015). An extensive protocol was developed for the extraction of these compounds and the level of variation was assessed in 34 PMiGAP lines by high performance liquid chromatography with photodiode array detection (HPLC-PDA). Identification of flavonone derivatives and other phenolic compounds was carried out by high performance liquid chromatography with photodiode array and tandem mass spectrometry detection (LC-PDA-MS^n). Compound structures were also elucidated from fragmentation patterns, obtained by drilling down into MS^n events.

In the pilot phase, HPLC-MS untargeted analysis of 57 PMiGAP lines revealed 16 compounds of interest including; hydroxycinnamic acids, phenolamides and flavones. Sugar moieties attached to aglycones were also tentatively identified by examination of MS^n events. Additionally, targeted analysis gave some insight into the chemical relationships between lines. However, it was found that the use of a larger population, an optimised methanol (MeOH) extraction protocol and the inclusion of an appropriate internal standard would be beneficial. Furthermore, MeOH extraction plus acid hydrolysis and comparison against standards enabled the identification of sugar moiety positions, when attached to core aglycones.

In the post-pilot phase, the extraction of phenolic compounds from 185 PMiGAP lines was achieved using an optimised MeOH extraction method, with biochanin A as an internal standard. After targeted analysis, it was possible to accurately identify and compare peak areas of apigenin/luteolin glycosides (as a measure of relative abundance) in UV chromatograms. Five flavonoid glycosides/aglycones were quantified using standard calibration curves, on a molar basis and it was found that the content of all compounds was highly variable between lines. The correlation between polyphenol content and micronutrients was also investigated. Although no significant correlations were observed, upon application of a statistical filter, some clustering was revealed between apigenin and luteolin derivatives.

5.2 Introduction

5.2.1 The Relationship between Polyphenolics and Micronutrient Bioavailability

The bioavailability of a substance refers to the proportion absorbed and entered into systemic circulation (Xu *et al.,* 1995). Pearl millet grains are rich sources of phytochemicals and micronutrients (Singh & Raghuvanshi, 2012). However, their benefits are often compromised by the presence of polyphenolic compounds, many of which are proven to have adverse effects on micronutrient bioavailability, via their metal chelating properties (Al-Sa'aidi, 2003, Cook & Samman, 1996). For example, some interfere with Fe absorption via the formation of insoluble complexes in the gastrointestinal lumen (Brune *et al.,* 1989). Plant polyphenols are secondary metabolites that are ubiquitous in monocotyledonous and dicotyledonous plants and are synthesised via complex non-reversible pathways (Croizer *et al.,* 2008). Whilst the literature reports several studies acknowledging their benefits, including improvement to gut microbiota, inflammation and diabetes symptoms (Cardoso *et al.,* 2015), this study is one of the first to investigate the flavonones; apigenin, luteolin and their derivatives as significant inhibitors of Fe/Zn uptake.

PMiGAP lines with elevated levels of Fe and Zn were previously identified in Chapter 3. However, increased grain Fe and Zn content may not necessarily translate into a proportional increase in absorbed Fe and Zn, if increased Fe and Zn correlates to increased concentrations of Fe and Zn absorption inhibitors (Tako *et al.,* 2015). In light of this, it is necessary to investigate potential phenolic inhibitors in the PMiGAP as a prerequisite for the selection of elite, micronutrient-rich lines, for biofortification. Polyphenolic compounds that have been reported as inhibitors of mineral bioavailability can be seen in Table 5.1 (Nestel *et al.,* 2006).

Table 5.1: Polyphenolic compounds in pearl millet showing a negative effect on mineral bioavailability/absorption (Tako *et al.*, 2015).

Class	Compound	Citation
Flavononones	Apigenin	Mira *et al.*, 2002
	Baicalein	Perez *et al.*, 2009
	Luteolin	Mira *et al.*, 2002
Flavonol	Galangin	Park *et al.*, 1989
	Kaempferol	Mira *et al.*, 2002
Isoflavones	Dihydrodaidzein	Mira *et al.*, 2002
	Genistein	Benheral *et al.*, 2008
Anthocyanins	Pelargonidin	Wang *et al.*, 2010

In a recent *in-vitro* study, Tako *et al.*, (2015) compared the capacity of Fe-biofortified pearl millet (ICTP-9203, high-Fe, 85µg/g) and standard pearl millet (DG-9444, low-Fe, 26µg/g) to deliver Fe for haemoglobin (Hb)-Synthesis, in Red Junglefowl (*Gallus gallus*). However, the high-Fe pearl millet provided less absorbable Fe than expected. It was later concluded that a likely reason why high-Fe pearl millet wasn't performing as well as expected was due to the presence of certain polyphenolic compounds and/or phytate. Therefore, polyphenolic and phytate profiles of pearl millet should be evaluated to further improve the bioavailability of Fe, and other mineral elements.

5.2.2 Generic Structure and Major Classifications

Polyphenols are a class of several thousand compounds found across all crops, fruits and vegetables. They are divided into two groups; flavonoids and non-flavonoids and their biological activity is directly linked with their chemical structure (Williamson & Manach, 2005). Flavonoids consist of the flavonols, flavones, isoflavones, flavononones and anthocyanidins. They represent the largest class of polyphenols, with a common structure of diphenylpropanes (C6-C3-C6), consisting of two aromatic rings linked through three carbons, as seen in Figure 5.1 (Ross & Kasum, 2002).

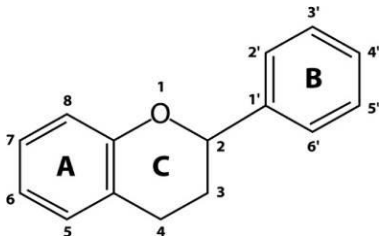


Figure 5.1: Basic flavonoid structure.

The non-flavonoids comprise the phenolic acids (hydrobenzoic and hydroxycinnamic acids), lignans, and stilbenes (Etcheverry *et al.,* 2012). Food polyphenols are typically bound to a sugar moiety, forming glycones; when the sugar moiety is absent, the core polyphenol structure is known as an aglycone. During gastrointestinal digestion, the aglycone may detach from the sugar moiety by acetolysis, resulting in a more absorbable compound (Etcheverry *et al.,* 2012).

5.2.3 Absorption of Flavonoids

In humans, flavonoids are absorbed from the gastrointestinal tract and excreted either unchanged or as flavonoid metabolites in urine or faeces (Cook & Samman, 1996). They are well known antioxidants, free radical scavengers and metal chelators. The structural requirements for metal chelating activity are between three potential coordination sites; (i) between the 5-hydroxy and 4 carbonyl group, (ii) between the 3-hydroxy and 4 carbonyl group and (iii) between the 3',4'-hydroxy group on the B ring (Figure 5.2) (Symonowicz & Kolanek., 2012).

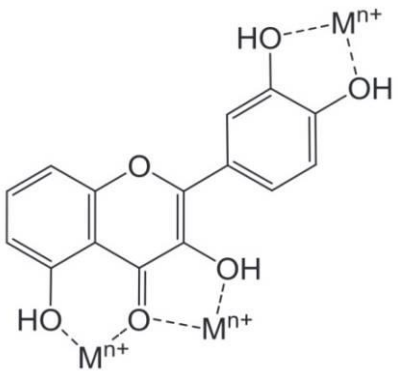


Figure 5.2: Potential metal binding sites on a flavonoid molecule.

Detailed information about absorption, metabolism and excretion of flavonoids in humans is scarce, which presents a gap in the literature. However, some studies suggest that they are not well-absorbed in humans and are unable to reach general

circulation unchanged/unmodified, in measurable concentrations (Gugler *et al.,* 1975). For example, in a study by Gugler *et al.,* (1975) the metabolism of the flavonoid quercetin was investigated in six humans, aged 21-32 years. After oral administration of a single 4g dose, no measurable concentrations of quercetin or its derivatives were detected in plasma/urine. However, 53% of the oral dose was recovered unchanged in faeces. 1% of the oral dose, which equates to approximately 40mg, was absorbed. Even though this result is significant, a 4g dose greatly exceeds what would normally be consumed in the diet and may not accurately reflect the metabolism of flavonoids from dietary sources.

5.2.4 Analysis of Soluble Phenolics

Analysis of soluble phenolics is greatly facilitated by tandem mass spectrometry (MS) coupled to liquid chromatography (LC), resulting in an important and widely used approach into the study of these compounds, without previous isolation or clean-up (Ferreres *et al.,* 2007). Studying the relative abundance of the main ions from the MS preferential fragmentation on the $^{-/+}MS^n$ events allows characterisation of the aglycone and other moieties that may be attached by *C* or *O* linkage. Factors that have previously hindered research are; (i) the inaccuracy of measuring compounds in biological samples, (ii) a dearth of information on their absorption/metabolism (Crozier *et al.,* 2000) and (iii), the occurrence of complex mixtures of *O*-glycosyl and *C*-glycosyl flavones in phytochemical extracts is frequent and isolation of different compounds prior to identification is challenging (Ferreres *et al.,* 2006).

5.2.5 Factors Influencing Flavonoid Absorption

The bioavailability of flavonoids (and attached micronutrients) depends on the glycoside moieties attached to the aglycone (Ross & Kasum, 2002; Gibson *et al.,* 2000). For example, a study by Felgines *et al.,* (2000) demonstrated that the flavonoid naringenin (a flavone commonly found in grapefruit) occurs predominantly in two glycoside forms; naringenin-7-rhamnoglucoside and naringenin-7-glucoside. When the absorption kinetics of the naringenin glycosides were studied in rats, it was found that even though the absorption kinetics were similar, naringenin-7-rhamnoglucoside exhibited a delay in intestinal absorption, resulting in a significant decrease in bioavailability (Felgines *et al.,* 2000).

Additionally, upon the study of flavonoid urine recovery in rats, Xu *et al.,* (1995) reported that the isoflavone daidzein had greater bioavailability than genistein, and apigenin had lower bioavailability than both isoflavones. This suggests that the core aglycone may also influence the rate of bioavailability (Xu *et al.,* 1995).

Flavonoid bioavailability may also be influenced by gut microflora (Klaasen & Rozman, 1991). For example, Xu *et al.,* (1995) demonstrated that the bioavailability of daidzein and genistein, at three doses per day was about 21 and 9%, respectively, as measured by urine recovery of a known dose. It was elucidated that the range in results was a result of flavonoid bioavailability being dependent on the relative ability of gut microflora to degrade these compounds. Some bacteria present in the human intestine have the ability to metabolise and degrade flavonoids, as well as to free aglycones. Several groups of bacteria are responsible for this, including: *Lactobacilli, Bacteroides* and *Bifidobacteria* (Hawksworth *et al.,* 1971). Xu *et al.,* (1995) also showed that flavonoid glycosides are poorly absorbed in the small intestine, as compared with their aglycones, owing to the glycosides' higher hydrophilicity and greater molecular weight.

5.2.6 Apigenin and Luteolin Pathways

Many flavonoids have been subjected to a wide range of applications in medicine, pharmacy and in the food industry as antioxidants, preservatives and flavouring agents (You *et al.,* 1993). However, flavonoids such as luteolin and apigenin (Figure 5.3) in grains are in need of considerable attention because of their adverse effect on nutritional quality (Salunkhe *et al.,* 1983).

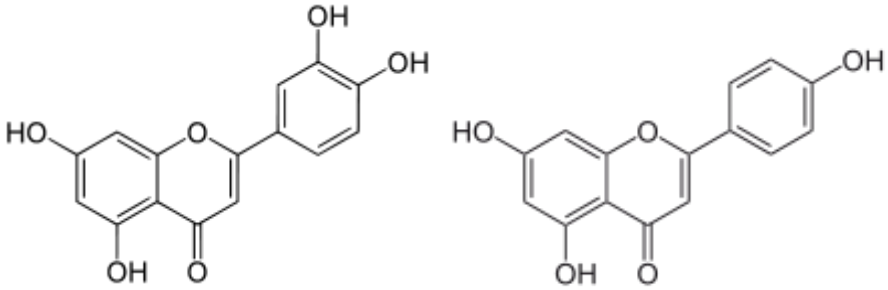


Figure 5.3: Luteolin (left) and apigenin (right).

Their anti-nutritional activity in grains can be reduced by; soaking in water, fermenting or by removal of the pericarp/testa (outer seed coat) by pearling. Treatment with alkaline reagents and ammonia can also remove up to 90% of polyphenols (Salunkhe *et al.*, 1983).

According to Figure 5.4, apigenin/luteolin biosynthesis begins with the condensation of three malonyl-CoA molecules with *p*-coumaroyl-CoA by chalcone synthase to form naringenin chalcone (Casas *et al.*, 2014). Chalcone isomerase (CHI) then converts narigenin chalcone into the flavanone naringenin (Winkel-Shirley, 2001). Naringenin then acts as the substrate for the flavanone-3′-hydroxylase enzyme, which is encoded by the locus *Pr1*, to synthesise eriodictyol. Naringenin is then converted into various flavones which involve the enzymes flavanone-2-hydroxylase (F2H) and *C*-glycosyl transferase (CGT), to generate apigenin-6-*C*-glucoside (isovitexin) or its isomer, apigenin-8-*C*-glucoside (vitexin). In the case of eriodictyol, the flavones generated after these conversions steps are luteolin-6-*C*-glucoside (isoorientin) or its isomer, luteolin-8-*C*-glucoside (orientin) (Winkel-Shirley, 2001; Morohashi *et al.*, 2012). Flavones are often *O* or *C* glycosylated by glycosyl transferases, to generate, for example, apigenin-7-*O*-glucoside (Casas *et al.*, 2014).

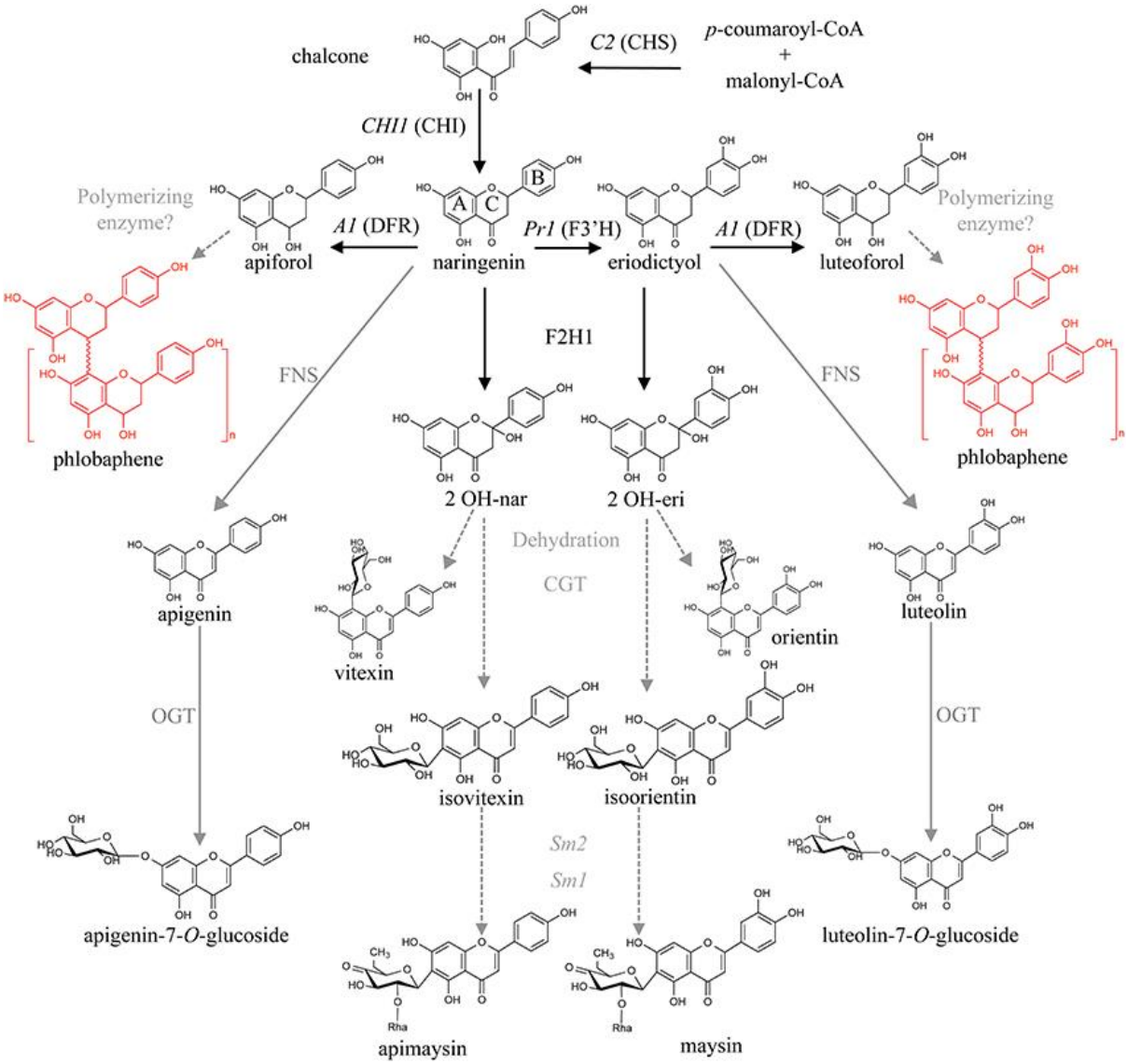


Figure 5.4: Proposed flavone and 3-deoxy flavonoid biosynthetic pathways. Condensation of *p*-coumaroyl-CoA and malonyl-CoA by chalcone synthase (CHS, encoded by the locus C2) produces naringenin chalcone, which is then converted to naringenin by chalcone isomerase (CHI). Naringenin is converted to apiforol (flavan-4-ol) by a dihydroflavonol reductase (DFR, A1) and polymerised into phlobaphenes. It may also be converted to isovitexin (*C*-glycosylflavone) by a flavanone-2-hydrohylase (F2H) and a *C*-glycosyl transferase (CGT). A flavone synthase (FNS) may also catalyse this step, followed by an *O*-glycosyl transferase (OGT). Naringenin may also be converted to eriodictyol by a flavanone-3′-hydroxylase (F3′H, Pr1). The proposed steps for the conversion of apigenin and luteolin into the *C*-glycosylflavones (apimaysin and maysin) is thought to involve at least three enzymatic conversions: glycosylation at C6, followed by rhamnosylation and dehydration, mediated by Sm2 and Sm1. Enzymes are identified in black, those proposed in grey (Casas *et al.*, 2014, Morohashi *et al.*, 2012).

5.2.7 Preferential Metal Chelation Sites

The metal chelating property of flavonoids is primarily due the presence of –OH groups, which chelate a free metal ion. The site where the removal of H atoms (deprotonation) from the -OH groups occur is where the metal atom is bound (Primikyri *et al.,* 2014), thus preventing the participation of the metal ion in systemic circulation. When flavonoids bind Fe, they inhibit the catalysis of the Fenton reaction, which produces hydroxyl radicals. These are known to cause damage to cells (Primikyri *et al.,* 2014). Iwahashi *et al.,* (2004) observed from the measurement of visible high performance liquid chromatography-electron spin resonance-mass spectrometry (HPLC-ESR-MS) spectra that luteolin-7-*O*-glucoside chelates Fe ions (Iwahashi *et al.,* 2004) and luteolin can form a luteolin–Fe(III) complex with a ratio of 1:1 and coordinate to the Fe(III) ion at the 3′,4′-dihydroxyl group in ring B of the luteolin molecule (Yang *et al.,* 2014). This is the most favourable site, whereas the 4-5 group is considered a less favourable complexation site (Primikyri *et al.,* 2014). There are a number of studies where electrospray MS has been used to study metal ion interactions in flavonoids from different classes (Fernandez *et al.,* 2002; Mira *et al.,* 2002). For example, Fernandez *et al.,* (2002) found that a spectrum given by luteolin, in the presence of Fe revealed two peaks at *m/z* 341 and 626, corresponding to the ions $[Fe(II)+(M-H)]^+$ and $[Fe(III)+2(M-H)]^+$. Complexation of Zn(II) ions with luteolin has also been verified with Nuclear Magnetic Resonance (NMR) spectroscopy and first-principle calculations, as demonstrated by Primikyri *et al.,* (2014). The study suggests that the combined use of NMR spectroscopy with emphasis on the phenolic OH resonances, with *ab-initio* calculations (quantum chemistry using computational chemistry methods) provides a valuable tool for accurate structural and electronic description of flavonoid-metal diamagnetic complexes (Primikyri *et al.,* 2014). The study also reports that even though it is possible for luteolin to bind to Zn^{2+} ions through two possible sites, (4-5_H and 3_H'-4_H'), the 4-5_H site appears to be the most favourable (Figure 5.5).

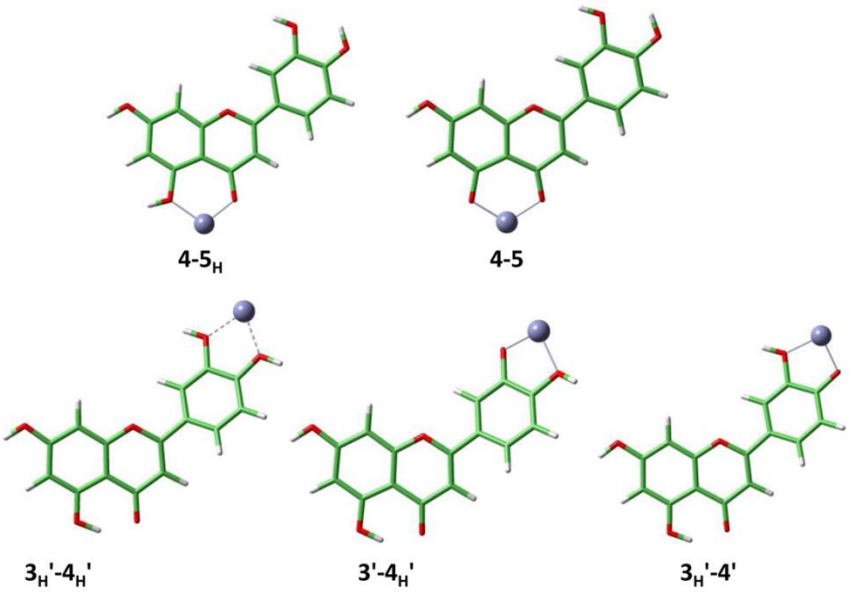


Figure 5.5: Potential chelation sites of metals (Image derived from Primikyri *et al.*, 2014).

5.2.8 Other Uses for Flavonoids

Although finding appropriate methods to reduce metal-chelating polyphenols in plants is important, research should equally focus on exploiting the pharmacological/medicinal properties of these compounds, especially in light of the on-going challenges associated with the high cost of drug development and the threat of drug resistance. Techniques such as CRISPR gene editing, tissue culture, GM and other modern plant breeding tools will play a vital role in exploring the beneficial effects of these compounds (Soetan, 2008). Certain flavonoids groups are reported to possess antibiotic (Soetan *et al.,* 2006), antifungal (Jun *et al.,* 1989) and antiviral activities (Okubo *et al.,* 1994). Their beneficial traits are also of great interest owing to their antioxidant and free-radical scavenging abilities, observed *in-vitro* (Ross & Kasum, 2002). Luteolin is thought to reduce high blood cholesterol levels by inhibiting intestinal cholesterol absorption, mediated by the gene Niemann–Pick C1-like 1 (Nekohashi *et al,* 2014). Flavones are also receiving considerable attention due to their potential role in the prevention/treatment of cancer (Messina, 1999). For example, tricin is proven to inhibit the growth of human-derived malignant MDA-MB-468 breast tumour cells at sub-micromolar concentrations (Cai *et al.,* 2004). Similarly, chrysoeriol, which is known for its antioxidant and anti-inflammatory

properties, may be used for the prevention/treatment of vascular diseases (Cha *et al.*, 2009). For example, a study by Cha *et al.*, (2009) demonstrated that, upon investigation of the effect of chrysoeriol on the proliferation of human aortic smooth muscle cells, chrysoeriol significantly inhibited platelet-derived growth factor migration, which is one of the most potent factors in the development and progression of a variety of vascular disorders.

5.3 Method Development

Method development trials were conducted with the aim of extracting a wide range of polyphenolic compounds in pearl millet seed, using an optimised extraction protocol.

5.3.1 Plant Material

See Chapter 2, Table 2,3.

5.3.2 Extraction, Clean-up and Separation

Extraction

HPLC analysis of grain begins with sample preparation, which involves grinding the sample into a fine powder, to obtain homogeneity of the matrix; this is usually followed by a solvent extraction to remove target compounds from other components in the matrix. In general, selecting a suitable solvent is based on the chemical properties of the compounds of interest. For polyphenols, MeOH (polarity index 5.1) is typically used as it is inexpensive and a wide range of polar (and some non-polar) compounds dissolve readily in it. It has a low boiling point of 65°C; therefore it is easily evaporated, thus reducing the length of time to complete a sample preparation protocol.

Clean-up

Sample clean-up is needed to reduce the detection limit and to limit interference from other compounds that may affect the identification and quantification of target compounds. This is often achieved by solid phase extraction (SPE), which is based on binding target compounds to a matrix, with bonded chemical groups which have appropriate chemical properties, such as C18 or an ion-exchange moiety. This is

usually carried out using cartridges based on silica particles for larger sample sizes (exceeding 1mL) or membrane discs for smaller sample sizes. Target compounds are separated from crude extracts by binding to the SPE matrix, followed by subsequent elution with an appropriate solvent. An advantage of this method is that it minimises the content of non-target compounds, which may interfere with the HPLC analysis and MS detection e.g. sugars.

Separation

Reversed Phase (RP) is routinely used for the separation of phenolic compounds and is characterised by hydrophobic interactions with the stationary phase and hydrophilic interactions with the mobile phase, which is dependent on the proportion of organic solvent in the mobile phase (Snyder, 2012).

5.3.3 Protocol Optimisation

A summary of the method development for protocol optimisation can be seen in Figure 5.6.

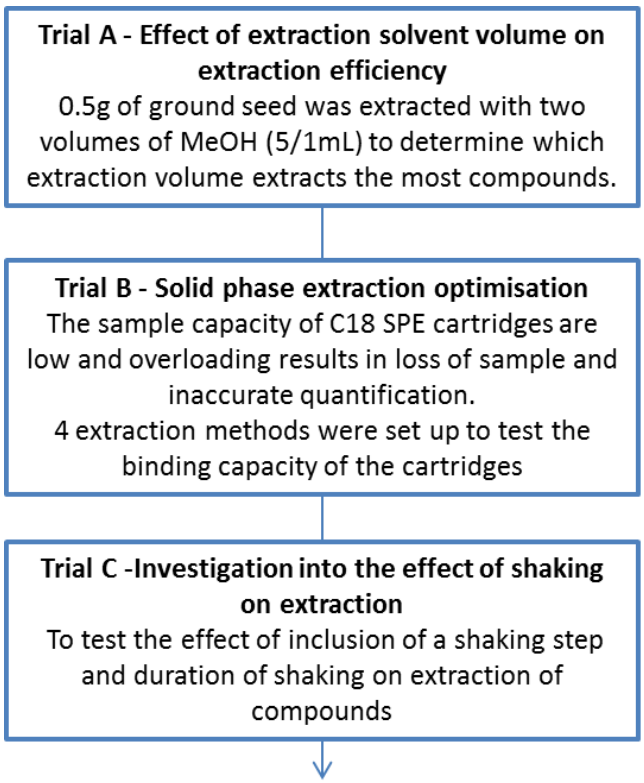


Figure 5.6: Summary of the protocol optimisation process.

Trial A: Effect of Extraction Solvent Volume on Extraction Efficiency

Materials and Methods

Using an electric powered grinding mill (Retsch Mortar Grinder Mill, 110V/60Hz OY-04181-10), 0.5g of whole pearl millet seed from each genotype was ground to obtain a fine powder, able to pass through a 1mm sieve. The effect of extraction solvent (MeOH) volume was tested with two extracts from the same PMiGAP line. Material was extracted initially with either 1/5mL 70% MeOH. Samples were vortexed for 1 minute and centrifuged at 13,000 rpm for 10 minutes. The supernatant was transferred to clean tubes. The pellets were then washed with 500µL 70% MeOH and centrifuged again at 13,000 rpm for 10 minutes. The second supernatant was added to the first supernatant fraction to maximise recovery of target compounds. The MeOH was then removed using a heated centrifugal evaporator (Jouan RC10.22) set to 70°C, at 13,000 rpm and 500µL distilled water was added to the residual aqueous fraction. A Waters Sep-Pak C18 500mg cartridge (WAT036945) was prepared by passing through 5mL 100% MeOH followed by 5mL 5% acetic acid. Either 25 or 100% of the sample was then loaded into the cartridge (Table 5.2) and washed with 3mL water, and then eluted into a clean vial.

Table 5.2: Sample extraction volumes and amount loaded into the SPE C18 cartridges.

Genotype	Initial extraction volume (mL)	% of sample added to SPE C18
4036	1	100%
4036	5	100%
1037	1	*25%
1037	5	*25%

*Samples were made up to 4mL with purified water. 25% of the diluted sample was added onto the SPE C18 cartridge.

The samples were then fully dried using a heated centrifugal evaporator at 13,000 rpm, at 70°C and reconstituted in 70µL 70% MeOH. HPLC analysis was carried out on a Waters system with a 996 PDA detector and a 8 mm×100mm i.d., 4µm, C18 Nova-Pak radial compression column (Waters), equilibrated with 100% solvent A (5% acetic acid) at a flow rate of 2 mL/min. The injection volume was 10µL. Compounds were eluted by linear gradient to 100% solvent B (100% MeOH) over 50 minutes and monitored from 240 to 400 nm.

Results and Discussion

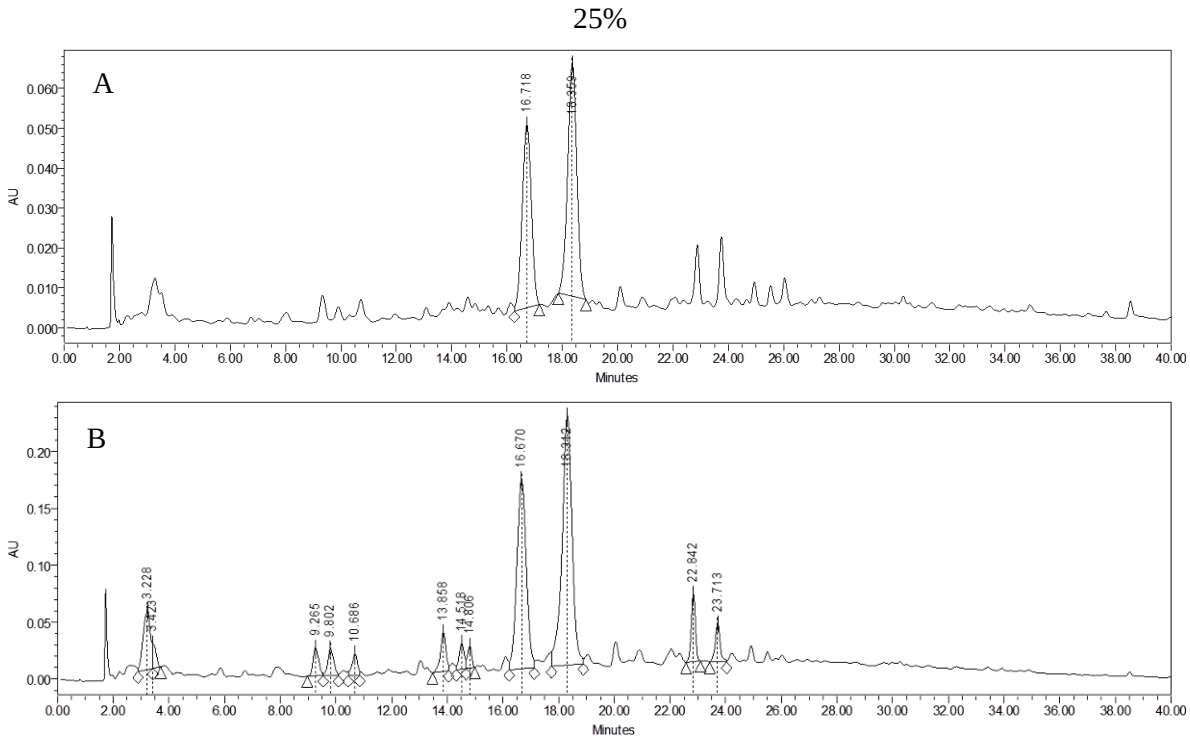


Figure 5.7: HPLC chromatogram showing separation of soluble phenolic compounds in extracts of pearl millet (genotype 1037) detected by PDA at 280nm following purification of 25% of the extract by SPE. Samples were extracted with either 1mL MeOH (A) or 5mL MeOH (B).

Upon extraction method A, two main compounds were detected by HPLC-PDA. The two peaks at t_r 16.7 and 18.3 minutes, which correspond to a luteolin derivative and an apigenin derivative, respectively were very weak according to the absorbance units. This indicates that extraction method A does not fully extract phenolic compounds compared to B, where stronger peak signals for 12 compounds were detected.

100%

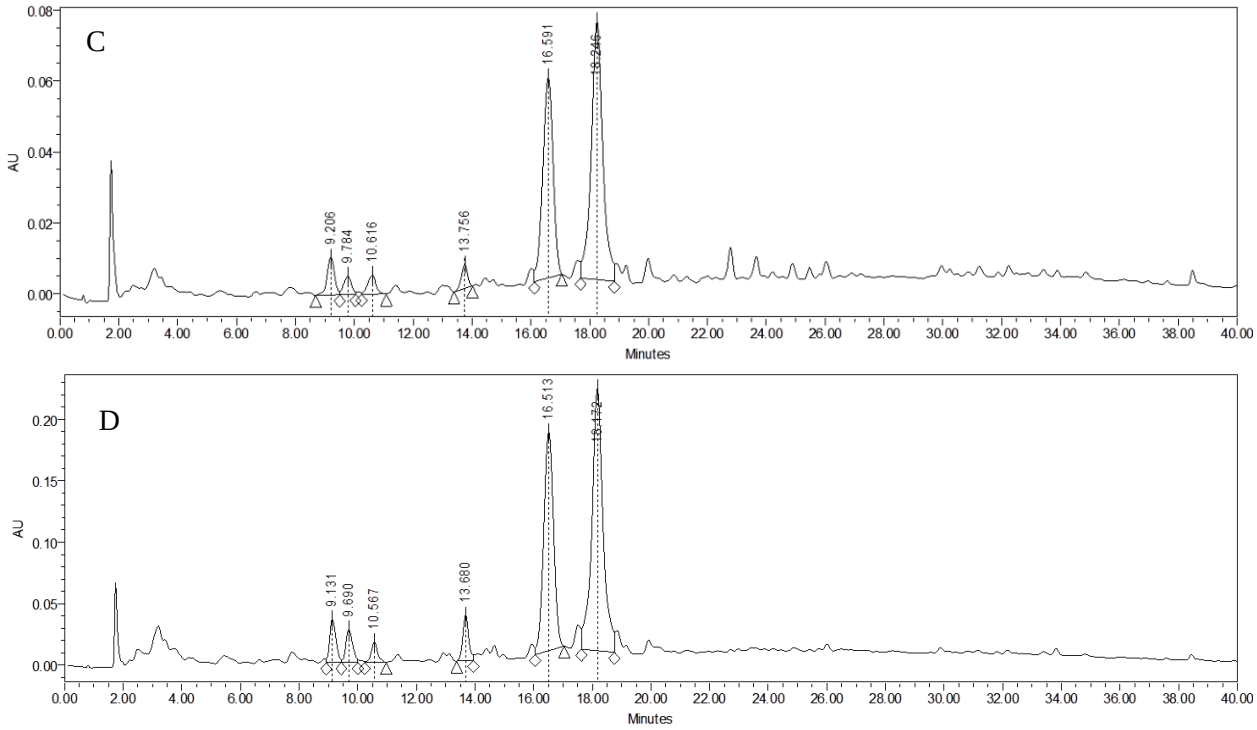


Figure 5.8: HPLC chromatogram showing separation of soluble phenolic compounds in extracts of pearl millet (genotype 4036) detected by PDA at 280nm following purification of 100% of the extract by SPE. Samples were extracted with either 1mL MeOH (C) or 5mL MeOH (D).

Upon comparison of extraction method C to D, the same 6 peaks were detected. However, the peaks were much weaker where a 1mL MeOH extraction was used (C), as observed by the reduced UV absorbance signal. In light of this, a 5mL extraction solvent volume is deemed more appropriate for future work, as this resulted in greater extraction of compounds.

Trial B – Solid Phase Extraction Optimisation

The initial experiment on extraction volume efficiency described above indicated that samples loaded onto SPE cartridges may have exceeded the binding capacity of the columns when the total extract volume was loaded. An experiment was designed to test both the efficiency of extraction volume and binding capacity of SPE columns, with extracts from the same material.

Materials and Method

Samples of genotype 2052 were prepared and analysed by HPLC, as described above.

Table 5.3: SPE optimisation of extracts of genotype 2052 including: MeOH extraction volume, quantity added to SPE C18 cartridge and quantity dried down and resuspended in the final volume.

Extraction method*	Initial extraction volume (mL)	% of sample added to SPE C18	% of SPE C18 eluent in final volume
A	1	100%	25%
B	5	100%	25%
C	5	25%	100%
D	1	25%	100%

*All treatments were carried out in triplicate.

Treatments A and D were extracted with 1mL 70% MeOH and B and C were extracted with 5mL 70% MeOH. B and C determined overloading of SPE C18 cartridges.

Results and Discussion

Extraction method C was deemed the most replicable, upon comparison of % peak area (Table 5.4), as indicated by smaller differences between replicate readings, as compared to B. Therefore, it is likely that the SPE C18 cartridge was overloaded with extraction method B compared with method C. Methods A and D produced weak peaks indicating that the extraction with 1mL was not efficient, as observed above. In light of the above, out of the four extraction methods tested, C is deemed the most accurate and reliable.

Table 5.4: A comparison of the peak area (%) of compounds detected from all extraction methods. Method C was deemed 100% efficient and peak area % from other methods were worked out as a percentage of the average peak area of method C, per peak at each respective t_r.

Method/Replicate	Peak Areas %									
t_r (Mins)	3	9	9.5	10.3	13.5	15.8	16.3	17.3	18	18.8
A1	-	18	-	-	17	40	20	22	21	20
A2	-	19	-	-	-	27	18	17	16	15
A3	-	10	-	-	6	10	13	10	10	10
B1	-	27	-	-	30	-	29	33	34	25
B2	-	49	38	-	28	73	44	50	53	46
B3	-	72	-	-	66	68	93	75	69	58
C1	99	86	87	-	74	86	91	90	91	92
C2	95	100	100	-	100	100	100	100	100	100
C3	100	100	100	-	100	100	100	100	100	100
D1	-	-	-	-	-	-	31	28	31	29
D2	-	-	-	-	-	26	25	27	29	26
D3	-	-	-	-	-	-	29	29	31	28

Conditional formatting has been used to colour the values in terms of high (green) and low (red) % peak area.

Trial C: Investigation into the Effect of Shaking on Compound Extraction

An experiment was designed to test if the inclusion/duration of a shaking step resulted in differences in the quantity of compounds extracted.

Materials and Methods

To test the effect of the inclusion of a shaking step and duration of shaking on the extraction of compounds, extracts were treated as described in Table 5.5

Table 5.5: Duration of shaking step during extraction (genotype 1030)

Extraction method*	Duration of shaking step (minutes)
X	0
A	10
B	30
C	1440

*All treatments were carried out in triplicate

The extraction volume was 5mL 70% MeOH. Extraction methods A, B and C included a shaking step of 10, 30 and 1440 minutes, respectively (170rpm using a KSSOL digital IKA Labortechnik Shaker). Samples of genotype 1030 were extracted and analysed by HPLC, as described above.

Results and Discussion

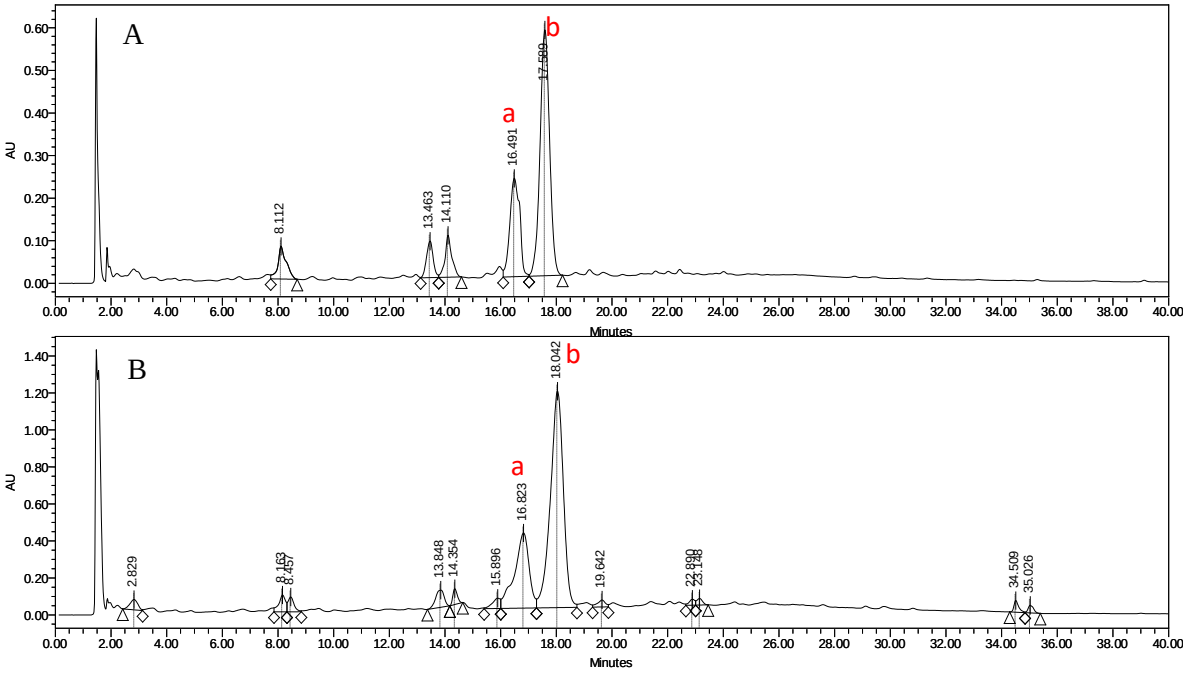


Figure 5.9: HPLC chromatograms showing differences in phenol profiles detected by PDA at 280.0nm. For A, samples were extracted with no shaking step and B,

samples were extracted with a 30 minute shaking step. a- a luteolin derivative; b- an apigenin derivative.

The results in Figure 5.9 suggest that the inclusion of a 30 minute shaking step caused a vast increase in the quantity of detectable compounds, when compared to a 0 minute control. Upon comparison of a luteolin derivative and an apigenin derivative extracted with no shaking (A) and the same compounds detected with 30 minutes shaking (B), it was observed from the peak area absorbance units (as a measure of relative abundance) that the quantity of compounds extracted increased up to twice as much.

Additionally, there were no major differences between the peak areas of compounds detected after 30 minutes shaking, as compared to 1440 minutes shaking (Table 5.6). Therefore, 30 minutes was deemed sufficient for the extraction of all phenolic compounds present in pearl millet grain.

Table 5.6: Table to show differences in peak area (%) when shaking duration was manipulated. The average values of replicates at 1440 minutes were deemed the most efficient and peak area % from other methods were worked out as a percentage of these values, per peak at each respective t_r.

Shaking duration/Replicate	Peak Areas %									
t_r (Mins)	3	8.3	8.7	14.1	14.3	14.7	15	16.4	17.5	18.6
0minsA	-	100	-	96	-	-	-	100	100	-
0minsB	-	100	-	90	-	-	-	100	100	-
0minsC	-	100	-	100	-	-	-	100	100	-
10minsA	86	100	74	38	-	100	100	-	87	88
10minsB	100	100	100	60	-	100	100	100	100	100
10minsC	100	100	89	46	100	100	100	100	81	81
30minsA	96	100	85	56	100	100	91	75	92	94
30minsB	93	100	85	48	100	100	100	-	89	92
30minsC	100	100	94	55	100	100	100	74	99	96
1440mins A	96	76	88	100	100	100	-	100	83	82
1440mins B	91	100	88	86	79	-	-	-	98	97
1440mins C	100	100	100	100	93	-	-	-	100	100

Conditional formatting has been used to colour the values in terms of high (green) and low (red) % peak areas.

5.4 Experimental Aims

34 PMiGAP lines were selected for analysis of phenolic compounds, based on high grain Fe and Zn content. Differences in the composition of phenolic compounds were determined by HPLC-PDA-MS^n analysis and the dependent mode provided information on the fragmentation patterns of compounds detected.

The initial study indicated differences between samples and revealed compounds that could potentially affect metal ion bioavailability. Based on these findings, a more detailed study was carried out with 57 PMiGAP lines. The untargeted HPLC-PDA-MS^n analysis provided data for the tentative identification and detailed analysis of phenolic compounds of matured raw grain. Additionally, sugar moieties attached to core aglycones were identified by analysing neutral losses (MS^2 events). UV absorption profiles also contributed to the tentative identification of the aglycone. A targeted analysis facilitated the identification and comparison of the peak areas of identified phenolic compounds that may affect Fe/Zn bioavailability. This involved identifying peaks of interest in PDA chromatograms, including apigenin and luteolin glycosides, and measuring areas below the peak as a measure of the relative abundance of the compounds, for comparison between samples. In addition, data were analysed statistically for correlations between abundance of targeted compounds and micronutrient content, as well as for clustering between lines.

MS^2 fragmentation patterns for luteolin and apigenin glycosides identified in grains clearly showed that compounds contained both *O* and *C* linked sugars. For the purpose of structural elucidation, a sample of the line *HHB67 improved*, was acid hydrolysed and analysed to help identify what position the *C*-linked sugar moieties were attached to core aglycones. When a glycoside is subjected to acidic conditions, acetolysis occurs, which breaks the attached *O*-linkage, releasing the *O*-glycoside. The resulting *C*-linked glycosides remain attached. Products can then be compared to iso-vitexin/vitexin for apigenin and iso-orientin/orientin for luteolin. On the basis that the position of the sugars on apigenin can only occur in two positions (C6 or C8); these were compared to iso-vitexin (where the sugar is attached on position C6) and vitexin (sugar on position C8) standards. For luteolin the same position principles occurs for iso-orientin (sugar on position C6)/ orientin (sugar on position C8).

A further study was carried out with the aim of developing a more high through-put method for the analysis of pearl millet phenolic compounds, compared with the previous two studies. The method used for the previous targeted analysis was optimised to accurately identify and compare the abundance of phenolic compounds that may affect the bioavailability of absorbable Fe/Zn, in 185 PMiGAP lines. This involved the isolation of peaks and comparing all data together by analysing the differences in compound relative abundance between each line. Flavonoid glycosides were also quantified using aglycone standard calibration curves, on a molar basis. Furthermore, the correlation between abundance of phenolic compounds and micronutrient content was investigated, as well as the clustering between lines.

5.5 Materials and Methods

5.5.1 Plant Material

See Chapter 2, Table 2.3.

5.5.2 Sample Preparation

34 and 57 × PMiGAP lines

Using an electric powered grinding mill (Retsch Mortar Grinder Mill, 110V/60Hz OY-04181-10), 0.5g of whole pearl millet seed from each genotype was ground to obtain a fine powder, able to pass through a 1mm sieve. The extraction volume was 5mL 70% MeOH with added internal standard (5mg of ethyl ferulate added to 50mL 70% MeOH, this created a 100× stock solution. 2mL of the stock solution was added to to 198mL 70% MeOH to create the working solution). Samples were left to shake for 30 minutes (170 rpm using a KSSOL digital IKA Labortechnik Shaker), then centrifuged at 13,000 rpm for 10 minutes. The supernatant was transferred to clean tubes. The pellets were then washed with 500µL 70% MeOH and centrifuged again at 13,000 rpm for 10 minutes. The second supernatant was added to the first supernatant fraction to maximise recovery of target compounds. The MeOH was then removed using a heated centrifugal evaporator (Jouan RC10.22) set to 70°C, at 13,000 rpm and 500µL distilled water was added to the residual aqueous fraction. A Waters Sep-Pak C18 500mg cartridge (WAT036945) was prepared by passing through 5mL 100% MeOH, followed by 5mL 5% acetic acid. 25% of the sample was then loaded into the cartridge and washed with 3mL water, and then eluted into a

clean vial. The samples were then fully dried using a heated centrifugal evaporator at 13,000 rpm, at 70°C and reconstituted in 70µL 70% MeOH.

Analysis of Hydrolysed and Non-Hydrolysed *HHB67 Improved*

Non-hydrolysed

10g of pearl millet seed was prepared as above and extracted in 100mL 70% MeOH. After centrifigation, the pellet was washed with 5mL 70% MeOH. The sample was fully dried using a heated centrifugal evaporator (Jouan RC10.22) set to 70°C, at 13,000 rpm, then reconstituted in 1mL 70% MeOH plus 9mL distilled water. 1mL at a time was loaded into a Waters Sep-Pak C18 500mg cartridge (WAT036945), prepared as above. The sample was then taken down to 2mL using a heated centrifugal evaporator at 13,000 rpm, set to 70°C.

Hydrolysed

10g of pearl millet seed was prepared as above. 8mL distilled water was added to the 2mL concentrated sample, along with an equal volume of 2M HCL. The sample was boiled using a water bath, set to 90°C for 1hr then neutralised with 1M NaOH, checking the pH regularly until pH3 was reached. The sample was loaded 1mL at a time through a Waters Sep-Pak C18 500mg Cartridge and taken to dryness using a heated centrifugal evaporator at 13,000rpm, set to 70°C and reconstituted in 2mL 70% MeOH.

185 × OP PMiGAP lines

0.1g of pearl millet seed was prepared as above and extracted in 1mL 70% MeOH with added internal standard (5mg of biochanin A added to 50mL 70% MeOH, this created a 100× stock solution. 2mL of the stock solution was added to to 198mL 70% MeOH to create the working solution). Samples were prepared as described above, minus the use of a Waters Sep-Pak C18 Cartridge, then taken to dryness using a heated centrifugal evaporator at 13,000rpm, at 70°C and reconstituted in 70µL 70% MeOH, then diluted with 70% MeOH, 10 fold.

Control and Blank

A control sample was prepared without pearl millet material, as per the method described above. The blank consisted of 1mL working solution.

5.5.3 HPLC Analysis

HPLC-PDA 34 × PMiGAP lines

HPLC analysis was carried out as described in section 5.3.3.

HPLC-PDA-MS^n 8 × PMiGAP lines

Samples were fully dried using a heated centrifugal evaporator at 13,000 rpm, set to 70°C and reconstituted in 50µL 70% MeOH, then diluted with 70% MeOH, 10 fold. HPLC-MS analysis was performed on a Thermo Finnigan LC-MS system (Thermo Electron Corp., Waltham, MA, USA) comprising a Finnigan Surveyor PDA Plus detector and a Finnigan LTQ linear ion trap with ESI source, and the column used was a 3.9mm×100mm i.d., 4µm, C18 Nova-Pak (Waters). The autosampler tray temperature was maintained at 5°C and the column temperature at 30°C. The sample injection volume was 20µL, the detection wavelength was set to 240−400nm, and the flow rate was 1 mL/min, with 100 µL/min going to the mass spectrometer. The mobile phase consisted of water/formic acid (A; 100:0.1, v/v) and MeOH/formic acid (B;100:0.1,v/v). The column was equilibrated with 95% solvent A and the percentage of B increased linearly to 60%, over 65 min. Ionisation parameters were optimised by infusion of chlorogenic acid standard at a constant rate into the LC flow. Mass spectra were acquired in negative and positive ionisation mode with the following interface and MS parameters: sheath gas 30 and auxiliary gas 15 (both arbitrary units), spray voltage -4.0 kV in negative and 4.8 kV in positive ionisation mode, capillary temperature 320 °C, capillary voltage -1.0 and 45 V, respectively, and tube lens voltage -68 and 110 V, respectively

HPLC-PDA-MS^n 57 × PMiGAP Lines for Untargeted Analysis

As described above.

HPLC-MS^n 55 × PMiGAP Lines for Targeted Analysis

A Waters Xevo TQ-S ultra performance liquid chromatography - tandem mass spectrometer (UPLC-MS/MS) system was used to target specific ions in negative ionisation mode. Targeted ions can be seen in Table 5.7.

Table 5.7: Targeted parent and daughter ions in 55 lines.

Compound	t_r (mins)	Parent Ion	Daughter Ions
Apigenin7-glycoside	3.5	431	-
Ethyl Ferulate	6.2	221	134
P-Coumaric Acid	3.4	163	119
Apigenin	5.2	269	179
Luteolin	4.6	285	217
Vitexin	3.0	431	311
Orientin	2.8	447	327
Luteolin7-glycoside	3.1	449	287
Luteolin + Caffeic	1.9	771	609
Dicaffeoyl Spermidine[a]	2.1	468	332
2”-*O*-Hex-*C*-Hex-Apigenin[a]	2.7	593	413
C-*O*-Dihexosyl-Luteolin[a]	2.5	609	489
C-Hex-*C*-Pent-Apigenin[a]	2.5	563	473

[a] Putative identifications.

A 5µL aliquot was injected onto a BEH C18 column (150×2.1mm, 1.7µm, Waters) on a Waters Acquity UPLC with a Xevo Triple Quadrupole Mass spectrometer. The mobile phase consisted of an acetonitrile (0.1% v/v formic acid)/water (0.1% v/v formic acid) gradient (10:90 to 60:40 over 6 minutes; to 98:2 over 0.5 minutes; hold for 2 minutes; to 10:90 over 0.5 minutes; hold for 1 minute) at a flow rate of 0.2 mL/min and a column temperature of 40 °C. Data were analysed using the Waters MassLynx 4.1 software. Quantities of each metabolite were established by monitoring specific transitions in multiple reaction monitoring (MRM) mode (Table 5.7) and normalised to the internal standard (ethyl ferulate, ethyl 4-hydroxy-3-methoxycinnamate).

Analysis of Hydrolysed and Non-Hydrolysed *HHB67 Improved*

UV-HPLC

A 5µL aliquot was injected onto a Kinetex 5u Biphenyl 100A column (Phenomenex, 150 × 4.6 mm) with a ZORBAX Eclipse XDB guard column (1.0 mm × 17 mm). The mobile phase consisted of a MeOH: acetic acid (100:0.2 v/v) gradient (25% to 50% over 20 minutes; hold for 4 minutes; to 95% over 1 minute; hold for 4 minutes) at a flow rate of 1 mL/min and a column temperature of 35 °C. Signals were recorded at 280 nm for putative flavonoid diglycosides.

Fraction collecting

Putative flavonoid diglycosides were collected using the fraction collecting function. Fractions were collected on a time basis, 6.6-7.2 minutes (Peak 1) and 8.5-9.5 minutes (Peak 2). Collected fractions were then fully dried under nitrogen and reconstituted in 50µL 70% MeOH, for further analysis by MS. The purpose of this was to verify compounds in collected fractions. Fractions were subsequently analysed by HPLC-MS^n, as described in above.

MS/MS

As described above.

HPLC-MS^n Targeted Analysis of 185 × OP PMiGAP Lines

A Waters Xevo TQ-S UPLC-MS/MS system was used to target specific ions in negative ionisation mode. Targeted ions can be seen in Table 5.8.

Table 5.8: Targeted parent and daughter ions in 185 lines.

Compound	t_r (mins)	Parent Ion	Daughter Ions
Apigenin7-glycoside	3.5	431	-
Biochanin A	6.98	283	268
P-Coumaric Acid	3.4	163	119
Apigenin	5.2	269	179
Luteolin	4.6	285	217
Vitexin	3.0	431	311
Orientin	2.8	447	327
Luteolin7-glycoside	3.1	449	287
Luteolin + Caffeic	1.9	771	609
Dicaffeoyl Spermidine[a]	2.1	468	332
2”-*O*-Hex-*C*-Hex-Apigenin[a]	2.7	593	413
C-*O*-Dihexosyl-Luteolin[a]	2.5	609	489
C-Hex-*C*-Pent-Apigenin[a]	2.5	563	473

[a] Putative identifications.

A 10μL aliquot was injected onto a BEH C18 column, as described above. Amounts of each metabolite were established by monitoring specific transitions in MRM mode (Table 5.8) and normalised to the internal standard (biochanin A, 5,7-dihydroxy-4′-methoxyisoflavone).

5.5.4 Standards

Standards: Orientin (O9765-1MG), iso-Orientin (I1536-1MG), vitexin (49513-10MG-F), iso-vitexin (17804-1MG), apigenin (A3145-5MG), luteolin (L9283-10MG), luteolin7-*O*-β-D-glucoside (449968), apigenin7-glucoside (44692), ethyl 4-hydroxy-3-methoxycinnamate (320617-1G), *p*-coumaric acid (C9008-1G), biochanin A, 5,7-dihydroxy-4′-methoxyisoflavone (D2016-100MG) were obtained from Sigma-Aldrich (Gillingham, UK).

5.5.5 Statistical Analysis

Data were analysed under the Qlucore Omics Explorer, v3.2 environment, using hierarchical cluster analysis and principal component analysis. $P<0.05$ filters were applied where appropriate.

5.6 Results and Discussion

5.6.1 Analysis of 34 × PMiGAP lines

HPLC-PDA

Initially, the separation of soluble phenolic compounds in 34 PMiGAP lines were detected by HPLC with PDA at 280.0nm, to identify compounds that could potentially affect metal ion bioavailability and any differences between lines. Results indicated that the majority of samples presented similar composition and a typical chromatogram is shown in Figure 5.10. This is consistent with the strong tendency for plants of the same species to produce similar compounds (Cuyckens & Claeys, 2004). Compounds were tentatively identified based on spectral scans, 240-400nm (Table 5.9).

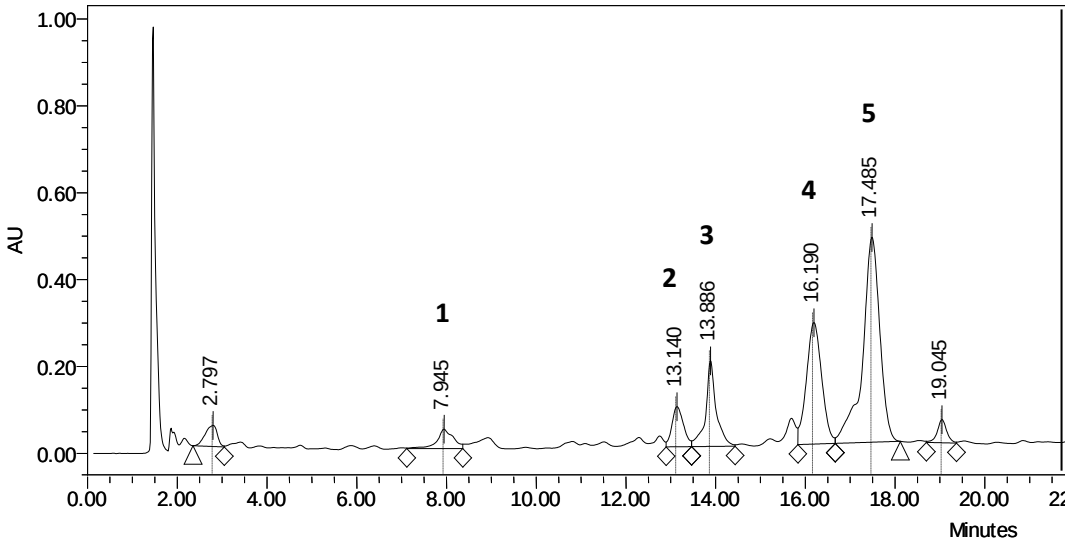


Figure 5.10: HPLC chromatogram showing a typical profile of soluble phenolic compounds in extracts of 34 PMiGAP lines detected by PDA at 280 nm. Compounds were separated with a Waters HPLC system on a 8 mm×100 mm i.d. 4µm, C18 Nova-Pak radial compression column. Peak numbering corresponds with Table 5.9.

Table 5.9: Tentative identifications of phenolic compounds in pearl millet based on UV spectra information

Peak No.	t_r (mins)	UV Max	Aglycone tentative ID
1	7.9	288	-*
2	13.1	310	*P*-Coumaric Acid derivative
3	13.9	319	-*
4	16.2	256, 349	Luteolin derivative
5	17.5	268, 337	Apigenin derivative

*Could not be identified

8 lines were selected based on a small amount of variation detected within the population, as per the quantity (absorbance units as per peak area) of compounds detected (Figure 5.11).

HPLC-PDA-MS^n Untargeted Analysis

In the 8 lines where variation was present, a more detailed analysis was carried out by HPLC-PDA-MS^n. Figure 5.12 shows an example of the data obtained by this method. The dependent mode provided information on the fragmentation patterns of compounds (Figure 5.11). Additionally, sugar moieties attached to aglycones were identified by analysis of neutral losses (MS^2 events). Peaks of interest were identified from the full scan PDA chromatograms (Figure 5.12, A). Full MS chromatograms in both negative (Figure 5.12, B) and positive ionisation modes (Figure 5.12, C) were obtained and MS peaks were matched with UV peaks of interest.

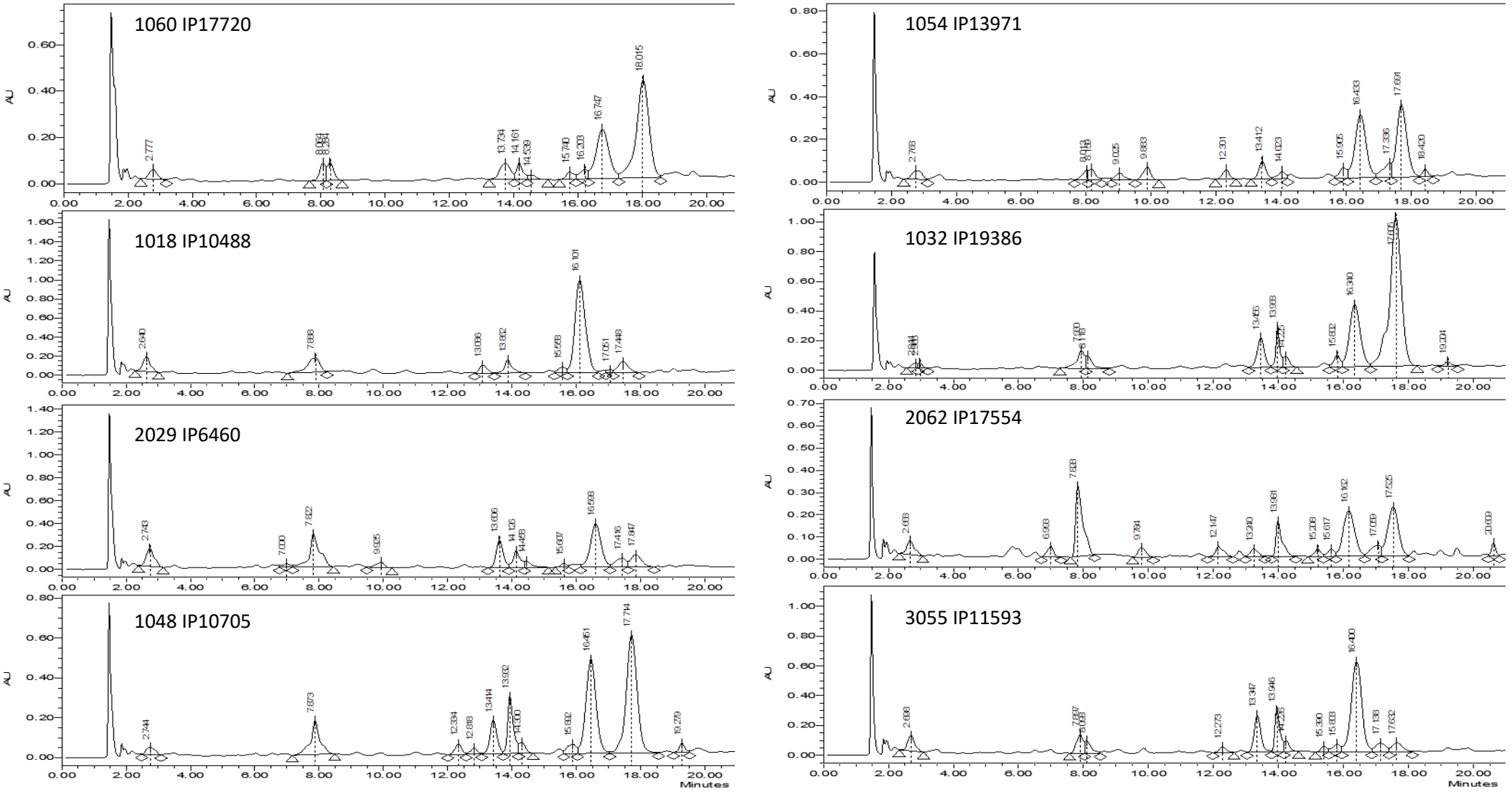


Figure 5.11: Differences in peak area as a measure of relative abundance seen for 8 PMiGAP genotypes.

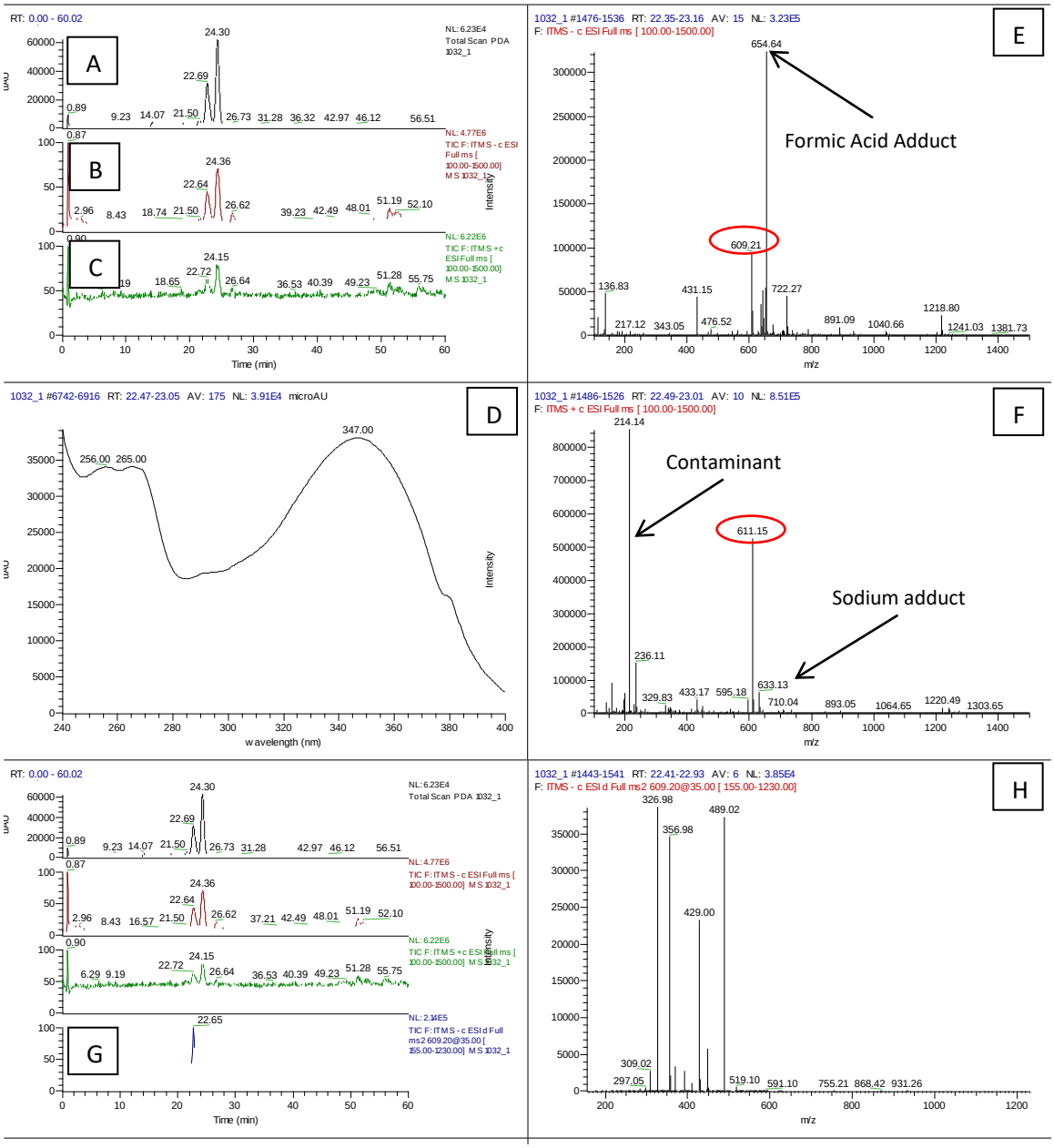


Figure 5.12: HPLC-PDA-MS chromatograms and spectra for an extract of a selected genotype. Individual figures show: A total PDA scan at 280-400nm (A); full MS scan in negative mode (B); full MS scan in positive mode (C); UV spectrum of the 22.7 min peak (D); full MS negative mode spectrum of the 22.7 min peak (E); full MS positive mode spectrum of the 22.7 min peak (F); MS^2 chromatogram for 609 negative ion (G); MS^2 spectrum for the 609 negative ion at 22.7 min (H).

For peaks of interest, the UV spectral scans (240-400nm) were recorded (Figure 5.12, D). The observed profile and UV maxima (256, 265 and 347nm) are consistent with a luteolin glycoside. Parent ions were identified by analysing the full MS data in both negative and positive ionisation modes (Figure 5.12, E and F). In this example, E and F show the nominal mass of parent ions present at 22.69 minutes.

From the data, it was possible to match the parent ions with the phenolic compound eluting at the time. The most abundant matching ions in negative ionisation mode have masses of 654.64 and 609.21 units, respectively. The former is a formic acid adduct (Figure 5.13) $[M–H + 46]^-$ of the latter $[M-H]^-$ where M, the nominal mass, is 610 units. This is supported by the dominant ion observed in positive mode, with a mass of 611.15 $[M+1]^+$.

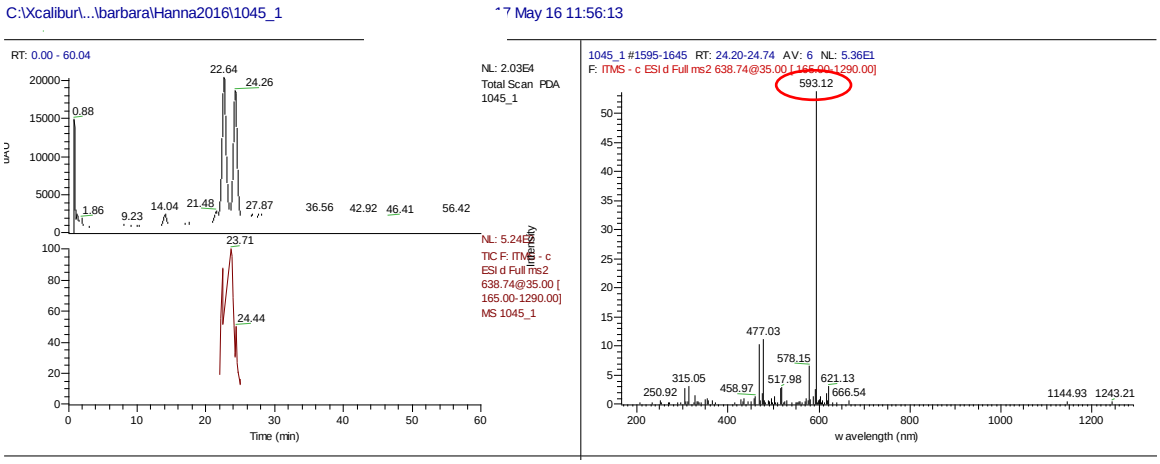


Figure 5.13: Formic artefact.

Figure 5.14 shows a typical UV profile of the soluble phenolics extracted from pearl millet grains of the 8 selected lines following analysis with the Thermo-Finnegan LC-MS system. The profile shows some differences with profiles obtained with the Waters HPLC system shown in Figure 5.10. This is probably due to differences between columns and gradients used within the two systems. Analysis by HPLC-MS with PDA detection revealed approximately 4 compounds that could be easily distinguished (Table 5.10). Several aglycones and their sugar moieties were readily identified by comparison of retention time and fragmentation pattern with known standards. Others were tentatively identified by the similarity of UV absorbance characteristics and comparison with compounds reported in the literature.

Four main compounds were identified (Table 5.10), with a luteolin glycoside being the most prominent compound detected in the majority of cases.

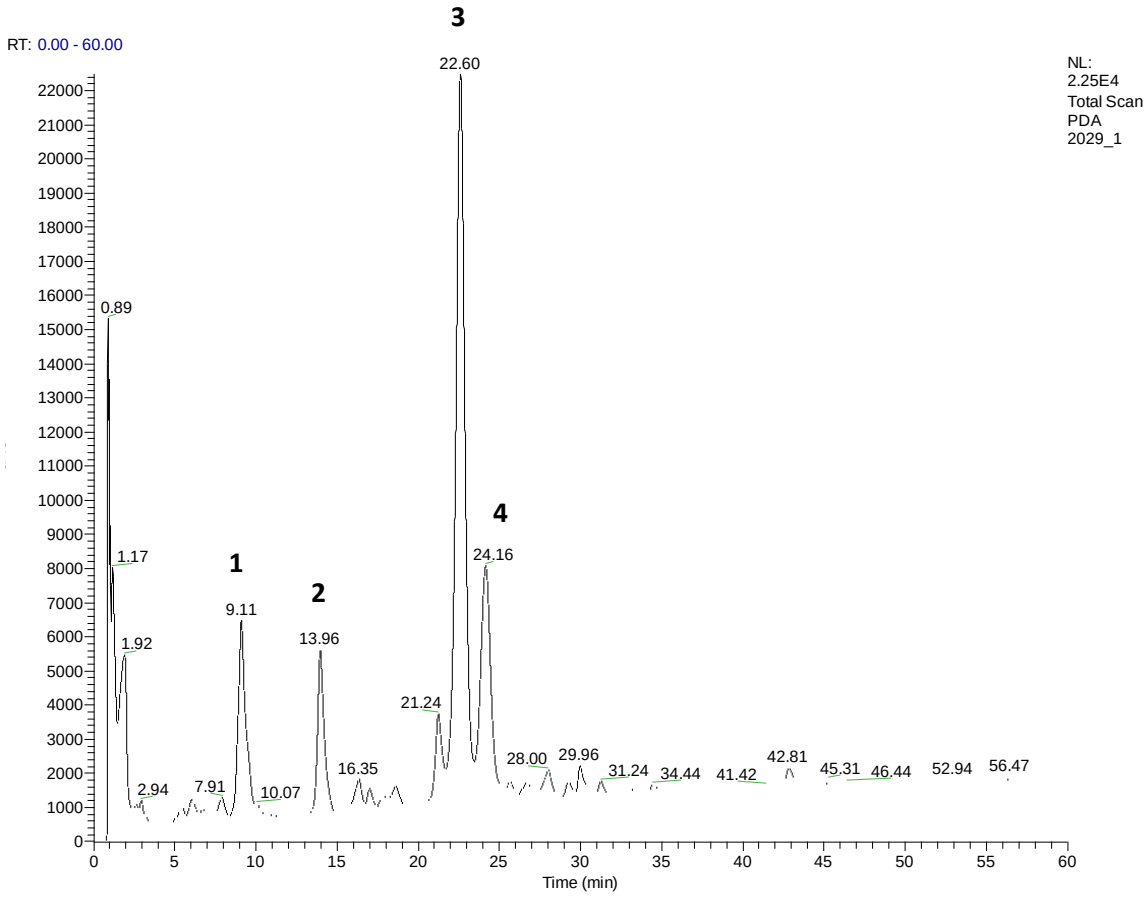


Figure 5.14: HPLC-MS chromatogram showing a typical profile of soluble phenolic compounds in extracts of pearl millet detected by PDA at 280 nm. Compounds were separated with a Thermo Finnigan LC-MS system on a 3.9 mm × 100 mm i.d. 4 µm, C18 Nova-Pak column. Peak numbering corresponds with Table 5.10.

Table 5.10: UV maxima, full MS ions and MS^2 events for compounds detected in extracts of selected PMiGAP genotypes.

Peak No.	t_r (mins)	UV Max	Nominal Mass, M_r Da	Parent Ions	MS^2 Fragments (in order of intensity)	Neutral loss ($MS - MS^2$)	Tentative identification
1	9.11	289.0 301.0	234	(-ve) Not detected (+ve) 235.00	(-ve) Not detected (+ve) 218 88	(-ve) Not detected (+ve) 17 147	Coumaroyl Putrescine
2	13.96	290.0 303.0	164	163.02 (-ve) (+ve) Not detected	(-ve) 119 (+ve) Not detected	(-ve) 44 (+ve) Not detected	*P*-Coumaric Acid
3	22.60	265.0 347.0	610	609.21 (-ve) 611.15(+ve)	(-ve) 327 489 357	(-ve) 282 120 252	Luteolin glycoside

4	24.16	268.0 336.0	594	593.23 (-ve) 595.18 (+ve)	429 (+ve) 449 329 491 413 (-ve) *Rerun of ion 593 (1:1 dilution) – 413 293 473 (+ve) 433 415 313 475 337	180 (+ve) 162 282 120 198 (-ve) 180 300 120 (+ve) 162 180 282 120 258	Apigenin glycoside

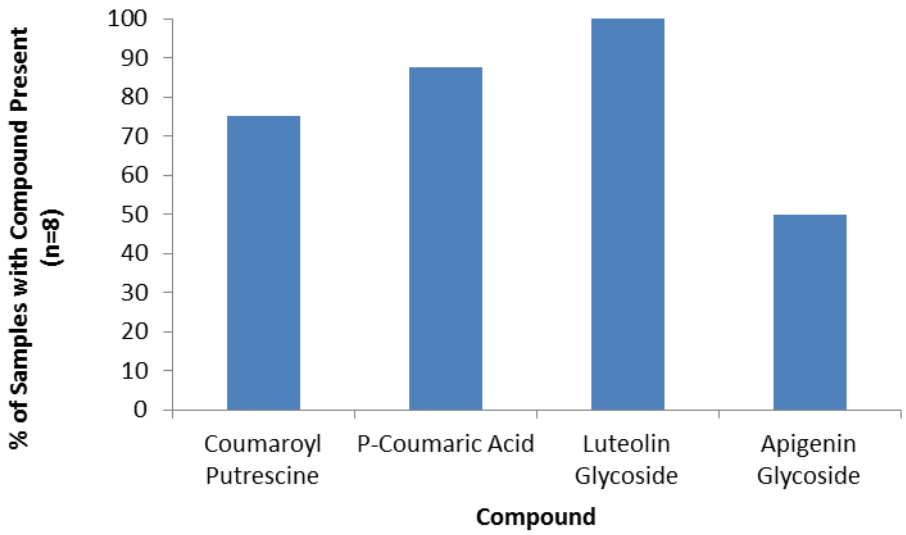


Figure 5.15: Frequency (%) of compounds detected within 8 × PMiGAP lines.

*

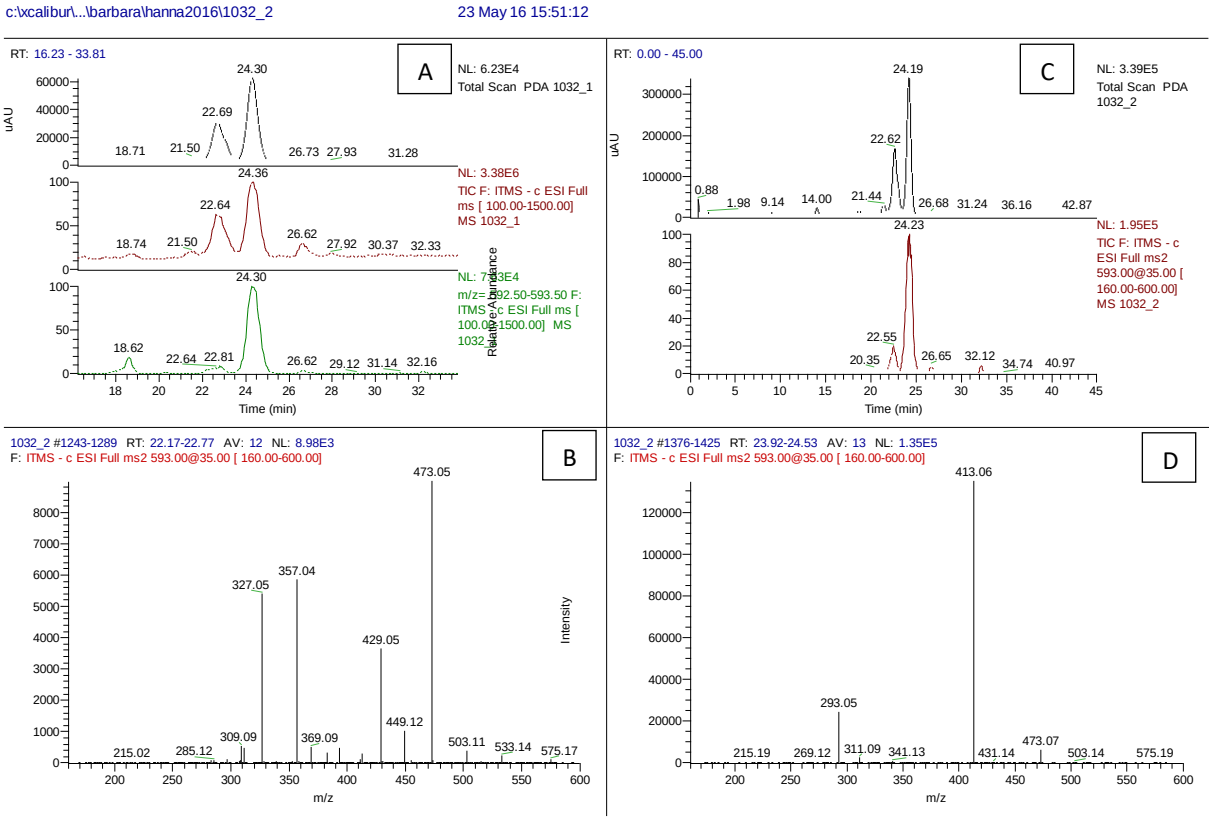


Figure 5.16: Detailed analysis of negative ion 593 in genotype 1032 (1:1 dilution-targeted). Individual figures show: 1:10 dilution, PDA, full MS negative mode (red), full MS positive mode (green) (A); 1:1 dilution MS^2 of targeted ion 593 at 22.2 minutes (B); 1:1 dilution, PDA, full MS negative mode (red) (C); 1:1 dilution MS^2 retention time 23.9 minutes (D).

Elucidation of compound structure

Two samples, where the parent ions were known (1032, IP19386 and 3055, IP16638) underwent further fragmentation by targeting specific ions from the MS^2 to help elucidate structure. Where compounds could not be identified using standards, their structure was elucidated based on UV profiles and co-eluting parent ions. UV spectra gave an indication of the phenol class and fragments observed and $^{+/-}MS^2$ events allowed confirmation of the core aglycone. Once this was established, the nature of sugar linkages in glycosylated compounds was determined. Two types were observed: *O*-linked glycosylation (the attachment of a sugar molecule to an *O* atom on the phenol ring) and *C*-glycosylation (the attachment of a sugar molecule to the phenol via a carbon-carbon linkage) (Ferreres *et al.*, 2007). The sugar fragment that is lost from the core can be determined by the fragmentation patterns observed in mass spectra in the case of hexoses, deoxyhexoses and pentoses (Vukics &

Guttman, 2010). Loss of 180/162 or 150/132 units indicates the loss of an *O*-linked hexose or pentose sugar respectively, while losses of 120/90 or 90/60 indicate the loss of a *C*-linked hexose or pentose sugar, respectively. A fragment corresponding to the aglycone +41/71 units also indicates a *C*-linked sugar. According to Ferreres *et al.,* (2007), three groups can be determined according to fragmentation pattern: Y_0^- (loss of 162), Z_1^- (loss of 180) and $X_{0/1}^-$ (loss of 120) (Figure 5.17). Cleavage at $X_{0/3}^-$ will result in a loss of 90 units and this is also observed in some cases. Z^- is the most common fragmentation pattern of *O*-glycosyl-*C*-glycosyl flavones (Ferreres *et al.,* 2007). Table 5.11 shows the diagnostic mass losses observed from mass spectral data for pearl millet compounds. This leads to determination of fragmentation pattern X_0^- Y_0^- and Z_1^-, respectively.

It is also possible to distinguish between *O*-glycosylation at the 2' and 6' positions on flavonoid core molecules (Ferreres *et al.,* 2007).

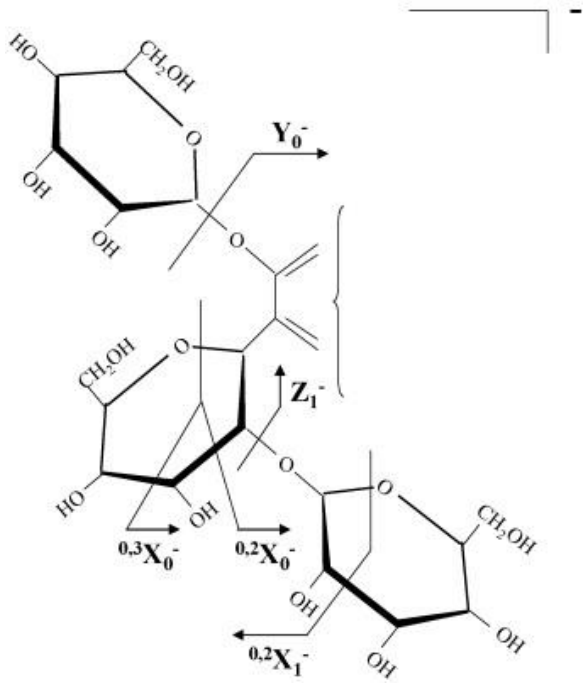


Figure 5.17: General fragmentation of *O*-glycosyl-*C*-glycosyl flavones (Ferreres *et al,* 2007).

Table 5.11: Aglycones (Ag) and diagnostic mass losses of main compounds detected in PMiGAP lines, full MS shows parent ion in negative ionisation mode: $[M-H]^-$ and MS^2 data show main fragment ions.

Aglycone	(Ag) Molecular Weight	Full MS $[M-H]^-$	MS^2 ions and mass losses
3 Luteolin	286	$[M-H]^-$ 609	$[M-H]^-$ -120 -180 Ag+41 Ag+71 489 429 327 357
4 Apigenin	270	$[M-H]^-$ 593	$[M-H]^-$ -120 -180 Ag+41-18 473 413 293
2 *P*-Coumaric Acid	164	$[M-H]^-$ 163	$[M-H]^-$ -44 -28 119 135

Z_1^-(2" –*O*-Glycosyl-*C*-glycosyl derivatives) -180(-162-18,hexose).

Compound 3: UV maxima (265, 347 nm) and MS fragments indicate that the aglycone is luteolin (mass 286 units). The fragmentation patterns from the MS^2 events in negative ionisation mode indicate a *C*-linkage to the aglycone with a hexose sugar dimer (-120, -180). The presence of $[Ag+41]^-$ and $[Ag+71]^-$ are indicative of mono-*C*-glycosyl flavones (a single *C*-*C* sugar linkage). In light of this, the compound has been tentatively identified as *C*-*O*-dihexosyl-luteolin.

Compound 4: UV maxima (268, 336 nm) and MS fragments indicate that the aglycone is apigenin (mass 270 units). The fragmentation patterns from the MS^2 events in negative ionisation mode indicate a *C*-linkage to the aglycone with a hexose sugar dimer (-120, -180). The fragmentation patterns indicate one *C*-linked glycoside plus another sugar moiety that is attached by a sugar linkage to the first sugar moiety and that the *O*-glycosylation is at the 2" position. In light of this, the compound has been tentatively identified as 2"-*O*-hex-*C*-hex-apigenin

Compound 2: This compound has been identified as *p*-coumaric acid (mass 164 units), the fragmentation patterns from the MS^2 events in negative ionisation mode indicate a loss of CO_2 (-44) and a loss of CO (-28). A positive identification was later provided by comparison with a standard.

5.6.2 HPLC-PDA-MS^n Untargeted Analysis of 57 × PMiGAP lines

The previous study revealed compounds that could potentially affect metal ion bioavailability in all PMiGAP lines tested. Based on these findings, a more detailed analysis was carried out with 57 PMiGAP lines, for the tentative identification and analysis of phenolic compounds present in pearl millet grain. Figure 5.18 shows 4 UV chromatograms representing different profiles of soluble phenolics extracted from 57 PMiGAP lines. Analysis by HPLC-MS with PDA detection revealed approximately 16 compounds that could be easily distinguished (Table 5.12). Several compounds were readily identified by comparison of retention times and fragmentation patterns with known standards. Others were tentatively identified by the similarity of UV absorbance characteristics and comparison with compounds reported in the literature. Compounds identified by HPLC-MS include hydroxycinnamic acids, phenolamides and flavones. Relatively abundant flavonoids, previously reported by Chandrasekara & Shahidi, (2011) and Shekhar *et al.,* (2016) in pearl millet were identified in extracts.

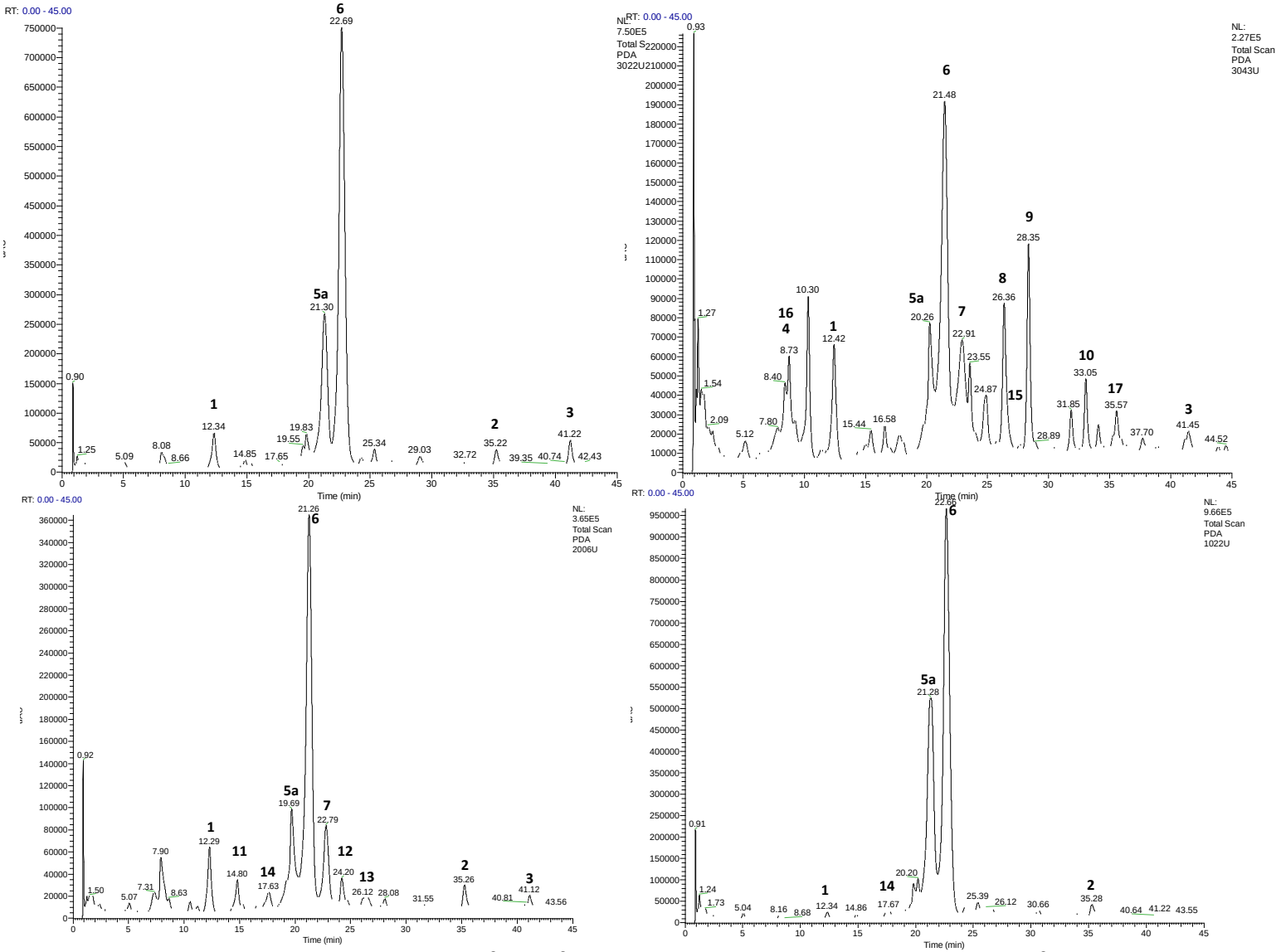


Figure 5.18: 4 UV chromatograms showing distinct profiles of soluble phenolic compounds in extracts of pearl millet detected by PDA at 280nm. Peak numbering corresponds with Table 5.12.

Table 5.12: UV maxima, full MS ions and MS^2 events for compounds detected in extracts of samples selected from 57 PMiGAP genotypes.

Peak No.	t_r (mins)	UV Max	Nominal Mass, M_r Da	Parent Ions	MS^2 Fragments (in order of intensity)	Neutral loss ($MS - MS^2$)	Tentative identification
1	12.34	309.0	164	163.20 (-ve) 165.27 (+ve)	(-ve) 119 (+ve) 147	(-ve) 44 (+ve) 18	*P*-Coumaric Acid
2	35.26	299.0 322.0	222	(-ve) Not Detected 223.20 (+ve)	-	-	Ethyl 4-hydroxy-3-methoxycinnamate (98%) (IS)
3	41.22	268.0 342.0	330	329.28 (-ve) 331.25 (+ve)	(-ve) 314 (+ve) 315 270 287	(-ve) 15 (+ve) 16 61 44	Tricin
4	8.73	322.0	354	353.22 (-ve) 355.08 (+ve)	(-ve) 191 (+ve) 163	(-ve) 162 (+ve) 192	Chlorogenic Acid (5-Caffeoyl quinate)
5a	20.26	295.0 319.0	469	468.40 (-ve) 470.46 (+ve)	(-ve) 332 (+ve) 220 308 453 291 163 145	(-ve) 136 (+ve) 250 162 17 179 307 325	Dicaffeoyl Spermidine
5b		Not Detected*	564	563.15 (-ve) 565.08 (+ve)	(-ve) 473 443 503 353 383 545 (+ve) 547 529 427 511	(-ve) 90 120 60 210 180 18 (+ve) 18 36 138 54	Apigenin (*C*-Hex-*C*-Pent-Apigenin)
6	21.48	267.0 348.0	610	609.44 (-ve) 611.27 (+ve)	(-ve) 489 429 327 357 449 (+ve)	(-ve) 120 180 282 252 160 (+ve)	Luteolin (*C*-*O*-Dihexosyl-Luteolin)

					449 329 431 287 491 (-ve) Not Detected (+ve)	162 282 180 324 120	
7	22.91	267.0 337.0	594	593.32 (-ve) 595.38 (+ve)	433 449 475 415 313 271 (-ve)	162 146 120 180 282 324 (-ve)	Apigenin (2”-*O*-Hex-*C*-Hex-Apigenin)
8	26.36	292.0	283	282.39 (-ve) 284.32 (+ve)	119 145 167 162 134 (+ve)	163 137 115 120 148 (+ve)	Coumaroyl Tyramine
9	28.35	sh 294.0 318.0	313	312.38 (-ve) 314.38 (+ve)	147 (-ve) 178 297 135 (+ve)	137 (-ve) 134 15 177 (+ve)	Feruloyl Tyramine
10	33.05	293.0 309.0	410	409.43 (-ve) 411.33 (+ve)	177 145 (-ve) 289 259 135 (+ve) 265 235	137 169 (-ve) 120 150 274 (+ve) 146 176	Feruloyl Coumaroyl Putrescine[1]
11	14.80	310.0	238	237.20 (-ve) 239.12 (+ve)	291 261 177 147 (-ve) 119 163 145 (+ve)	120 150 234 264 (-ve) 118 174 192 (+ve)	Coumaroyl glycerate
12	24.20	sh 296.0 320.0	483	482.44 (-ve) 484.37 (+ve)	147 (-ve) 346 332 (+ve) 234 322 147 220 177	92 (-ve) 136 150 (+ve) 250 162 17 264 307	Caffeoyl Feruloyl Spermidine

13	26.12	296.0 320.0	624	623.29 (-ve) 625.22 (+ve)	(-ve) 443 (+ve) 463 445 343 301 305	(-ve) 180 (+ve) 162 180 282 324 320	Chrysoeriol-*C,O*-Diglycoside
14	17.67	270.0 333.0	772	771.32 (-ve) 773.26 (+ve)	(-ve) 609 (+ve) 611 449 491	(-ve) 162 (+ve) 162 324 282	Luteolin + Caffeic
15	27.64	288.0 295.0	467	466.00 (-ve) 468.00 (+ve)	(-ve) Not Detected (+ve) 147 177 204 234 292	 321 291 264 234 176	*P*-Coumaroyl Feruoyl Spermidine
16 (Co-eluted with 4)	8.73	299.0 308.0	410	409.00 (-ve) 411.00 (+ve)	(-ve) 177 265 (+ve) 89 264	(-ve) 232 144 (+ve) 322 147	Feruloyl Coumaroyl Putrescine[2]
17	35.57	298.0 322.0	424	423.00 (-ve) 425.00 (+ve)	(-ve) Not Detected (+ve) 177 147 279 249	 (+ve) 248 278 146 176	Coumaroyl Feruoyl Cadaverine

*Not detected due to coeluting compounds (5a and 5b).

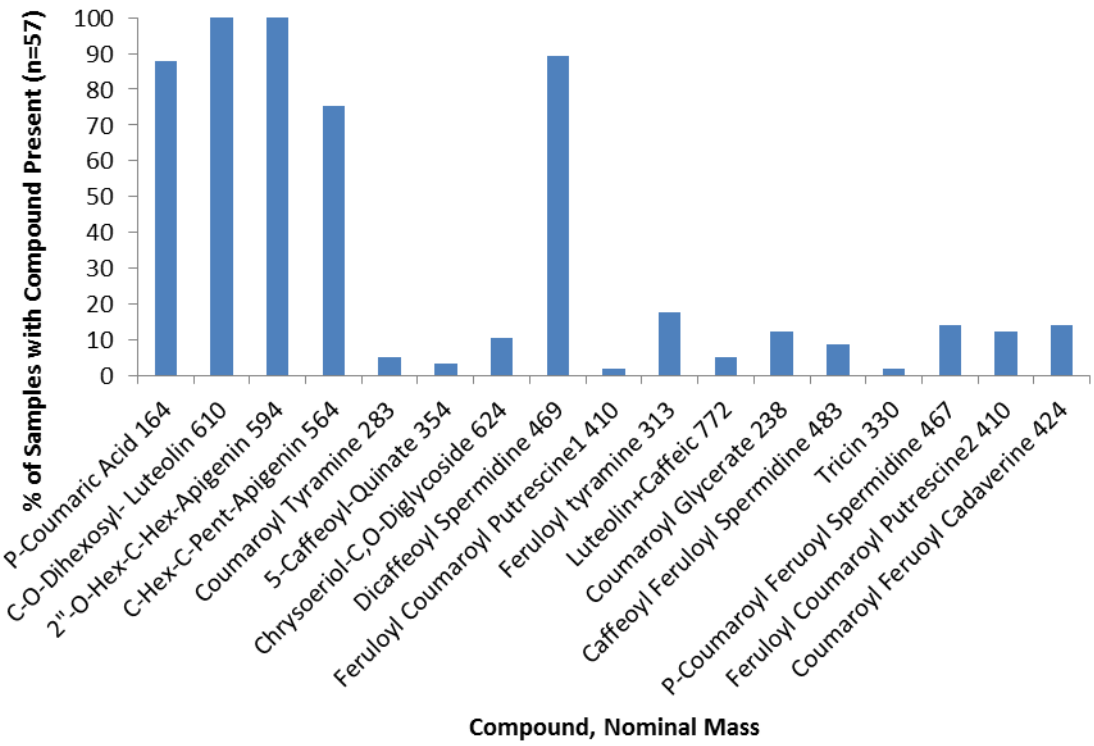


Figure 5.19: Frequency (%) of compounds detected within 57 × PMiGAP lines (Names as in table).

Hydroxycinnamic acids

A number of hydroxycinnamic acids, previously identified in pearl millet were identified in this study (Chandrasekara & Shahidi, 2011). For example, compound 1 with M_r 164 Da and t_r 12.34 mins showed characteristics consistent with *p*-coumaric acid with a UV-maxima at 309.0. The parent ion at *m/z* 163 in negative ionisation mode produced a base peak in the MS^2 at *m/z* 119 [coumaric acid – CO_2]. Additionally, the parent ion at *m/z* 165 in positive ionisation mode produced a base peak in the MS^2 at *m/z* 47 [coumaric acid – H_2O].

Although many medicinal characteristics are associated with this group, such as antioxidant, anti-carcinogenic and anti-inflammatory activities, very few studies have addressed the bioavailability of hydroxycinnamic acids and their effect on micronutrient uptake. A study by Manach *et al.*, (2005) suggests that when ingested in the free form, hydroxycinnamic acids are rapidly absorbed from the small intestine. However, these compounds are naturally esterified in plant products, and this impairs their absorption (Manach *et al.*, 2005).

Compound 4 with M_r 354 Da and t_r 8.73 mins showed characteristics consistent with chlorogenic acid (5-caffeoyl-quinate), with a UV maxima at 322.0. The parent ion at *m/z* 353 in negative ionisation mode produced a base peak in the MS^2 at *m/z* 191 [5-caffeoyl-quinate]. Chlorogenic acids are some of the most abundant phenolics in the human diet and have been reported in a wide variety of plants, including pearl millet (Meng *et al.*, 2013). Despite extensive research into the bio-pharmacological properties attributed to these compounds (antibacterial, antioxidant, and anticarcinogenic activities), little is known about their bioavailability in humans (Monteiro *et al.,* 2007), or their interactions with other micronutrients. Health benefits attributed to chlorogenic acid include anti-diabetic properties. For example, a study by McCarty (2005) found that when chlorogenic acid was used as an insulin sensitiser, it acted in a similar way to the drug Metformin, which is used to control blood sugar levels in diabetic patients. This has been confirmed in other studies. For example, Bassoli *et al.,* (2008) researched the effect of chlorogenic acid on hepatic glucose output, blood glucose levels, and glucose tolerance and found that chlorogenic acid promoted a reduction in the plasma glucose peak in the oral glucose tolerance test, most likely by reducing intestinal glucose absorption. This finding indicates a possible role for chlorogenic acid as a glycaemic index lowering agent for reducing the risk of developing type 2 diabetes (Bassoli *et al*., 2008).

Compounds 8 (M_r 283 Da, t_r 26.36 mins and UV maximum 292.0) and 11 (M_r 238 Da, t_r 14.80 mins and UV maximum 310.0) have been tentatively identified as coumaroyl conjugates. In the case of compound 8, this is due to the parent ion at *m/z* 282 in negative ionisation mode producing a base peak in the MS^2 at *m/z* 145 and 119 ($[\text{coumaric} - H_2O]^-$ and $[\text{coumaric} - CO_2]^-$ respectively) and the parent ion at *m/z* 284 in positive ionisation mode producing a base peak in the MS^2 at *m/z* 147 ($[\text{coumaric} - H_2O]^+$). This is also the case for compound 11, with a UV maximum of 309 nm (characteristic of coumaric) and where the parent ion at *m/z* 237 in negative ionisation mode produced a base peak in the MS^2 at *m/z* 163 $[\text{coumaric}]^-$, 145 and 119. Additionally, the parent ion at *m/z* 239 in positive ionisation mode produced a base peak in the MS^2 at *m/z* 147. Compound 8 could be further characterised as coumaroyl tyramine, due to the parent ion in negative and positive ionisation modes producing peaks resulting from the neutral loss of 137 Da (*m/z* 145 and 147 respectively), consistent with the presence of a tyramine moiety. Compound 11

could also be further characterised as coumaroyl glycerate. This is supported by the parent ion at *m/z* 237 in positive ionisation mode producing a peak in the MS^2 resulting from the loss of 92 Da (*m/z* 147).

Compound 9 (M_r 313 Da, t_r 28.35 mins and UV maxima 294.0sh, 318.0) has been tentatively identified as feruloyl tryamine due to the nominal mass of 313 M_r Da. Additionally, the parent ion in negative ionisation mode produced a peak in the MS^2 at *m/z* 178, indicative of [ferulic acid - CH_3]$^-$. Furthermore, the parent ion in positive ionisation mode produced a peak in the MS^2 at *m/z* 177, which is indicative of [ferulic - H_2O]$^+$. The neutral loss of 137 can also indicate tyramine.

Compound 12 (M_r 483 Da, t_r 24.20 mins and UV maxima 296.0sh, 320.0) has been tentatively identified as caffeic ferulic spermidine. This was concluded due to the UV data being consistent with caffeic acid (318.0 and 320.0, respectively) and also due to the parent ion at *m/z* 484 in positive ionisation mode producing a base peak in the MS^2 at *m/z* 177, indicative of ferulic acid. Additionally the peak at *m/z* 322 is indicative of a ferulic acid unit (176+H^+) plus spermidine (molecular weight 145units).

Flavones

Compound 3 with M_r 330 Da and t_r 41.22 mins showed characteristics consistent with tricin. The parent ion at *m/z* 329 in negative ionisation mode produced a base peak in the MS^2 at *m/z* 314 [tricin – methyl group]$^-$. However, the UV maxima of 268.0, 342.0 was not indicative of tricin and may be due to coeluting compounds. This requires further targeted analysis of the MS^2 fragments.

Compound 6 with M_r 610 Da and t_r 21.48 mins showed characteristics consistent with *C*-*O*-dihexosyl-luteolin with a UV maxima at 267.0, 348.0. The parent ion at *m/z* 609 in negative ionisation mode produced dominant base peaks in the MS^2 at *m/z* 489 and 429, thus giving neutral losses of 120 and 180, respectively, indicative of *C*-linkage with a hexose sugar dimer. The parent ion at *m/z* 611 in positive ionisation mode produced a base peak in the MS^2 at 449, which produced a neutral loss of 162, consistent with the loss of the terminal hexose in an attached sugar dimer.

Compound 5b with M_r 564 Da and t_r 20.26 mins showed characteristics consistent with *C*-hex-*C*-pent-apigenin. The UV maxima could not be detected due to coelution with compound 5a. The parent ion at *m/z* 563 in negative ionisation mode produced peaks in the MS^2 at *m/z* 353 [apigenin+83] and *m/z* 383 [apigenin+113]. The *m/z* 353 is indicative of 2 *C*-linked sugars attached to apigenin, whilst *m/z* 383 indicates one *C*-hexose sugar and one *C*-pentose sugar.

Compound 7 with M_r 594 Da and t_r 22.91 mins showed characteristics consistent with 2"-*O*-hex-*C*-hex-apigenin, with a UV maxima at 267.0, 337.0. Although peaks from the parent ion in negative ionisation mode could not be detected, the fragmentation patterns from the MS^2 events in positive ionisation mode (*m/z* 595) indicate a *C*-link with a hexose sugar dimer (due to losses of -120, -180). The fragmentation patterns indicate 1 *C*-linked glycoside plus another sugar moiety that is attached to it via a sugar linkage at the 2"position. The *O*-link is indicated by the parent ion at *m/z* 595 in positive ionisation mode producing a base peak in the MS^2 at *m/z* 433, which is produced by a neutral loss of 162 (hexose).

Compound 13 with M_r 624 Da and t_r 26.12 mins showed characteristics consistent with chrysoeriol-*C,O*-diglycoside, with a UV maxima at 296.0, 320.0. The parent ion at *m/z* 623 in negative ionisation mode produced a single base peak in the MS^2 at *m/z* 443, which indicated a loss of 180. This is indicative of the loss of a terminal hexose on the sugar chain of two or more. However, the parent ion at *m/z* 625 in positive ionisation mode produced a base peak in the MS^2 at *m/z* 301, which indicated a loss of 324. This could be interpreted as two *O*-linked hexose sugars.

Compound 14 with M_r 772 Da and t_r 17.67 mins showed characteristics consistent with *C*-*O*-dihexosyl-luteolin (described in compound 6) plus a caffeic acid unit attached. The parent ion at *m/z* 771 in negative ionisation mode produced a single base peak in the MS^2 at *m/z* 609 (which matches the molecular weight of *C*-*O*-dihexosyl-luteolin), this also indicated a loss of 162, which represents the loss of a caffeic acid unit.

Phenolamides

Phenolamides comprise a diverse class of secondary metabolites that are found ubiquitously in plants and are known to play important roles in a wide range of biological processes, including plant development and defense (Dong *et al.*, 2014).

Compound 5a with M_r 469 Da and t_r 20.26 mins is present in 90% of samples (Figure 5.13) and showed characteristics consistent with dicaffeoyl-spermidine. The parent ion at *m/z* 470 in positive ionisation mode produced a base peak in the MS^2 at *m/z* 453, which produced a neutral loss of 17, indicating a loss of NH_3 (amine group). Additionally, the peak at *m/z* 308, which produced a neutral loss of 162, indicates the loss of a caffeic unit (amine group) and the peak at *m/z* 145 is indicative of spermidine, as it matches the molecular weight.

Compound 15 with M_r 467 Da and t_r 27.64 mins showed characteristics consistent with *p*-coumaroyl feruloyl spermidine, with a UV maxima at 288.0, 295.0. The parent ion at *m/z* 468 in positive ionisation mode produced peaks in the MS^2 at *m/z* 147, 177, 204, 234 and 292. All of these are consistent with the findings of Dong *et al.*, 2014.

Compound 16 (coeluted with compound 4), with M_r 410 Da and t_r 8.73 mins showed characteristics consistent with feruloyl coumaroyl putrescine, with a UV maxima at 299.0, 308.0. The parent ion at *m/z* 411 in positive ionisation mode produced peaks in the MS^2 at *m/z* 264, indicating the loss of a coumaric acid unit (147) and *m/z* 89, which matched the size of putrescine. This compound could be further verified with putrescine and coumaric acid standards. Additionally, the parent ion at *m/z* 409 in negative ionisation mode produced base peaks in the MS^2 at *m/z* 177 (indicative of a ferulic unit) and 265 (indicative of a ferulic unit attached to putrescine). This compound also matched compound 10 with M_r 410 Da, t_r 33.05 mins and UV maxima 293.0, 309.0. In the case of compound 10, the parent ion at *m/z* 411 in positive ionisation mode produced a base peak in the MS^2 at *m/z* 147, indicative of the loss of a coumaric acid unit and a peak at *m/z* 177, indicative of the loss of a ferulic acid unit. Furthermore, when losses of 146 (coumaric acid unit) and 176 (ferulic acid unit) from the parent ion at *m/z* 411 in positive ionisation mode are added together, the value is 322, which when subtracted from 410 is 88. This is

indicative of a putrescine unit, thus a further characterisation of feruoyl coumaroyl peutrescine could be made.

Compound 17 with M_r 424 Da and t_r 35.57 mins showed characteristics consistent with coumaroyl feruoyl cadaverine, with a UV maxima at 298.0, 322.0. The parent ion at *m/z* 425 in positive ionisation mode produced peaks in the MS^2 at *m/z* 147, 177, 279 (indicative of a ferulic unit + cadaverine) and 249 (indicative of a coumaric unit + cadaverine).

In addition to verification with standards, there are a number of techniques that could also be used to confirm the identification of compounds, including $^{1}H/^{13}C$ NMR and Infrared (IR) spectroscopy. They are powerful tools often used for qualitative analysis of unknown compounds (Liotta & James-Pederson, 2008) and the resulting spectra can be used to determine the compound's structure. In the case of ^{1}H NMR, the spectrometer will be tuned to H^{-1} nuclei within the molecules of a substance. The elucidation of a compound's structure is dependent on a magnetic field, generated as a result of an element's characteristic spin (Reusch, 2013). Different parts of the compound will absorb radio frequency energy, which will subsequently reveal a resonance signal. Additionally, ^{13}C NMR may be used when significant portions of a molecule lack C-H bonds (Reusch, 2013). With respect to IR spectroscopy, a spectra is generated based on the amount and frequency of light absorbed, depending on different functional groups and the structure of the compound. Thus, the existence of hydroxyl groups, double bonds, and other active parts of a molecule can be easily deduced (Stuart, 2005).

5.6.3 Pilot-Phase HPLC-MS Targeted Analysis of 55 × PMiGAP lines

The previous untargeted analysis identified 16 phenolic compounds of interest, present in pearl millet grain. Based on these findings, a targeted analysis of the same lines was conducted to compare the relative abundance of typical compounds between lines (Table 5.7) and to analyse data for correlations between abundance of targeted compounds and micronutrient content, as well as for clustering between lines. A summary of the analytical method development process can be seen in Figure 5.20.

Analytical Method Development

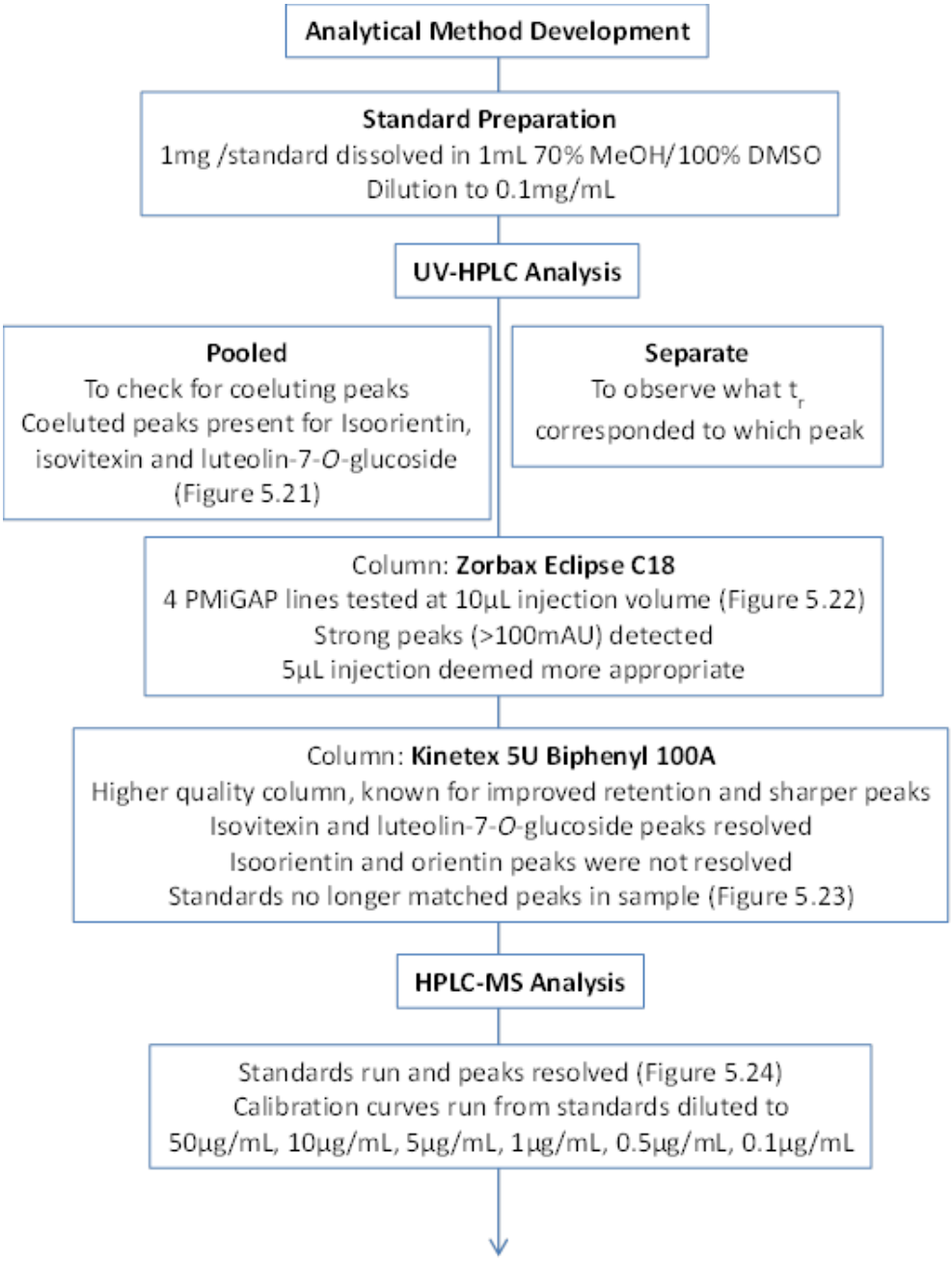


Figure 5.20: Analytical method development process for targeted analysis

The analytical method development process included an attempt to the separate co-eluted peaks found during UV-HPLC analysis. When the peaks for isoorientin and orientin could not be resolved, an HPLC-MS system was deemed more appropriate, which resulted in successful separation of peaks.

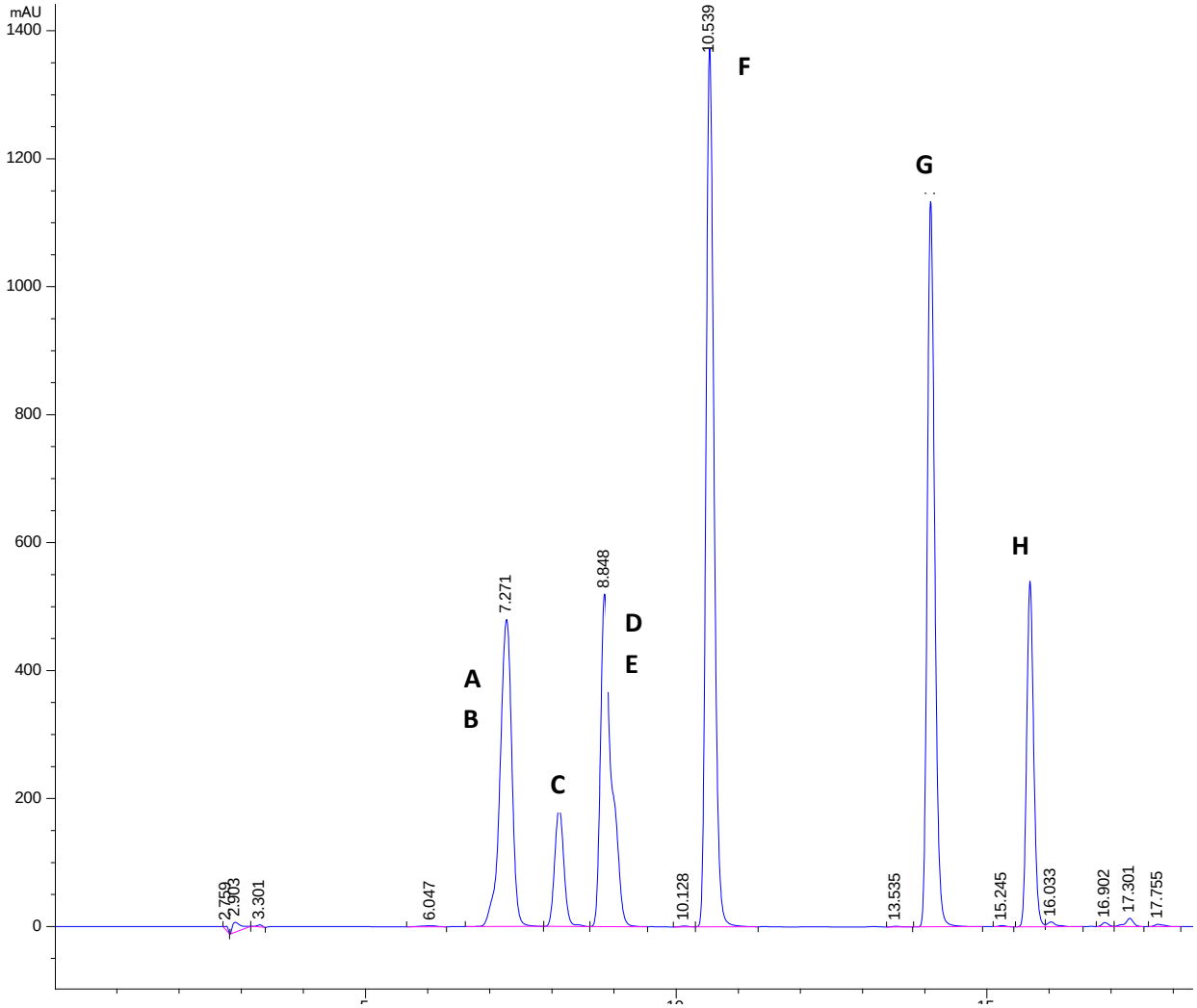


Figure 5.21: HPLC chromatogram showing standards at 0.1mg/mL. A- orientin, B- iso-orientin, C-vitexin, D- iso-vitexin, E- luteolin7-*O*-β-D-glucoside, F- apigenin7-glucoside, G- luteolin, H- ethyl 4-hydroxy-3-methoxycinnamate. Compounds A, B, D and E are Coeluted (Zorbax Eclipse C18 Column).

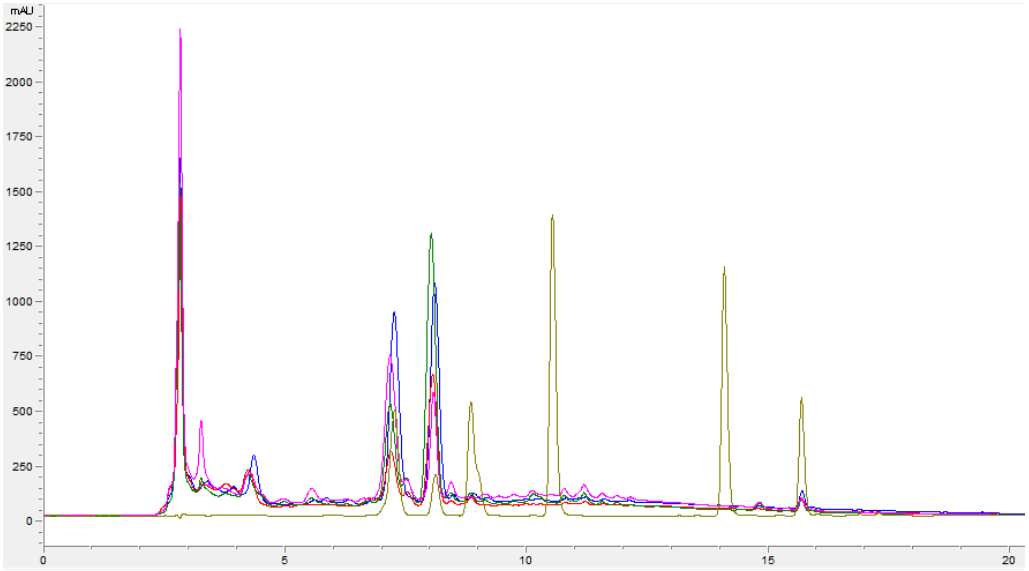


Figure 5.22: Overlay of 4 PMiGAP lines; Olive- initial standard mix in Figure 5.21, Pink, Blue, Green, Red- 4 PMiGAP lines.

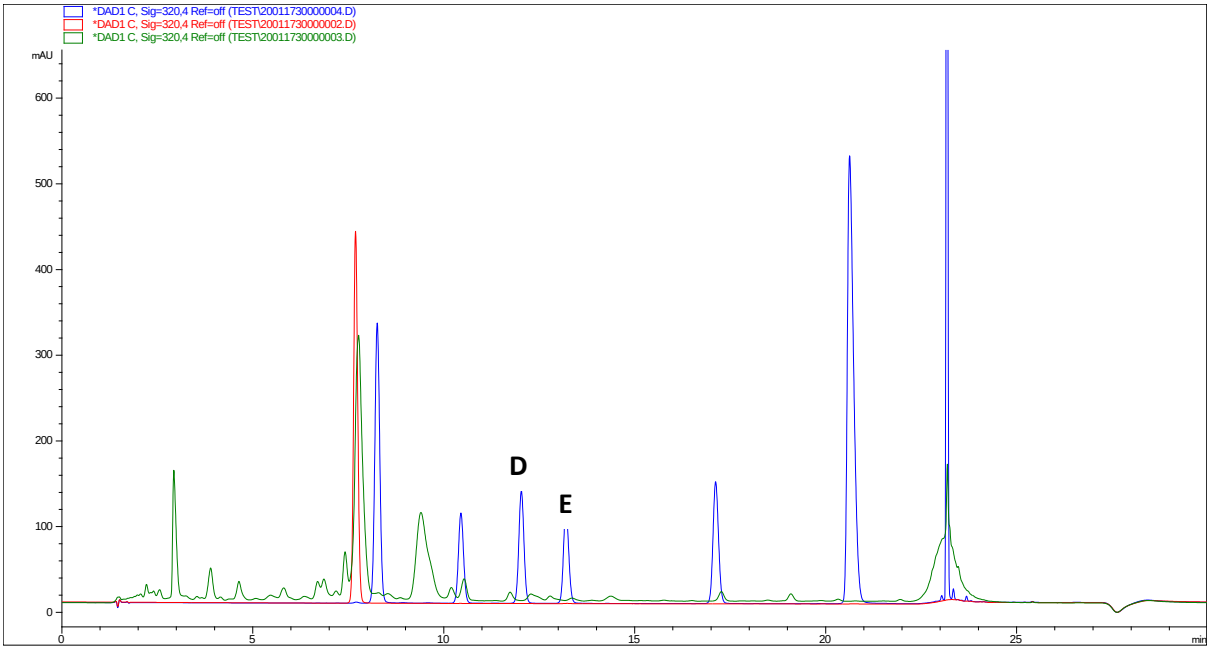


Figure 5.23: Overlay of HPLC data: Blue- standards with D and E resolved, D- iso-vitexin and E- luteolin7-*O*-β-D-glucoside, Red- *P*-coumaric acid standard, Olive- PMiGAP sample (Kinetex 5U Biphenyl 100A Column).

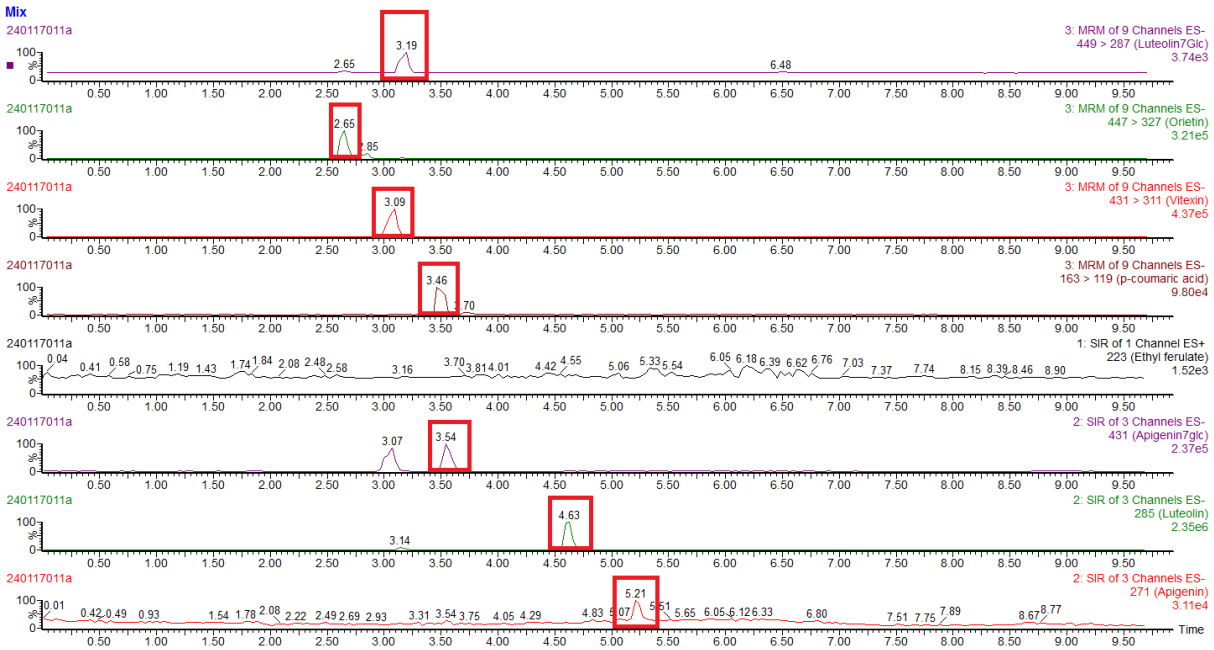


Figure 5.24: HPLC-MS chromatograms showing standards, Waters Xevo TQ-S UPLC-MS/MS. Top – bottom = luteolin7-*O*-β-D-glucoside; orientin, vitexin, *P*-coumaric acid, ethyl-ferulate, apigenin7-glucoside, luteolin and apigenin.

HPLC-MS Targeted Analysis

High-resolution metabolite profiling by UPLC-MS/MS facilitated a detailed analysis of the secondary metabolite composition in 55 PMiGAP lines. The profiling data and applied multivariate data analysis provided insight into the chemical relationships

between lines. This data could contribute to the selection of elite genotypes, for breeding pearl millet varieties with increased levels of bioavailable Fe/Zn, via lower amounts of metal-chelating compounds. The data also contributes to the current chemical knowledge of luteolin and apigenin compounds in pearl millet, and the hierarchical clustering analyses of the results provided an efficient method to investigate cluster relationships between lines.

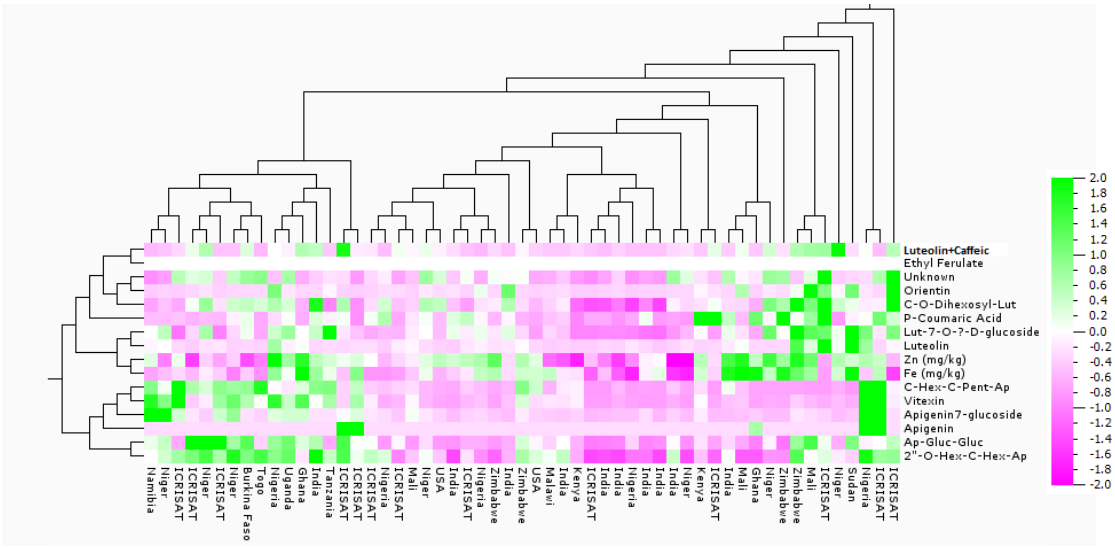


Figure 5.25: Heatmap with hierarchical cluster analysis of 55 PMiGAP lines built under the Qlucore Omics Explorer, v3.2 environment. Data has been normalised to the internal standard, ethyl ferulate. Clusters on the heatmap are identified by groups of common colour as indicated by similar values. As a result of the hierarchical clustering, the rows in the heat map have been reordered to correspond to the cluster calculation. Those placed in the same cluster are indicative of results that are similar, although not statistically significant.

The relationship between compounds in the extracts under investigation and a comparison of data can be seen in Figure 5.25. When a $P<0.05$ statistical significance threshold filter was applied (Figure 5.26), significant clustering for levels of Fe, Zn and luteolin7-*O*-β-D-glucoside were revealed. These findings were confirmed by Pearson's product moment correlation tests, where a strong correlation of 0.71 ($P<0.001$) between Fe and Zn occurred. When the Pearson's product moment correlation test was applied to Fe/Zn and luteolin7-*O*-β-D-glucoside, slight positive correlations of 0.42 and 0.34 were revealed, respectively. Significant clustering for high and low levels of Fe and Zn were also verified by PCA (Figure 5.28).

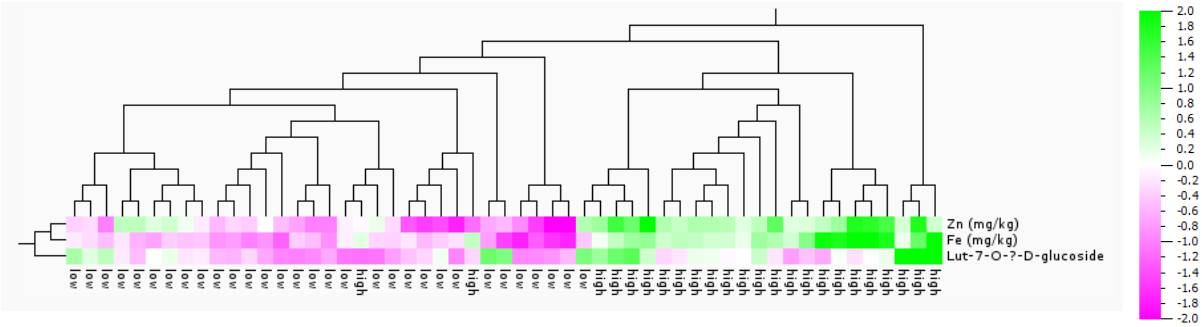


Figure 5.26: Heatmap with hierarchical cluster analysis of 55 PMiGAP lines with P<0.05 statistical significance threshold filter lines, built under the Qlucore Omics Explorer, v3.2 environment.

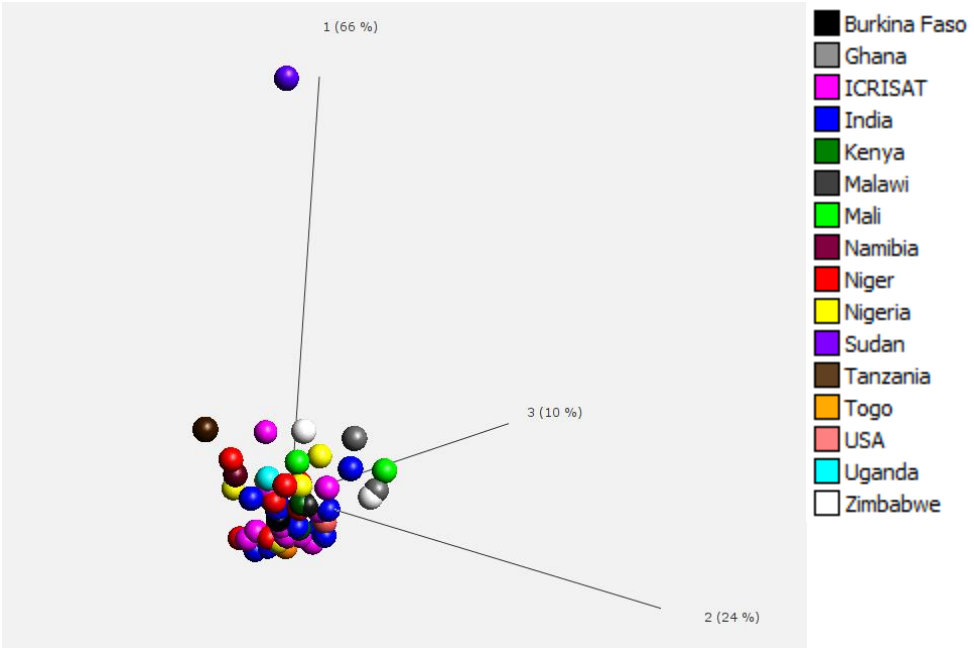


Figure 5.27: 3D-PCA plot to show variation between 55 PMiGAP lines based on origin and phenotype data (P<0.05) lines, built under the Qlucore Omics Explorer, v3.2 environment.

The PCA plot in Figure 5.27 suggests that when the PMiGAP lines were studied based on origin, there was one significant outlier, identified as genotype 1001 (IP10820) from Sudan. This line presented significantly higher levels of luteolin7-*O*-β-D-glucoside, luteolin, orientin, Fe, *C*-*O*-dihexosyl-luteolin and *C*-hex-*C*-pent-apigenin than all other PMiGAP lines.

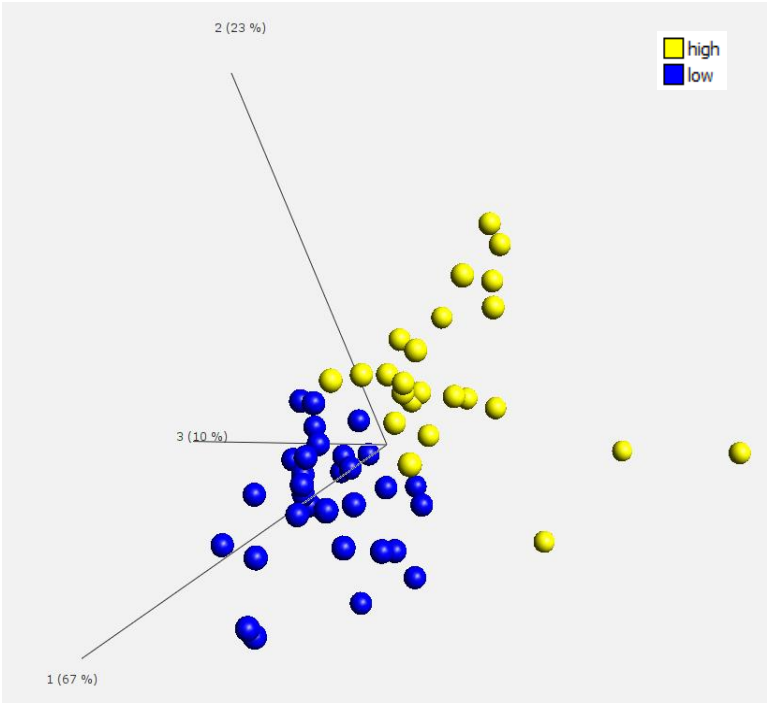


Figure 5.28: 3D-PCA plot to show variation between 55 PMiGAP lines based on high/low Fe and Zn levels, and phenotype data (P<0.05) lines, built under the Qlucore Omics Explorer, v3.2 environment. Data has been normalised to the internal standard, ethyl ferulate.

Figure 5.28 suggests that there was significant clustering between lines with high/low levels of Fe and Zn (mg/kg).

On the basis of low total phenolics content, a set of lines were identified (Table 5.13), which can be used in crossing programs for the development of elite lines with low apigenin/luteolin content and high Fe/Zn levels. Interestingly, all lines are of African origin which suggests that typically, lines from this region may be naturally lower in apigenin/luteolin compounds, as compared to the Indian lines.

Table 5.13: Proposed lines for the creation of elite lines. These were selected based on a phenolic content below the average total phenolic content of 55 lines analysed (average content = 536385.35AU), coupled with high Fe/Zn content. High Fe/Zn content was defined as above an average of 113.78mg/kg, for the total 55 lines analysed.

Line (IP No.)	Origin	Total Phenolics content*(% of average)	Combined Fe/Zn (mg/kg)
IP 10471	Zimbabwe	38	138.30
IP 6460	Mali	46	170.99
IP 19584	Niger	49	139.50
IP 5713	Nigeria	58	129.28
IP 13964	Zimbabwe	68	152.54

*(Total phenolics content per line / 536385.35×100)

For a more significant comparison, a larger population would have been favourable and with minor changes, this approach is fully compatible with the study of a larger population. Other aspects of the sample preparation method also required optimisation. For example, it was found that the internal standard, ethyl ferulate was barely detectable in negative ionisation mode. This is likely due the SPE clean-up step removing significant amounts of this compound. The peak area range for the internal standard was between 54-792 AU, this is highly variable, and indicative of inaccurate sample preparation. In addition, when amounts of each targeted metabolite were established by normalising to the internal standard, this could have produced inaccurate data in the downstream analysis as it is highly probably that many peaks, especially the ones with peak areas below 200 AU were within the level of background noise. For this reason, results must be approached with caution and the analysis should be repeated with; i) a larger population, ii) changes to the sample preparation method to allow for increased concentrations of internal standard to be detected in the final solutions, at invariable levels and iii), the use of a different internal standard, that ionises well in both negative and positive ionisation modes and does not co-elute with compounds of interest.

5.6.4 Untargeted/Targeted Analysis of Non-Hydrolysed and Acid Hydrolysed *HHB67 Improved*

MS^2 fragmentation patterns for luteolin and apigenin glycosides, identified in PMiGAP lines clearly showed that compounds contained both *O* and *C* linked sugars. *HHB67 improved* was subjected to untargeted (data dependant mode, which selects the most abundant ions for MS^2) and targeted analysis to confirm the presence of *O/C* linked apigenin/luteolin glycosides. Phenolic extracts were treated with 1M HCL, as described above to acid hydrolyse phenol glycosides and analysed by HPLC-MS analysis for the determination of *C*-linked sugar moiety position, when attached to the core aglycone.

HPLC-PDA-MS^n Untargeted Analysis

Figure 5.29 shows the profile of soluble phenolics extracted from *HHB67 Improved.* Analysis by HPLC-MS with PDA detection revealed approximately 6 compounds that could be easily distinguished (Table 5.14), which included hydroxycinnamic acids, phenolamides and flavones. These were identified as described above.

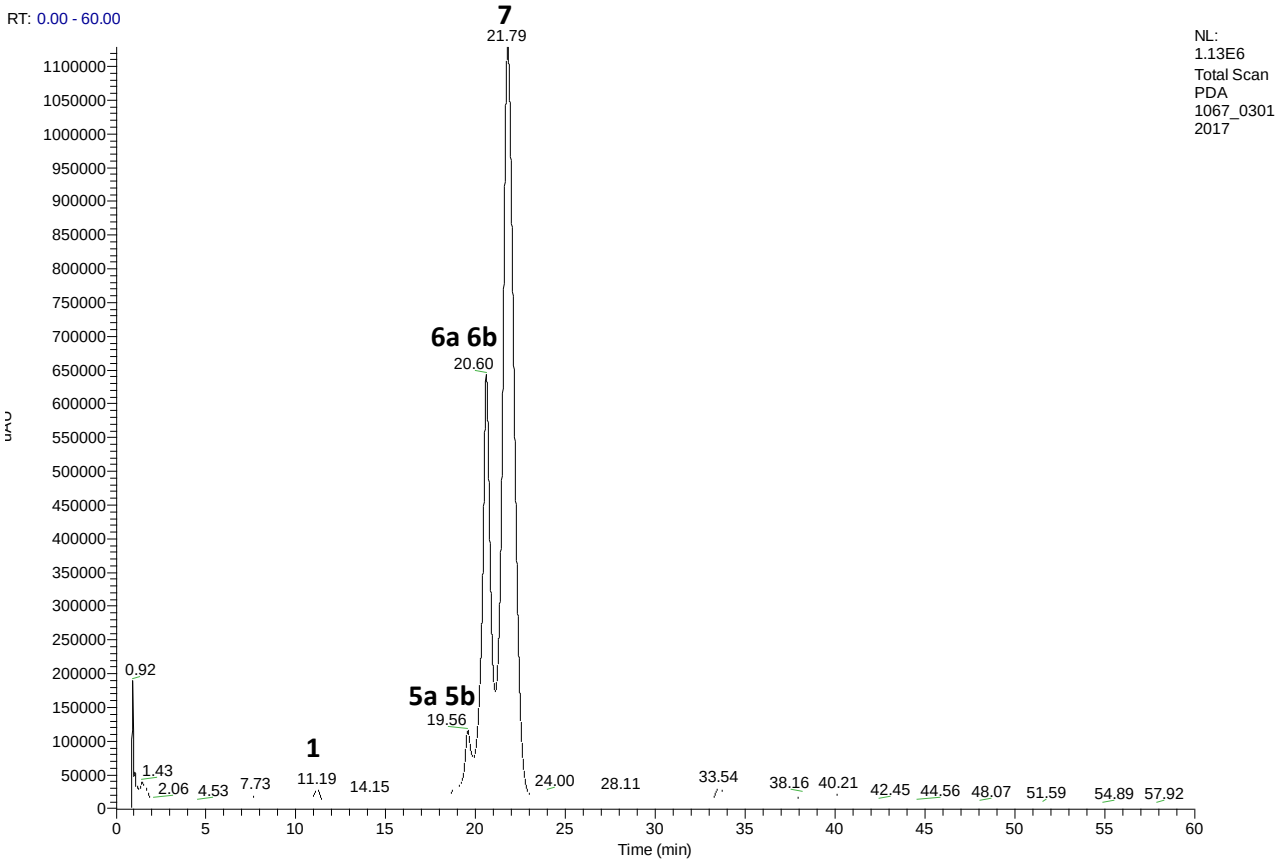


Figure 5.29: HPLC-MS chromatogram showing the separation of 6 soluble phenolic compounds in *HHB67 Improved* detected by PDA at 280nm. Peak numbering corresponds with Table 5.14.

Table 5.14: UV maxima, full MS ions and MS^2 events for compounds detected in an extract of *HHB67 Improved.*

Peak No.	t_r (mins)	UV Max	Nominal Mass, M_r Da	Parent Ions	MS^2 Fragments (in order of intensity)	Neutral loss ($MS - MS^2$)	Tentative identification
1	11.19	309.0	164	163.20 (-ve) 165.27 (+ve)	(-ve) 119 (+ve) 147	(-ve) 44 (+ve) 18	*P*-Coumaric Acid
5a	19.56	Not Detected*	469	468.22 (-ve) 470.27 (+ve)	(-ve) 332 (+ve) 308 220 453 291 163 145	(-ve) 136 (+ve) 162 250 17 179 307 325	Dicaffeoyl Spermidine
5b		Not Detected*	564	563.15 (-ve) 565.08 (+ve)	(-ve) 473 443 503 353	(-ve) 90 120 60 210	Apigenin (*C*-Hex-*C*-Pent-Apigenin)

6a	20.60	268.0 339.0	432	431.17 (-ve) 433.17 (+ve)	383 545 (+ve) 547 529 427 511 (-ve) 311 (+ve) 415 397 367 313 (-ve)	180 18 (+ve) 18 36 138 54 (-ve) 120 (+ve) 18 36 66 120 (-ve)	Vitexin
6b		Not Detected[*]	610	609.44 (-ve) 611.27 (+ve)	489 429 327 357 449 (+ve) 449 329 431 287 491 (-ve)	120 180 282 252 160 (+ve) 162 282 180 324 120	Luteolin (*C*-*O*-Dihexosyl-Luteolin)
7	21.79	268.0 336.0	594	593.32 (-ve) 595.38 (+ve)	Not Detected (+ve) 433 449 475 415 313 271	 162 146 120 180 282 324	Apigenin (2”-*O*-Hex-*C*-Hex-Apigenin)

[*]Not detected due to coeluting compounds

All compounds are discussed in section 5.6.2, with the exception of vitexin which was unique to *HHB67 Improved*.

Compound 6 with M_r 432 Da and t_r 20.60 mins showed characteristics consistent with vitexin (apigenin-*C*-hexoside), with a UV maxima at 268.0, 339.0. The parent ion at *m/z* 431 in negative ionisation mode produced a dominant base peak in the MS^2 at *m/z* 311, thus giving a neutral loss of 120. This is indicative of a *C*-linkage with a hexose sugar. This compound was also identified with a standard for confirmation.

HPLC-PDA-MS^n Targeted Analysis

Table 5.15: Targeted MS^2 events in negative ionisation mode, in *HHB67 Improved*.

Compound	t_r (mins)	MS^2 target ion (negative mode)	MS^2 Fragments (in order of intensity)	Neutral loss (MS – MS^2)	Identification
6a	20.61	431	311 341	120 90	Vitexin
5a	19.71	468	332	136	Dicaffeoyl Spermidine
5b	19.37	563	473 443 503 383 353	90 120 60 180 210	*C*-Hex-*C*-Pent-Apigenin
7	21.79	593	545 413 293	18 180 300	2”-*O*-Hex-*C*-Hex-Apigenin
6b	20.51	609	489 327 357 429 449 369 309	120 282 252 180 160 240 300	*C*-*O*-Dihexosyl-Luteolin

The MS^2 fragmentation pattern data obtained from the targeted analysis (Table 5.15) confirmed the findings from the untargeted (data dependant) analysis.

Acid Hydrolysis of Flavone Glycosides

Acid hydrolysis facilitated the identification of *C*-linked sugar moiety position, when attached to the aglycones luteolin and apigenin as described above. Additionally, an *O*-glucoside standard for both apigenin and luteolin (apigenin7-glucoside and luteolin7*O*-β-D-glucoside) was used as a control to confirm the *O*-linkages were hydrolysed by the treatment. Predicted flavones as a result of acid hydrolysis can be seen in Figure 5.30.

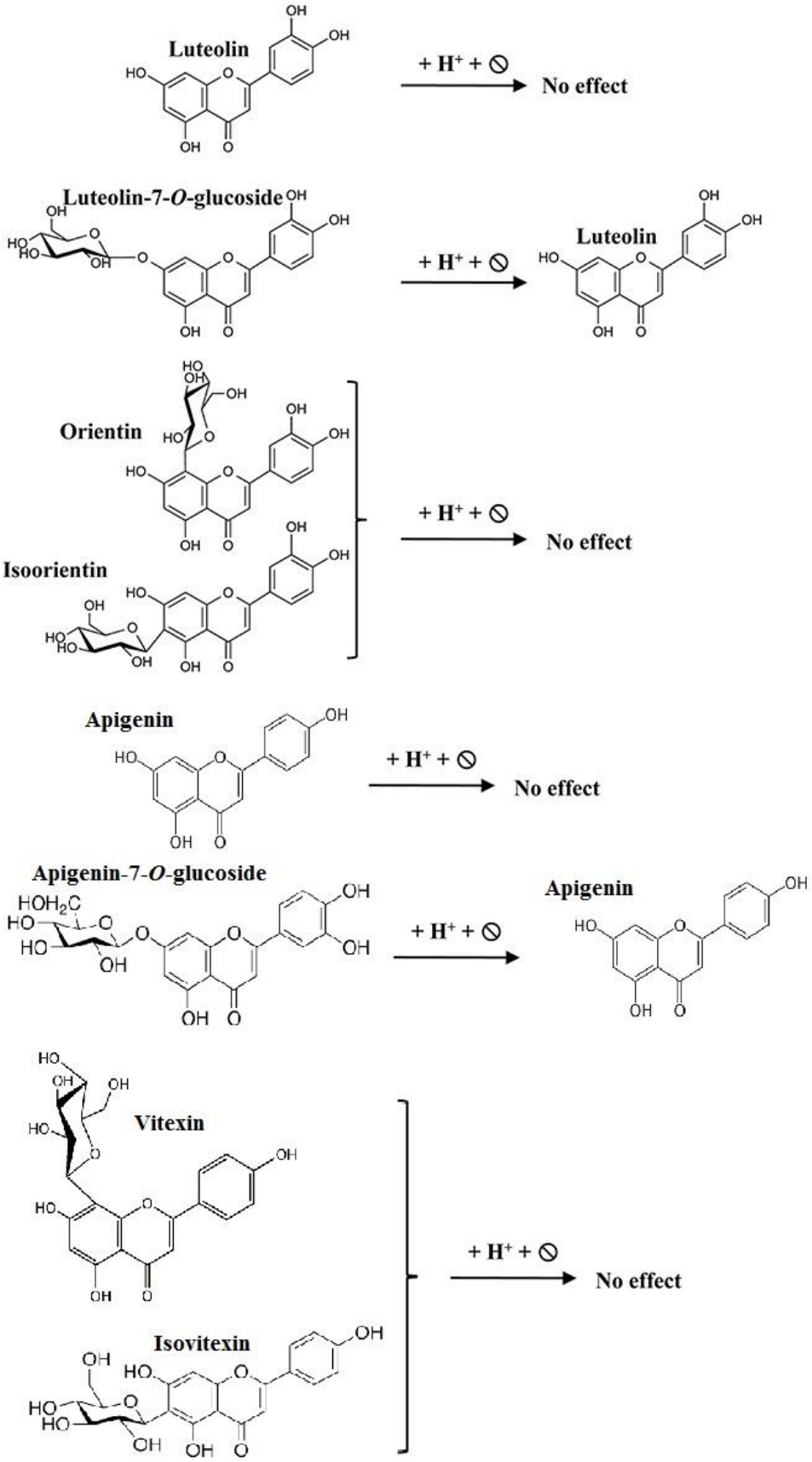


Figure 5.30: Predicted flavones resulting from acid hydrolysis, adapted from Casas *et al*., 2014. Orientin and isoorientin are luteolin-8/6-*C*-glucosides and vitexin and isovitexin are apigenin-8/6-*C*-glucosides.

UVHPLC Analysis

Peak 2 – Apigenin *C-O/C-C* diglycosides

The results from the UVHPLC and MS data analysis (Figures 5.31 and 5.32, respectively) revealed that the peak at t_r 8.93 minutes, in the non-hydrolysed extract was composed of several apigenin diglycosides coeluting together. One of the apigenin diglycosides contained an *O*- and a *C*-link. This conclusion was drawn from the UVHPLC data, where upon hydrolysis, the *O*-link was cleaved off and an iso-vitexin peak remained at t_r 11.6 minutes, in the hydrolysed extract. This suggests that the location of the *C*-linked sugar moiety was on the C6 position. The other apigenin diglycoside present in the same peak contained two *C*-links, indicating a C6 and C8 linkage. This conclusion was drawn because the peak did not disappear completely upon hydrolysis, and was still present at t_r 9.03 minutes in the hydrolysed extract, indicating an apigenin compound with two sugars attached, both linked by C-linkage.

Vitexin

According to Figure 5.31, vitexin was present in both extracts (non-hydrolysed t_r 10.04 minutes, hydrolysed t_r 10.06 minutes) and was not affected by acid hydrolysis because there was no *O*-linkage to hydrolyse. Therefore, it is concluded that the sugar was on the C8 position and *C*-linkage was confirmed.

Luteolin diglycoside

A luteolin diglycoside was present in the non-hydrolysed extract, although it cannot be seen on the PDA in Figure 5.31 (possibly due to coeluting peaks); it was predicted to contain an *O*- and *C*-linkage because upon hydrolysis, the product of this was isoorientin/orientin, which presented a peak at t_r 7.9 minutes. This suggests that when the *O*-link was removed, a monoglycoside remained. However, the position of the sugar whether it was at C6 or C8 could not be distinguished as isoorientin and orientin gave rise to coeluting peaks (as seen in the standard mix, Figure 5.31). Furthermore, in the hydrolysed extract, there was no obvious formation of a luteolin aglycone peak, which proves that this was not a diglycoside where the sugars were attached by two *O*-linkages.

MS/MS Analysis

Peak 1

The MS results for Peak 1 (Figure 5.32) revealed that dicaffeoyl spermidine, with a mass of 468 Da (in negative ionisation mode) was a major constituent. It also coeluted with an apigenin diglycoside with a mass of 593 Da, in negative ionisation mode (Figure 5.32).

Peak 2

Peak 2 (Figure 5.32) represents two coeluting apigenin diglycosides with masses of 593 and 563 Da, respectively. The results verified the predictions from the UVHPLC data (Figure 5.31), that one apigenin diglycoside contained a *C*- and *O*-linkage and the other contained two *C*-linkages.

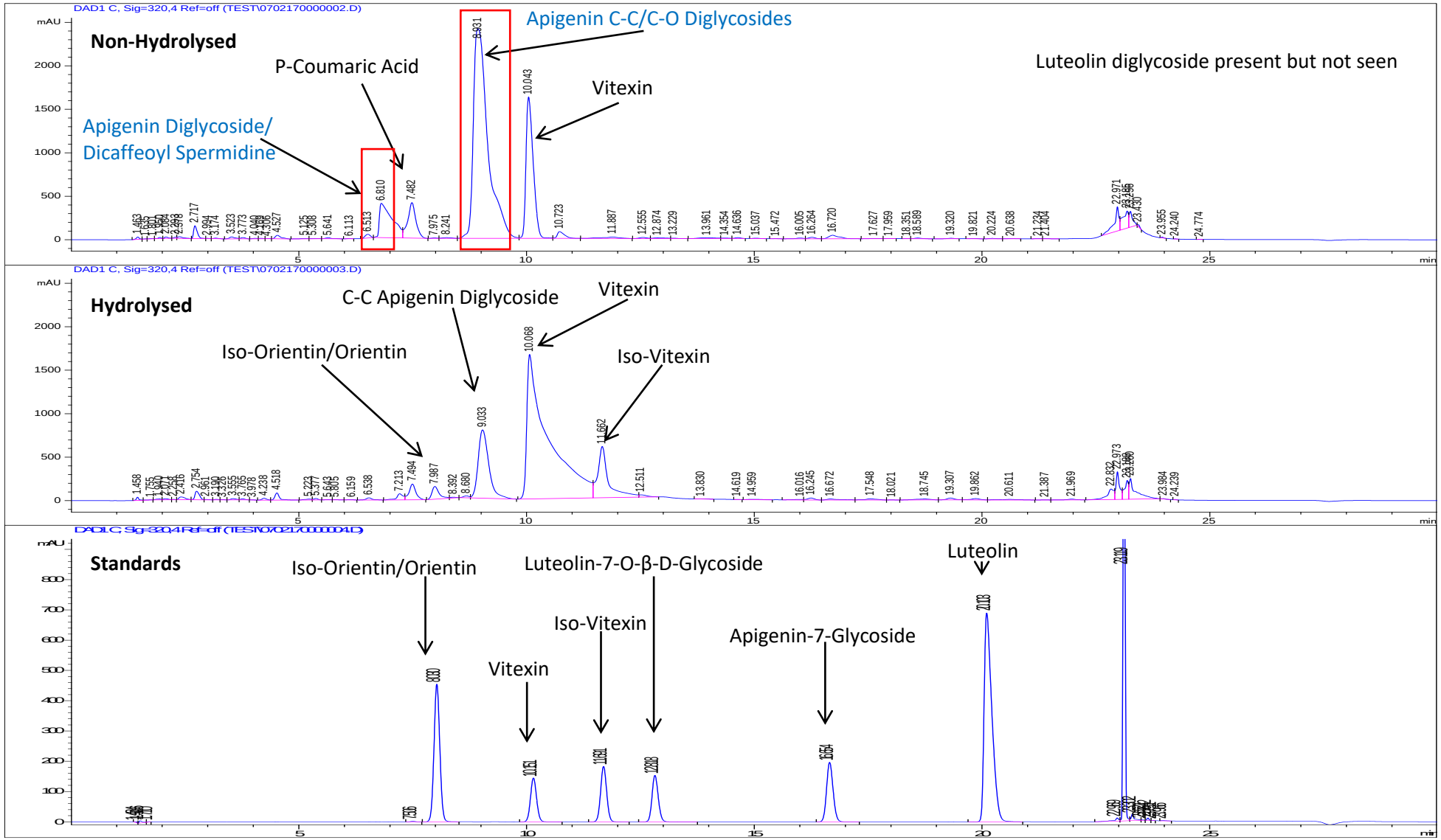


Figure 5.31: UV-HPLC chromatograms showing non-hydrolysed and hydrolysed extracts of *HHB67 Improved* as compared to standards, at 320nm PDA. Peaks in red boxes were collected and run on the MS for verification purposes, the peak on the left is referred to as Peak 1 in the text and the one on the right is referred to as Peak 2. Blue text represents peaks verified by MS

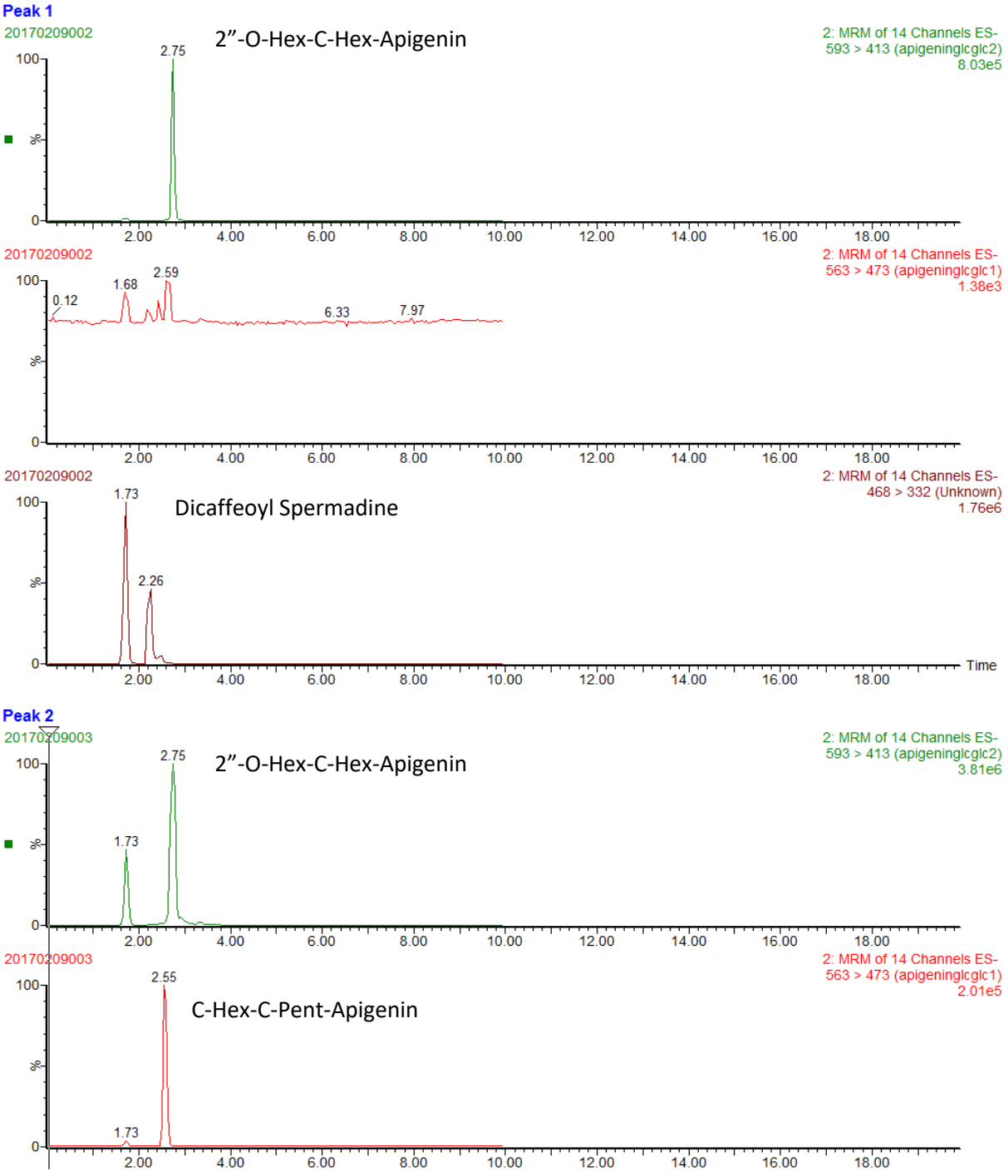


Figure 5.32: MS data for Peak 1 and Peak 2, corresponding to peaks in the red boxes in Figure 5.31, from the UV-HPLC data.

5.6.5 Post Pilot-Phase HPLC-MS Targeted Analysis of 185 × PMiGAP lines

The results in section 5.6.3 revealed that significant changes to the phenolic compound extraction method would benefit future work, in terms of increased reliability/accuracy. Implemented changes included; scaling down the MeOH extraction protocol to 0.1g seed + 1mL MeOH to avoid compound loss due to pipetting errors and to facilitate higher though-put, a larger PMiGAP population, the use of a more consistent internal standard (biochanin A) and the removal of the SPE C18 clean-up step. As a result, an improved targeted analysis of typical compounds present in 185 PMiGAP lines (Table 5.8) facilitated the identification and comparison of the relative abundance of phenolic compounds, which may affect the bioavailability of absorbable Fe/Zn. Additionally, 5 flavonoid glycosides/aglycones were quantified using standard calibration curves on a molar basis and the correlation between polyphenol content and micronutrient content was also investigated. A summary of the analytical method development process can be seen in Figure 5.33.

Analytical Method Development

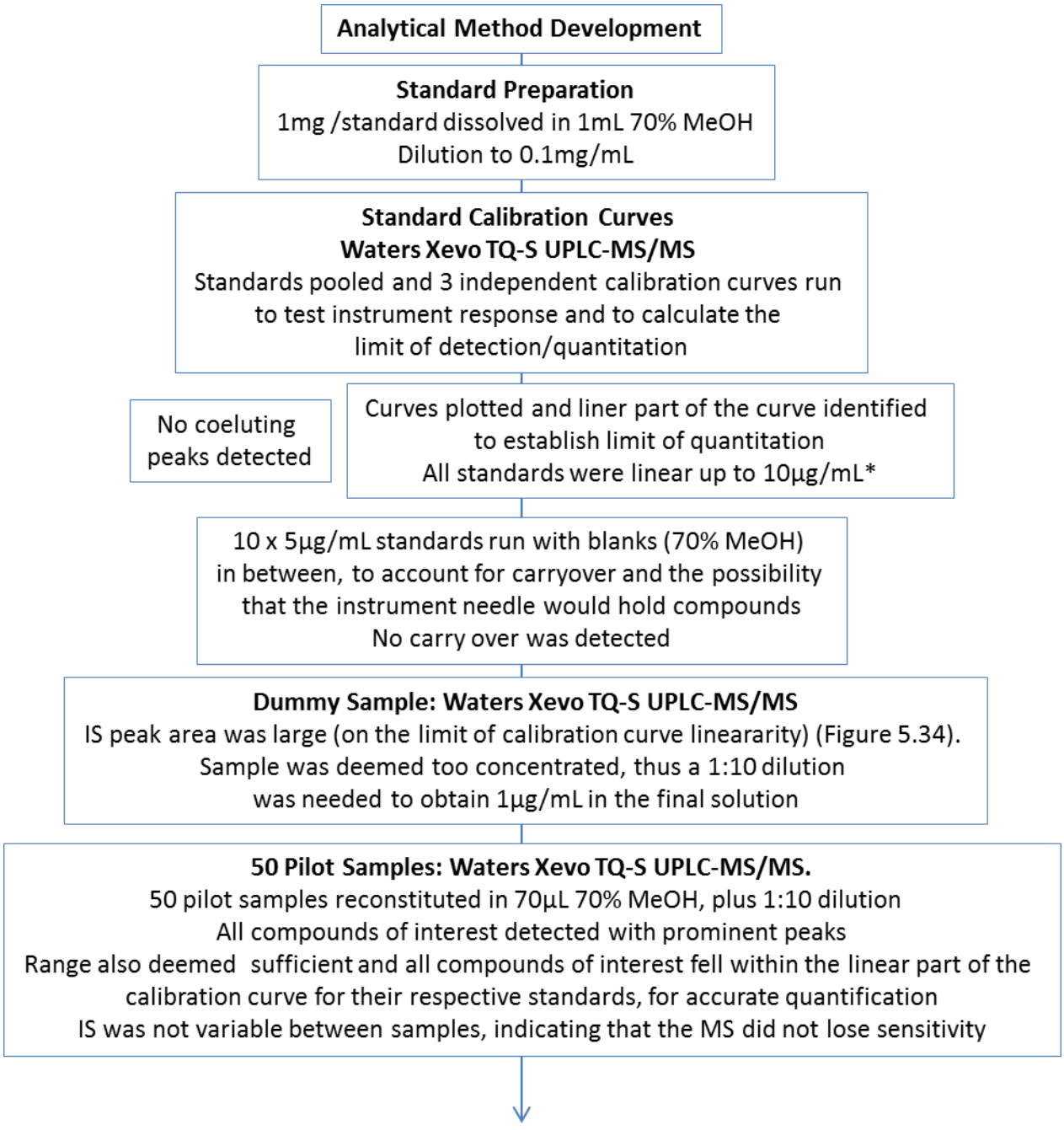
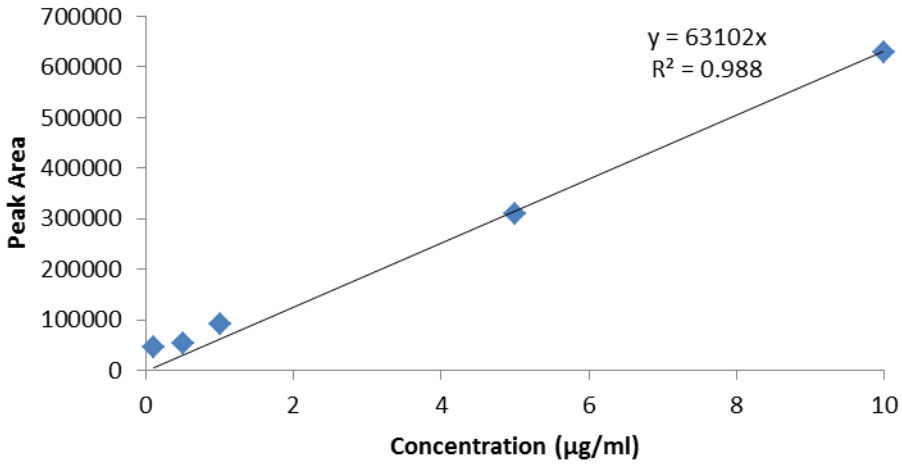
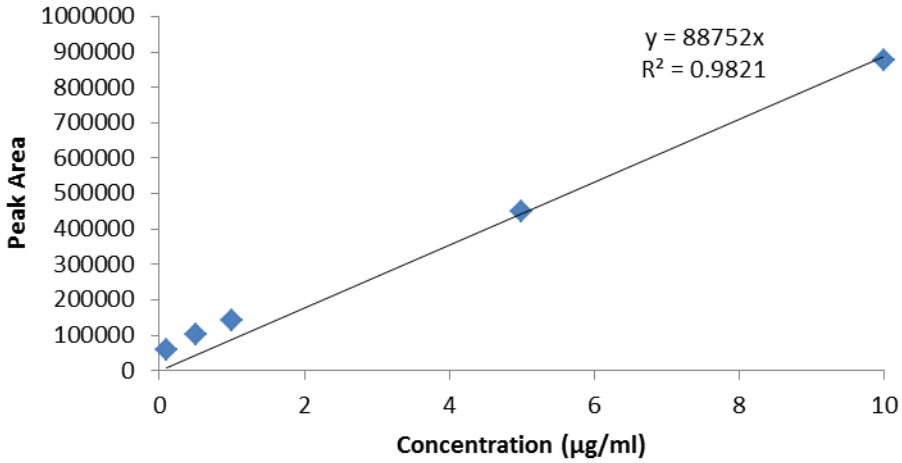


Figure 5.33: Method development process for targeted analysis of 185 PMiGAP samples *p-coumaric-acid linear up to 100µg/mL. IS = Internal Standard, biochanin A.

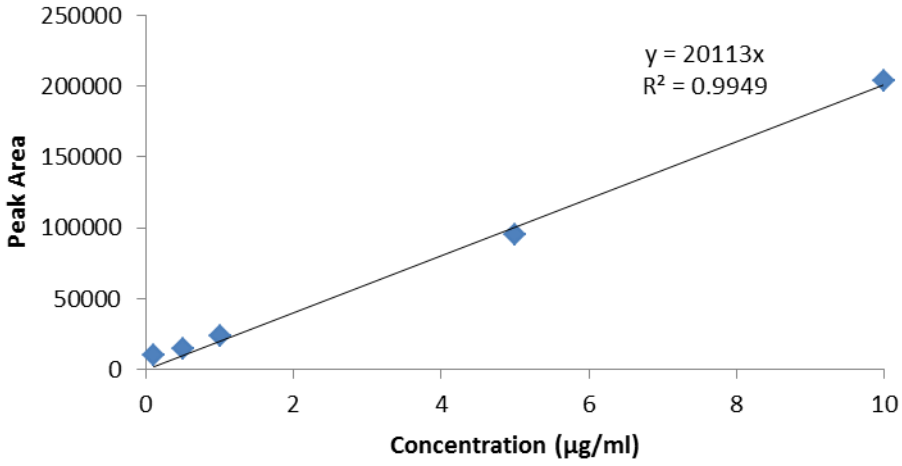
Apigenin 7 Glucoside



Biochanin A



P Coumaric Acid



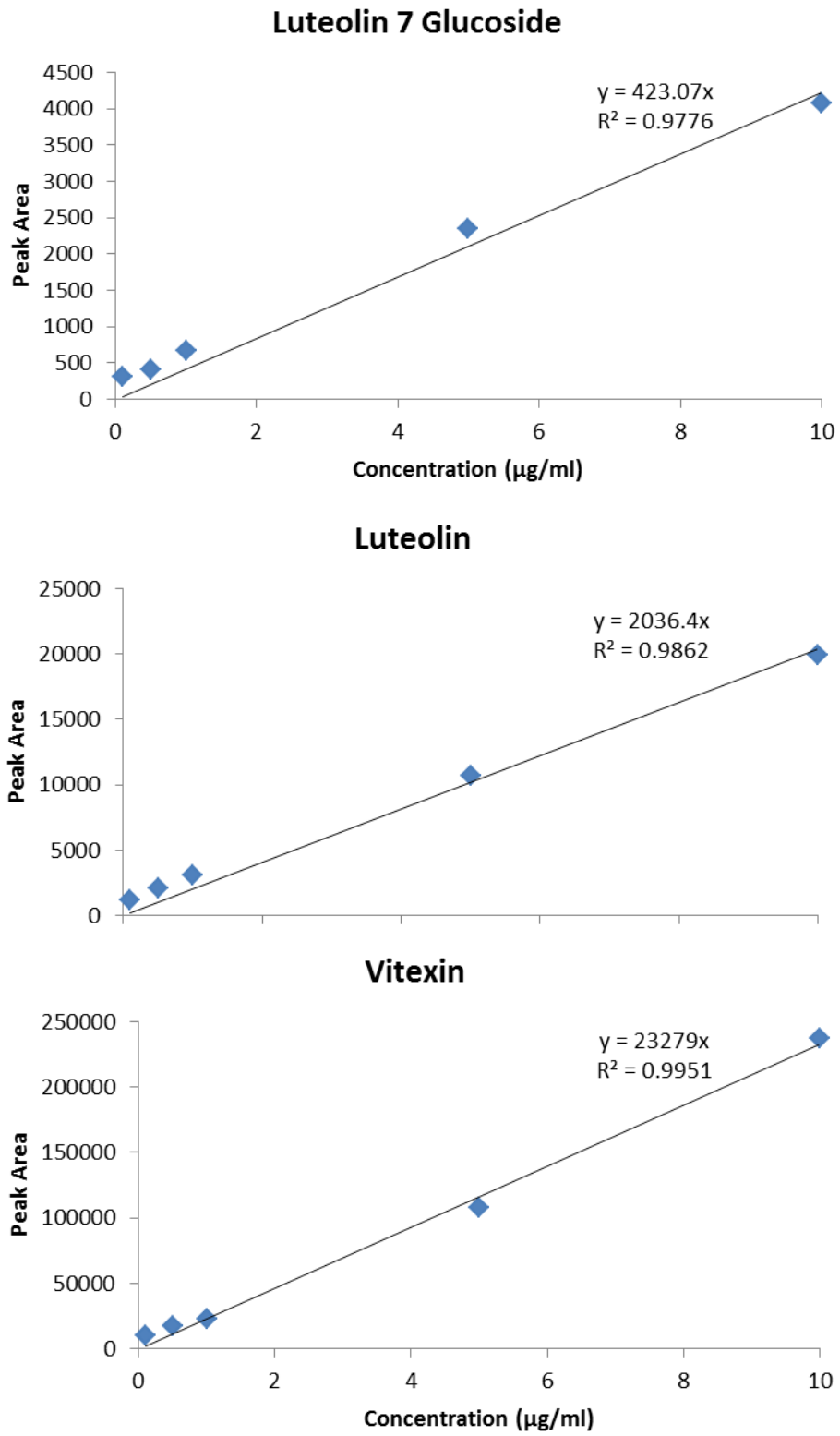


Figure 5.34: Standard calibrations curves of apigenin-7-glucoside, biochanin A, *p*-coumaric acid, luteolin-7-glucoside, luteolin and vitexin based on UV absorption.

HPLC-MS Targeted Analysis

Free phenols were obtained from the MeOH extraction of ground pearl millet seed. An initial screening of the extracts identified 16 phenolic compounds (Section 5.6.2), 6 of which were highly prominent in the majority of samples and, with the exception of dicaffeoyl spermidine, contained the aglycones apigenin/luteolin. The mass spectra and retention times of the phenols were compared with those of standards including; vitexin, luteolin, luteolin7-*O*-β-D-glucoside, apigenin7-glucoside and *p*-coumaric acid. UPLC-MS/MS facilitated a detailed analysis of the secondary metabolite composition of 185 PMiGAP lines and flavonoid glycosides were quantified using aglycone standard calibration curves on a molar basis (Figure 5.36). The profiling data and applied multivariate data analysis provided valuable insight into the chemical relationships between lines, with increased reliability than the previous analysis due to the implemented changes, as discussed.

Data produced can be approached with increased confidence. For example, the larger sample set (n=185) is more representative of the full diversity of the PMiGAP and changes made to the sample preparation method, which included scaling down the extraction method 5 fold and not using SPE filtering decreased the chances of compounds being lost, either by transfer between tubes/pipetting errors or by binding to the SPE filter. The resulting accuracy was reflected in the low levels of internal standard variability between samples, as compared to that of the previous targeted analysis (Figure 3.35). The internal standard used in this analysis was biochanin A, as it was found that ethyl ferulate (used in section 5.6.3) did not ionise well in negative ionisation mode. From the data in Figure 5.35, it can also be seen that biochanin A yielded more intense peaks, therefore this reduced the probability that the readings could be mistaken for ‘background noise’, which was the case for ethyl ferulate.

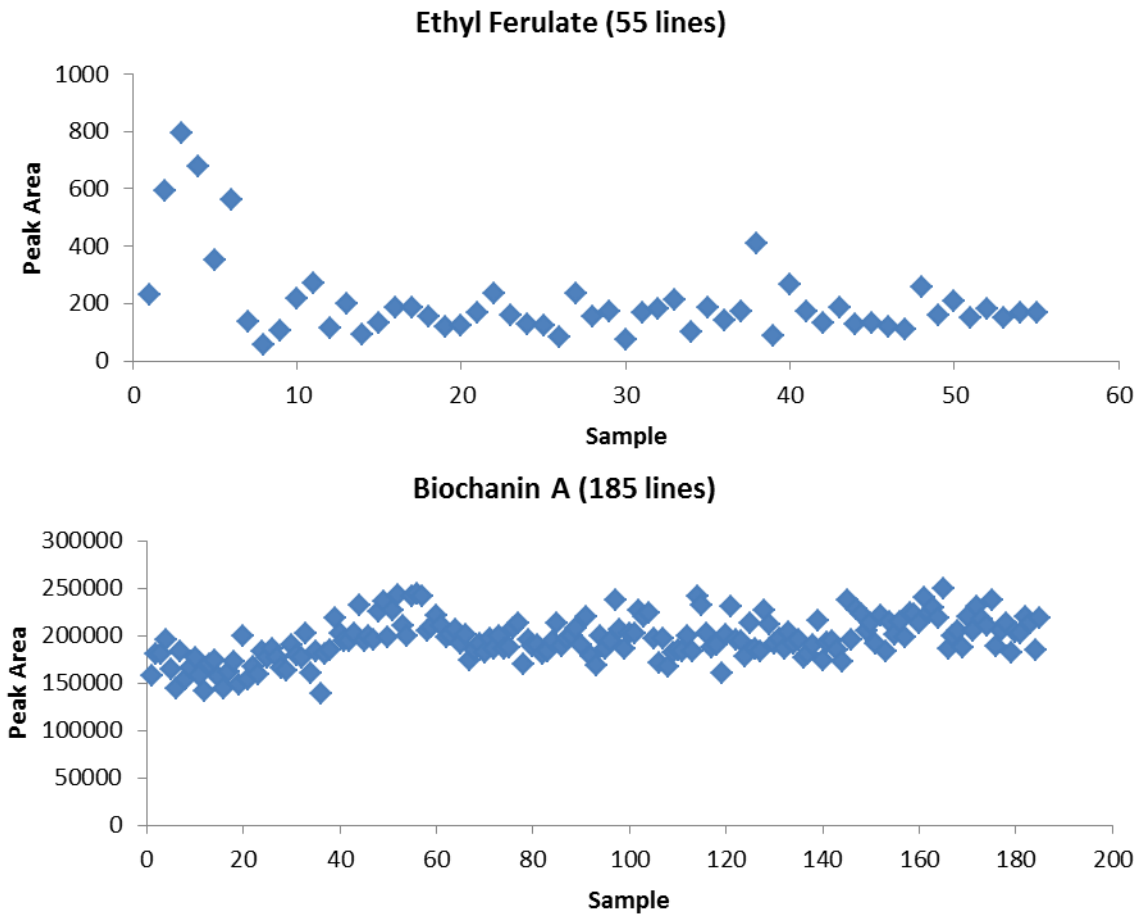


Figure 5.35: A comparison of the peak areas, absorbance units; 240-400nm range (as a measure of relative abundance) detected for two internal standards; ethyl ferulate and biochanin A.

According to the data in Figure 5.36, there were extremely large amounts of variation for all 5 compounds, when quantified using standard calibration curves. The average content of apigenin-7-glucoside found in 185 PMiGAP lines was 58.5 µg/g, with values varying from 1.5 – 757.5 µg/g. The average content of *p*-coumaric acid was 1609.0 µg/g, with values varying from 483.2 – 4105.2 µg/g. The average content of luteolin was 60.8 µg/g, with values varying from 7.68 – 910.2 µg/g. The average content of vitexin was 2797.8 µg/g, with values varying from 175.2 – 16194.8 µg/g. The average content of luteolin-7-glucoside was 587.7 µg/g, with values varying from 68 – 15480.8 µg/g

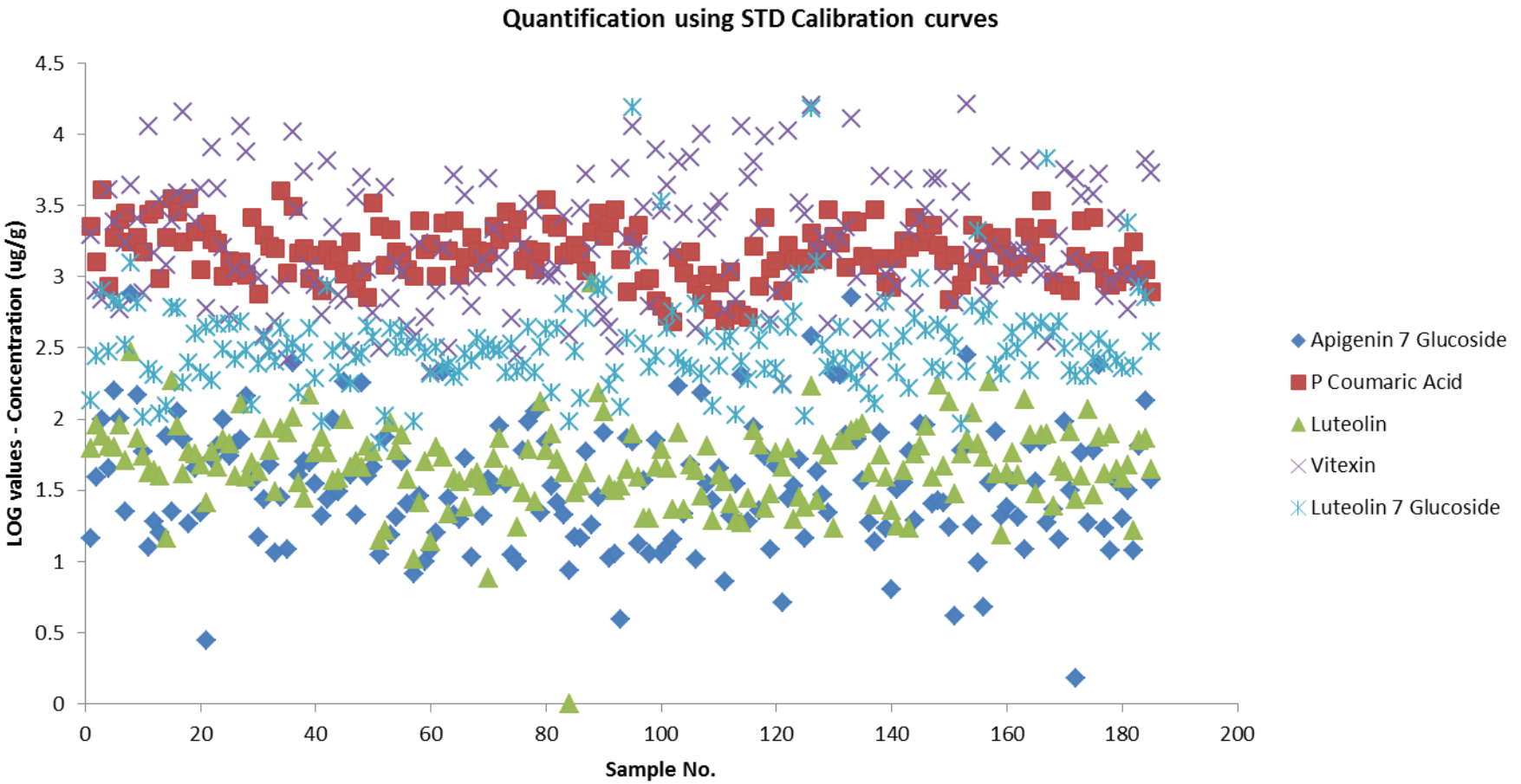


Figure 5.36: 6 compounds quantified (LOG base 10µg/g) using aglycone standard calibration curves. All compounds have been normalised to the internal standard, biochanin A.

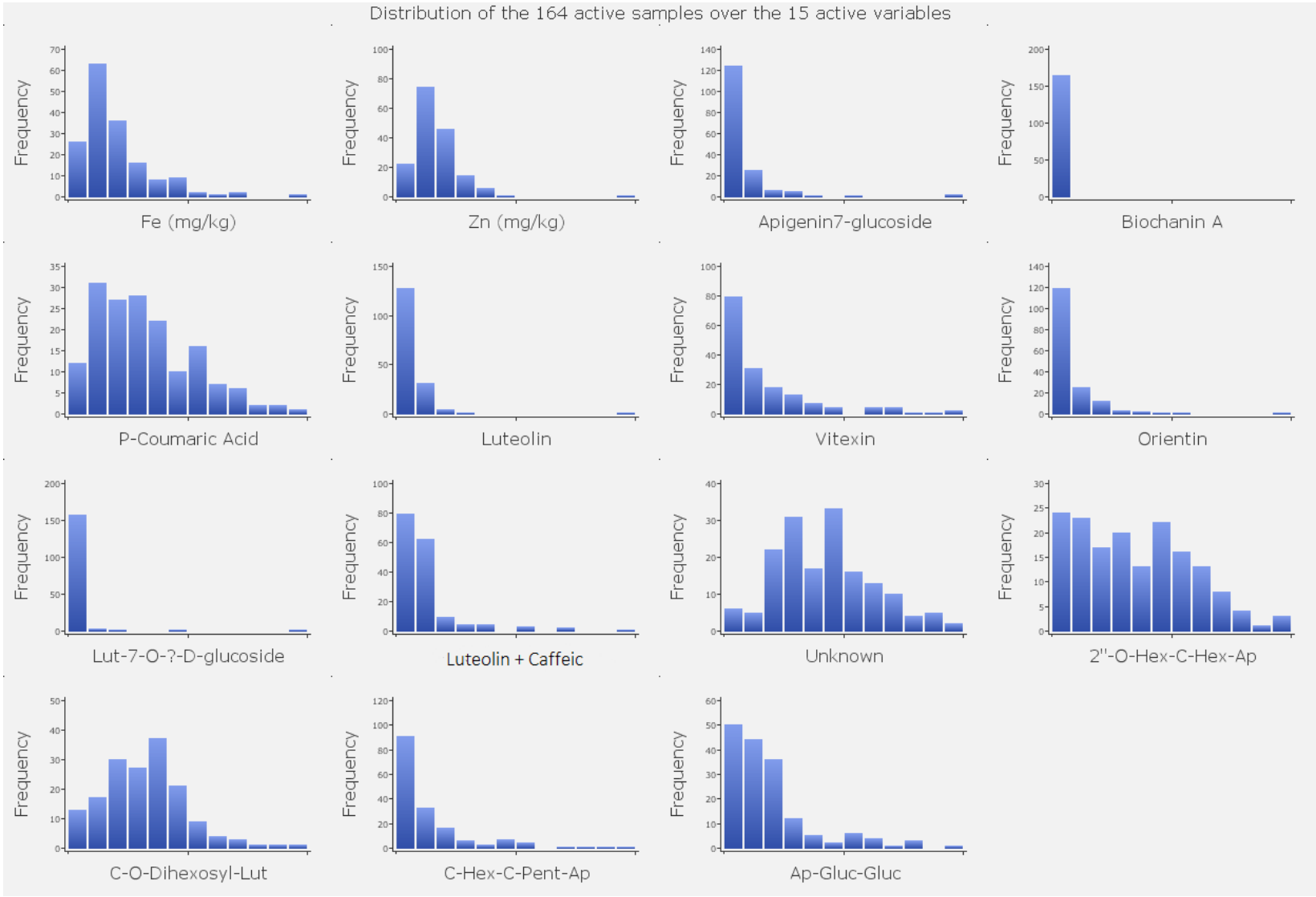


Figure 5.37: Histograms, 185 PMiGAP lines, frequency over variables measured in absorbance units in all cases with the exception of Fe and Zn (mg/kg). The distribution of samples is split into 12 equal bars calculated by dividing the highest value per metabolite by 12. Readings that fell within the 12 range values represent each bar (Unknown = Dicaffeoyl Spermidine).

The histograms in Figure 5.37 show the sample distribution between variables. The data for biochanin A further strengthens the conclusions from the results in Figure 5.35 because all samples are represented in one bar, showing low levels of variability and that the method used in this experiment was robust. The highest levels of variability can be seen for 2”-*O*-hex-*C*-hex-apigenin, dicaffeoyl spermidine, *p*-coumaric acid and *C*-*O*-di-hexosyl-luteolin; whereas moderate levels of variability were detected for Fe, Zn, vitexin, *C*-hex-*C*-pent-apigenin and apigenin-gluc-gluc. Low levels of variability are seen for apigenin-7-glucoside, luteolin, orientin, luteolin-7-glucoside and luteolin+caffeic.

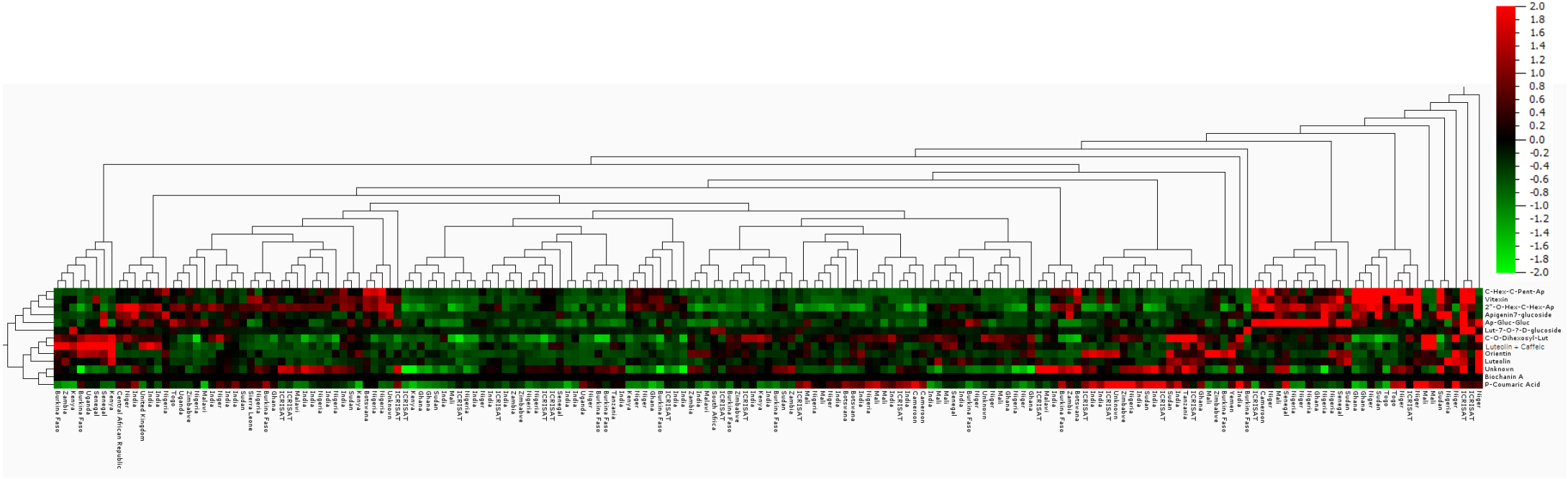


Figure 5.38: Heatmap with hierarchical cluster analysis of 185 PMiGAP lines built under the Qlucore Omics Explorer, v3.2 environment. Data has been normalised to the internal standard, biochanin A. Clusters on the heatmap are identified by groups of common colour as indicated by similar values. As a result of the hierarchical clustering, the rows in the heat map have been reordered to correspond to the cluster calculation. Those placed in the same cluster are indicative of results that are similar, although not statistically significant.

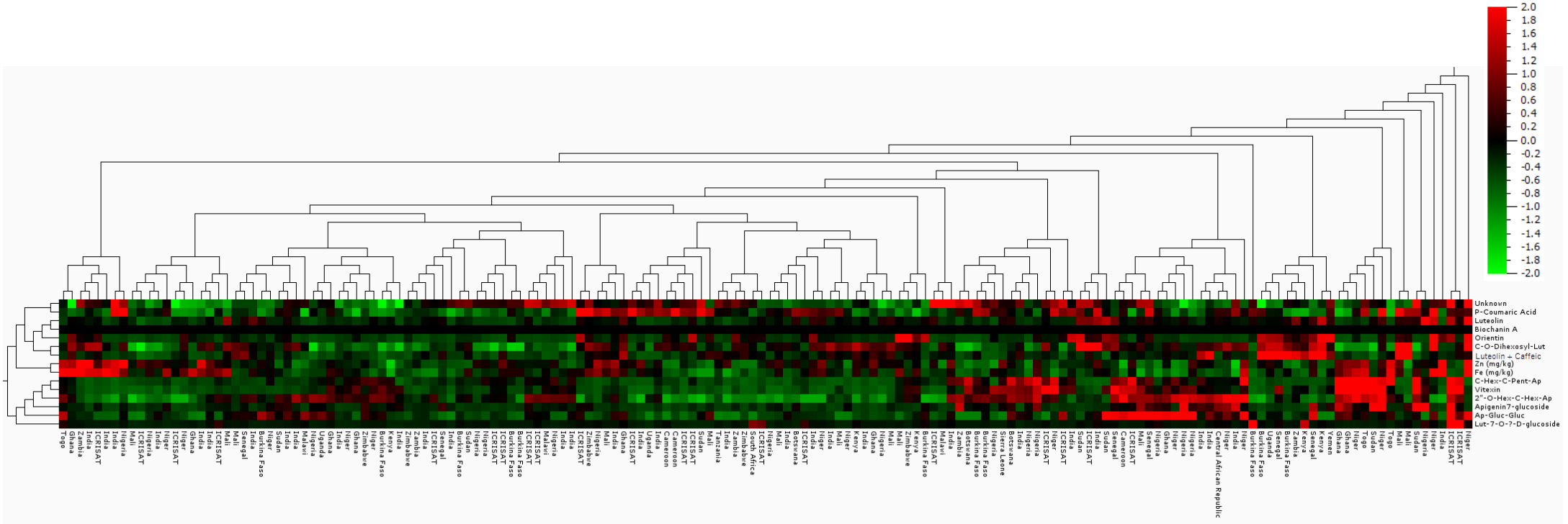


Figure 5.39: Heatmap with hierarchical cluster analysis of 164 PMiGAP lines plus mineral data built under the Qlucore Omics Explorer, v3.2 environment. 21 (of the total of 185) samples have been removed due to lack of mineral data because leaving them in for analysis would have skewed the statistics. Data has been normalised to the internal standard, biochanin A. Clusters on the heatmap are identified by groups of common colour as indicated by similar values. As a result of the hierarchical clustering, the rows in the heat map have been reordered to correspond to the cluster calculation. Those placed in the same cluster are indicative of results that are similar, although not statistically significant.

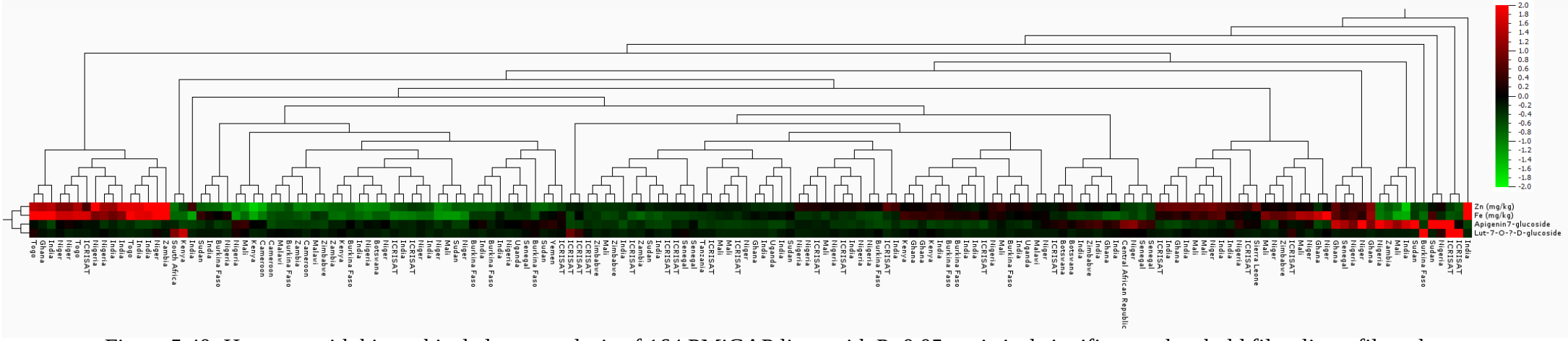


Figure 5.40: Heatmap with hierarchical cluster analysis of 164 PMiGAP lines with P<0.05 statistical significance threshold filter lines, filtered by genotype built under the Qlucore Omics Explorer, v3.2 environment.

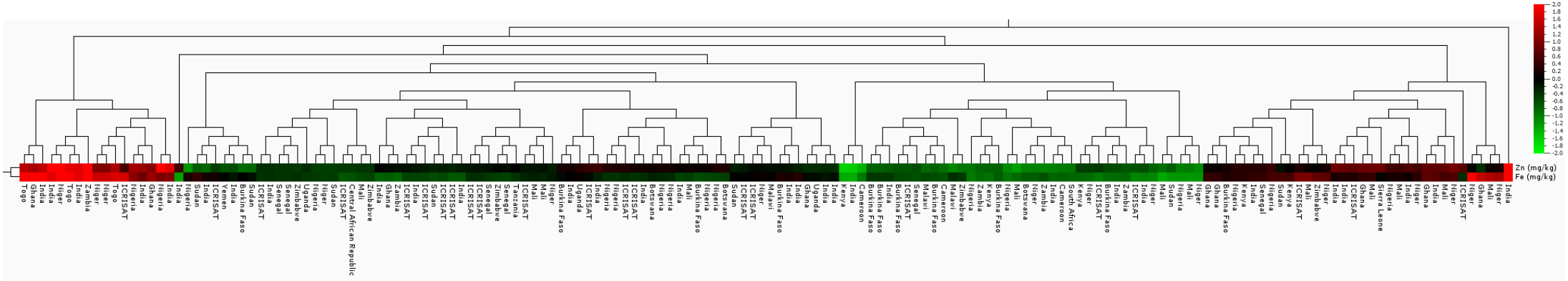


Figure 5.41: Heatmap with hierarchical cluster analysis of 164 PMiGAP lines with P<0.05 statistical significance threshold filter lines built under the Qlucore Omics Explorer, v3.2 environment.

Figures 5.38 and 5.39 provide information about the relationships between the occurrence of compounds in the extracts under investigation and compare all data from all PMiGAP lines tested together. Even though no statistical filtering was applied in Figure 5.38, findings revealed that in some cases apigenin compounds clustered together and luteolin compounds also clustered together with respect to abundance, when filtered by genotype. When a $P<0.05$ statistical significance threshold filter was applied (Figures 5.40 and 5.41), the data suggests that there was significant clustering for levels of Fe, Zn, apigenin-7-glucoside and luteolin-7-glucoside, when the data was filtered by genotype. Data also suggests that luteolin-7-glucoside was negatively correlated with Fe and Zn, which is interesting as the opposite was revealed from the 55 PMiGAP lines analysed previously (Section 5.6.3). In addition, apigenin-7-glucoside was negatively correlated with Fe/Zn.

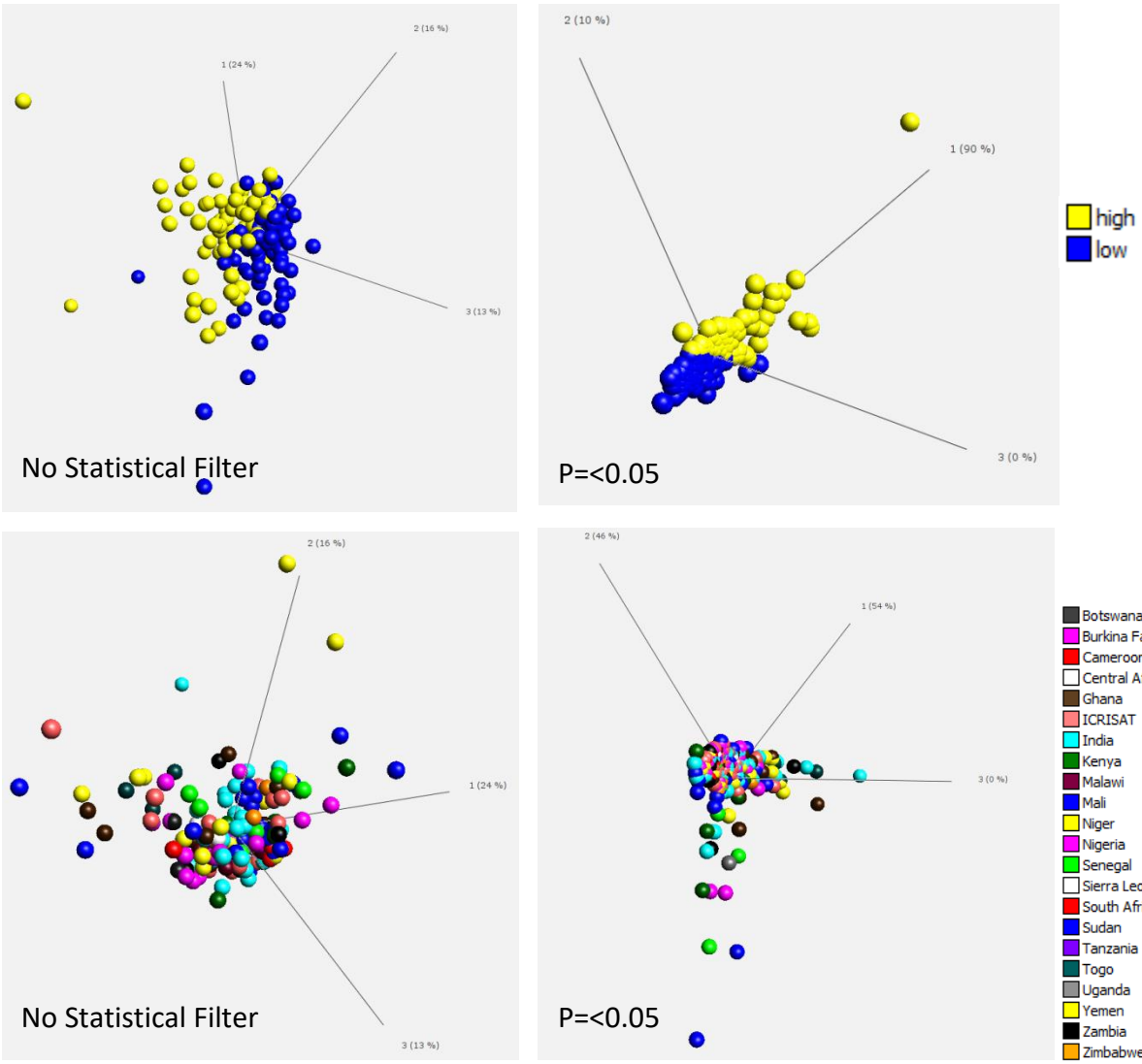


Figure 5.42: Top- 3D-PCA plots to show variation between 164 PMiGAP lines based on high/low Fe and Zn, and phenotype data lines. Bottom- 3D-PCA plot to show variation between 164 PMiGAP lines based on origin and phenotype data lines. All PCA's were built under the Qlucore Omics Explorer, v3.2 environment.

Figure 5.42 suggests that when the data was normalised to the internal standard, there was significant clustering between lines that had high/low Fe and Zn. This was confirmed by a Pearson's product moment correlation of 0.8 ($P<0.05$). However, there was no significant clustering when the data was filtered by origin ($P<0.05$).

On the basis of low total phenolics content, a set of lines were identified (Table 5.16), which can be used in crossing programs for the creation of elite lines with low apigenin/luteolin content and high Fe/Zn levels. Interestingly, 70% of lines are of African origin, which may suggest that lines from this region may be naturally lower in apigenin/luteolin compounds, as compared to the Indian lines. This finding

coincides with the result of the targeted analysis of 55 PMiGAP lines (Section 5.6.3, Table 5.13).

Table 5.16: Proposed lines for the creation of elite lines. These were selected based on a phenolic content below the average total phenolic content of 185 lines analysed (average content = 564385.35AU), coupled with high Fe/Zn content. High Fe/Zn content was defined as above an average of 96.3mg/kg, for the total 185 lines analysed.

Line (IP No.)	Origin	Total Phenolics content* (% of average)	Combined Fe/Zn (mg/kg)
IP 6869	Kenya	40	105.52
IP 7941	ICRISAT	40	99.8
IP 6102	Niger	41	111.05
IP 10701	Mali	41	106.13
IP 9446	Ghana	44	118.26
IP 5272	Niger	45	119.93
IP 9692	Nigeria	48	105.99
IP 4828	India	48	111.57
IP 9532	Ghana	49	114.39
IP 3732	India	49	99.1

*(Total phenolics content per line / 564385.35×100).

5.7 Conclusions

Over many years, flavonoids have been investigated and quantified in a vast number of crops. Similar to this study, major crops including rice, wheat, barley and maize were also found to be rich in apigenin/luteolin aglycones/glycosides (Kim *et al.*, 2008; Olenichenko *et al.*, 2006; Markham *et al.*, 1998; Casas *et al.*, 2014). Apigenin/luteloin compounds have also been studied in a variety of lesser known edible plants. These were characterised in a study by Miean & Mohammed (2001), where out of 62 edible plants, apigenin and/or luteolin was present in 21.

Due to their abundance in crop species, it is important to increase our understanding of how apigenin/luteolin compounds interact with micronutrients, for food and nutrition security purposes. It has been documented that they interact with micronutrients in the gut lumen, as they form stable complexes with non-heme dietry Fe and Zn, thus limiting the absoption of these important micronutrents into the gut. Therefore, it is advisable that population groups that are susceptible to Fe/Zn deficiency should avoid the intake of flavonoid rich foods (Scalbert *et al.*, 2002). Naturally, this cannot be avoided in communities dependent on pearl millet, therefore decreasing flavonoid content via biofortification is an appropriate solution to subsequently increase Fe/Zn intake and to enchance the nutritional credentials of

pearl millet. This research will contribute to the development of new pearl millet lines to incorporate in diets with optimal health benefits.

A major finding from this chapter is that: apigenin and luteolin compounds were found in large abundance, as compared to all other phenolic compounds. Furthermore, most of the flavonoids present in extracts were attached to sugars (glycosides), although occasionally they were found as aglycones. The presence of glycosides may further decrease mineral absorption, as compared to aglycones due to their increased molecular weight, as previously discussed.

There are limited research studies acknowledging the antinutrient qualities of apigenin and luteolin flavonone compounds, present in pearl millet. Many studies instead focus on the benefits of such compounds, in relation to their antioxidant properties as they may protect against many degenerative diseases including; heart disease, many types of cancer, gastrointestinal problems and inflammation. This is achieved in part by reactive oxygen species (Dykes & Rooney, 2007).

Although it is clear that polyphenolic compounds are of great importance in all vegetables, fruits and crops, their relationship with Fe and Zn must be acknowledged as detrimental. This is important when improving the micronutrient bioavailability status of pearl millet for human nutritional benefits. One of the few studies that recognises the adverse effects of phenolic compounds on Fe/Zn has concluded that in order to ensure a food secure future, appropriate processing methods (including soaking, fermenting and pearling) are necessary for the improvement of micronutrient bioavailability and storage stability of millet flour (Rani *et al.*, 2018). As previously discussed, polyphenolic compounds predominantly exist in the pericarp, alurone and endosperm layers of the seed, thus targeting these layers would be beneficial. Whilst this is a suitable method for improvement, another path to explore would be to decrease apigenin/luteolin compounds in the grain, whilst keeping Fe/Zn levels high using modern plant breeding methods. As a result of the work presented in this chapter, individual PMiGAP lines for the improvement of Fe/Zn bioavailability in pearl millet have been identified and may be taken forward for this purpose. Although the research is in its preliminary stages, an extremely robust method for the extraction of such compounds, from milled raw grain has been

developed for phenotyping purposes. Results are promising thus far and will contribute to pearl millet improvement and pave the way for future endeavors.

If this work were to be followed up, GWAS could be conducted and markers associated with reliable and robust trait data could be utilised for the mining of candidate genes associated with apigenin/luteolin reduction. In the longer term, to test the bioavailability of Fe/Zn in subsequent elite hybrids, resulting from plant breeding methods, simulations of gastro-intestinal digestion may be used to estimate *in-vitro* Fe/Zn availability from seed, taking into account the complex physiological conditions of gastric digestion, followed by the simulation of intestinal absorption, using the Caco-2 cell culture model. This could be performed to confirm apigenin and luteolin as antinutritional factors and also their mineral interactions in more detail.

Chapter 6: Population Structure, Linkage Disequilibrium (LD) and Genome Wide Association Studies (GWAS) to Identify Genomic Regions Associated with Increased Mineral Content

6.1 Summary

The key to understanding the genetics behind important phenotypic traits in pearl millet is genome wide characterisation of diverse germplasm panels, with high marker density. Genetic variations were identified in PMiGAP lines by GBS. These variations acted as genetic markers (SNPs) – the tools needed for GWAS. Prior to GWAS, population structure and LD were accounted for to control confounding factors and false positives that may arise as a result.

Three sets of markers were generated for this study. The first resulted from GBS libraries prepared in house from PMiGAP DNA extractions. Next generation DNA sequencing and preliminary data analysis was performed to generate GBS sequence reads, which were mapped to a foxtail millet (*Setaria italica*) genome because at the time, no pearl millet genome assembly was available. Subsequently, 663 SNPs were generated for GWAS. Although the number of resulting SNPs was small, the analysis still provided important insights into the methods used for the dissection of mineral accumulation at a molecular level.

In September 2017, a pearl millet genome assembly and 32,901,665 SNPs became available for public use (Varshney *et al.,* 2017). The SNPs were downloaded and filtered into the two remaining data sets, which contained >37,000 SNPs (as a result of stringent filtering) and >3,000,000 SNPs (as a result of less stringent filtering), respectively.

Population structure was insignificant in all cases, which was the pattern expected from inbred lines derived from a highly allogamous species. The extent of LD was also assessed among all pairs of loci. In the 663 SNP data set 2.08% of the total marker pairs were in LD based on r^2 values. LD also decayed extremely rapidly, which may be attributable to the relatively low number of markers. LD was more prominent in the >37,000 SNP data set, especially on chromosomes 3 and 5, and decayed at a slower rate, indicating an adequate number of markers for GWAS.

Three GWAS were conducted to identify markers associated with mineral uptake using ICP-AES data generated in Chapter 3. From the 663 SNP data set, just 12 marker trait associations (MTA's) were identified, whereas 199 and >1000 were identified from the >37,000 SNP data set and the >3,000,000 SNP data set, respectively with much lower P-values. 35 SNPs that had the lowest P-values from the 3 GWAS were located by position and chromosome number on their respective genome assemblies. A 4kb region surrounding each SNP was BLASTed against the most annotated *Setaria italica* reference genome, for the identification of candidate genes.

A BLAST search of the NCBI database revealed a large number of hits, representing a vast number of candidate genes of which functions included; defense against biotic/abiotic stress factors, growth, development and regulation of bioactive metabolites. Many of these contained metal binding sites including, Zn fingers and heme groups. This suggests that Fe and Zn may be used as co-factors in these pathways or they may regulate genes through cellular Fe/Zn ion changes. Furthermore, some candidate genes were found to be directly associated with mineral uptake across all SNP data sets including: V-type proton ATPase subunit D, DETOXIFICATION 16, ERECTA and YUCCA11. YUCCA 11 was detected from the 4kb region of the MTA that had the lowest P-value (P = 6.84E-06) and will be taken forward for verification by haplotyping.

6.2 Introduction

Rapid human population growth on a global scale is boosting the demand for a corresponding increase in crop grain yield, coupled with better nutrition credentials for a food secure future. Understanding the molecular genetic control of useful traits, such as yield and nutrition quality, remains a major challenge in the genetic study of staple cereal crops (Jin *et al.,* 2010). Historically, genetic characterisation of pearl millet has lagged behind other cereals due to its orphan crop status and lack of annotated reference genome assembly, prior to September 2017 (Varshney *et al.,* 2012). To address this challenge, the literature reports several robust linkage maps and QTL mapping studies conducted on grain and stover yield, height and biotic/abiotic stress traits (Jones *et al.,* 2002; Kannan *et al.,* 2014; Sehgal *et al.,* 2015). High density markers are needed for QTL fine mapping, molecular breeding

and MAS, as previously discussed in Chapter 1 (Wu *et al.,* 2014). Several studies have been conducted with the aim of understanding the genetic diversity of pearl millet germplasm panels using: Restriction Fragment Length Polymorphism (RFLP), Amplified Fragment Length Polymorphism (AFLP), SNP and SSR markers (Bhattacharjee *et al.,* 2002; Vom Brocke *et al.,* 2003; Bertin *et al.,* 2005; Sehgal *et al.,* 2015). Despite the lack of a reference genome, these studies provided valuable insights into population structure and diversity. However, it was often the case that these were targeted to germplasm from specific regions (Stich *et al.,* 2010), or were conducted on global germplasm but with relatively low marker density (Oumar *et al.,* 2008) Therefore, genome wide characterisation of worldwide pearl millet germplasm panels with increased marker density is required to gain a more holistic understanding of genomic resources in this important crop.

As previously discussed, NGS platforms have been used in many staples, including rice, wheat and maize to identify useful genomic variations and to develop genetic maps (Huang *et al.,* 2012; Jiao *et al.,* 2012). In order to facilitate sequencing to genotype vast populations, reduced representation sequencing approaches have been developed, including restriction site-associated DNA sequencing (RAD-seq) and GBS (Hu *et al.,* 2015). Prior to the sequencing of the pearl millet reference genome, the development of new markers required a procedure that did not depend on sequence information. GBS is one approach that results in genome-wide SNP markers, even in species without a reference genome. This method involves; the digestion of DNA with restriction enzymes, ligation of unique barcoded adapters, PCR amplification and the sequencing of resulting pooled libraries (Elshire *et al.,* 2011) (Figure 6.1).

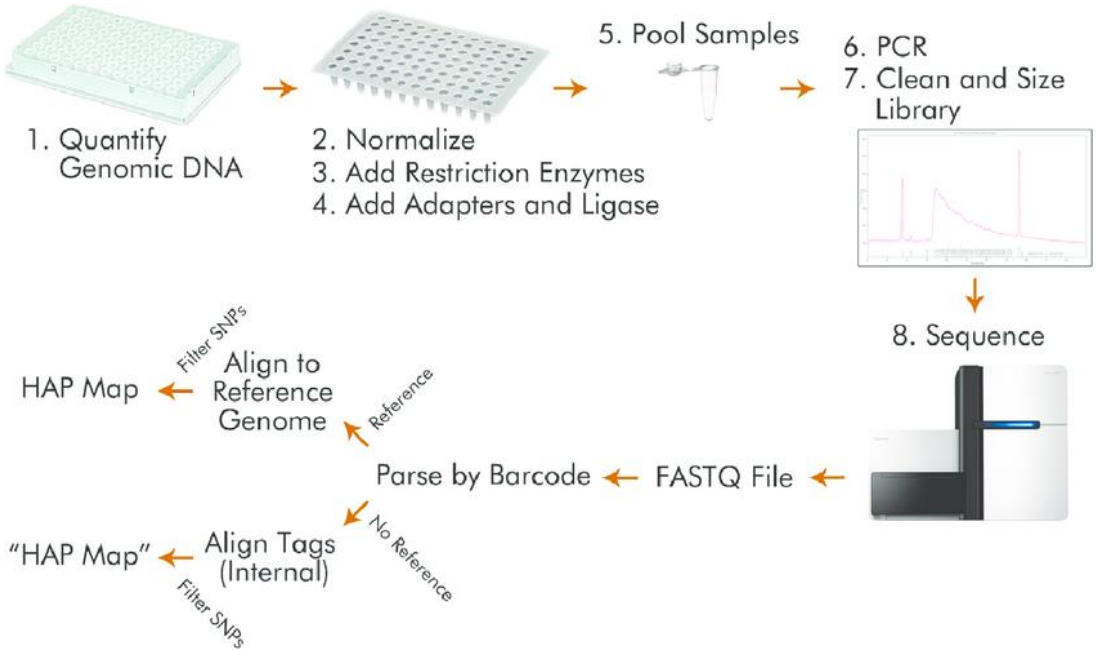


Figure 6.1: Schematic overview of GBS library preparation, sequencing and analysis. (1) Genomic DNA is quantified. (2) Genomic DNA is normalised in a new 96 well plate to ensure equal representation of samples and equal molarity of DNA/adapters. (3) A master mix with a restriction enzyme and buffer is added to the plate and incubated. (4) The DNA barcoded adapters, ligase and ligation buffers are added. (5) Samples are pooled and AMPure cleaned. (6) The GBS library is PCR amplified. (7) The amplified library is AMPure cleaned. (8) Libraries are sequenced. (9) Data analysis: FASTQ files containing raw data are used to parse sequencing reads to samples using the DNA barcode sequence. Once assigned to individual samples, the reads are aligned to a reference genome. In the case of species without a reference genome, reads are internally aligned (alignment of all sequence reads with all other reads from that library) and SNPs are identified from 2 bp sequence mismatch. Various filtering methods can then be applied to distinguish true SNPs from sequencing errors (Poland and Rife, 2012).

GBS facilitates high density SNP discovery and genotyping for vast numbers of diverse lines at relatively low cost and has been successfully used to genotype diverse germplasm panels and facilitate GWAS in many crops. For example, a high-density genetic map for pearl millet was constructed using GBS on a bi-parental population (Moumouni *et al.*, 2015) and a GBS survey of a diverse pearl millet germplasm has also been described by Hu *et al.*, (2015), where GBS was applied to 500 pearl millet accessions, including Senegalese landraces and diverse global accessions. As a result, a total of 83,875 SNPs were identified and used to characterise genomic diversity and population structure in pearl millet (Hu *et al.*, 2015).

6.2.1 Why Analyse Population Structure and LD prior to GWAS?

Understanding the genomic diversity and population structure of cereal germplasms is necessary for conservation, cultivar development, QTL mapping, and lays the foundation for association mapping (Hu *et al.,* 2015; Jin *et al.,* 2010). A key part of GWAS is controlling confounding factors and false positives that may arise as a result of population structure and family relatedness.

Understanding the extent of LD also provides information to guide GWAS and is defined as "a property of SNPs on a contiguous stretch of genomic sequence that describes the degree to which an allele of one SNP is inherited or correlated with an allele of another SNP within a population" (Bush & Moore, 2012). LD analysis determines the extent to which association mapping can be used in a species and is a sensitive indicator of the population genetic factors that structure a genome. It is used to understand the joint evolution of linked sets of genes (Slatkin, 2008). For example, the number of SNPs required for GWAS is justified by the distance at which LD has decayed, i.e. how quickly LD breaks down. The more rapidly LD breaks down, the more SNPs that are needed (Vot *et al.,* 2017). LD decay is usually determined by estimating the decay of r^2 with distance or estimating the decay of D' with distance. The r^2 value is the most typically used and is defined as the square of the correlation coefficient between two loci. It has more reliable sampling properties than D' within cases of low allele frequencies because D' is strongly inflated in populations of small sample size and with low-allele frequencies (Abdurakhmonov & Abdukarimov, 2008)

6.2.2 STRUCTURE

The software STRUCTURE (https://web.stanford.edu/group/pritchardlab/structure.htmL) is a well-established population analysis tool that allows the assessment of patterns of genetic structure in a set of samples (Porras-Hurtado *et al.,* 2013). Developed by Pritchard *et al.,* (2000), it works by detecting differences in allele frequency within data sets and assigns individuals to sub-population clusters based on analysis of likelihoods. The differences in the distribution of genetic variants within populations coupled with a complex Bayesian iterative algorithm allows individuals to be assigned to clusters and members of which share patterns of variation similar to one another (Porras-

Hurtado *et al.*, 2013). A Markov-Chain-Monte-Carlo (MCMC) estimation is applied, which assigns individuals to a pre-determined number of subpopulations. Variant frequency estimates per group then allow individuals to be reassigned. This is repeated, typically with 50,000 iterations (chosen by the user). The burnin process, which is performed separately from the MCMC estimation results in the assignment of reliable allele frequency estimation per population and membership probabilities of individuals to a population (Porras-Hurtado *et al.*, 2013). Individual analysis is performed for each assumed population number, from one to an appropriate number of K (subpopulations). Usually the range of K is unknown by the user, although this parameter must be pre-selected. This is usually rationalised by calculating the likelihood of the data for a range of K values. As K is not an absolute value, user-defined values should be carefully considered and must account for characteristics of the sampled population. This can be extrapolated by analysing passport/origin data in many cases. It has been observed that better clusters may be generated when the most accurate K values are applied without over-estimation; therefore it is crucial to obtain the smallest K value that maximises the likelihood of data (Kalinowski, 2011).

The two ancestry models that can be applied in STRUCTURE are the no admixture and admixture models. If the origin of the population under study is unknown, the no admixture model is considered appropriate. However, admixture between populations is typical and a large proportion of individuals, particularly in plant population panels may have recent ancestors from multiple populations.

6.2.3 TASSEL (Trait Analysis by aSSociation, Evolution and Linkage)

The last decade has given rise to significant advances in genotyping technology, including a rapid increase in the number of genetic markers available for QTL mapping and MAS studies. Because of this, association analysis has become a viable platform for the dissection and visualisation of complex traits (Bradbury *et al.*, 2007). TASSEL (http://www.maizegenetics.net) (Bradbury *et al.*, 2007) provides a user friendly platform for such analysis, where confounding factors including population and family structure can be accounted for, in order to minimise the risk of false positives. This is achieved by implementing the general linear model (GLM) and mixed linear model (MLM) functions. A structured association method may

partially correct for type 1 errors (i.e. false positives) using a Q-matrix of population membership estimate, the Q method is implemented as the GLM function. The average relationship between individuals is estimated by kinship (K) and is calculated from pedigrees or an appropriate number of random markers across the genome. A combined approach, that uses the information from both Q + K, is implemented as a MLM function.

LD statistics may also be calculated and visualised graphically in triangle plots, which are useful for showing block-like LD structures. This simplifies LD mapping efforts of complex traits and can be estimated by the standardised disequilibrium coefficient, D′, as well as r^2 and P-values. Other features of TASSEL include; analysing/calculating diversity statistics, integration of trait and genotypic data, imputing missing data and calculating principal components (Bradbury *et al.*, 2007).

6.2.4 Identifying Genomic Regions for Grain Fe/Zn Content

Considerable global efforts are under way to improve the Fe and Zn content of pearl millet, through genetic enhancement. GWAS are a useful tool to decipher genomic regions for these important traits, and progress will increase intensely during the next few years, at a faster rate thanks to the recent pearl millet genome assembly (Varshney *et al.*, 2017). Some studies have already identified favourable genomic regions for grain Fe and Zn through association mapping. For example, an association mapping panel of 130 diverse lines was assessed at three pearl millet growing agro-climatic environments in India: Delhi, Jodhpur and Dharwad (Anuradha *et al.*, 2017). MTA's were analysed with 267 markers (250 SSRs + 17 genic markers). As a result, a total of 16 MTA's for both grain Fe and Zn content were identified (6 Fe + 10 Zn). 3 markers in particular; *Xpsmp* 2261 belonging to LG5 (R^2-value of 13.34%), *Xipes* 0180 belonging to LG3 (R^2-value of 11.40%) and *Xipes* 0096 belonging to LG7 (R^2-value of 11.38%) were consistently associated with both grain Fe and Zn content for the 3 locations. Most notably, the *Xipes* 0180 amplicon sequence matched a segment of the pearl millet reference genome, which was annotated as the aspartic proteinase (Asp1) gene. Although the Asp1 gene is not documented to play a direct role in grain Fe and Zn metabolism, it may be indirectly involved via other gene networks/pathways. Furthermore, stably expressing alleles linked to these three markers are promising targets for MAS.

Even though research associated with the identification of markers, and in turn genes linked with grain Fe/Zn content in pearl millet is still in its pioneering stages, this has been achieved extensively in other staples, including maize. For example, in a study by Hindu *et al.,* (2018) GWAS was conducted on 923 diverse maize lines for the identification of genomic regions associated with increased kernel Fe/Zn content, using 347,765 SNPs obtained by GBS. Findings revealed 46 SNPs associated with kernel Fe/Zn content (20 Zn + 26 Fe). Some of the MTA's identified were co-located within genes, which were previously documented to be associated with Fe/Zn uptake, transport or localisation in plants. Among the genes, one stood out and was common for both Fe and Zn - No Apical Meristem (NAC) domain transcriptional regulator super family protein. Research has shown that NAC family Transcription Factors (TFs) regulate Fe and Zn remobilisation from source organs to developing seeds associated with senescence in wheat and rice (Ricachenevsky *et al.* 2013).

6.2.5 Experimental Aims

Using SNP data from GBS libraries prepared in house (which generated 663 SNPs), and the SNP files from the recent pearl millet genome assembly, which after filtering resulted in >37,000 and >3,000,000 SNPs, respectively, the aim of this work was to conduct GWAS to identify pearl millet genomic regions associated with increased mineral content. The ICPAES data described in Chapter 3, Section 3.4.2 was used as the trait data in all cases.

Population structure was assessed as well as the extent of LD between all pairs of loci at the genome level and at the chromosome level, as measured by r^2. Differences between GLM and MLM model based approaches were also measured to determine which approach was superior at accounting for confounding effects.

Using the MLM model based approach, GWAS was conducted which resulted in the identification of SNPs associated with mineral content. Once the SNPs associated with the trait data with the lowest P-values were identified, they were located on their respective genome assemblies (*Setaria italica* and *Pennisetum glaucum*), a 4kb region surrounding the selected SNP was used to identify candidate genes. This was facilitated by a NCBI nucleotide/protein BLAST search. Hits consisting of >500 identities, found within the region of each

SNP were literature searched and function/association with Fe and Zn pathways were determined where possible.

6.3 Materials and Methods

6.3.1 Qubit – DNA Quantification

DNA from 251 PMiGAP DNA extractions (prepared at ICRISAT in 2012) was quantified using a Qubit fluorometer. The Qubit fluorometer uses fluorescent dyes to determine the concentration of nucleic acids and proteins in a sample. The purpose of this was to generate readings for the dilution calculations for the preparation of GBS libraries. DNA was quantified using a Qubit DS DNAbr assay kit. Standards and samples were prepared according to Table 6.1.

Table 6.1, Sample and standard preparation using the Qubit DS DNAbr assay kit.

	Standard Assay Tubes**	DNA Sample Assay Tubes**
Volume of **Working Solution***	190µL	195µL
Volume of **Standard** from kit	10µL	-
Volume of **DNA Sample** to add	-	5µL
Total Volume	**200µL**	**200µL**

*Working Solution = 200µL/sample Qubit buffer plus 1µL/sample fluorescent dye.
** Thin wall clear 0.5mL PCR tubes (Qubit assay tubes, Cat no. Q32856).

Samples were vortexed for 3 seconds and incubated, at room temperature for 3 minutes. Tubes were then inserted into the Qubit Fluorometer and readings were taken after the stock concentration of the initial sample was determined.

6.3.2 GBS Library Preparation

Each DNA sample was Qubit quantified and the concentration was adjusted to 10ngμL^{-1}. 10µL (100ng) of each DNA sample and 3µL of barcoded adapter was added to a 96 well plate (barcoded adapters are at 0.6 ngμL^{-1}). A different barcoded adaptor was used in each well, and these are Pst-1 adaptors. To each well, 7µL of digest mix was added. The digest mix consisted of: 3µL common adapter at 0.6 ngμL^{-1}, 2µL Cut-Smart digestion buffer and 1µL Pst-1 enzyme. This was vortexed for 5 seconds and digested at 37°C overnight in a PCR machine. 30µL of master ligation mix was added to each well (master ligation mix consisted of: 5µL ligase

buffer, 0.3µL T4 DNA ligase and 24.7µL H_2O). This was mixed by pipetting and incubated at 22°C for 1 hour. After 1 hour, the ligase was heat inactivated at 65°C for 30 mins. 10µL of each sample from two rows of the plate were combined via multichannel pipetting into a PCR 8-strip. This was repeated for each two rows into a separate 8-strip until the plate was completely pooled in this manner. Each 8-strip was then further pooled into a 1.5mL high-recovery microfuge tube. The sample was AMPure cleaned and resuspended in 50µL Qiagen EB buffer. The digestion/ligation product was Qubit quantified after AMPure clean-up and 40ng was set aside for an initial test PCR using the following: -µL pooled cleaned DNA, 25µL 2× Phusion Master Mix, 2µL of PCR Primer mix (12.5ngμL^{-1} final concentration of primer A and B mixed together at 25ngμL^{-1} each) and -µL of H_2O to get a total of 50µL. The PCR cycle was set to:

72 ºC 5 minutes
98ºC 30 seconds
18 ×
98 ºC 10 seconds
65ºC 30 seconds
72 ºC 30 seconds
72ºC 5 minutes
4ºC hold

If the PCR was judged to have worked well on the basis of agarose gel electrophoresis (bright smear ~300-100bp), then the number of cycles were reduced to 15. This is to avoid over-amplification and potential for introducing PCR-based errors into the library. The PCR product was AMPure cleaned with 100µL 80% EtOH and resuspended in 30µL Qiagen EB buffer. A 1% agarose gel was run to check that there was no adapter contamination. The concentration of the PCR was quantified by Qubit, mixed together such that the entire 96-well plate was now pooled, and the final concentration was re-Qubited.

6.3.3 Next Generation Sequencing at Floragenex, CA

GBS libraries were sent to Floragenex, CA where next-generation DNA sequencing and preliminary data analysis was performed to generate GBS sequence reads. A Quality Control (QC) report was also generated (Figure 10.1). Dr Jason Kam, IBERS, Aberystwyth University then mapped the reads to a foxtail millet (*Setaria*

italica) reference genome (https://www.ncbi.nlm.nih.gov/assembly/GCA_001652605.1). The foxtail millet genome was used because, as of August 2016, a pearl millet reference genome was not available and the literature reports a comparison of a pearl millet GBS linkage map with the foxtail millet genome, which indicated extensive regions of synteny (Hu *et al.*, 2015). Foxtail millet is also taxonomically most closely related to pearl millet (Devos *et al.*, 2000). As a result, a set of 663 polymorphic SNP markers were generated. The low number of markers could be attributable to poor quality DNA in plate 2, as indicated by the QC report in Figure 10.1.

6.3.4 Model-based population STRUCTURE analysis and PCA

Population structure in the PMiGAP was assessed using Bayesian model-based population structure analysis implemented in the software STRUCTURE v2.3, using the 663 SNP data set. K (cluster groups) values from 1 - 15 were assessed and the 'admixture' and 'correlated allele frequency' models were applied. Three independent runs were achieved for each K and the replication number was set to 20,000 for the burn-in and 50,000 for the MCMC periods. Once all the runs were finished, a zip-file containing the results was created and used as an input in the STRUCTURE HARVESTER online software programme (http://taylor0.biology.ucla.edu/structureHarvester/index.php) in order to estimate ΔK, which is an *ad hoc* measure that identifies the number of subpopulations via the estimation of the rate of change in the log probability of data between successive K values.

PCA was conducted as an alternative method of determining population structure and was performed using TASSEL v5.2.38 (http://www.maizegenetics.net) (Bradbury *et al.*, 2007). The PCA method was used to analyse population structure in the 663 and >37,000 SNP data sets. SNP markers were filtered, selecting 0.9 as the minimum proportion of sites present in order to remove data with ≥ 10% missing data. A matrix, which consisted of principle component vectors accounting for population structure was generated. Population stratification was visualised by plotting the first three PCs.

6.3.5 Linkage disequilibrium analysis

LD was estimated in the 663 and >37,000 SNP data sets only. LD was not analysed using the >3,000,000 SNP data set as this analysis was already completed by Varshney *et al.,* (2017). Using the software TASSEL v5.2.38, (http://www.maizegenetics.net), LD significance was determined with 100,000 permutations for each locus. The squared correlation coefficients values, r^2 were used to quantify LD. P-values for each r^2 estimate were obtained with a two-sided Fisher's exact test and the LD values between all pairs of marker loci were plotted as triangle LD plots to estimate the general view of genome-wide LD patterns and to evaluate 'block-like' LD structures, which represent regions of high LD.

For the LD decay plots, LD data generated for both the 663 and the >37,000 SNP data set in TASSEL v5.2.38 was loaded into RStudio (RStudio Team, 2015 http://www.rstudio.com) and r^2 values were plotted against physical distance (bp) between markers, generated from the *Setaria italica* and *Penisetum glaucum* genomes, respectively. Using a script adapted from Marroni *et al.*'s study (2011), LD decay plots were generated for each linkage group per data set, and then combined to produce genome wide LD decay plots. The decay of r^2 with distance was fitted using the Hill and Weir expectation of r^2 between adjacent sites (Hill and Weir 1988).

6.3.6 GWAS, Marker Trait Associations

MTA's were determined using TASSEL v5.2.38 (http://www.maizegenetics.net), employing both the GLM and MLM functions (both based on the kinship matrix). It must be noted that only the MLM results are described in this study, due to their superiority over GLM.

6.3.7 Analysis of 37,296 SNPs from the Pearl Millet Genome Assembly

A genome sequence of reference genotype Tift $23D_2B_1$-P1-P5 became available for public use (http://ceg.icrisat.org/ipmgsc/genome.htmL) (Varshney *et al.,* 2017). At ICRISAT in 2014, WGS and BAC sequencing was used to create libraries (10 small insert and 13 large inserts). The libraries were sequenced on an Illumina HiSeq 2000 and 520 Gbp data was generated with paired end reads ranging from 49-150 bp, representing 296× genome coverage. Two BAC libraries were constructed from the pearl millet line: Tift $23D_2B_1$-P1-P5 with average insert size ~120kb. As a result,

972 Gbp data was created by sequencing 100,608 BAC clones at 80× coverage. After filtering and correction parameters were applied, 1.49Tb sequence data was assembled into 1.58 Gbp contigs (sequences without gaps or missing data) and 1.82 Gb scaffolds (contigs with estimated gaps filled in). As a result, >90% of the genome was assembled. For assembling and ordering genomic scaffolds into pseudomolecules, three biparental mapping populations and the PMiGAP coupled with collinearity with the foxtail millet genome were used. 92.8% of scaffolds were anchored into 7 pseudomolecules (Pg 1 – 7). Each of the seven pseudomolecules was known to physically be located on one DNA molecule i.e. chromosome, and so correspond to chromosomes. Sequence data from multiple genomes (963 inbred lines of cultivated pearl millet, including the PMiGAP and 31 heterozygous wild individuals) provided a vast resource of genome wide variations including millions of SNPs, which will help in allele mining of genes with significant MTA's (Varshney *et al.,* 2017).

The pearl millet genome assembly and SNP files were downloaded from http://ceg.icrisat.org/ipmgsc/. However, upon extraction, the SNP files contained 32,901,665 SNPs, too many to realistically analyse in the TASSEL v5.2.38 environment. Under Dr Matthew Hegarty's instruction, Dr Dan Smith and Dr Vasilis Lenis (Bioinformatics division, Aberystwyth University) filtered out any missing data or what was suspected to be indels, as represented by "-". Since there were many cases where individuals were labelled as "-", it was unclear whether this meant data was missing or represented an indel. Therefore, these were treated as missing data. From these assumptions, a first round filter for loci with <10% missing data was applied. After filtering, the number of SNPs was still too large, as seen in Table 6.2, (most GUI population genetics/GWAS software including TASSEL is designed for a few thousand or tens of thousands of SNPs) therefore a further filtering step was applied based on a minimum allele frequency of <5%. As a result, rare cases (<5%) of polymorphism were filtered out based on the parameter that any SNP not matching the reference base in <5% of samples was removed. Based on the results of this, a large number of SNP were still retained so the stringency was increased to <10%.

Table 6.2 Number of SNPs retained as a result of each filtering step.

Chromosome	original calls	filtered ("-")	filtered (5%)	filtered (10%)
1	5682988	78538	73117	44599
2	5014786	61430	58586	37190
3	6321154	64870	61339	39732
4	4362915	54651	49782	31575
5	3172895	44726	42493	26570
6	5045470	51717	49579	32129
7	3301457	44181	41548	26558

Dr Matthew Hegarty (Translational Genomics Facility, Aberystwyth University) then took the <10% files and filtered the resulting SNPs to further reduce the number of SNPs and to restrict the data set to include PMiGAP genotypes only. SNPs were retained only where the number of alleles were called as identical to the reference genome within 1 standard deviation of the mean across the 376 plants sequenced at ICRISAT (typically ~30-70% of calls). This resulted in the exclusion of rarer alleles. For example, 20% for chromosome 1, all markers with 30%-70% of calls matching the reference base were used, this resulted in >6600 SNPs. For all other chromosomes, this reduced it to 5000-6000 SNPs on average. Multi-allelic SNPs were also removed as these would have complicated the TASSEL analysis, resulting in a further 2000 SNPs being filtered. In total, 37,296 SNPs from PMiGAP entries were generated in HAPMAP format. Chromosome 1 contained 6647 SNPs, chromosome 2 contained 6199 SNPs, chromosome 3 contained 6455 SNPs, chromosome 4 contained 5583 SNPs, chromosome 5 contained 5408 SNPs, chromosome 6 contained 4587 SNPs and chromosome 7 contained 2417 SNPs.

6.3.8 Analysis of 3,150,286 SNPs from the Pearl Millet Genome Assembly

Dr. Sarah Beynon (Genome and Diversity, Aberystwyth University) applied more relaxed filtering stringencies to the 32,901,665 SNPs downloaded from (ftp://cegresources.icrisat.org/). This was achieved by applying a 0.05 minor allele frequency filter and a “minimum count’ threshold, where a call must be present in 80% of individuals at a given loci. This was done in the TASSEL v5.2 command line environment. As a result, 3,150,286 SNPs were generated for GWAS on a greater scale than previously. Chromosome 1 contained 517989 SNPs, chromosome 2 contained 531723 SNPs, chromosome 3 contained 624454 SNPs, chromosome 4 contained 423250 SNPs, chromosome 5 contained 329516 SNPs, chromosome 6

contained 415923 SNPs and chromosome 7 contained 307431 SNPs. Phenotype and genotype data were analysed in a High Performance Computing (HPC) Linux environment. MLM based associations are reported due to the clean MLM QQ plot as compared to GLM for this dataset. For MLM, Multi-Dimensional Scaling (MDS) was applied for the GWAS, taking into account 5 components. Outputs of the GWAS included the results table, including P-Values and Marker r^2 values of each significant marker, per trait. The 'grep' and 'cat' functions were used to break up the large results file in the HPC Linux environment by trait (see appendices). The table was then used as an input in TASSEL v5.2.38 (GUI version) to generate QQ and Manhattan plots.

6.3.9 Bonferroni Corrected Threshold

Multiple testing was accounted for by adjusting the threshold below which a p value is considered significant. This was achieved using the Bonferroni threshold correction method at the 5% significance level. In cases where the Bonferroni corrected threshold has not been met, the lowest P-values possible have been used in further analysis.

6.3.10 NCBI BLAST

Nucleotide (n) and protein (x) Basic Local Alignment Search Tool (BLAST) searches were performed using the NCBI (National Centre for Biotechnology Information) website and the *Setaria italica* genome (in the case of BLASTn), using the standard BLAST settings to find regions of similarity between biological sequences. A hit was deemed significant if the number of identities was >500 for the BLASTn search and >40 for the BLASTx search. The software compares nucleotide sequences to sequence databases and calculates the statistical significance to detect whether the SNP falls within a gene.

6.3.11 Plant Material and Phenotype Data

See Chapter 2, Table 2.3.

6.4 Results and Discussion

6.4.1 Population Structure, LD and GWAS Using 663 SNPs Mapped to Foxtail Millet

Population Structure

The most likely number of subpopulations in the PMiGAP was determined to account for population structure. The clustering program STRUCTURE (v.2.3.4) (Pritchard *et al.*, 2000) was used to estimate the membership probability of each PMiGAP accession to a number of hypothetical subpopulations (K) and the STRUCTURE HARVESTER program was used to determine the ΔK value and to generate the Evanno table.

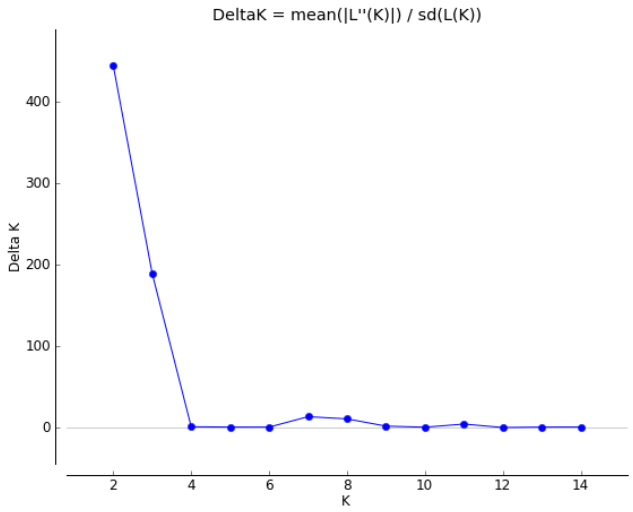


Figure 6.2, Plot of mean likelihood L(K) and variance per K value.

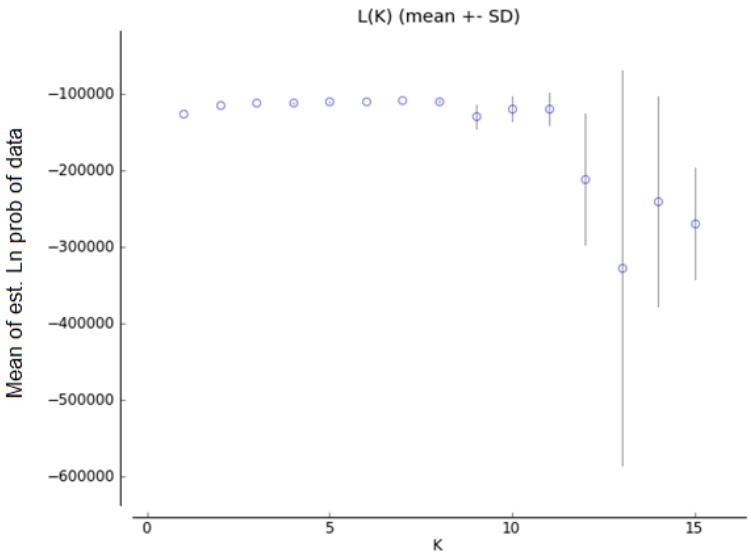


Figure 6.3, Ad-hoc statistic ΔK for K values ranging from 1 to 15.

K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	|Ln''(K)|	Delta K
1	3	-125575.600000	2.851315	—	—	—
2	3	-115428.366667	15.617405	10147.233333	6930.266667	443.752771
3	3	-112211.400000	15.463182	3216.966667	2919.800000	188.822714
4	3	-111914.233333	882.841494	297.166667	1053.066667	1.192815
5	3	-110564.000000	723.615112	1350.233333	583.066667	0.805769
6	3	-109796.833333	415.510930	767.166667	315.300000	0.758825
7	3	-108714.366667	129.046516	1082.466667	1746.700000	13.535429
8	3	-109378.600000	1824.080681	-664.233333	19742.333333	10.823169
9	3	-129785.166667	15735.193628	-20406.566667	30878.700000	1.962397
10	3	-119313.033333	16435.345972	10472.133333	10676.833333	0.649626
11	3	-119517.733333	20552.896436	-204.700000	91998.333333	4.476174
12	3	-211720.766667	85718.505855	-92203.033333	23379.000000	0.272742
13	3	-327302.800000	257645.744984	-115582.033333	203019.300000	0.787978
14	3	-239865.533333	136800.303802	87437.266667	116955.066667	0.854933
15	3	-269383.333333	72776.237737	-29517.800000	—	—

Figure 6.4, Table output of the Evanno method results. Data highlighted in yellow shows the largest value in the ΔK column.

The results from the model-based STRUCTURE analysis can be seen in Figures 6.2, 6.3 and 6.4. Analysis was performed for K populations varying from 1 to 15 and no distinct population structure was identified in 251 PMiGAP lines. The peak at K=2, as seen in Figure 6.3 and highlighted in Figure 6.4 is explained by the fact that there was enough variation to divide the population between the two clusters, but the effect is random. STRUCTURE divides the data into genetic components, meaning that if there is at least some differentiation between samples, it will divide them by the extremes and group them into 2. Figure 6.2 shows the rate of change within values of K. According to Porras-Hurtado *et al.*, (2013), "plots of K values typically progress to a plateau for levels of K beyond the most applicable number of detected populations, so the smallest stable K value represents the optimum value". Thus, the data points should creep up and plateau if population structure was present and where the data would plateau may indicate the number of K that is most likely. In Figure 6.2, the data is fairly constant, indicating no rate of change and therefore

indicative of no population structure. The lack of population structure in PMiGAP accessions is ideal for the genetic dissection of complex traits with GWAS.

The results in this study do not coincide with those of Sehgal *et al.*, (2015), who used STRUCTURE to analyse population structure in 250 PMiGAP lines, using 37 microsatellite SSR and Conserved Intron-Scanning Primer (CISP) markers and found that topologically meaningful clustering was captured at K = 6. According to Porras-Hurtado *et al.,* (2013), SNPs are binary markers that have lower variability than multiple-allele loci. Therefore, small sample sizes may be used to obtain accurate allele frequency estimates. Microsatellites require larger samples sizes than SNPs to reliably assess patterns of variability within a population. This could explain the differences in the results generated here, as compared to that of Sehgal *et al.*, (2015), in that there may not have been enough markers in Sehgal *et al.,'*s (2015) study to accurately determine population structure.

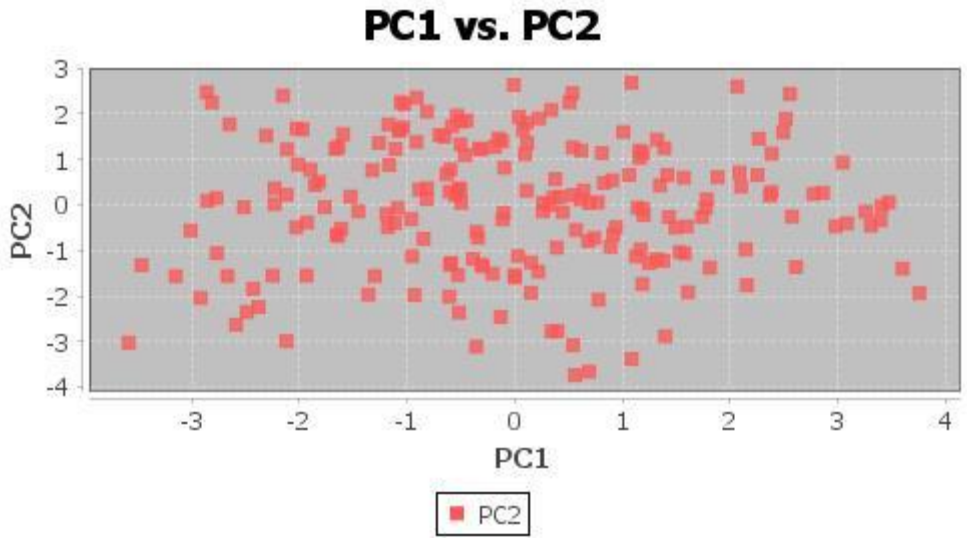


Figure 6.5, PCA plot for 212 PMiGAP lines generated in the TASSEL v5.2.38 environment.

Similarly, the PCA plot in Figure 6.5, based on SNP marker data from 212 PMiGAP lines (after filtering for 10% missing data) showed no clear population structure. The plot shows two principal components displaying no separation into sub-populations, therefore the results from the different statistical analyses were quite consistent.

The results from this study also contrast with the results from Hu *et al.,*'s (2015) study, where population structure in 500 global pearl millet accessions was

characterised using 8,377 SNPs, from which extensive geographic structure was reported and topologically meaningful clustering was captured at K = 3. Upon comparison of this study with Hu *et al.,'s* study, the larger number of SNPs (8,377 as compared to 663) from a larger population (500 as compared to the 251) is obviously more favourable. However, when comparing the origins of the samples from this study to Hu *et al.,'s* study it was found that Hu *et al*.,'s study held accessions with particularly large numbers from Senegal as compared to all other origins. Even though a smaller sample set was generated from this study, global diversity is better represented here, as seen in Figure 6.6.

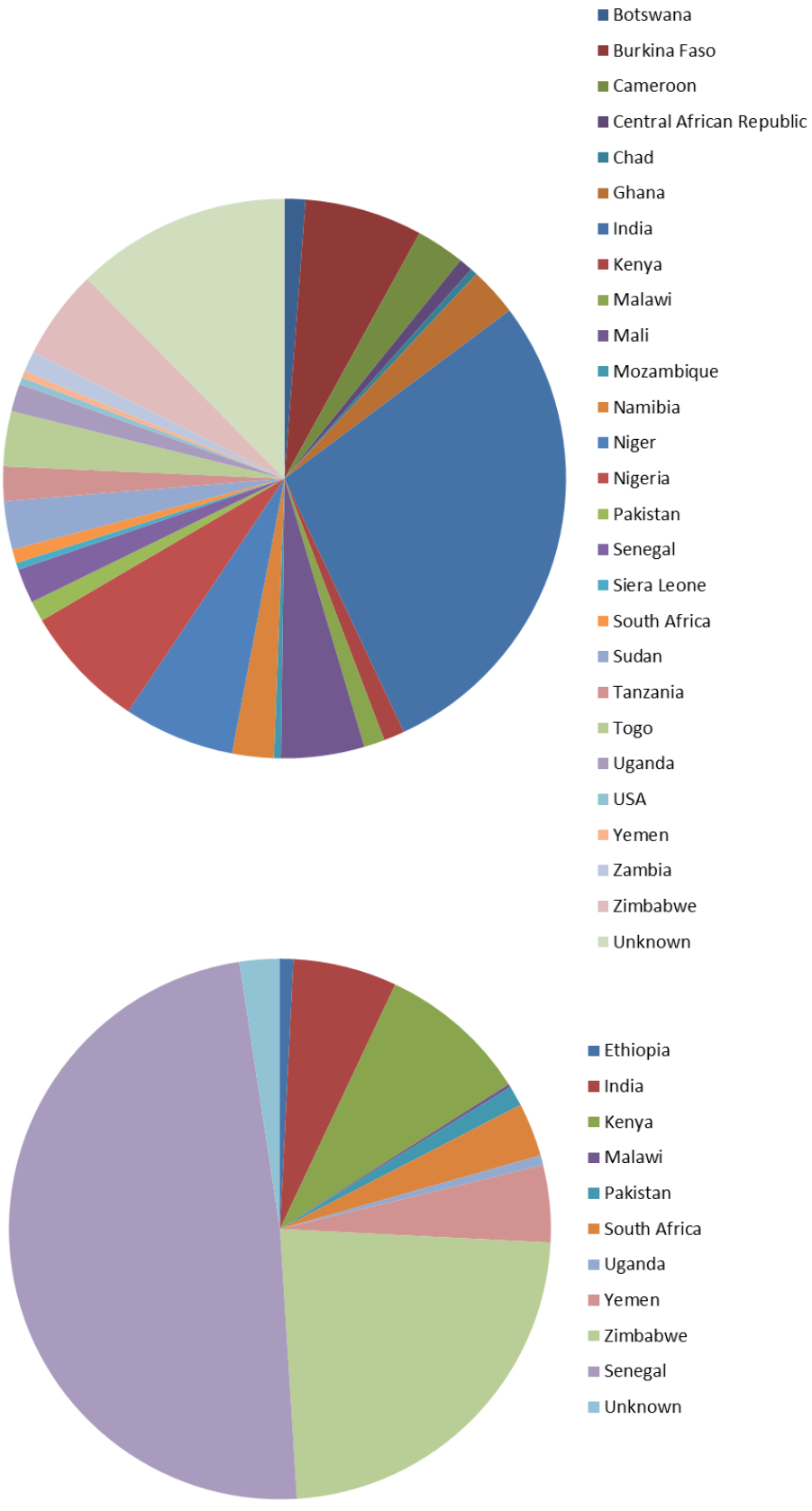


Figure 6.6, Top: The distribution of pearl millet accessions used in this study. Bottom: The distribution of pearl millet accessions used in Hu *et al.'s* (2015) study.

Linkage Disequilibrium

A LD plot shows whether any two given SNPs are inherited together. The extent of LD can be observed from chromosomes 1 – 9 (Figure 6.7), as indicated by the r^2 values and the areas of blue, green and red, which show levels of significance. Within the areas of significance, markers are significantly associated. In general, the triangle plot allows visualisation of blocks of loci (red blocks) in significant LD and any prominent block-like LD structures are of interest in association mapping studies and may simplify LD mapping efforts of complex traits.

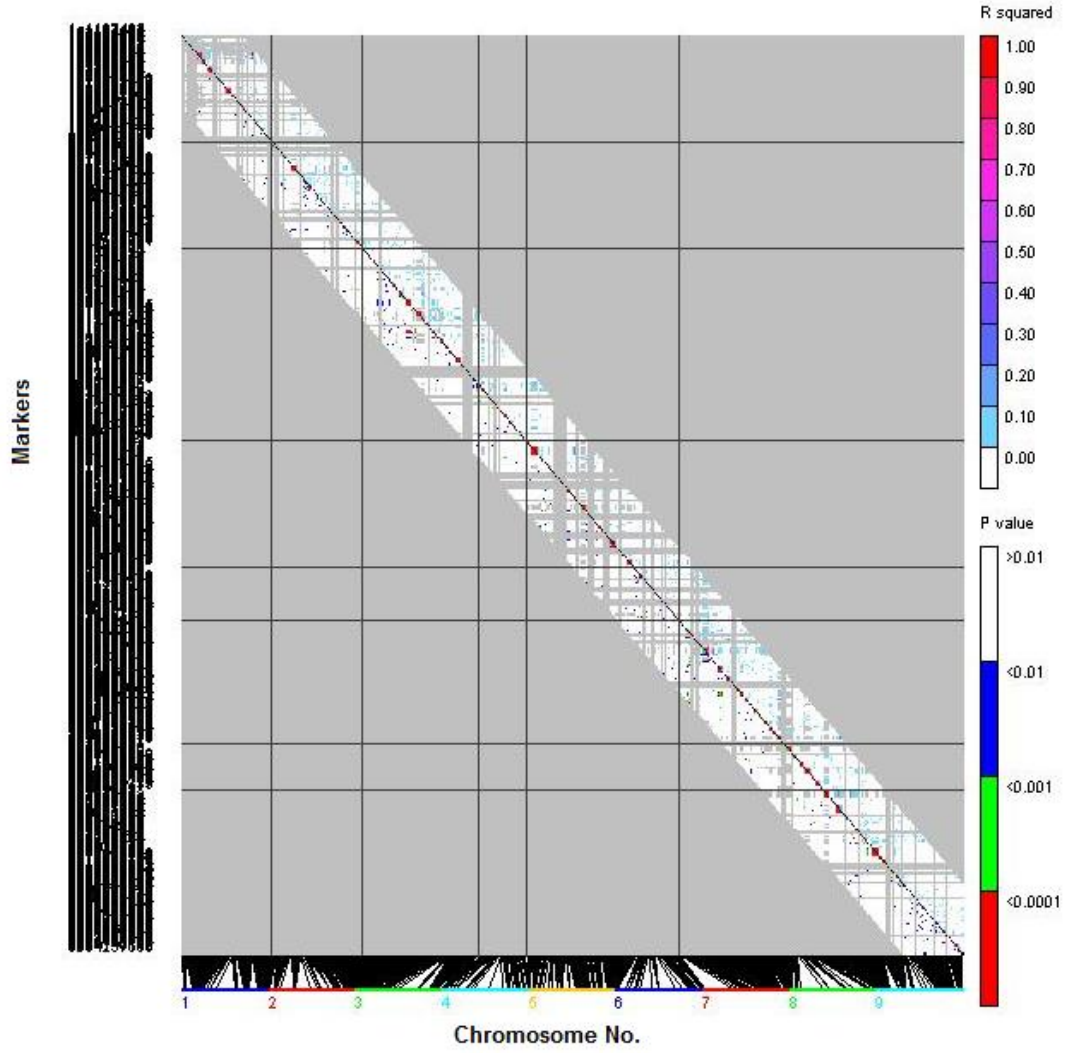


Figure 6.7, LD plot built in the TASSEL v5.2.38 environment. The squared correlation coefficient (r^2) values are denoted by a colour scale from white (0.0) to red (1.0) in the upper triangle. The p values ranging from non-significant (0.01; white) to highly significant (<0.0001; red) are shown in the lower triangle. R^2 represents the correlation between alleles at two loci, which is informative for evaluating the resolution of association approaches. The white area with coloured

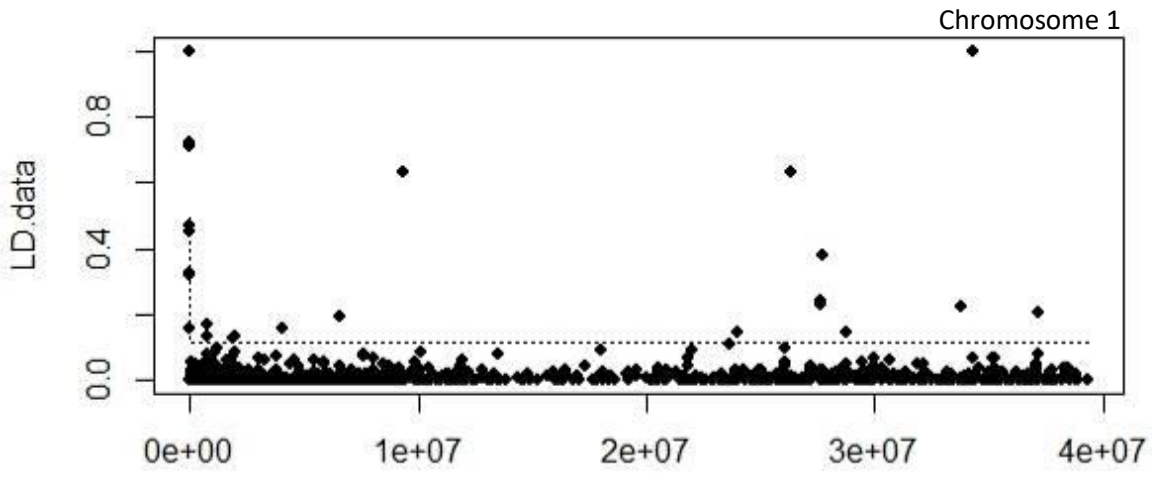
pixels represent the area that applies to chromosomes 1-9 and the lines represent the division of chromosomes.

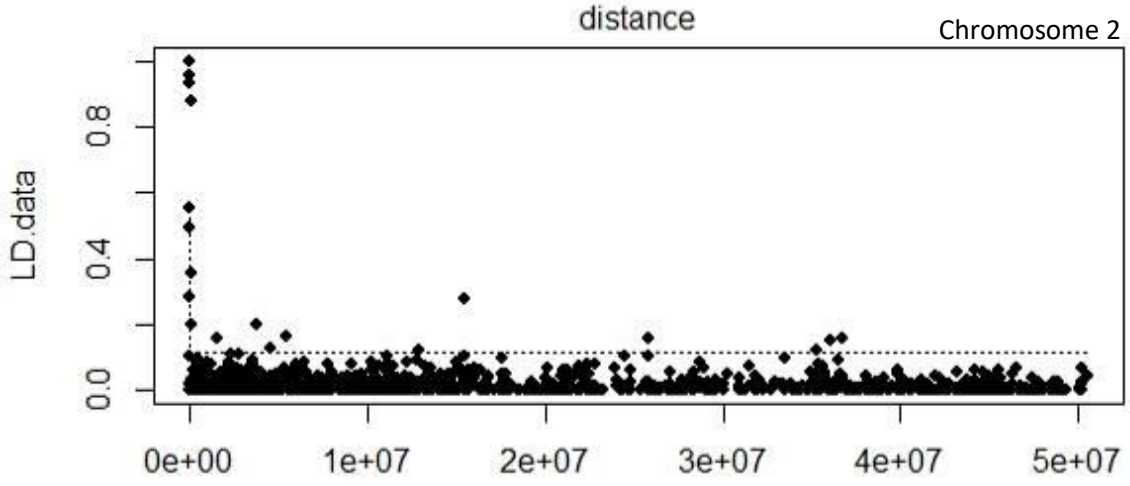
The extent of LD was assessed among 31,425 pairs of loci (Figure 6.7). Across all accessions, 2.08% of the total marker pairs were in LD (based on r^2), and no large block-like structures were observed. This may be attributable to low levels of markers. Counts for individual r^2 values can be seen in Table 6.3.

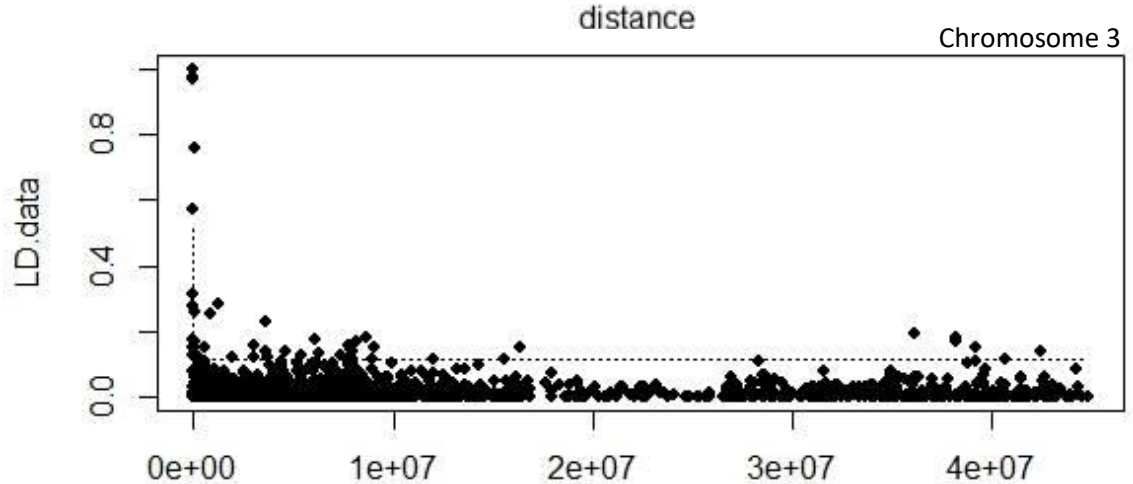
Table 6.3, r^2 results among 31,425 pairs of loci.

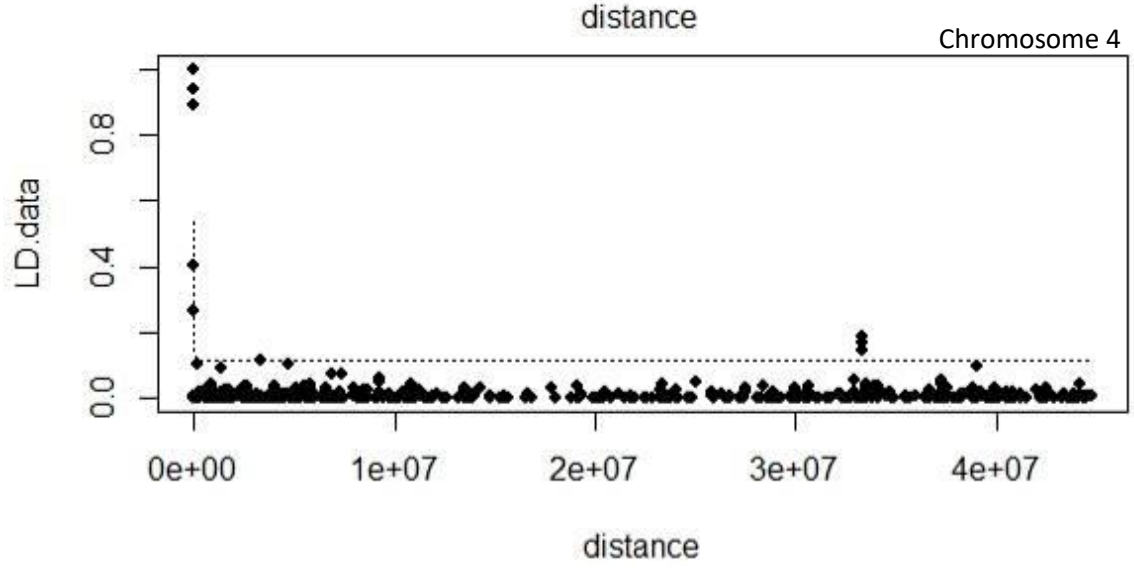
r^2 Min	r^2 Max	Count	r^2 Min	r^2 Max	Count
0	0.01	11467	**0.5**	**0.51**	**0**
0.01	0.02	3070	**0.51**	**0.52**	**1**
0.02	0.03	1512	**0.52**	**0.53**	**1**
0.03	0.04	990	**0.53**	**0.54**	**0**
0.04	0.05	591	**0.54**	**0.55**	**4**
0.05	0.06	337	**0.55**	**0.56**	**1**
0.06	0.07	239	**0.56**	**0.57**	**1**
0.07	0.08	170	**0.57**	**0.58**	**3**
0.08	0.09	160	**0.58**	**0.59**	**0**
0.09	0.1	109	**0.59**	**0.6**	**0**
0.1	**0.11**	**82**	**0.6**	**0.61**	**0**
0.11	**0.12**	**38**	**0.61**	**0.62**	**0**
0.12	**0.13**	**45**	**0.62**	**0.63**	**0**
0.13	**0.14**	**44**	**0.63**	**0.64**	**10**
0.14	**0.15**	**40**	**0.64**	**0.65**	**0**
0.15	**0.16**	**36**	**0.65**	**0.66**	**1**
0.16	**0.17**	**27**	**0.66**	**0.67**	**0**
0.17	**0.18**	**19**	**0.67**	**0.68**	**2**
0.18	**0.19**	**19**	**0.68**	**0.69**	**1**
0.19	**0.2**	**6**	**0.69**	**0.7**	**0**
0.2	**0.21**	**4**	**0.7**	**0.71**	**2**
0.21	**0.22**	**7**	**0.71**	**0.72**	**4**
0.22	**0.23**	**14**	**0.72**	**0.73**	**4**
0.23	**0.24**	**4**	**0.73**	**0.74**	**0**
0.24	**0.25**	**11**	**0.74**	**0.75**	**0**
0.25	**0.26**	**4**	**0.75**	**0.76**	**1**
0.26	**0.27**	**1**	**0.76**	**0.77**	**0**
0.27	**0.28**	**10**	**0.77**	**0.78**	**0**
0.28	**0.29**	**3**	**0.78**	**0.79**	**0**
0.29	**0.3**	**1**	**0.79**	**0.8**	**0**
0.3	**0.31**	**0**	**0.8**	**0.81**	**0**
0.31	**0.32**	**4**	**0.81**	**0.82**	**0**
0.32	**0.33**	**5**	**0.82**	**0.83**	**0**
0.33	**0.34**	**8**	**0.83**	**0.84**	**0**
0.34	**0.35**	**0**	**0.84**	**0.85**	**2**

0.35	**0.36**	**5**	**0.85**	**0.86**	**1**
0.36	**0.37**	**0**	**0.86**	**0.87**	**0**
0.37	**0.38**	**1**	**0.87**	**0.88**	**1**
0.38	**0.39**	**0**	**0.88**	**0.89**	**3**
0.39	**0.4**	**2**	**0.89**	**0.9**	**1**
0.4	**0.41**	**3**	**0.9**	**0.91**	**0**
0.41	**0.42**	**0**	**0.91**	**0.92**	**0**
0.42	**0.43**	**0**	**0.92**	**0.93**	**0**
0.43	**0.44**	**1**	**0.93**	**0.94**	**3**
0.44	**0.45**	**1**	**0.94**	**0.95**	**2**
0.45	**0.46**	**4**	**0.95**	**0.96**	**1**
0.46	**0.47**	**3**	**0.96**	**0.97**	**3**
0.47	**0.48**	**0**	**0.97**	**0.98**	**10**
0.48	**0.49**	**1**	**0.98**	**0.99**	**0**
0.49	**0.5**	**2**	**0.99**	**1**	**138**
			NaN	NaN	12124

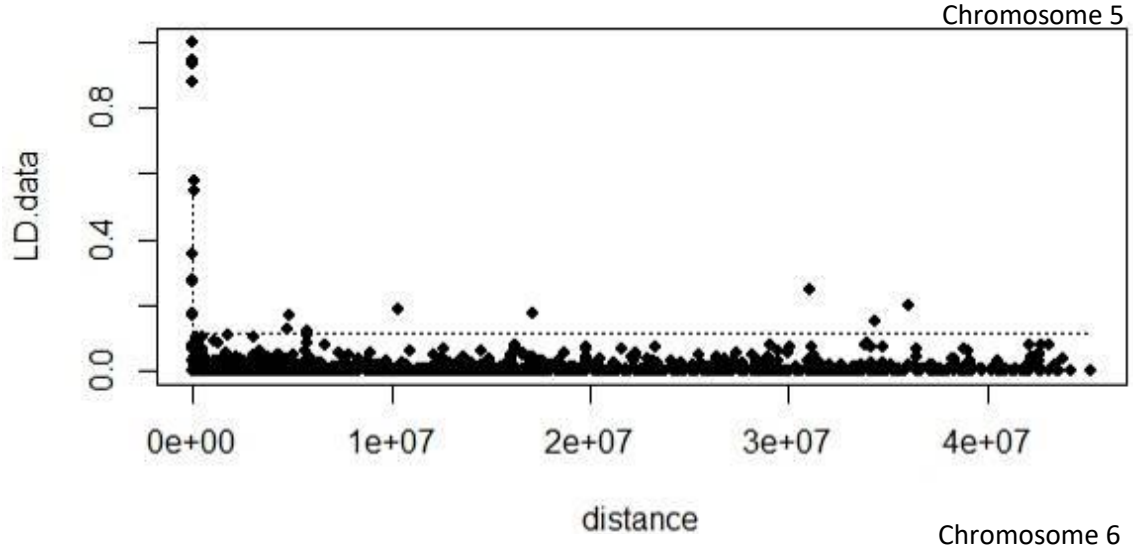

Chromosome 1
LD.data
0.0
0.4
0.8
0e+00
1e+07
2e+07
3e+07
4e+07
distance


Chromosome 2
LD.data
0.0
0.4
0.8
0e+00
1e+07
2e+07
3e+07
4e+07
5e+07
distance

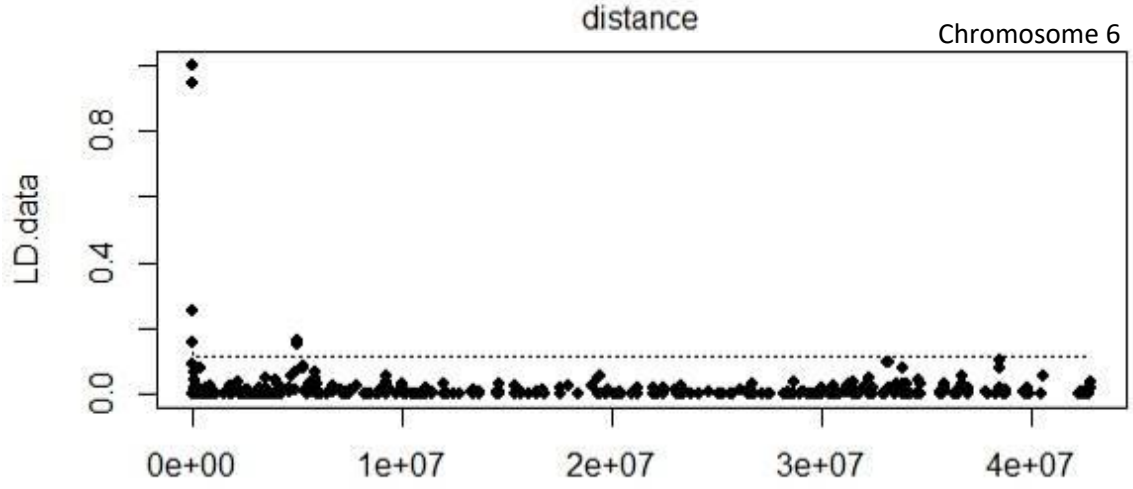

Chromosome 3
LD.data
0.0
0.4
0.8
0e+00
1e+07
2e+07
3e+07
4e+07
distance


Chromosome 4
LD.data
0.0
0.4
0.8
0e+00
1e+07
2e+07
3e+07
4e+07
distance

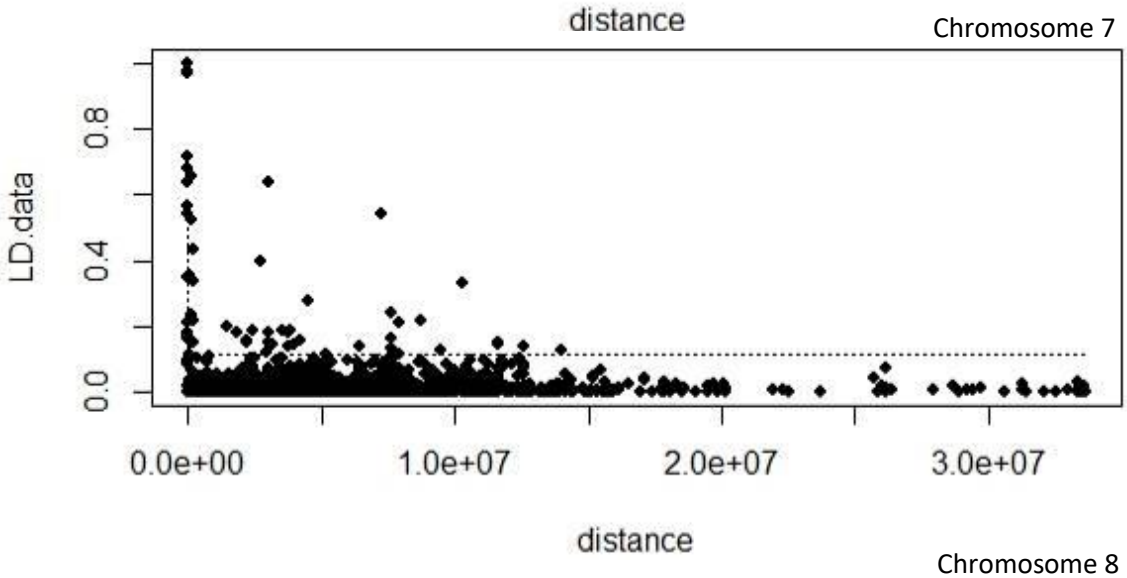
Chromosome 5



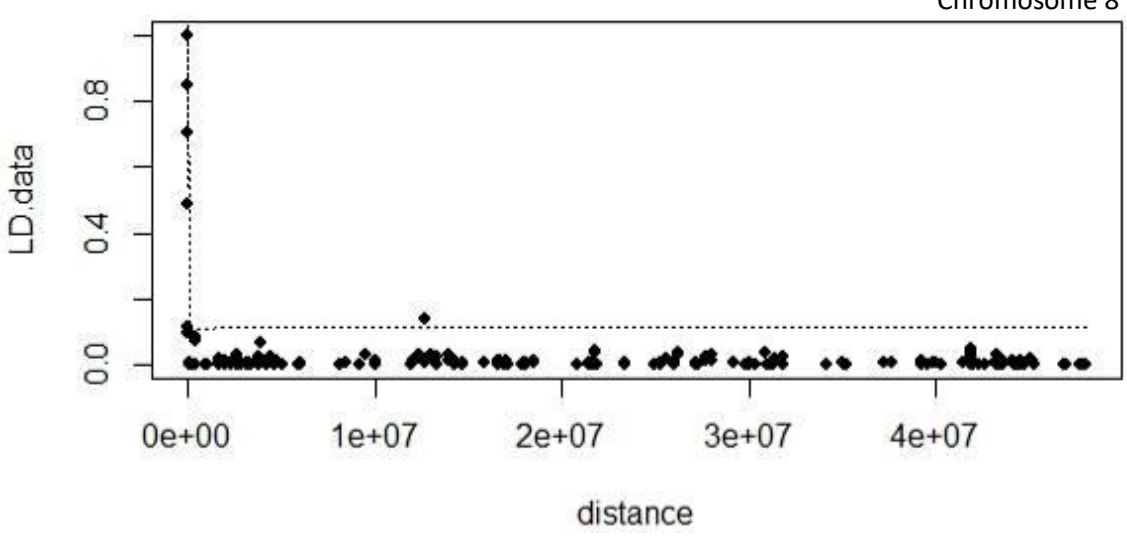
Chromosome 6



Chromosome 7



Chromosome 8



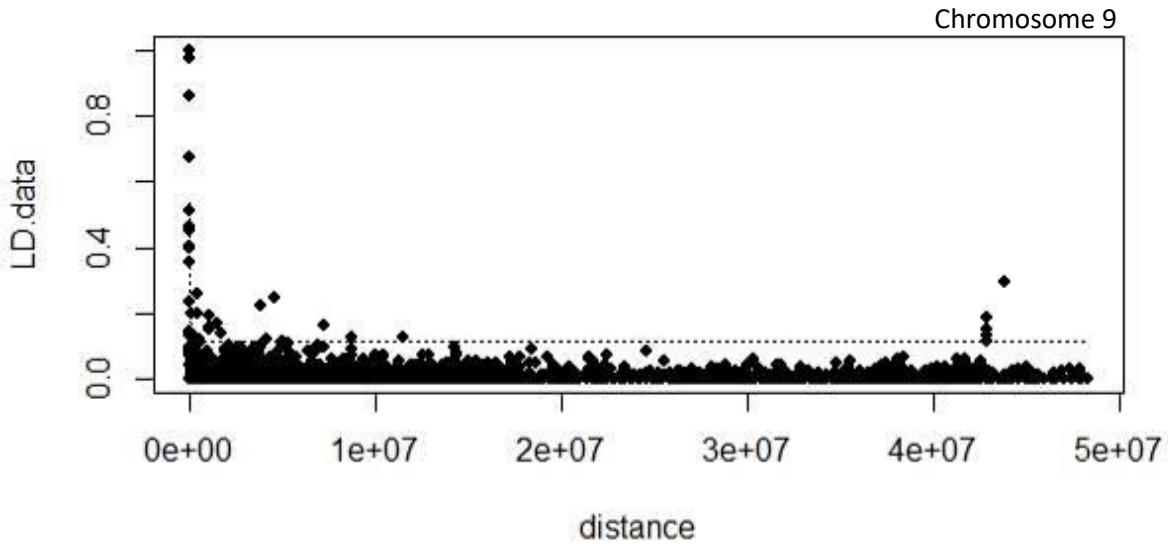


Figure 6.8, LD decay per chromosome for the *Seratia Italica* genome built in the RStudio environment. Regression line is based on Hill and Weir (1988). LD data is represented by r^2 values and distance is measured in bp. The positions of pearl millet SNPs are mapped to *Setaria italica*.

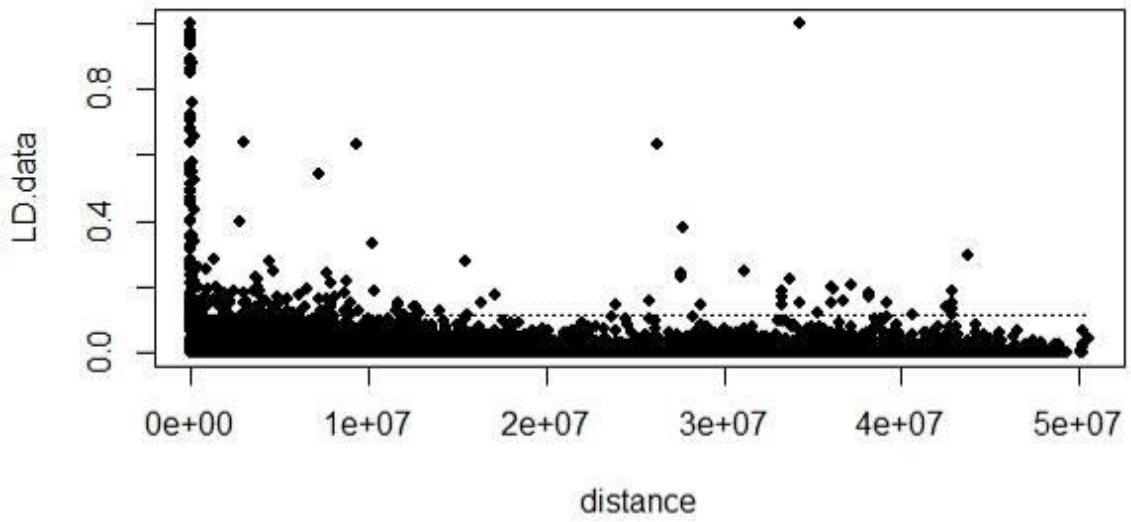


Figure 6.9, LD decay across the entire genome of *Seratia Italica* built in the RStudio environment. Regression line is based on Hill and Weir (1988). LD data is represented by r^2 values and distance is measured in bp. The positions of pearl millet SNPs are mapped to *Setaria italica*.

In general, LD decays slowly in inbred species and more rapidly in natural populations of outcrossing species, such as pearl millet. In species where LD declines rapidly, genome scans require a larger marker density; however the testing of candidate genes is still feasible (Garris *et al.*, 2003). In this study, r^2 values were plotted against distance (bp) to generate LD decay plots. A regression line based on Hill and Weir (1988) was included in each plot. From the data, it was possible to

estimate average genome-wide decay of LD by plotting r^2 values from the entire genome against distance (Figure 6.9). Alternatively, the extent of LD for each chromosome was also estimated from the LD decay plots generated, using the 'subset' function in RStudio to separate r^2 values per chromosome (Figure 6.8). When LD decay plots are generated, it is typical to observe the distance point where the LD value (r^2) decreases below 0.1 based on the nonlinear logarithmic trend line (regression line). This provides an estimate of the extent of LD for association analysis. The decrease of LD with genetic distance is indicative of the portion of LD that is conserved with linkage and proportional to recombination.

The extent of LD was investigated using 31,425 loci pairs across the genome and loci pairs from each separate chromosome (Figure 6.8 and 6.9). Pairwise LD, estimated using r^2, was found to decay extremely rapidly with genetic distance. This suggests the importance of an adequate number of markers for GWAS and that there were not nearly enough markers in this case. Some differences were observed in the extent of LD at the chromosome level. LD across the *Setaria italica* genome was seen to rapidly breakdown in all chromosomes (regression line did not extend past $r^2 = 0.1$). However, the greatest amount of LD, although minimal was seen on chromosomes 3 and 7. For chromosome 3, the LD decay distance was 0.8e+07bp for locus pairs with r^2 0.1. For chromosome 7, the LD decay distance was 1.4e+07bp for locus pairs with r^2 0.1.

Marker Trait Associations

Two approaches, GLM and MLM, were compared for all traits using the kinship matrix in both models. The QQ (quantile-quantile) plot is commonly used for GWAS to show that confounding factors aren't at work and shows the expected distribution of association test statistics (X-axis) across the SNPs as compared to the observed values (Y-axis). A clean QQ plot should indicate a solid line matching X=Y until it curves sharply at the end (representing true associations among SNPs). The QQ plots of traits, shown in Figure 6.10 suggest that the MLM model is superior at accounting for confounding effects.

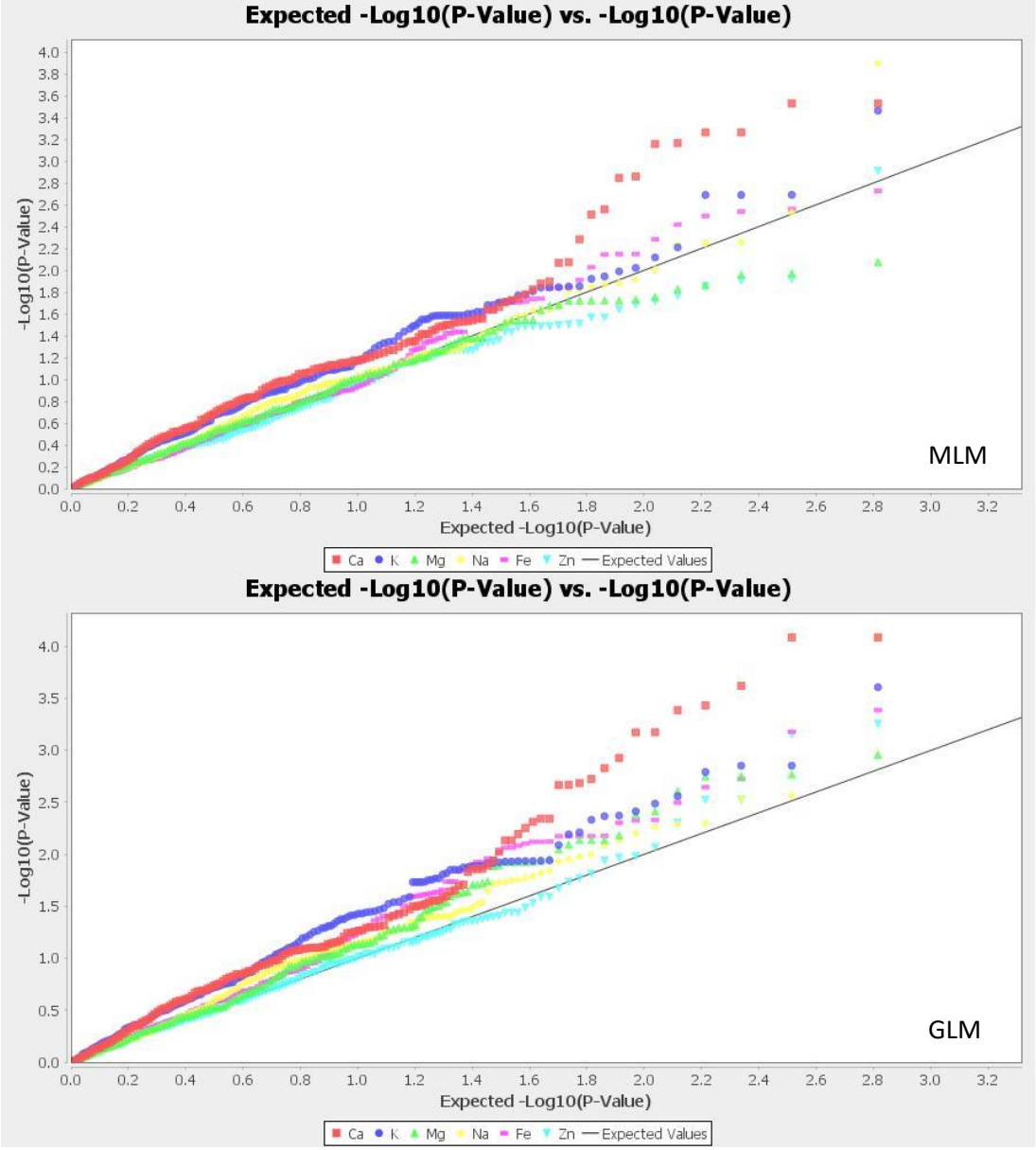
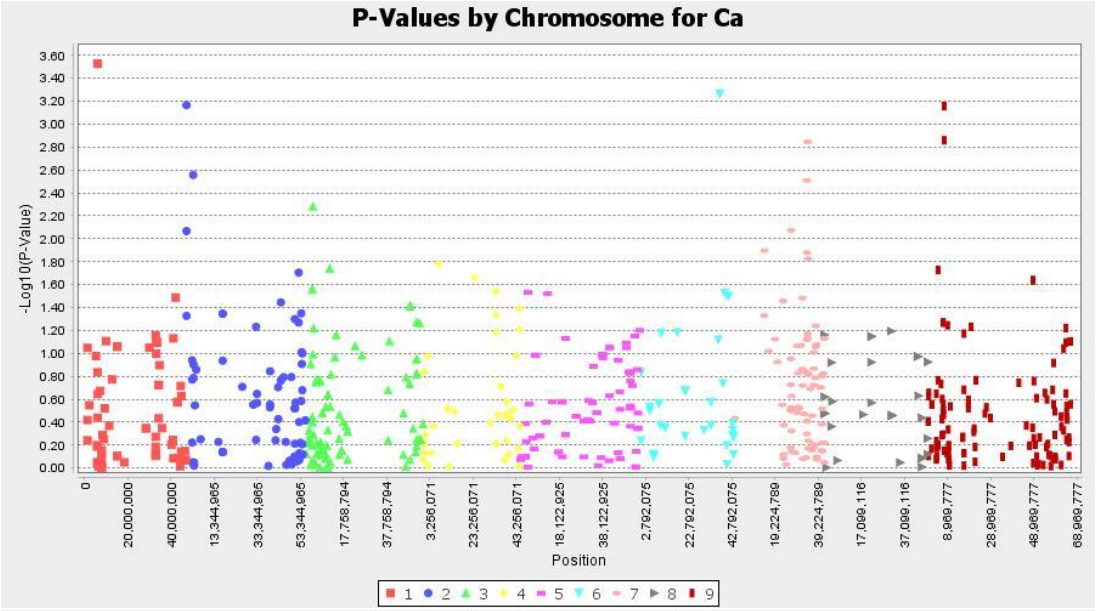
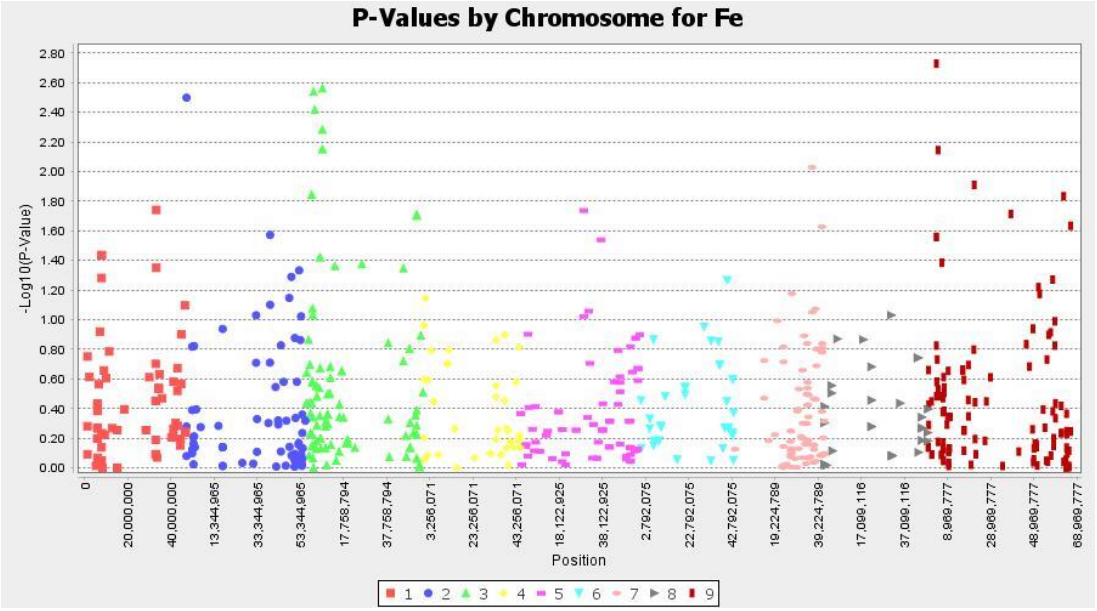
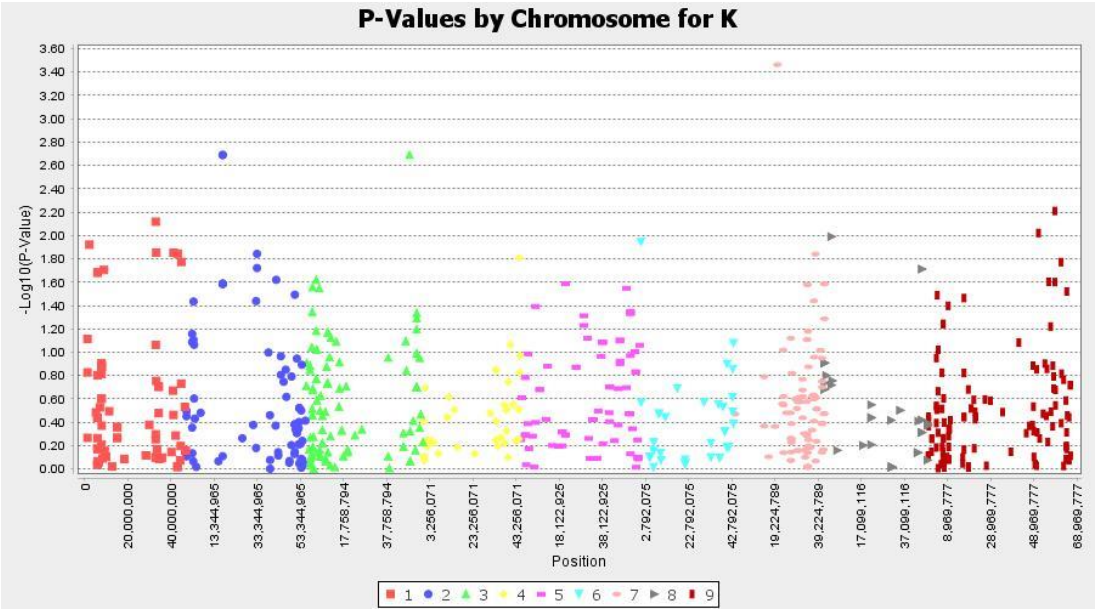


Figure 6.10, Quantile-quantile plots of Ca, K, Mg, Na, Fe and Zn using MLM and GLM models built in the TASSEL v5.2.38 environment.

Here we present results of only MLM model-based associations due to the clean MLM QQ plot in Figure 6.10 as compared to the GLM associations. When the Bonferroni threshold correction at the 5% significance level was applied, none of the MTA's exceeded the threshold, thus associations with candidate genes are explored based on the lowest P-values possible, eventhough these fall below the Bonferroni threshold for significance. Some Low P-value MTA's can be seen in Figure 6.11 and Table 6.4.


P-Values by Chromosome for Ca
-Log10(P-Value)
Position
1 2 3 4 5 6 7 8 9


P-Values by Chromosome for Fe
-Log10(P-Value)
Position
1 2 3 4 5 6 7 8 9


P-Values by Chromosome for K
-Log10(P-Value)
Position
1 2 3 4 5 6 7 8 9

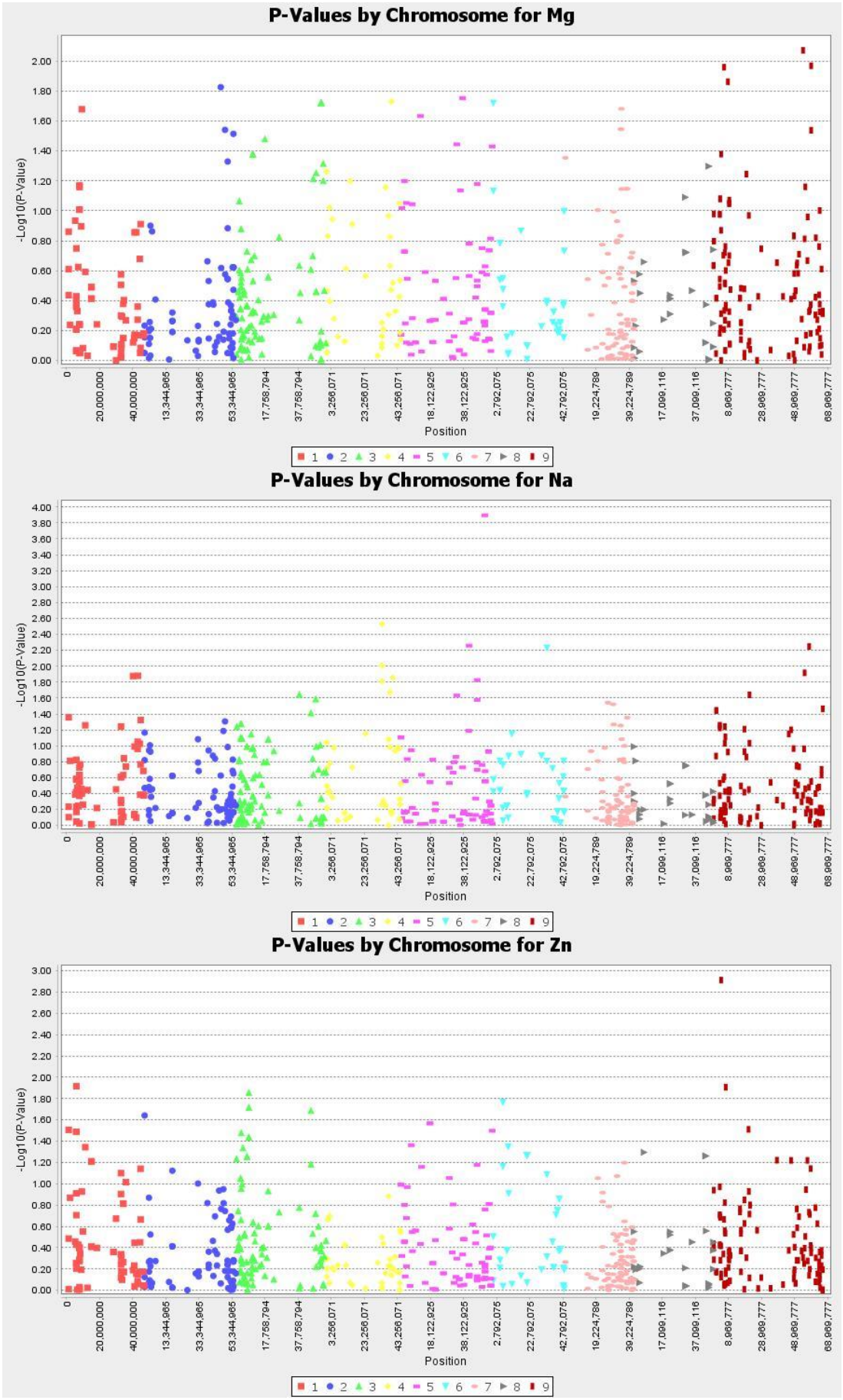


Figure 6.11, GWAS-based Manhattan plots built in the TASSEL v5.2.38

environment exhibiting the lowest P-values (measured by the MLM model) associated with seed mineral concentrations using 663 genome-wide GBS SNPs in pearl millet. The x-axis illustrates the relative density of *Setaria italica* reference genome-based SNPs physically mapped on 9 chromosomes. The y-axis displays the -log10 (P)-value for the degree of association of SNP loci with seed-mineral concentrations.

Table 6.4, SNPs associated with minerals, P=<0.002. Bonferroni corrected threshold = 1.26E-05.

Trait	Marker	Chromosome	Position	Df	F	P-Value	Marker r^2	$\%r^{2a}$
Na	5_50971121	5	50971121	2	9.416687	**1.27E-04**	0.099236	**9.92**
Ca	1_5882761	1	5882761	2	8.484362	**2.97E-04**	0.089309	**8.93**
Ca	1_5882763	1	5882763	2	8.484362	**2.97E-04**	0.089309	**8.93**
K	7_20384756	7	20384756	2	8.324543	**3.45E-04**	0.087953	**8.80**
Ca	6_37189739	6	37189739	2	7.824584	**5.44E-04**	0.082364	**8.24**
Ca	6_37189769	6	37189769	2	7.824584	**5.44E-04**	0.082364	**8.24**
Ca	2_467521	2	467521	2	7.582021	**6.82E-04**	0.079811	**7.98**
Ca	9_7373490	9	7373490	2	7.557771	**6.97E-04**	0.079555	**7.96**
Zn	9_4623914	9	4623914	2	6.952239	**0.001225**	0.073464	**7.35**
Ca	9_7373520	9	7373520	2	6.821473	**0.00138**	0.071805	**7.18**
Ca	7_34304700	7	34304700	2	6.786889	**0.001427**	0.071513	**7.15**
Fe	9_3820759	9	3820759	2	6.497722	**0.001874**	0.070313	**7.03**

[a] Percentage of phenotypic variation explained.
Phenotypic variance (cumulative r^2) explained by the genetic effects of all associated SNPs can only be calculated for Ca because only Ca has multiple MTA's (8), the value is 64.6%.

A GWAS was conducted on grain samples from field grown plants (ICRISAT, Patencheru, 2010). The MLM model-based association mapping approach identified 12 gene-based SNPs exhibiting an association (P =<0.002) with seed-mineral concentrations (Table 6.4). The percentage of phenotypic variation explained is the Marker $r^2 \times 100$ for each individual associated marker and is also summarised in Table 6.4. For the 12 MTA's, the percentage of the phenotypic variation explained ranged from 7.03% - 9.92% and the marker 50972232 on chromosome 5 for Na uptake had the highest r^2 value. According to the data in Figure 6.11, the log10 P-value scores in all cases were relatively low (the cut-off point was P=0.002), which may have been attributable to a number of factors. For example, in this study, the sample size was relatively small as compared to many GWAS in other crops reported in the literature. GWAS are generally aimed at finding very small effects; therefore they need large numbers of samples to confirm small differences with statistical confidence. For example, in barley, 298 landraces were characterised for micronutrient concentration and then genotyped with 7,842 SNPs. Results yielded

increased statistical significance as compared to this study where 191 PMiGAP lines were genotyped with 663 SNPs (Mamo *et al.*, 2014). This was also the case for chickpea, where increased statistical significance was also observed when 369 individuals were characterised for Fe and Zn concentrations and genotyped with 24,620 SNPs generated from genome-wide GBS (Upadhyaya *et al.*, 2016).

The biggest challenge of successfully carrying out a reliable GWAS is obtaining robust genotype data; this was accounted for by filtering out 10% missing data, implemented in TASSEL v5.2.38 . The original QC report by Floragenex can also be seen in the appendix.

As discussed previously, next-generation DNA sequencing and preliminary data analysis of pearl millet DNA was performed to generate GBS sequence reads. The reads were then mapped to a foxtail millet reference genome due to their close taxonomic relationship. A foxtail millet genome was used because a pearl millet reference genome was not available at the time of this research and the literature reports extensive regions of synteny between foxtail millet and pearl millet (Hu *et al.*, 2015). However, some large-scale rearrangements in the pearl millet lineage has been detected, as reported by Devos *et al.*, (2000). This may have affected LD data and the quality and accuracy of the downstream analysis, thus findings from the GWAS may be less valid than in if a pearl millet reference genome was used. A different reference genome may also be one of the reasons for the low levels of SNPs detected, simply because tags may not have matched the reference genome due to sequence variation. If too much sequence variation is present, the read will not map and therefore may not be detected as a tag.

Identifying Candidate Genes from Low P-Value SNPs

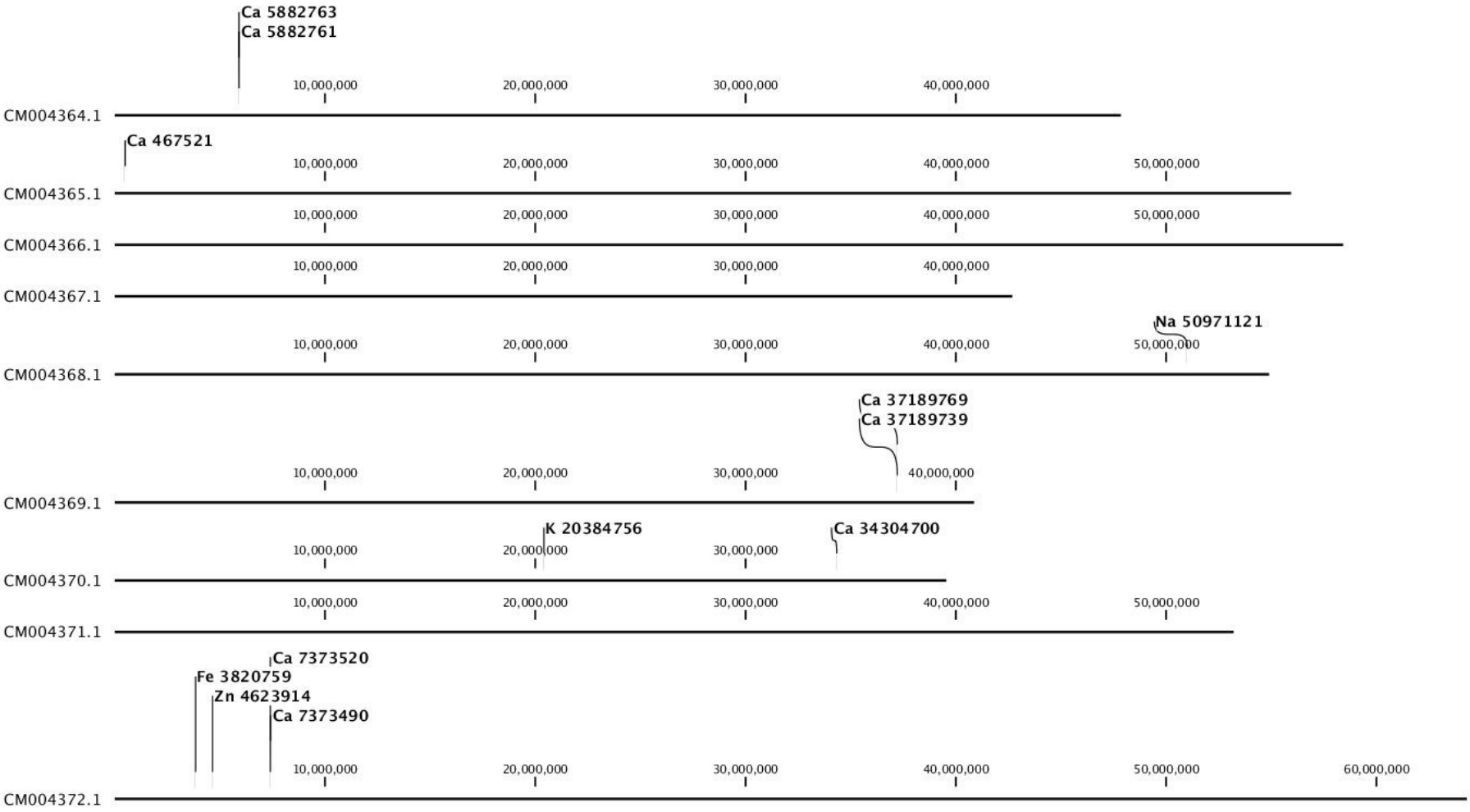


Figure 6.12, SNPs with the lowest P-values from GWAS mapped onto the *Setaria italica* genome. Chromosome numbers 1 – 9 correspond to CM004364.1 – CM004372.1, respectively. Image built in the CLC Genomic Workbench environment (v6.5).

SNPs with the lowest P-values from the GWAS were located by position and chromosome number on the *Setaria italica* reference genome (https://www.ncbi.nlm.nih.gov/genome/10982?genome_assembly_id=276542), as seen in Figure 6.12, using the CLC Genomics Workbench v.6.5 (CLC Bio, Aarhus, Denmark). A 4kb region surrounding each SNP was selected and a BLAST search of the NCBI database was conducted on each region of interest using the *Setaria italica* reference genome. Blast alignments can be seen in Table 6.5 and 6.6.

Table 6.5, BLASTn alignments of low P-value SNPs mapped onto the *Setaria italica* reference genome assembly + 4kb – 663SNP dataset.

SNP (Chromosome no._Position)	Score	E Value	Identities	Gaps	Range	Candidate Gene(s)
Na 5_50971121	2931 bits(1587)	0.0	1641/1665(99 %)	12/1665(0%)	44102337 to 44103995	1681 bp at 5' side: LOW QUALITY PROTEIN: probable carboxylesterase 16 5182 bp at 3' side: probable histone acetyltransferase HAC-like 3
Ca 1_5882761/5882763	6986 bits(3783)	0.0	3942/4014(98 %)	29/4014(0%)	5593345 to 5597342	premnaspirodiene oxygenase-like
	712 bits(385)	0.0	1074/1408(76 %)	42/1408(2%)	5585865 to 5587258	6301 bp at 5' side: citron Rho-interacting kinase-like 3313 bp at 3' side: premnaspirodiene oxygenase-like
K 7_20384756	3834 bits(2076)	0.0	2105/2119(99 %)	2/2119(0%)	20283626 to 20285743	U-box domain-containing protein 33-like
Ca 6_37189739/37189769 However, BLAST hits only detected on chromosome 5	6953 bits(3765)	0.0	3950/4031(98 %)	45/4031(1%)	43912019 to 43916034	transcription factor bHLH128-like
Ca 2_467521	7269 bits(3936)	0.0	3982/4003(99 %)	7/4003(0%)	147809 to 151806	60S ribosomal protein L13a-4-like probable LRR receptor-like serine/threonine-protein kinase RPK1
Ca 9_7373490	5081 bits(2751)	0.0	2751/2751(100%)	0/2751(0%)	6263550 to 6266300	folate transporter 1, chloroplastic
Zn 9_4623914	2959 bits(1602)	0.0	1606/1608(99 %)	0/1608(0%)	4486437 to 4488044	3839 bp at 5' side: BTB/POZ domain-containing protein At3g50780 1298 bp at 3' side: uncharacterized protein

						LOC101772810
Ca 9_7373520	5136 bits(2781)	0.0	2781/2781(100%)	0/2781(0%)	6263550 to 6266330	folate transporter 1, chloroplastic
Ca 7_34304700	7315 bits(3961)	0.0	3988/4001(99%)	1/4001(0%)	31286727 to 31290726	992 bp at 5' side: microtubule-associated protein RP/EB family member 1A-like 20506 bp at 3' side: V-type proton ATPase subunit D
Fe 9_3820759	2392 bits(1295)	0.0	1331/1347(99%)	8/1347(0%)	3704917 to 3706255	homeobox protein Hox-A13-like
	1315 bits(712)	0.0	726/732(99%)	3/732(0%)	3704010 to 3704739	6066 bp at 5' side: DNA-binding protein HEXBP-like 179 bp at 3' side: homeobox protein Hox-A13-like

Table 6.6, BLASTx alignments of low-P-value SNPs mapped onto the *Setaria italica* reference genome assembly + 4kb – 663SNP dataset.

SNP (Chromosome no._Position)	**Score**	**E Value**	**Identities**	**Positives**	**Gaps**	**Range**	**Sequences producing significant alignments**
Na 5_50971121	370 bits(950)	2e-114	240/379(63%)	259/379(68%)	49/379(12%)	64 to 399	hypothetical protein SETIT_004765mg [*Setaria italica*]
Ca 1_5882761/5882763	526 bits(1354)	0.0	286/286(100%)	286/286(100%)	0/286(0%)	26 to 311	PREDICTED: premnaspirodiene oxygenase-like [*Setaria italica*]
K 7_20384756	463 bits(1192)	2e-142	238/341(70%)	264/341(77%)	37/341(10%)	516 to 819	hypothetical protein SORBI_3006G086700 [*Sorghum bicolor*]
	415 bits(1067)	4e-125	196/200(98%)	198/200(99%)	0/200(0%)	497 to 696	PREDICTED: U-box domain-containing protein 33-like [*Setaria italica*]
Ca 6_37189739/37189769	339 bits(870)	2e-101	211/249(85%)	213/249(85%)	33/249(13%)	69 to 284	hypothetical protein SETIT_001123mg [*Setaria italica*]
	178	2e-63	125/148(84%)	127/148(85%)	13/148(8%)	113 to	PREDICTED: transcription factor

	bits(452)					247	bHLH128-like [*Setaria italica*]
Ca 2_467521	779 bits(2012)	0.0	462/464(99%)	463/464(99%)	0/464(0%)	259 to 722	PREDICTED: probable LRR receptor-like serine/threonine-protein kinase RPK1 [*Setaria italica*]
Ca 9_7373490	74.7 bits(182)	2e-10	70/100(70%)	70/100(70%)	30/100(30%)	1 to 70	PREDICTED: folate transporter 1, chloroplastic [*Setaria italica*]
Zn 9_4623914	42.0 bits(97)	0.69	17/18(94%)	18/18(100%)	0/18(0%)	81 to 98	hypothetical protein SETIT_033505mg [*Setaria italica*]
Ca 9_7373520	74.7 bits(182)	2e-10	70/100(70%)	70/100(70%)	30/100(30%)	1 to 70	PREDICTED: folate transporter 1, chloroplastic [*Setaria italica*]
Ca 7_34304700	138 bits(347)	4e-39	77/139(55%)	87/139(62%)	25/139(17%)	52 to 165	uncharacterized protein LOC8079218 isoform X1 [*Sorghum bicolor*]
	93.6 bits(231)	7e-24	45/97(46%)	63/97(64%)	4/97(4%)	416 to 512	Os03g0643433 [*Oryza sativa Japonica* Group]
	87.0 bits(214)	1e-18	40/80(50%)	52/80(65%)	0/80(0%)	523 to 602	Nitrate transporter *[Aegilops tauschii*]
Fe 9_3820759	165 bits(418)	5e-44	102/102(100%)	102/102(100%)	0/102(0%)	12 to 113	hypothetical protein SETIT_040692mg [*Setaria italica*]
	109 bits(272)	2e-22	45/49(92%)	49/49(100%)	0/49(0%)	1 to 39	PREDICTED: zinc finger protein GIS-like [*Brachypodium distachyon*]
	82.8 bits(203)	4e-15	36/43(84%)	39/43(90%)	1/43(2%)	46 to 88	PREDICTED: zinc finger protein 7-like [*Setaria italica*]

Na 5_50971121

The Carboxylesterase 16 (CXE16) candidate gene was located 1681bp upstream at the 5' side of the Na 50971121 SNP. CXE family members are short chain fatty acids and although their roles influencing signal transduction and xenobiotic detoxification in animals has been well defined for decades, possibly due to their importance in detoxifying drugs and pesticides (Gershater & Edwards, 2007), much less is known about their roles *in planta*, even though CXE activities *in planta* have been reported for over 40 years. Specifically, CXE16 is expressed in roots, leaves, stems, flowers and siliques (Marshall *et al.,* 2003). In enzymology, CXE's catalyses the following reaction:

Carboxylic ester + H_2O ⇆ Alcohol + Carboxylate

The making and breaking of carboxylic ester bonds are fundamental reactions in organic chemistry, with esters being more hydrophobic and typically less reactive than the respective alcohol and acid components. This may protect carboxylic acids from undergoing unwanted side reactions. Additionally, the hydrolysis of carboxylic esters in biological systems may have potentially important roles in regulating the synthesis and release/transport of bioactive metabolites. Members of the CXE family have also been proven to hydrolyse soluble low molecular weight natural products and xenobiotics in plants (Gershater & Edwards, 2007).

A probable histone acetyltransferase HAC-like 3 protein was located 5182 bp downstream at the 3' side of this Na associated SNP. Structurally, HAC-like 3 proteins contain 4 zinc finger regions including plant homeodomain (PHD)-type, ZZ-type 1 + 2 and transcriptional adaptor zinc (TAZ)-type, where Zn plays a critical structural role for protein stabilisation (Grotz *et al.,* 1998). HAC family genes regulate flowering time in *Arabidopsis thaliana* and a study by Han *et al.,* (2007) revealed that the transcript level of FLOWERING LOCUS C was significantly higher in *hac1*-involved mutants than in wild-type plants (Han *et al.,* 2007). Ethylene is an important plant hormone that regulates various growth dynamics and developmental events, including seed germination, seedling growth, fruit ripening and organ senescence. The function of the HAC family genes in the ethylene pathway has been investigated by Li *et al.,* (2014), where it was proven that *hac1*-

involved mutants were hypersensitive to ethylene and they also influenced the expression levels of ethylene-responsive genes.

Ca 1_5882761/5882763

Premnaspirodiene oxygenase-like was found to be associated with the 2kb region surrounding the Ca 5882761/5882763 SNPs. Premnaspirodiene oxygenase is involved in the biosynthesis of solavetivone, a potent antifungal phytoalexin (Takahashi *et al.,* 2007) by catalysing the reaction below:

Vetispiradiene + 2 NADPH + 2 O_2 ⇆ solavetivone + 2 NADP+ + 3 H_2O

Structurally, it contains a metal binding site for Fe (heme axial ligand), which acts as an inorganic cofactor, required for a protein to be catalytically active. Solavetivone is a well-known phytoalexin that is produced by potato plants in response to various stress factors, including the application of arachidonic acid (Desjardins *et al.,* 1995) and infection with *Erwinia carotovora* (Engström *et al.,* 1999). It also reduces the growth of the potato pathogen *Phytophthora infestans* (Engström *et al.,* 1999).

Additionally, a citron Rho-interacting kinase-like protein was located 6301 bp at the 5' side and is known to play an important role in the regulation of cytokinesis in both animals and plants. Citron is a 183 kDa protein that contains a C6H2 zinc finger, a PH domain, and a long coiled-coil forming region including 4 leucine zippers and a rho/rac binding site (Madaule *et al.,* 1995). Plant cytokinesis differs from animal cytokinesis, partly because of the rigidity of plant cell walls. The regulation of cytokinesis is facilitated by recently identified endolysosmal ion channels known as Two-Pore Channels (TPC). There are three known TPC's: TPC1, TPC2 and TPC3. TPC1 is proven to interact with citron kinase, regulating completion of cytokinesis (Horton *et al,* 2015). In plants, studies of Ca^{2+} release in *Arabidopsis thaliana* identified AtTPC1 as a channel that mediates the slow vacuolar current, regulating germination and stomatal movement (Peiter *et al.,* 2005). This may provide some insight as to why this candidate gene was associated with a Ca SNP. Additionally, the literature documents other studies that prove that proton-permeable ion channels are activated by NAADP or Ca^{2+} (Pitt *et al.,* 2014).

K 7_20384756

A U-box domain-containing protein 33-like was located within the 4 kb region surrounding the K associated SNP, 20384756. This candidate gene functions as an E3 ubiquitin ligase in *Arabidopsis thaliana*, which facilitates protein ubiquitination. The importance of E3 ubiquitin ligasess is highlighted by the vast number of cellular processes they regulate (such as hormone metabolism and photomorphogenesis in plants), and the number of diseases associated with their loss of function or inappropriate targeting. U-box proteins account for only 3 - 4% of the annotated ubiquitin ligases in humans, which is far less than those found in plants (Patterson, 2002).

Ca 6_37189739/ 37189769

A TF basic helix-loop-helix (bHLH) 128-like was located in relation to these SNPs (which are within 30bp of each other), associated with Ca uptake. Functions associated with this TF that have been characterised in *Arabidopsis thaliana*, include: DNA binding, regulation of stomatal movement and protein dimerisation activity. At present, few plant bHLH TFs have been studied in detail, but among the few that have, some provided insights into the central roles of TFs in plants, and into their biochemical function. For example, genetic analysis of the anthocyanin biosynthetic pathway in *Zea mays* resulted in the identification of a group of bHLH TFs required for synthesis of the purple anthocyanin pigments (Heim *et al.*, 2003, Neuffer *et al.*, 1997).

Ca 2_467521

A probable LRR receptor-like serine/threonine-protein kinase RPK1 gene was located within the 4 kb region surrounding the SNP, 467521. This gene is involved in the main abscisic acid-mediated (ABA) signaling pathway and in early ABA perception. Coupled with RPK2, it is needed for pattern formation along the radial axis (e.g. the apical embryonic domain cell types that generate cotyledon primordia), and the apical-basal axis (e.g. differentiation of the basal pole during early embryogenesis) (Nodine *et al.*, 2007). Microarray studies have proven that hundred of genes respond to water deficiency via a specific temporal and spatial expression pattern, including signalling cascades involving protein kinases/phosphatases (e.g.,

RPK1) and the upregualtion of chaperones and molecules concerned with osmoprotectant metabolism. ABA plays a key role in cellular signaling during drought and salt stress (Reddy *et al.,* 2011). Synthesis of ABA is induced under drought stress and elevated levels of ABA signals for plant guard cell stomatal closure, which inturn induces the expression of drought stress–related genes that encode proteins which facilitate dehydration tolerance mechanisms. Promoters of many ABA-responsive genes contain cis-acting elements such as ABRE (PyACGTGGC) (Uno *et al.,* 2000). ABA-inducible transcription generally requires the existence of >2 ABREs or the combination of 1ABRE with a coupling element at an appropriate position within the promoter region (Uno *et al.,* 2000). Many of the Ca^{2+} regulated genes contain these elements, which suggests that ABA may regulate ABA-responsive genes through cellular Ca^{2+} changes (Kaplan *et al.,* 2006), hence the association with a Ca SNP.

Ca 9_7373490/7373520

Folate transporter 1 (FOT1), chloroplastic was the candidate gene associated with the SNPs, 7373490 and 7373520. It facilitates the movement of folate into chloroplasts and has been extensively characterised in *Arabidopsis thaliana,* where it is expressed throughout development (Bedhomme *et al.,* 2005). A study by Bedhomme *et al.,* (2005) indicated that even though AtFOT1 belongs to the mitochondrial carrier family, green fluorescent proteins (GFP)-tagging experiments and Western blot analyses indicate that it is targeted to the envelope of chloroplasts.

Zn 9_4623914

A BTB/POZ domain-containing the protein At3g50780 was located 3839 bp at the 5' side. The BTB domain (Broad-Complex, Tramtrack and Bric a brac) is also referred to as the POZ domain (POxvirus and Zinc finger). It is a homodimerisation domain occurring at the N terminus of proteins which contains several copies of either C2H2 zinc fingers or Kelch repeats comprised of around 50 amino acid residues which form the structure of a four stranded beta-sheet "blade" for multiple potential protein-protein contact sites (Zollman *et al.,* 1994). Many BTB proteins are transcription regulators and may act via the control of the chromatin structure. C2H2 zinc fingers are a common type of DNA binding domain. The motif typically occurs in tandem repeats and consists of 2 cysteine and 2 histidine residues that coordinate a

Zn ion and fold the domain into a finger-like protein that is able to interact with DNA (Klug & Schwabe, 1995) – hence the association with a Zn SNP. In terms of function, this candidate gene may act as a substrate-specific adapter of an E3 ubiquitin-protein ligase complex (CUL3-RBX1-BTB), which mediates the ubiquitination and proteasomal degradation of target proteins, this has been characterised in *Arabidopsis thaliana* (Gingerich *et al.,* 2005).

Ca 7_34304700

A microtubule-associated protein RP/EB family member 1A-like (MAPRE1) was located 992 bp at the 5' side. Microtubles represent tracks for the transport of material within the cell by means of molecular motor proteins and EB1 proteins are evolutionarily conserved plus-end-tracking proteins that localise to growing microtubule plus ends where they regulate microtubule dynamics and interactions with intracellular targets (Komaki *et al.,* 2010). Microtubules are dynamic polar polymers that establish cell shape, facilitate cell motility, organise organelles within cells, and assist with cell division. Studies suggest that the EB1 genes are expressed in various cell types of *Arabidopsis thaliana* and GFPs fused to EB family members have been shown to track growing plus ends of microtubules in plant cells.

The role of microtubule-associated proteins and their association with Ca is poorly understood in plants. However, it has been recently discovered in animals that binding at the plus end is regulated by an EF-hand motif, which contains a helix-loop-helix topology in which Ca^{2+} ions are coordinated by ligands within the loop. This regulation is EF-hand and Ca^{2+} dependent. Alteration of Ca^{2+} responsiveness, resulting from mutations in the EF-hands, renders the rapid switch redundant, causing permanent binding with the microtubule lattice or the plus end. This recently identified, Ca^{2+} dependent regulatory mechanism may play a critical role in a number of diverse microtubule associated processes and Ca^{2+} can directly and rapidly regulate the dynamic interaction of microtubules with key regulatory proteins (Kapur *et al.,* 2012).

A V-type proton ATPase subunit D was located 20506 bp at the 3' side. It is associated with ATPase activity, coupled to transmembrane movement of substances, including metal ions (Dietz *et al.,* 2001). V-ATPase functions as a dominant vacuolar electrogenic H^{+}-pump in the majority of plant cells and plays a

vital role in plant growth due to its role in energising secondary transport, maintenance of solute homeostasis and in facilitating vesicle fusion. Under abiotic stress conditions including; high salinity, drought, cold, acid stress, anoxia, and excess heavy metals in the soil, the survival of cells is highly dependant on maintaining/adjusting the activity of V-ATPase (Dietz *et al.*, 2001). In terms of heavy metal stress, metal tolerance appears to depend on additional membrane transporters (as well as other cellular mechanisms, such as the synthesis of phytochelatins and metallothioneins) (Dietz *et al.*, 2001). Therefore, V-ATPase is likely to be affected under excess heavy metal exposure. At present, the influence of heavy metals on either the structure or the activity of V-ATPase is incompletely understood. However, there is some experimental data in the literature that highlights the effect of V-ATPase activity on metal transport (Chardonnens *et al.*, 1999). Although all heavy metals are toxic when present in abundance, plants utilise essential metals, such as Fe, Zn and Ni in low quantities, as discussed in Chapter 1. Based on research on yeast metal transporters, any particular metal has both a high- and a low-affinity uptake mechanism, the use of which is dependent on the amount of the metal, respectively (Guerinot, 2000). Intracellular transport systems are essential in understanding membrane transport processes involved in metal ion homeostasis. This may explain how plants may have adapted to high metal concentrations, including high levels of Ca in the soil, by evolving cellular tolerance mechanisms which involve vacuolar transporters such as V-ATPase. The literature reports on proton gradient-dependent transport of metals across the tonoplast and this type of antiporter activity is dependent on the presence of a proton gradient across the vacuolar membrane and thus, indirectly, on V-ATPase (Dietz *et al.*, 2001). Other studies suggest that proton ATPase's accumulate in response to Fe deficiency, this is best characterised in *Arabidopsis thaliana*; protons are released into the rhizosphere, by proton ATPases expressed in the epidermis. This lowers the soil pH, making Fe more soluble (Morrissey & Guerinot, 2009).

A BLASTx search of the 4kb region surrounding this Ca associated SNP revealed a nitrate transporter in *Aegilops tauschii* (Tausch's goats grass) within the 523 to 602 range. The purpose of the BLASTx search against other proteins was to see if there was a coding sequence within this region. Even though components involved in nitrogen signalling pathways are poorly understood, Ca is a known second

messenger in signal transduction pathways in plants, and it has been indirectly implicated in nitrate responses. This has been characterised in *Arabidopsis thaliana* (Riveras *et al.,* 2015). A study by Riveras *et al.,* (2015) revealed that gene expression of nitrate-responsive genes were severely affected by pre-treatments with Ca^{2+} channel blockers, thus proving that Ca^{2+} acts as a second messenger in the nitrate signaling pathway – hence the association of this region with a Ca SNP.

Fe 9_3820759

Hox-A13-like was located 179 bp at the 3' side. The HOX-A13 gene provides instructions for the production of proteins that bind to specific regions of DNA and regulate the activity of other genes. On the basis of this general-purpose role, the HOX-A13 gene is referred to as a TF. The HOXA13 gene is part of a larger family of TFs called homeobox (HOX) genes. HOX genes are considered general-purpose control genes. Even though HOX genes are found in plants, fungi, and animals, there are few known direct HOX target genes and their mechanisms of regulation are currently poorly understood (McCabe & Innis, 2005).

A DNA-binding protein HEXBP (hexamer-binding protein) -like protein was located 6066 bp at the 5' side in association with Fe uptake. HEXBP encodes a sequence-specific DNA-binding protein that contains nine 'CCHC' zinc finger motifs. This motif is present in a number of nucleic acid-binding proteins and functions by binding to single-stranded nucleic acids. Researchers investigated the relationship between this TF and Fe-deficiency in rice plants, as demonstrated in a study by Sperotto *et al.,* (2011). The study aimed to provide further insight into the pathways induced during Fe-deficiency in rice seedlings which were grown for 3, 6 and 9 days in the presence/absence of Fe. Using Representational Difference Analysis, sequences of 32 induced genes/TFs in rice shoots under Fe-deficiency were identified, including HEXBP.

Additionally, a BLASTx search revealed two zinc finger proteins, GIS and 7-like in *Brachypodium distachyon* and *Setaria italica*, respectively. The zinc finger proteins family members are likely to be involved in numerous activities associated with plant growth and development and are also known to regulate resistance mechanisms for various biotic and abiotic stress factors (Gupta *et al.,* 2012). Additionally, the zinc finger binding domains are involved in sequence specific binding to DNA/RNA

and contribute in protein-protein recognitions. More specifically, zinc finger protein GIS functions as a probable TF, required for the initiation of inflorescence trichomes in response to gibberellin (Gan *et al.*, 2006). It contains a single C2H2-type zinc finger binding domain, whereas zinc finger 7 contains 4 CCHC-type binding domains, indicative of how many Zn ions are required to stabilise a protein fold. Several studies have reported that C2H2-type zinc finger proteins are responsible for the activation of some stress-related genes and enhanced tolerance to salt, dehydration, and/or cold stresses (Gupta *et al.*, 2012).

6.4.2 Population Structure, LD and GWAS Using 37,296 SNPs from the Pearl Millet Genome Assembly

Prior to September 2017, the limited range of genomics tools prevented full exploitation of modern breeding methods for the improvement of pearl millet. Recently, a genome sequence of reference genotype Tift $23D_2B_1$-P1-P5 and a vast resource of genome wide variations including >39 million SNPs, developed at ICRISAT, Patencheru has become available for public use (Varshney *et al.*, 2017).

Population Structure

The population structure within the PMiGAP was assessed using >37,000 high quality SNPs using PCA in TASSEL (Figure 6.13).

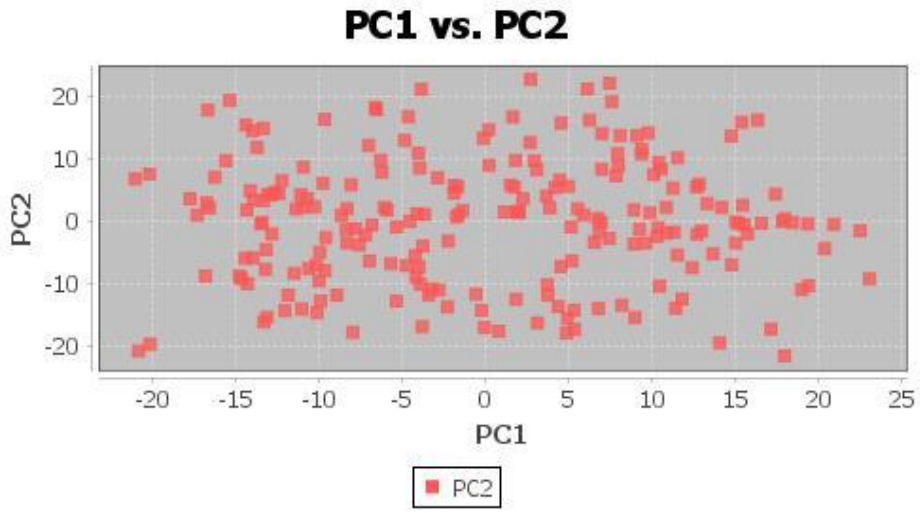


Figure 6.13, PCA plot for 223 PMiGAP lines generated in the TASSEL v5.2.38 environment.

Insignificant population structure was identified in the PMiGAP population studied; this coincides with the previous set of results using 663 SNPs and with that of Varshney *et al.*, 2017. This pattern is expected from inbred lines derived from a

highly allogamous species, from a collection of FAO-designated germplasm entries. The lack of population structure also points to homogenous genetic diversity on a large geographical scale and may be associated with a rapid spread of pearl millet agriculture in Africa and India without major bottlenecks during the migration (Varshney *et al.*, 2017).

Linkage Disequilibrium

The extent of LD was also assessed among all 1,863,525 pairs of loci (Figure 6.14). Across all accessions, the largest blocks of marker pairs that were in LD (based on r^2) can be seen on chromosomes 3 and 5, respectively and findings were highly significant.

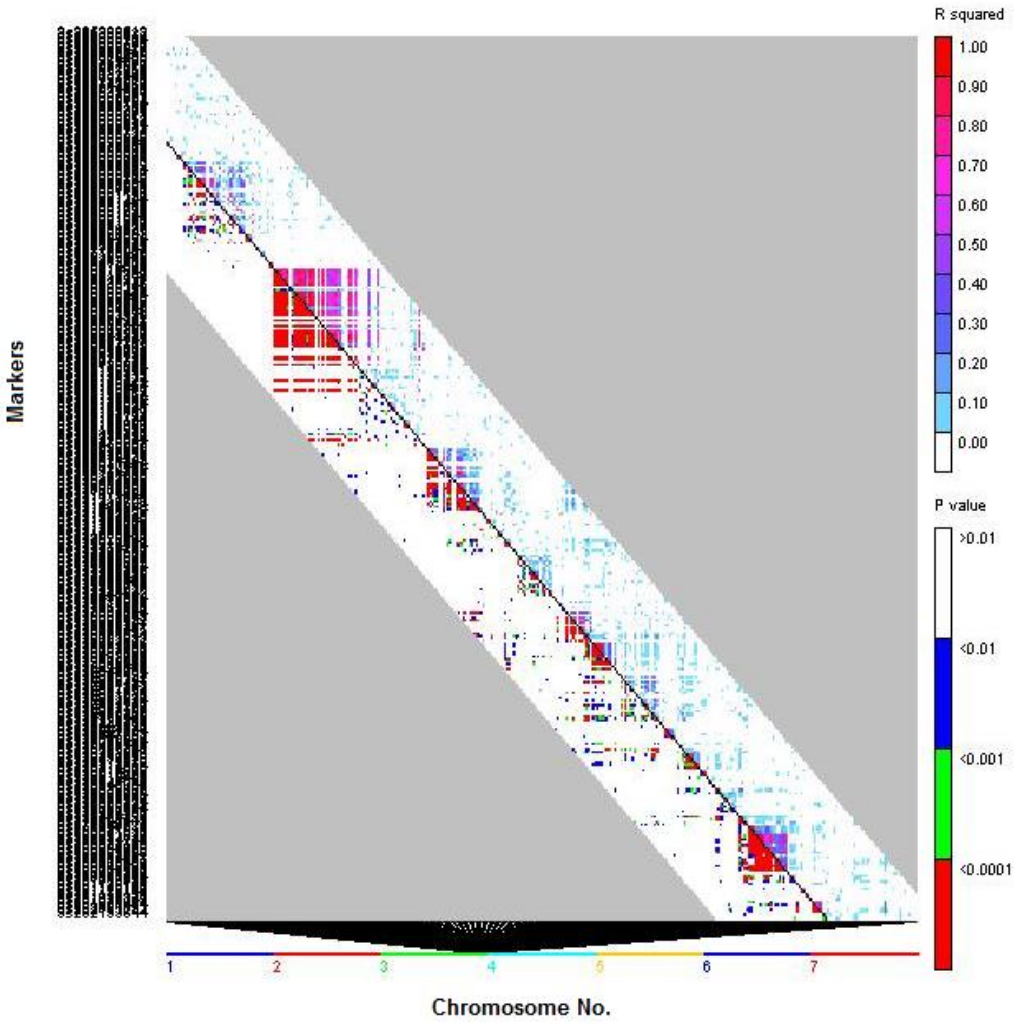


Figure 6.14, LD plot built in the TASSEL v5.2.38 environment zoomed in on regions of high LD on chromosomes 3 and 5. The squared correlation coefficient (r^2) values are denoted by a colour scale from white (0.0) to red (1.0) in the upper triangle. The p values ranging from non-significant (0.01; white) to highly

significant (<0.0001; red) are shown in the lower triangle. The white area with coloured pixels represent the area that applies to chromosomes 1-7

The extent of LD is greater within this data set, as compared to the 31,425 pairs of loci studied previously; this could be attributable to increased levels of markers.

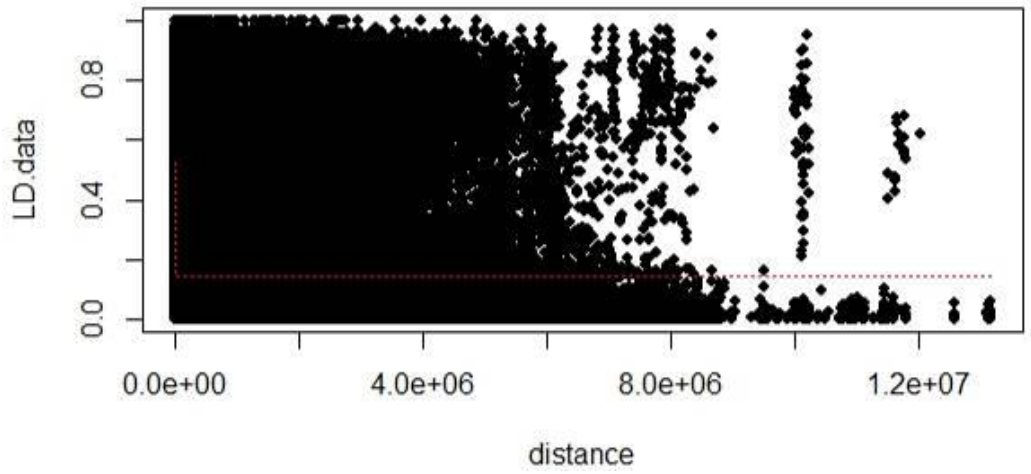
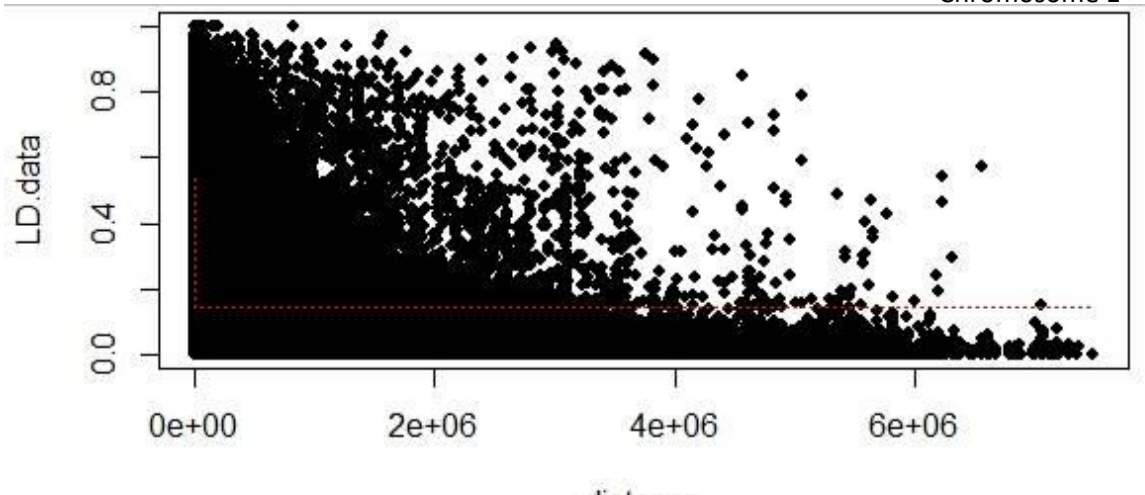
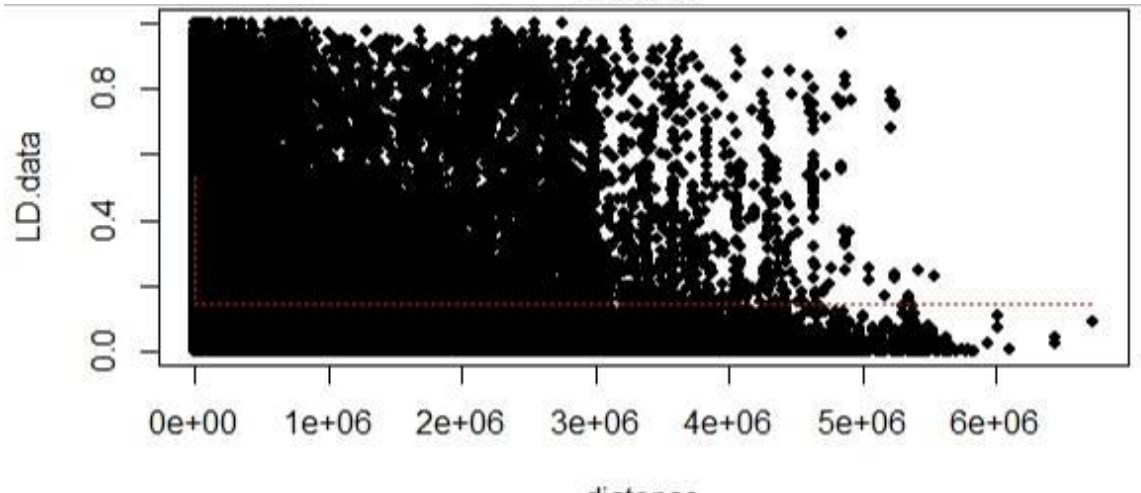


Figure 6.15, LD decay for the *Pennisetum glaucum* genome built in the RStudio environment. Regression line is based on Hill and Weir (1988). LD data is represented by r^2 values and distance is measured in bp.

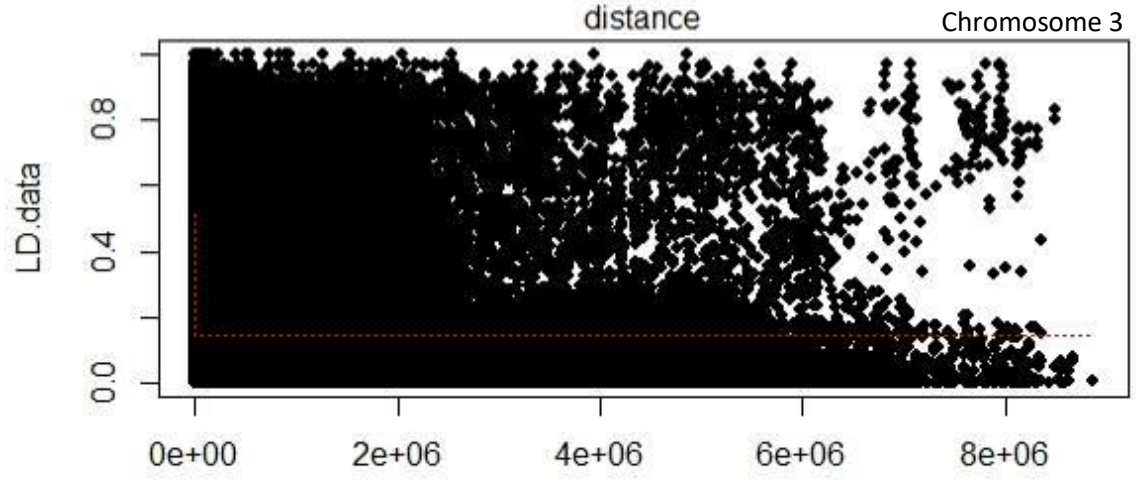
Chromosome 1



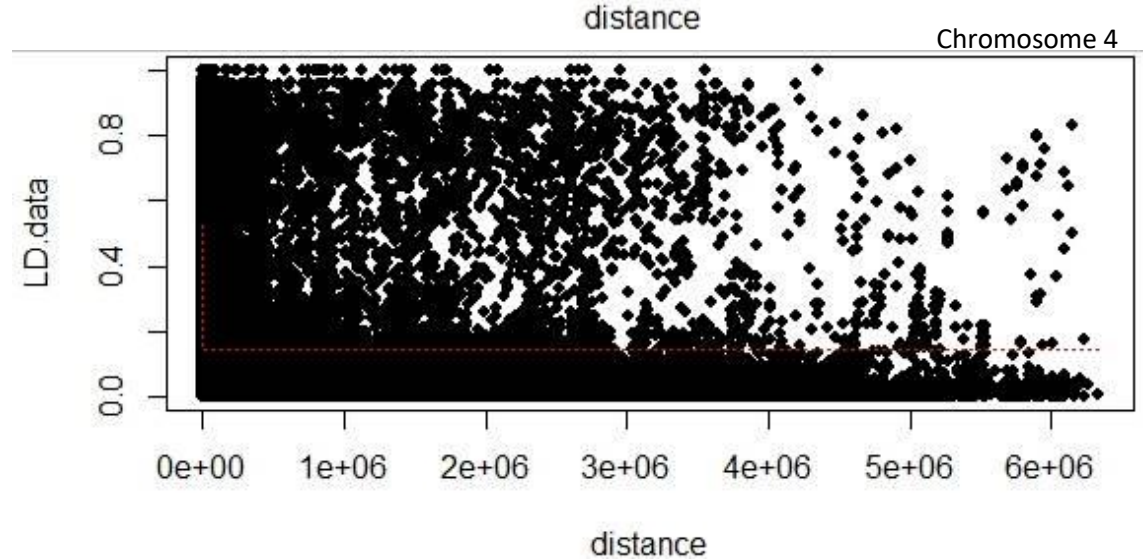
Chromosome 2



Chromosome 3



Chromosome 4



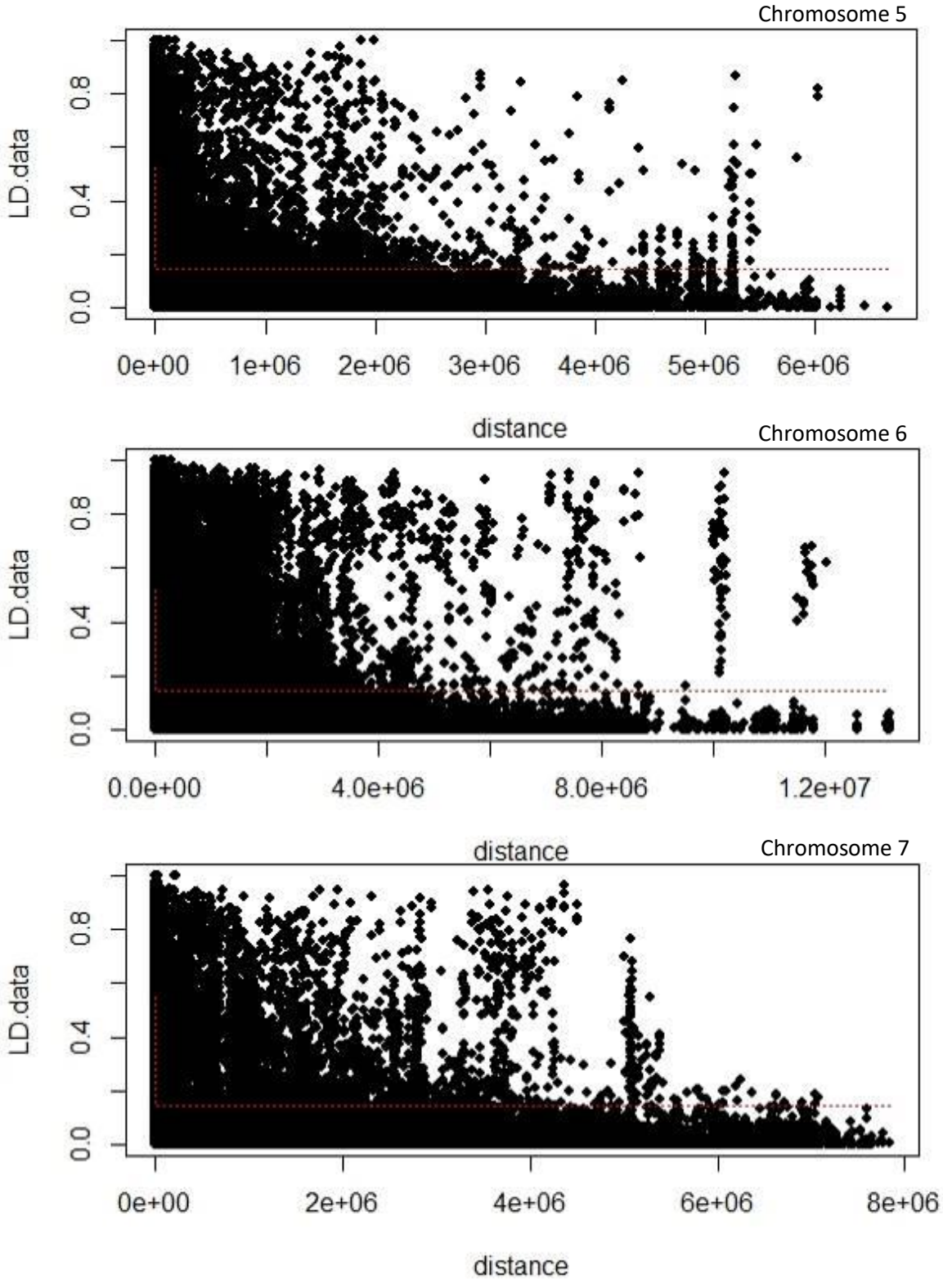


Figure 6.16, LD decay per chromosome for the *Pennisetum glaucum* genome built in the RStudio environment. Regression line is based on Hill and Weir (1988). LD data is represented by r^2 values and distance is measured in bp.

The extent of LD decay was investigated using 1,863,525 pairs of loci at the genome wide level and at the chromosome level (Figure 6.15 and 6.16). Pairwise LD,

estimated using the squared allele frequency correlation (r^2), was found to decay slowly, as compared to the 663 SNP dataset across all chromosomes with genetic distance. This also suggests that the number of markers for GWAS is adequate. Differences between chromosomes were also observed. For example, although LD declined slowly in all cases, it was most prominent on chromosome 3 (Figure 6.14), which is also seen in the LD plot in Figure 6.16.

Even though slow LD decay is reported here (from 37,000 PMiGAP SNPs), Varshney *et al.,* (2017) reported contrasting results, when 450,000 PMiGAP SNPs were used to compute LD decay. These differing results may be attributable to a number of differing factors between this study and Varshney *et al.*'s, (2017) study, including the marker selection process. For example, in this study, markers deemed 'better for GWAS' were selected, i.e. markers that worked across the bulk of the accessions, meaning rarer alleles, indels and monomorphic alleles were removed, which may have skewed the data, leading to a biased estimate of LD decay. Additionally, other studies have found that increased marker coverage may be the reason for rapid LD decay. including that of Liu *et al.,* (2015) and Blackmore *et al.,* (2016), who both found that accuracy of LD decay is highly dependent on sufficient marker coverage. Marroni *et al.*, (2001) also reported that small sample sizes may lead to biased estimates of LD.

Marker Trait Associations

The two model-based approaches, GLM and MLM, were compared for all traits using the kinship matrix in both models. The QQ plots of traits shown in Figure 6.17 suggest that the MLM model is superior at accounting for confounding effects, as previously demonstrated. Here we present results of only MLM model-based associations due to the clean QQ plots in Figure 6.17. When the Bonferroni threshold correction at the 5% significance level was applied, just 3 of the MTA's in Table 6.7 (all associated with Na) exceeded the threshold, thus for the remaining MTA's, associations with candidate genes are explored based on the lowest P-values possible, eventhough these fall below the Bonferroni threshold for significance. MTA's with the lowest P-values can be seen in Figure 6.18 and Table 6.7.

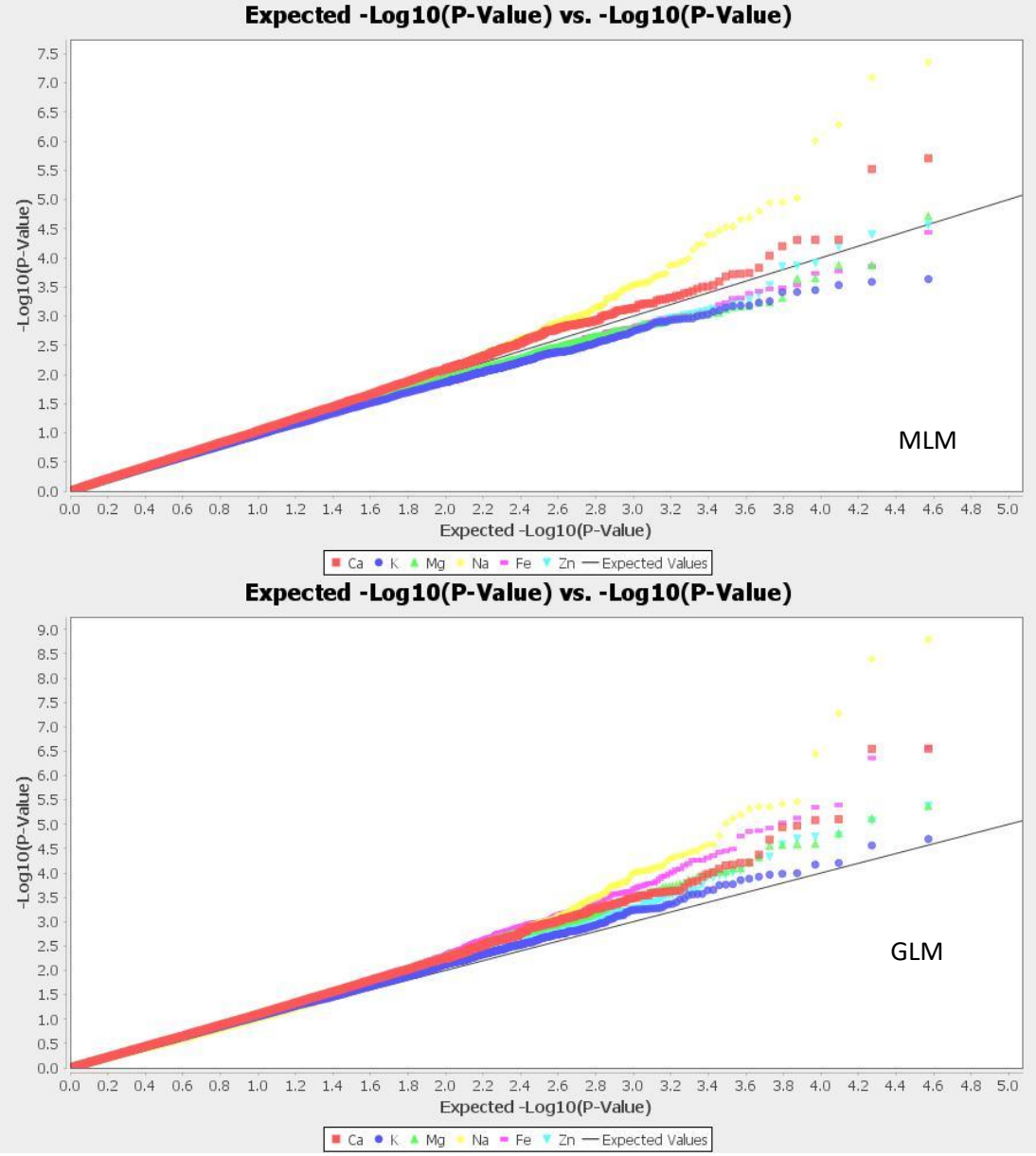
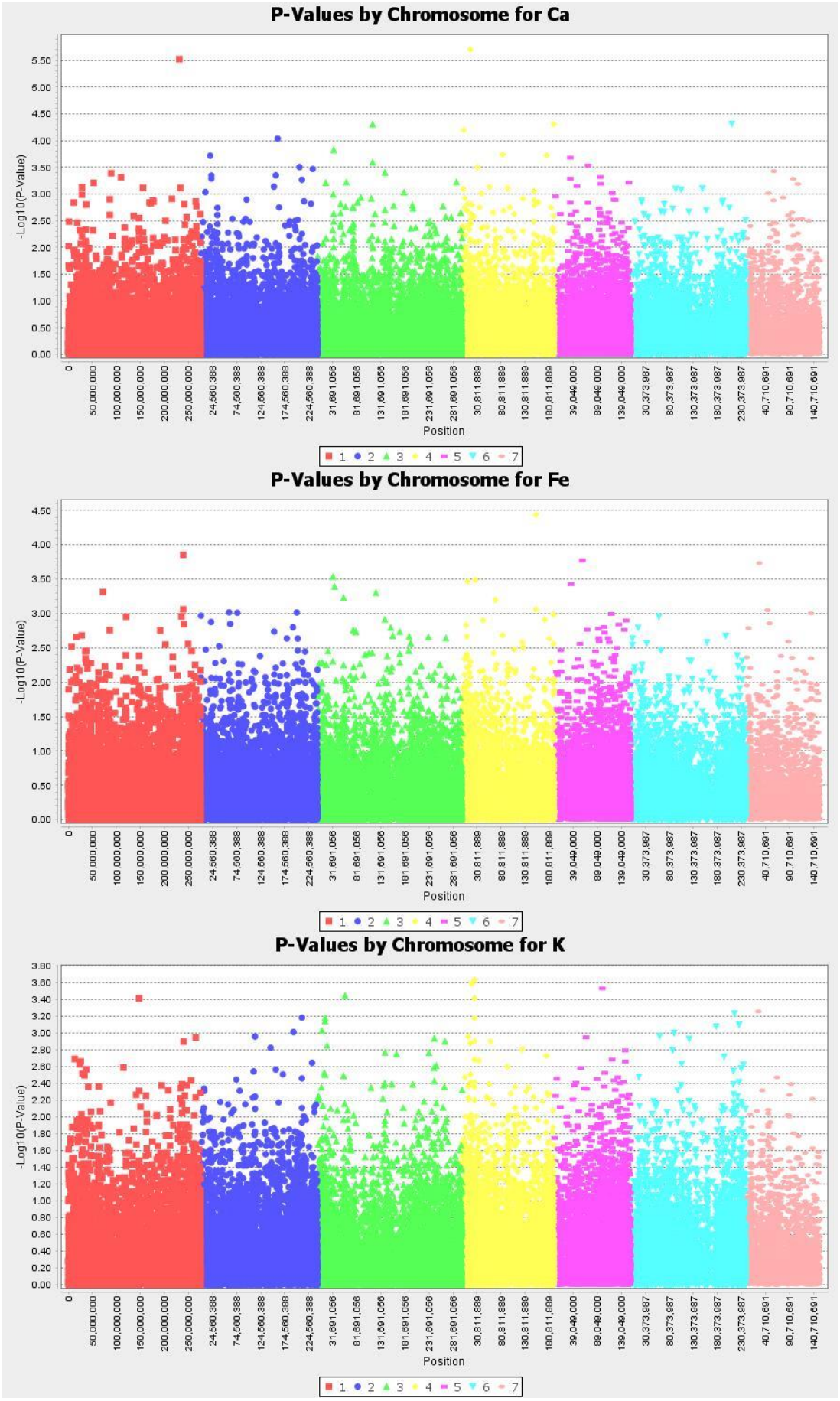


Figure 6.17, Quantile-quantile plots of Ca, K, Mg, Na, Fe and Zn using MLM and GLM models built in the TASSEL v5.2.38 environment


P-Values by Chromosome for Ca
-Log10(P-Value)
Position
1 2 3 4 5 6 7
P-Values by Chromosome for Fe
-Log10(P-Value)
Position
1 2 3 4 5 6 7
P-Values by Chromosome for K
-Log10(P-Value)
Position
1 2 3 4 5 6 7

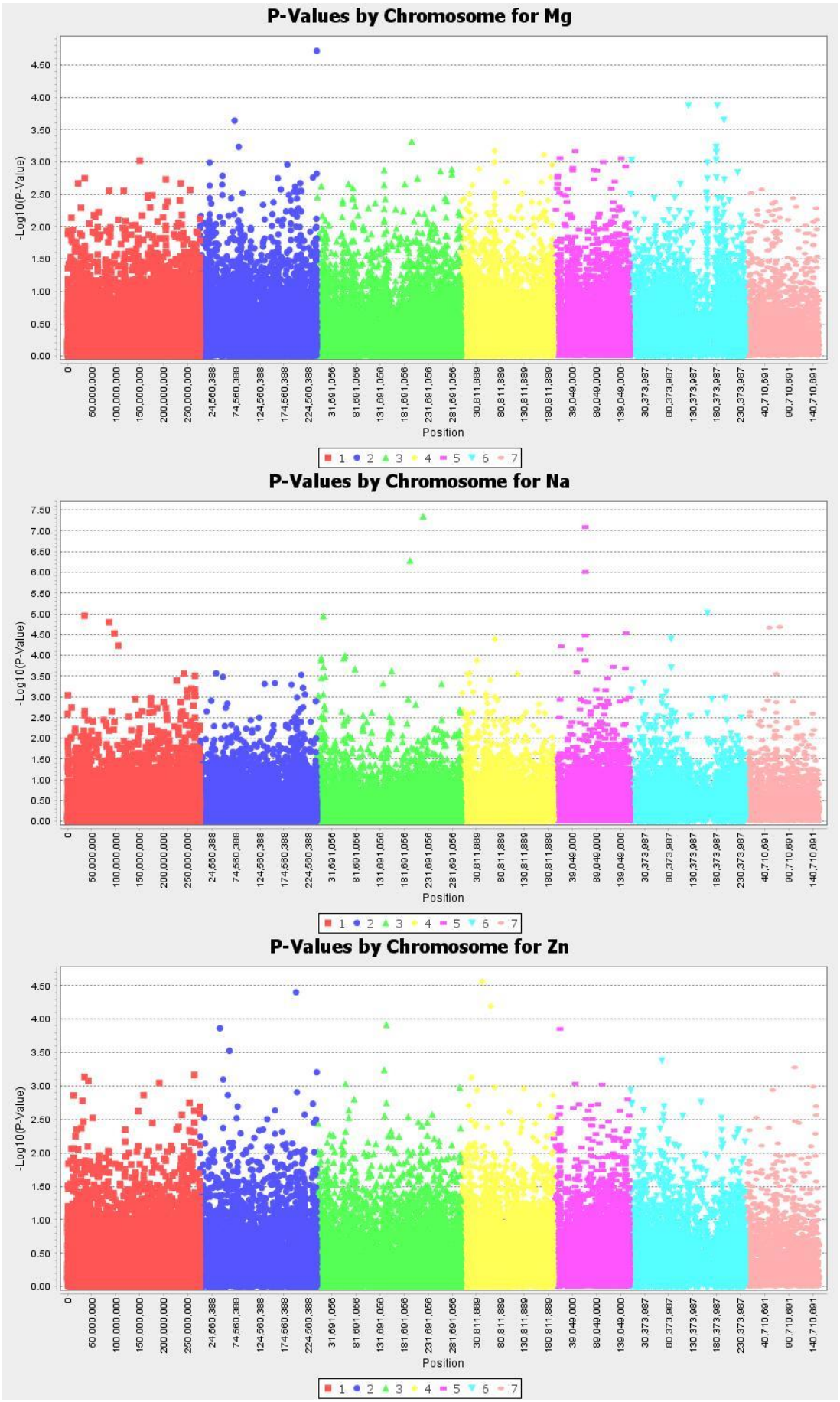


Figure 6.18, GWAS-based Manhattan plots built in the TASSEL v5.2.38 environment exhibiting the lowest P-values (measured by the MLM model)

associated with seed mineral concentrations using >37,000 genome-wide GBS SNPs in pearl millet. The x-axis illustrates the relative density of *Pennisetum glaucum* reference genome-based SNPs physically mapped on 7 chromosomes. The y-axis displays the -log10 (P)-value for the degree of association of SNP loci with seed-mineral concentrations

Table 6.7, Top 3 SNPs per minerals Ca, Na, Fe, Zn and Mg, P=<0.001. Those highlighted exceeded the Bonferroni corrected threshold (2.23 E-07).

Trait	Marker	Chromosome	Position	Df	F	P-Value	Marker r^2	$\%r^{2a}$	Cumulative $\%r^{2b}$
Na	chr3_221196763	3	221196763	3	13.4446	**4.46E-08**	0.184173	**18.42**	
Na	chr5_65686933	5	65686933	3	12.9662	**8.06E-08**	0.177619	**17.76**	51.89
Na	chr3_193997829	3	193997829	3	11.4699	**5.23E-07**	0.157122	**15.71**	
Ca	chr4_16960040	4	16960040	2	13.9676	**1.97E-06**	0.127558	**12.76**	
Ca	chr1_230986046	1	230986046	3	10.0906	**3.00E-06**	0.138228	**13.82**	37.33
Ca	chr6_210316981	6	210316981	3	7.9270	**4.86E-05**	0.108589	**10.86**	
Mg	chr2_242869315	2	242869315	3	8.6429	**1.92E-05**	0.118395	**11.84**	
Mg	chr6_182035399	6	182035399	3	7.1527	**1.33E-04**	0.097983	**9.80**	31.43
Mg	chr6_121772772	6	121772772	3	7.1489	**1.34E-04**	0.09793	**9.79**	
Zn	chr4_43134799	4	43134799	3	8.3700	**2.74E-05**	0.114658	**11.47**	
Zn	chr2_199712768	2	199712768	3	8.0853	**3.96E-05**	0.110758	**11.08**	33.12
Zn	chr4_60825590	4	60825590	3	7.7134	**6.42E-05**	0.105664	**10.57**	
Fe	chr4_153484722	4	153484722	3	8.1491	**3.65E-05**	0.111632	**11.16**	
Fe	chr1_239148556	1	239148556	3	7.115	**1.40E-04**	0.097474	**9.75**	30.46
Fe	chr5_58171748	5	58171748	3	6.9704	**1.69E-04**	0.095484	**9.55**	

[a] Percentage of phenotypic variation explained.
[b] Phenotypic variance (cumulative r^2) explained by the genetic effects of all associated SNPs.

Trait data from field grown plants (as discussed in Chapter 3) was used in the GWAS analysis, coupled with SNP genotype data from corresponding PMiGAP lines. The MLM model based association mapping approach identified 199 SNPs exhibiting associations below the P=<0.001 threshold with seed mineral concentrations. The top three markers for each mineral, with the exception of K (due to low Log P-values) were selected for a BLAST search of the NCBI database to identify potential candidate genes. These can be seen in Tables 6.7, 6.8 and 6.9, respectively. The MTA's explained a large proportion of observed phenotypic variation with individual marker $\%r^2$ values ranging from 9.55 – 18.42%. Cumulative $\%r^2$ values ranged from 30.46 - 51.89% for all SNPs found associated with each mineral.

The 3 SNPs that exceeded the Bonferroni corrected threshold and the low P-value SNPs from the GWAS were located by position and chromosome number on the

Pennisetum glaucum reference genome assembly (http://ceg.icrisat.org/ipmgsc/genome.htmL), as seen in Figure 6.19, using the software CLC Genomics Workbench v.6.5 (CLC Bio, Aarhus, Denmark). A 4kb region surrounding each SNP was selected and a BLAST search of the NCBI database was conducted on each region of interest. Where no significant hits were found, a 10kb region was selected. This was the case for 2 Fe associated SNPs-labelled in Table 6.8. BLAST alignments to *Setaria italica* can be seen in Table 6.8. *Setaria italica* was used due to the lack of annotated pearl millet genome assembly. If hits were located on different chromosomes, this was still considered significant due to the fusion of chromosomes (9 chromosomes to 7 chromosomes from *Setaria italica* to *Pennisetum glaucum).*

Identifying Candidate Genes from Significant SNPs

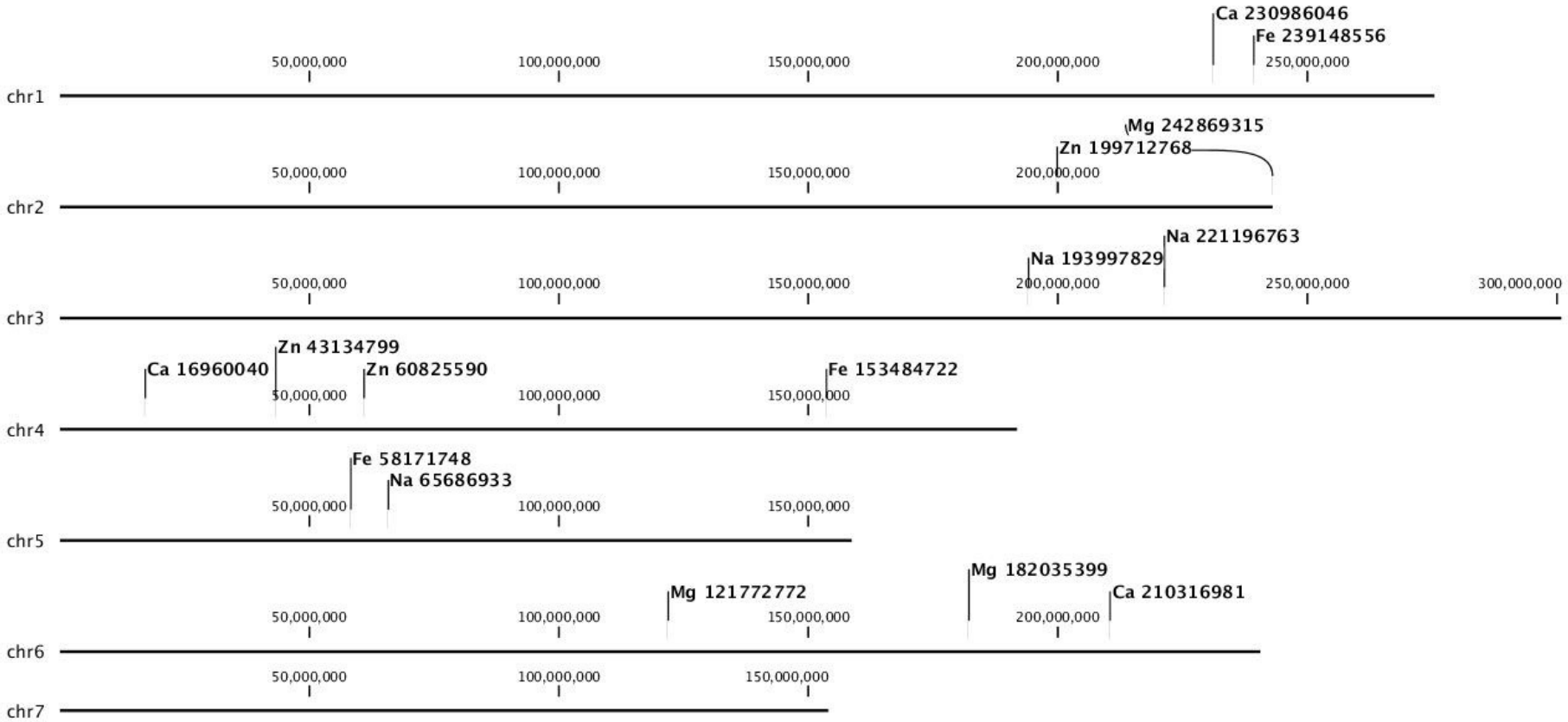


Figure 6.19, SNPs with the lowest P-values from GWAS mapped onto the *Pennisetum glaucum* genome. Chromosome numbers 1 – 7 correspond to chr1 – 7, respectively. Image built in the CLC Genomic Workbench environment.

Table 6.8, BLASTn alignments of low P-value SNPs mapped onto the *Pennisetum glaucum* reference genome assembly + 4kb. A 10kb region surrounding the Fe associated SNPs; 153484722 and 58171748 were analysed using BLASTn due to the lack of significant hits within the +4kb region. – >37,000 SNP dataset.

SNP (Chromosome no._Position)	Score	E Value	Identities	Gaps	Range	Candidate Gene(s)
Na 3_2211967637	1011 bits(547)	0.0	1021/1247(82 %)	43/1247(3%)	8586326 to 8587560	120400 bp at 5' side: cyclin-dependent kinase inhibitor 1C-like 59608 bp at 3' side: zinc finger BED domain-containing protein RICESLEEPER 2-like
Na 5_65686933	333 bits(180)	6e-89	499/651(77%)	29/651(4%)	12609251 to 12609888	206267 bp at 5' side: cytochrome P450 99A2-like 68119 bp at 3' side: putative receptor-like protein kinase At1g80870
	243 bits(131)	1e-61	388/508(76%)	34/508(6%)	26553639 to 26554135	15423 bp at 5' side: uncharacterized protein LOC101762765 13185 bp at 3' side: uncharacterized protein LOC101763173
	231 bits(125)	2e-58	275/346(79%)	16/346(4%)	3141364 to 3141703	47241 bp at 5' side: uncharacterized protein LOC101756405 47133 bp at 3' side: protein DETOXIFICATION 16-like
Na 3_193997829	407 bits(220)	4e-111	301/340(89%)	5/340(1%)	27527898 to 27528237	25917 bp at 5' side: E3 ubiquitin-protein ligase Os06g0535400-like 5336 bp at 3' side: RING-H2 finger protein ATL30-like
Ca 4_16960040	298 bits(161)	2e-78	395/510(77%)	8/510(1%)	19523743 to 19524248	231263 bp at 5' side: uncharacterized mitochondrial protein AtMg00810-like 210552 bp at 3' side: RNA-directed DNA polymerase homolog
	285 bits(154)	2e-74	395/510(77%)	9/510(1%)	23179946 to 23180450	5004 bp at 5' side: putative germin-like protein 2-2 49008 bp at 3' side: putative germin-like protein 2-2

	283 bits(153)	7e-74	433/568(76%)	20/568(3%)	13812042 to 13812593	34557 bp at 5' side: uncharacterized protein LOC101785864 41475 bp at 3' side: expansin-B11-like
Ca 1_230986046	119 bits(64)	2e-24	134/165(81%)	15/165(9%)	8609538 to 8609695	4770 bp at 5' side: uncharacterized protein LOC101778091 1203 bp at 3' side: RNA polymerase sigma factor sigE, chloroplastic/mitochondrial
Ca 6_210316981	1478 bits(800)	0.0	972/1054(92%)	15/1054(1%)	5693028 to 5694069	snurportin-1 snurportin-1
Mg 2_242869315	1495 bits(809)	0.0	1641/2026(81%)	124/2026(6%)	77570 to 79528	protein disulfide isomerase-like 1-4
Mg 6_182035399	634 bits(343)	2e-179	1167/1565(75%)	55/1565(3%)	32444027 to 32445557	2378 bp at 5' side: transcription factor bHLH130-like 8906 bp at 3' side: IRK-interacting protein-like
	545 bits(295)	8e-153	1209/1652(73%)	56/1652(3%)	8815877 to 8817502	5301 bp at 5' side: LOW QUALITY PROTEIN: pentatricopeptide repeat-containing protein At2g13420, mitochondrial-like 6328 bp at 3' side: TPD1 protein homolog 1A-like
	355 bits(192)	1e-95	1234/1744(71%)	43/1744(2%)	14063096 to 14064817	1585 bp at 5' side: 21 kDa protein-like 2613 bp at 3' side: uncharacterized protein LOC101781726
Mg 6_121772772	193 bits(104)	1e-46	208/255(82%)	20/255(7%)	2237547 to 2237799	uncharacterized protein LOC101778133
Zn 4_43134799	174 bits(94)	4e-41	243/309(79%)	34/309(11%)	7448256 to 7448547	637 bp at 5' side: ACT domain-containing protein ACR8-like 9319 bp at 3' side: CAX-interacting protein 4-like
Zn 2_199712768	429 bits(232)	8e-118	249/257(97%)	2/257(0%)	5084978 to 5085234	uncharacterized protein LOC105914899
	265 bits(143)	2e-68	306/383(80%)	18/383(4%)	37313929 to 37314305	128615 bp at 5' side: uncharacterized protein LOC101783484 4699 bp at 3' side: lipase-like isoform X2

	143 bits(77)	1e-31	566/788(72%)	90/788(11%)	17178404 to 17179142	151254 bp at 5' side: uncharacterized protein LOC105914943 61873 bp at 3' side: uncharacterized protein LOC105914944
	137 bits(74)	5e-30	301/401(75%)	53/401(13%)	34305577 to 34305943	9795 bp at 5' side: UDP-glycosyltransferase 91B1-like 2935 bp at 3' side: UDP-glycosyltransferase 91C1-like
	272 bits(147)	1e-70	476/627(76%)	53/627(8%)	32679972 to 32680579	16598 bp at 5' side: GDSL esterase/lipase EXL3-like 21727 bp at 3' side: vegetative cell wall protein gp1-like
Zn 4_60825590	3988 bits(2159)	0.0	3242/3732(87%)	205/3732(5%)	4853785 to 4857421	factor of DNA methylation 1-like isoform X1 factor of DNA methylation 1-like isoform X2
Fe 4_153484722 (10kb region)	353 bits(191)	1e-94	458/581(79%)	42/581(7%)	13570742 to 13571308	54260 bp at 5' side: cytochrome P450 71A1-like 29146 bp at 3' side: LOW QUALITY PROTEIN: gibberellin 2-beta-dioxygenase-like
	339 bits(183)	3e-90	485/621(78%)	60/621(9%)	196841 to 197426	S-formylglutathione hydrolase-like
	333 bits(180)	2e-88	403/505(80%)	37/505(7%)	21162266 to 21162744	1875 bp at 5' side: LEC14B protein-like 11388 bp at 3' side: LRR receptor-like serine/threonine-protein kinase ERECTA
	333 bits(180)	2e-88	459/586(78%)	49/586(8%)	43877345 to 43877914	67735 bp at 5' side: protein GOS9-like 64570 bp at 3' side: luminal-binding protein 3-like
Fe 1_239148556	1899 bits(1028)	0.0	1337/1477(91%)	57/1477(3%)	345380 to 346832	heat stress transcription factor A-1-like
	1635 bits(885)	0.0	1456/1711(85%)	122/1711(7%)	335485 to 337161	inactive poly [ADP-ribose] polymerase RCD1-like inactive poly [ADP-ribose] polymerase RCD1-like
Fe 5_58171748 (10kb region)	364 bits(197)	6e-98	608/796(76%)	70/796(8%)	17822474 to 17823254	169840 bp at 5' side: heterogeneous nuclear ribonucleoprotein 1-like 23415 bp at 3' side: uncharacterized protein LOC101770745 isoform X2
	344 bits(186)	7e-92	658/876(75%)	71/876(8%)	23491805 to 23492662	35844 bp at 5' side: cytochrome P450 76C2-like 11428 bp at 3' side: auxin-induced in root cultures

						protein 12-like
	329 bits(178)	2e-87	597/791(75%)	61/791(7%)	18120310 to 18121089	22102 bp at 5' side: uncharacterized protein LOC101772338 138099 bp at 3' side: glucan endo-1,3-beta-glucosidase 5-like
	305 bits(165)	4e-80	431/556(78%)	32/556(5%)	9275087 to 9275633	94870 bp at 5' side: probable purine permease 4 49451 bp at 3' side: protein ASPARTIC PROTEASE IN GUARD CELL 2-like

Table 6.9, BLASTx alignments of low P-value SNPs mapped onto the *Pennisetum glaucum* reference genome assembly + 4kb – >37,000 SNP dataset.

SNP (Chromosome no._Position)	**Score**	**E Value**	**Identities**	**Positives**	**Gaps**	**Range**	**Sequences producing significant alignments**
Na 3_2211967637	174 bits(442)	3e-66	84/117(72%)	95/117(81%)	0/117(0%)	185 to 301	hypothetical protein SETIT_015444mg, partial [*Setaria italica*]
	225 bits(573)	5e-62	111/187(59%)	136/187(72%)	1/187(0%)	90 to 276	PREDICTED: uncharacterized protein LOC9270230 isoform X4 [*Oryza sativa Japonica* Group]
Na 5_65686933	237 bits(605)	1e-59	118/207(57%)	147/207(71%)	11/207(5%)	292 to 490	gag-pol [*Zea mays*]
	186 bits(471)	4e-46	96/205(47%)	123/205(60%)	31/205(15%)	280 to 453	uncharacterized protein LOC110431876 [*Sorghum bicolor*]
	181 bits(460)	3e-42	91/208(44%)	128/208(61%)	19/208(9%)	629 to 829	retrotransposon protein, putative, Ty3-gypsy subclass [*Oryza sativa Japonica* Group]
Na 3_193997829	58.5 bits(140)	4e-05	30/37(81%)	34/37(91%)	0/37(0%)	66 to 102	PREDICTED: kinectin-like [*Setaria italica*]

Ca 4_16960040	80.1 bits(196)	7e-14	46/110(42%)	58/110(52%)	0/110(0%)	2 to 111	hypothetical protein LOC_Os11g22820 *[Oryza sativa Japonica* Group]
	70.5 bits(171)	2e-10	42/95(44%)	54/95(56%)	9/95(9%)	22 to 116	hypothetical protein LOC_Os10g39934 *[Oryza sativa Japonica* Group]
	70.9 bits(172)	2e-08	50/106(47%)	60/106(56%)	3/106(2%)	1119 to 1221	retrotransposon protein, putative, Ty3-gypsy sub-class [*Oryza sativa Japonica* Group]
Ca 1_230986046	112 bits(281)	2e-21	82/269(30%)	137/269(50%)	0/269(0%)	682 to 950	PREDICTED: uncharacterized protein LOC101770518 [*Setaria italica*]
Ca 6_210316981	267 bits(682)	3e-73	135/193(70%)	145/193(75%)	36/193(18%)	345 to 501	PREDICTED: LOW QUALITY PROTEIN: uncharacterized protein LOC101758266 [*Setaria italica*]
	252 bits(644)	4e-72	130/175(74%)	135/175(77%)	36/175(20%)	36 to 174	hypothetical protein SORBI_3003G108800, partial [*Sorghum bicolor*]
	241 bits(615)	6e-69	123/173(71%)	128/173(73%)	36/173(20%)	46 to 182	uncharacterized protein LOC100194257 [*Zea mays*]
	221 bits(563)	2e-58	107/176(61%)	128/176(72%)	36/176(20%)	258 to 397	PREDICTED: probable bifunctional methylthioribulose-1-phosphate dehydratase/enolase-phosphatase E1 [*Oryza brachyantha*]
Mg 2_242869315	182 bits(462)	4e-45	110/216(51%)	123/216(56%)	58/216(26%)	337 to 502	Protein disulfide isomerase-like 1-4 [*Dichanthelium oligosanthes*]
Mg 6_182035399	401 bits(1031)	0.0	194/244(80%)	214/244(87%)	0/244(0%)	672 to 915	retrotransposon protein, putative, Ty1-copia subclass [*Oryza sativa Japonica* Group]
	318 bits(815)	0.0	164/286(57%)	214/286(74%)	5/286(1%)	122 to 407	hypothetical protein VITISV_038102 [*Vitis vinifera*]
	317 bits(813)	0.0	155/246(63%)	195/246(79%)	2/246(0%)	946 to 1191	hypothetical protein VITISV_035070 [*Vitis vinifera*]

	298 bits(762)	3e-177	154/255(60%)	192/255(75%)	1/255(0%)	839 to 1093	Exostosin-like protein [*Corchorus capsularis*]
Mg 6_121772772	-	-	-	-	-	-	No significant similarity found
Zn 4_43134799	90.5 bits(223)	1e-15	39/52(75%)	48/52(92%)	0/52(0%)	268 to 319	uncharacterized protein LOC103645210 [*Zea mays*]
	90.9 bits(224)	1e-15	39/52(75%)	48/52(92%)	0/52(0%)	286 to 337	uncharacterized protein LOC103645715 [*Zea mays*]
	90.5 bits(223)	1e-15	39/52(75%)	48/52(92%)	0/52(0%)	286 to 337	uncharacterized protein LOC103628528 [*Zea mays*]
	90.5 bits(223)	1e-15	39/52(75%)	48/52(92%)	0/52(0%)	292 to 343	uncharacterized protein LOC109942017 isoform X2 [*Zea mays*]
	71.6 bits(174)	1e-09	33/51(65%)	40/51(78%)	0/51(0%)	220 to 270	PREDICTED: uncharacterized protein LOC105915029 [*Setaria italica*]
Zn 2_199712768	-	-	-	-	-	-	No significant similarity found
Zn 4_60825590	635 bits(1639)	0.0	310/363(85%)	316/363(87%)	36/363(9%)	29 to 355	PREDICTED: factor of DNA methylation 1-like isoform X1 [*Setaria italica*]
	473 bits(1218)	1e-148	231/366(63%)	278/366(75%)	42/366(11%)	18 to 344	PREDICTED: factor of DNA methylation 1-like [*Oryza brachyantha*]
Fe 4_153484722	229 bits(583)	1e-68	129/282(46%)	159/282(56%)	65/282(23%)	43 to 259	PREDICTED: uncharacterized protein LOC101760660 [*Setaria italica*]
	91.7 bits(226)	3e-15	55/120(46%)	72/120(60%)	9/120(7%)	320 to 439	protein FAR1-RELATED SEQUENCE 5-like [*Zea mays*]
Fe 1_239148556	539 bits(1388)	4e-175	276/293(94%)	283/293(96%)	0/293(0%)	232 to 524	PREDICTED: heat stress transcription factor A-1-like [*Setaria italica*]
	468 bits(1204)	5e-148	239/297(80%)	261/297(87%)	6/297(2%)	226 to 521	heat stress transcription factor A-1 [*Sorghum bicolor*]
Fe 5_58171748	43.5 bits(101)	3.0	23/82(28%)	38/82(46%)	0/82(0%)	138 to 219	hypothetical protein [*Paenibacillus beijingensis*]

Na 3_2211967637

A cyclin-dependent kinase (CDK) inhibitor 1C-like gene was located 120400 bp at the 5' side. The cell cycle is regulated by CDK's, and CDK inhibitors bind to CDKs and inhibit their activities, as explained by their name. At present, their molecular and cellular functions, regulation and cellular localisation are poorly understood. However, some studies suggest that their function may be related to the regulation of tissue senescence and stressful conditions may induce the expression of CDK inhibitors (Wang *et al.,* 1998). The literature also reports that CDK inhibitors have been expressed in transgenic *Arabidopsis thaliana* plants constitutively or tissue-specifically. For example, the phenotypic effects as a result of over-expression driven by the 35S promoter include; inhibition of cell division, dwarfism, leaf serrations and modified flower morphology (Wang *et al.,* 2000). Therefore, it is generally accepted that over-expression of CDK inhibitors affects plant growth and morphology. Despite the significant progress in understanding the role of plant CDK inhibitors, there are still many important factors that need to be considered. For example, it is clear that CDK inhibitors do not function alone, instead they interact with other factors. Currently, little is known about the dynamics of interactions between CDK inhibitors and other regulators in plants or cofactor minerals. When this is known, it may be clearer as to why this candidate gene was associated with a Na SNP.

A zinc finger BED domain-containing RICESLEEPER 2 was located 59608 bp at the 3' side, which functions as a transposase-like protein that is essential for normal plant growth and development (Knip *et al.,* 2012). The zinc finger associated is a BED-type and the RICESLEEPER2 gene has been characterised in *Arabidopsis thaliana*. A study by Knip *et al.,* (2012) revealed that rice plant lines with an insertion in the RICESLEEPER1 or 2 locus displayed phenotypic abnormalities, therefore these genes are functional and required for normal development in rice.

Na 5_65686933

A cytochrome P450 99A2-like protein was located 206267 bp at the 5' side. Cytochrome P450-dependent monooxygenases represent a large group of enzymes that contain heme as a cofactor. The majority of these catalyse NADPH- and O_2-dependent hydroxylation reactions (Chapple, 1998). More specifically cytochrome

P450 99A2 has been characterised in *Oryza sativa subsp. japonica* (rice) and was found to be involved in momilactone phytoalexin biosynthesis (Shimura *et al.*, 2007). Momilactone phytoalexins are known to accumulate in rice husks and function as (i) growth inhibitors involved in seed dormancy, (ii) play an important role in the rice plant defense system against pathogens/insects, (iii) contribute to allelopathy where they inhibit the growth of weeds and (iv) are induced by the jasmonic acid plant growth hormone (Kato-Noguchi, 2011). Additionally, a study by Colangelo & Guerino (2004) revealed a number of cytochrome P450 family members that accumulate in response to Fe deficiency by microarray gene expression analysis. Therefore, it would be interesting to investigate their role in Fe deficient plants for future analysis in pearl millet.

A putative receptor-like protein kinase At1g80870 was located 68119 bp at the 3' side, which is a protein kinase superfamily protein that is characterised by protein serine/threonine kinase activity, protein kinase activity, kinase activity and ATP binding in *Arabidopsis thaliana*. The functions of protein kinases in plants is currently incompletely understood, however protein-serine/threonine kinases in plant cells are thought to act as a "general central processor units", by accepting input information from receptors that sense environmental conditions, stress factors, phyto-hormones, and other external factors, and converting these signals into appropriate responses including; changes in metabolism, gene expression, and cell growth/division (Hardie, 1999).

The DETOXIFICATION 16 candidate gene was located 47133 bp at the 3' side, and is known to be involved in the transport of sugars, bile salts, organic acids, metal ions and amine compounds (Croft *et al.*, 2013). In relation to the transport of metal ions, members of the DETOXIFICATION family are capable of detoxifying Cd^{2+}, a toxic heavy metal and there is some evidence in the literature that suggests some family members serve as efflux carriers that extrude a number of toxic compounds and heavy metals from cells (Li *et al.*, 2002). Several mechanisms of detoxification include modification of toxic compounds by endogenous enzymes, sequestration into the vacuole and subsequent transport outside of the cell. However, at present the mechanism of heavy metal detoxification in plants is incompletely understood, suggesting a gap in the literature (Li *et al.*, 2002).

Na 3_193997829

The E3 ubiquitin-protein ligase Os06g0535400 was located 25917 bp at the 5' side, and contains a RING-type zinc finger binding site, characterised by the presence of a cysteine-rich domain that coordinates two Zn atoms (Stone *et al.*, 2005). It is involved in protein ubiquitination, which is part of protein modification. More specifically, the addition of ubiquitin to a protein affects proteins in many ways; it can mark them for degradation, alter their cellular location, affect their activity, and promote/prevent protein interactions. Target specific ubiquitination plays an important role in protein regulation in *Arabidopsis thaliana* (Stone *et al.*, 2005). Similarly, in terms of function, a RING-H2 finger protein ATL30 was located 5336 bp at the 3' side, which is involved in protein ubiquitination.

Additionally, a BLASTx search revealed a kinectin-like [*Setaria italica*] protein within the 4kb region of the SNP. Kinetin is a plant growth-promoting hormone that has an anti-aging effect on several different systems, including plant and human cells (Sheu *et al.*, 2003). For example, a study by Ray *et al.*, (1983) revealed that spraying 100-day-old rice plants with kinetin solution (100 μgmL^{-1}) resulted in a significant delay in leaf senescence as indicated by higher total chlorophyll and protein content in the three uppermost leaves, as compared with controls. Leaf senescence constitutes the final stage of leaf development and is critical for plants' fitness as nutrient relocation from leaves to reproducing seeds is achieved through this process. Both the accumulation of specific toxic ions (e.g. Na^+) and changes in leaf hormone interactions are involved in the regulation of this process (Ghanem *et al.*, 2008). The delay in leaf senescence could be explained by the interference with Na pathways to block the process at a molecular level. However, this is only speculation as to why this protein is associated with a Na SNP at this stage and further research is needed to investigate this hypothesis.

Ca 4_16960040

A putative germin-like protein 2-2 was located 5004 bp at the 5' side. Germin-like proteins constitute a ubiquitous family of plant proteins. All germins contain the germin motif that gives rise to a predicted β-barrel core involved in metal binding. Some germin family members are classically associated with defense against pathogens, based on gene regulation studies. For example, it has been proven that

infection with pathogens, insect feeding or the application of chemicals such as salicylic acid, H_2O_2 or ethylene increases the expression of germin-like proteins. The literature also reports that germin genes are involved in other, more general processes including development, osmotic regulation, photoperiodic oscillation, defense and apoptosis. There is also evidence that germin encodes an enzyme that degrades oxalate to CO_2 and H_2O_2 and also releases Ca^{2+} in some plant species. The degraded residual H_2O_2 acts a molecular signal for the induction of defence mechanisms (Dunwell & Gane, 1998). Despite the abundance of germin like proteins in many crops, they remain poorly understood. Future research could provide insights into understanding the function and elucidating the molecular mechanisms of germin genes in plant defense responses and development.

Additionally, an expansin-B11-like protein was found 41475 bp at the 3' side, which is concerned with cell wall regulation. It may cause loosening and extension of plant cell walls by disrupting the non-covalent bonds between cellulose microfibrils and matrix glucans. Expansin action is implicated in the growth responses of plants to hormones and to external stimuli including light, drought, salt stress and submergence (anoxia) and in morphogenetic processes such as root-hair formation (Sampedro & Cosgrove, 2005).

Mg 2_242869315

Protein disulfide isomerase (PDI)-like 1-4 was located within the 4kb region of the SNP. It acts as a protein-folding catalyst that interacts with nascent polypeptides to catalyse the formation, isomerisation, and reduction/oxidation of disulfide bonds. It may also play a role in storage protein biogenesis. It has a strong affinity for binding and ligands range from peptide/protein substrates to hormones and Ca/Mg. Mg, affects PDI-chaperone/anti-chaperone activity (Primm & Gilbert, 2001) – hence the possible association with this Mg associated SNP.

Mg 6_182035399

An inflorescence and root apices receptor-like kinase (IRK)-interacting protein was located 8906 bp at the 3' side. The IRK gene encodes a LRR (leucine rich repeat)-type RLK (receptor-like kinase) and is expressed in proliferating and expanding tissues, such as shoot meristems, floral buds, and root meristems. Despite extensive

genetic studies, the function of IRK still remains poorly understood. However a study by Hattan *et al.,* (2004) revealed that in *Arabidopsis thaliana,* IRK may be involved in lateral organ or lateral root formation under auxin regulation, via microarray gene expression analysis.

Additionally, a retrotransposon protein, Ty1-copia subclass [*Oryza sativa Japonica* Group] was located upon BLASTx search. Retrotransposons are ubiquitous among higher plant species and several studies have documented their activation by stress factors and external changes (Grandbastien, 1998). An Exostosin-like protein [*Corchorus capsularis*] was also located. Family members encode a xyloglucan galactosyltransferase located in the membrane of golgi stacks that is involved in the synthesis of fructose. It is also involved in endomembrane organisation and may play a role in actin organisation and the synthesis of cell wall materials (Tedman-Jones *et al.,* 2008).

Zn 4_43134799

An ACT domain-containing protein, ACR8 was located 637 bp at the 5' side. This gene facilitates the binding of amino acids, due to the ACT domain. However, at present little is known about proteins or regulatory domains involved in amino acid sensing and signalling in plants. A study by Hsieh & Goodman (2002), reports on the identification and molecular characterisation of the ARC Arabidopsis gene family encoding proteins with four copies of the ACT domain that extends throughout the whole polypeptide. The expression patterns of the ACR gene family was assessed by northern-blot analysis using total RNA which was extracted from roots, leaves, stems, flowers, and siliques of Arabidopsis plants. ACR8 mRNA was detected in all the organs tested with significantly higher levels in roots and siliques, thus indicative of specific roles in these particular organs. Interestingly, the steady-state levels of ACR8 mRNA were significantly increased by ABA and NaCl and moderately increased by cold stress, thus indicative of some role in defense. Since the pathways are currently poorly understood, it is unclear whether Zn may play a role as a cofactor or if there may be some Zn tolerance activity. This warrants future study and would involve *in vivo* experiments to establish their distinctive activities and biological roles.

A cation exchanger (CAX)-interacting protein 4 (CXIP4) was located 9319 bp at the 3' side. CXIP4 regulates the CAX1 cation transporter in *Arabidopsis thaliana* (Cheng *et al.,* 2004) and plays an important role in ion homeostasis. Given that CXIP4 contains a CCHC-type zinc finger motif at its N-terminus, it is possible that CXIP4 is regulated by Zn^{2+} in response to ion stress – hence the association with a Zn SNP. The expression of CXIP4 is induced by Ca^{2+}. Cytosolic Ca^{2+} acts as a secondary messenger, and is involved in many biological signalling pathways. To translate the generic signals to specific responses, the concentration of cytosolic Ca^{2+} is efficently regulated by influx systems, such as Ca^{2+} channels, and efflux systems, including Ca^{2+} pumps and anti-porters (Sanders *et al.,* 2002). CAX1 is a high capacity and low affinity Ca^{2+} transporter and is reported to be localised to the plant vacuole. CAX1 has also been found to increase Ca^{2+} levels in tobacco plants and causes numerous stress sensitivity phenotypes often associated with Ca^{2+} deficiencies (Hirschi, 1999). According to Cheng *et al.,* (2004), there are many possible mechanisms by which CXIP4 may regulate CAX1. For example, CXIP4 may activate CAX1-mediated Ca^{2+} transport by altering the transporter conformation. Alternatively, CXIP4 may bind with with additional proteins to alter the subcellular localisation of CAX1 and therefore mediate transport activity and finally, CXIP4 may change cytosolic Ca^{2+} levels and this environmental perturbation may modulate the N-terminus of CAX1 to activate the transporter.

Zn 2_199712768

UDP-glycosyltransferase 91B1 and UDP-glycosyltransferase 91C1 were located 9795 bp at the 5' side and 2935 bp at the 3' side, respectively. Uridine diphosphate (UDP) mediates the transfer of glycosyl residues from activated nucleotide sugars to a wide range of acceptor molecules (aglycones), thus regulating properties of the acceptors such as their bioactivity, solubility and transport within the cell and throughout the plant (Ross *et al.,* 2001). Very little is known about the regulation and pathways of plant UDP genes or the localisation of the enzymes they encode at the cellular and subcellular levels, thus further study is warranted, which may answer why this candidate gene was associated with a Zn SNP.

Fe 4_153484722

Plant cytochrome P450 71A1 was located 54260 bp at the 5' side. It contains an Fe (heme axial ligand) binding site, which may explain its association with an Fe SNP. Plant P450 family members are involved in a wide range of metabolic functions including use in synthetic pathways leading to phenylpropanoids, alkaloids, terpenes, lipids, cyanogenic glycosides and glucosinolates (Bak *et al.,* 2001). Many of these products influence important properties such as flavour and colour in plants. They are also classically associated with pathogen resistance. P450s are also essential in the synthetic pathways of some plant growth regulators including giberellins, brassinosteroids and jasmonic acid (Bundock *et al.,* 2003). More specifically, P450 71A1 has been characterised in *Sorghum bicolour*, which shares a large amount of synteny with pearl millet. It is involved with dhurrin synthesis, which is a cyanogenic glucoside. In response to external damage to the stem, some sorghum varieties release dhurrin at the damage site as a potent insect repellent (Busk *et al.,* 2002).

A gibberellin 2-beta-dioxygenase protein was located 29146 bp at the 3' side and is involved in the gibberellin biosynthesis pathway. Gibberellins are plant hormones that are involved in the regulation of growth in plants. They also influence various developmental processes, including stem elongation, germination, dormancy, flowering, sex expression, enzyme induction and leaf senescence. Fe^{2+} is a cofactor, and the protein contains 3 Fe binding sites, which could explain why this gene is associated with the Fe SNP.

The LRR ERECTA gene was detected 11388 bp at the 3' side. It is involved in the general regulation of aerial architecture, flowering time and stomatal patterning (e.g. density and clustering) (Torii *et al.,* 1996). More specifically, it may also be involved in the regulation of phytate and mineral uptake (Ghandilyan *et al.,* 2009). For example in a study by Ghandilyan *et al.,* (2009), researchers investigated genetic variations for the accumulation of minerals in seeds, rosettes, and roots of *Arabidopsis thaliana* plants grown on different media, to distinguish common QTLs involved in mineral homeostasis. Four regions of interest were located for co-locating QTLs associated with increased mineral uptake (including Fe and Zn) in the

region of chromosome 2, around the ERECTA gene. Thus, the ERECTA locus has been identified as a major QTL for mineral concentrations.

Fe 1_239148556

Heat stress TF A-1 (HSFA1) was located within the 4kb range of this Fe associated SNP, and plays a critical role in response to several abiotic stressors via regulation of the expression of stress-responsive genes, e.g. heat shock proteins (Guo *et al.*, 2016). Specifically, The Arabidopsis HSFA1's are involved in response and tolerance to salt/osmotic stress, and oxidative stresses during seedling establishment (Liu *et al.*, 2011). Additionally, an inactive poly [ADP-ribose] polymerase RCD1-like gene was located within the same region. The Radical-induced Cell Death1 (RCD1) protein is a key regulator of several ROS and abiotic stress related responses in *Arabidopsis thaliana* and rcd1 mutant plants display several phenotypes including salt sensitivity, UV-B and methyl viologen tolerance, early flowering and senescence (Jaspers *et al.*, 2010).

Fe 5_58171748

Heterogeneous nuclear ribonucleoprotein 1 was located 169840 bp at the 5' side. Family members are involved in the regulation of plant growth and environmental stress responses. For example, in *Arabidopsis thaliana*, AtRNP1 is highly expressed in rosette and cauline leaves, and induced under drought, salt, osmotic and ABA stress (Wang *et al.*, 2016).

Cytochrome P450 76C2 was located 35844 bp at the 5' side. As previously discussed, these belong to a large group of enzymes that contain heme as a cofactor, the majority of which catalyse NADPH- and O-dependent hydroxylation reactions (Chapple, 1998). Additionally, its gene expression is associated with various processes leading to cell death such as leaf senescence, ageing, wounding and treatment with the necrotising heavy metal salt, lead nitrate. This was studied in *Arabidopsis thaliana* (Godiard *et al.*, 1998).

Auxin-induced in root cultures protein 12 (AIR12) was located 11428 bp at the 3' side. It binds a single, highly axial low-spin heme, likely coordinated by methionine-91 and histidine-76, which are strongly conserved in AIR12 sequences (Preger *et al.*, 2009). AIR12 has been characterised in *Arabidopsis thaliana* as a single gene that

codes for a mono-heme cytochrome b and plays a role in the regulation of the apoplastic redox state and in the response to necrotrophic pathogens (Costa *et al.,* 2015). At present, little is known about its physiological role. However, a study by Gibson and Todd (2015) revealed an air12 Arabidopsis mutant line which demonstrated increased germination rates in the presence of many of abiotic stressors including high salt/acid and hormones. The same study also demonstrated that the disruption of AIR12 affected primary and lateral root development (Gibson & Todd, 2015).

Additionally, a glucan endo-1,3-beta-glucosidase 5 was located 138099 bp at the 3' side. Family members are associated with plant defence mechanisms against pathogens. For example, Lindthorst *et al.,* (1990) demonstrated that healthy tobacco plants accumulate beta-1,3-glucanases in their roots and in specific parts of their flowers. After infection with tobacco mosaic virus and salicylate treatment, beta-1,3-glucanases were induced in the inoculated and virus-free leaves. Also with respect to plant defense, the candidate gene ASPARTIC PROTEASE IN GUARD CELL 2 was located 49451 bp at the 3' side, which may be involved in drought avoidance through ABA signalling in guard cells (Yao *et al.,* 2012). The increase in ABA biosynthesis caused by dehydration indicates the importance of ABA signalling in response to drought stress in plants (Guerrero & Mullet, 1986). A study by Yao *et al.,* (2012) demonstrated that the overexpression of ASPG1 can enhance ABA sensitivity in guard cells, in turn promoting adaptive drought avoidance in *Arabidopsis thaliana* plants. Other studies suggest that aspartic proteases are involved the regulation of many general physiological processes during plant development including seed germination in wheat (Belozersky *et al.,* 1989), leaf senescence in tobacco (Kato *et al.,* 2004) and reproduction in rice (Chen *et al.,* 2008).

6.4.3 GWAS Using 3,150,286 SNPs from the Pearl Millet Genome Assembly

Genome-wide SNP data (3,150,286 SNPs) was used to compute MTA's in the PMiGAP for grain Fe and Zn uptake. As compared to research by Varshney *et al.,* (2017), for GWAS, a similar total of 3,117,056 SNPs, retained after filtering the minor alleles (MAF<0.05) and 20% missing data were used in their study. The number of SNPs generated for this GWAS is very similar and represents increased coverage of the PMiGAP population than previously used datasets. The data is also

deemed to be less biased i.e. markers that were 'good for GWAS' were selected previously, forming the >37,000 SNP data set, meaning that markers that worked across the bulk of the accessions were selected. Therefore, rarer alleles, indels and monomorphic alleles were removed, which may have skewed the data. QQ plots indicated that population stratification was appropriately corrected for (Figures 6.20 and 6.22).

As previously discussed, LD decay is reported here to be rapid (contradictory to what was observed in the >37,000 SNP dataset, perhaps due to more relaxed filtering stringencies in the marker selection process). According to Varshney *et al.*, (2017), when the r^2 threshold was set as 0.2, rapid LD decay of less than 0.5 kb in PMiGAP lines (84–444 bp) was observed. This is characteristic of allogamous species. GWAS was carried out across 221 PMiGAP lines for Fe and Zn, and the top 4 MTA's (Table 6.10, Figure 6.24) were selected for a NCBI BLAST search for candidate genes (Table 6.12). When the Bonferroni threshold correction at the 5% significance level was applied, none of the MTA's exceeded the threshold, thus associations with candidate genes are explored are based on the lowest P-values possible, eventhough these fall below the Bonferroni threshold for significance.

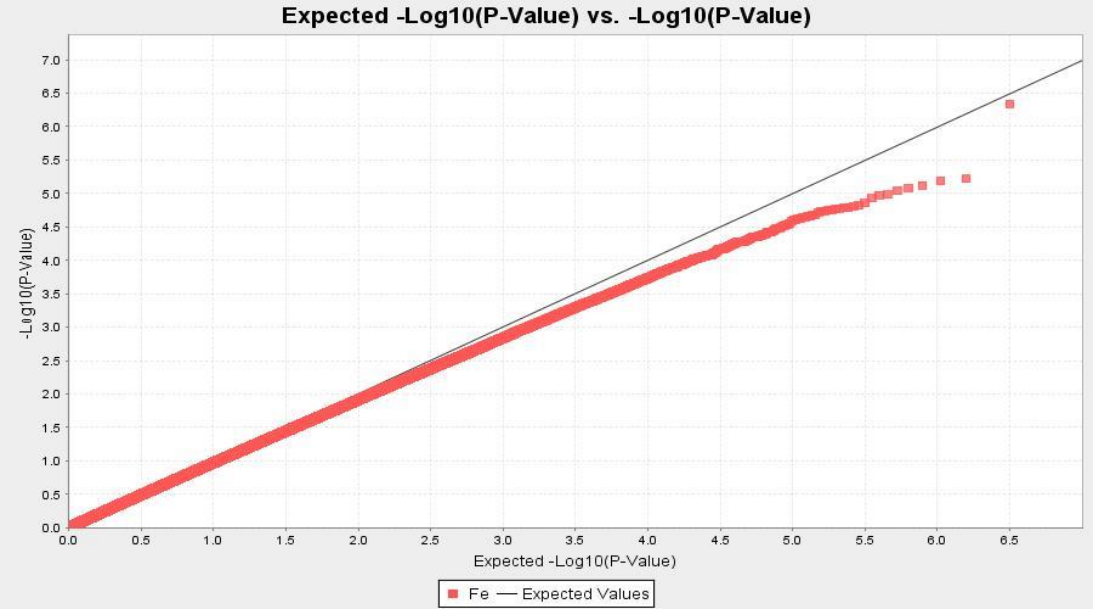


Figure 6.20, Quantile-quantile plot of Fe using the MLM model, built in the TASSEL v5.2.38 environment

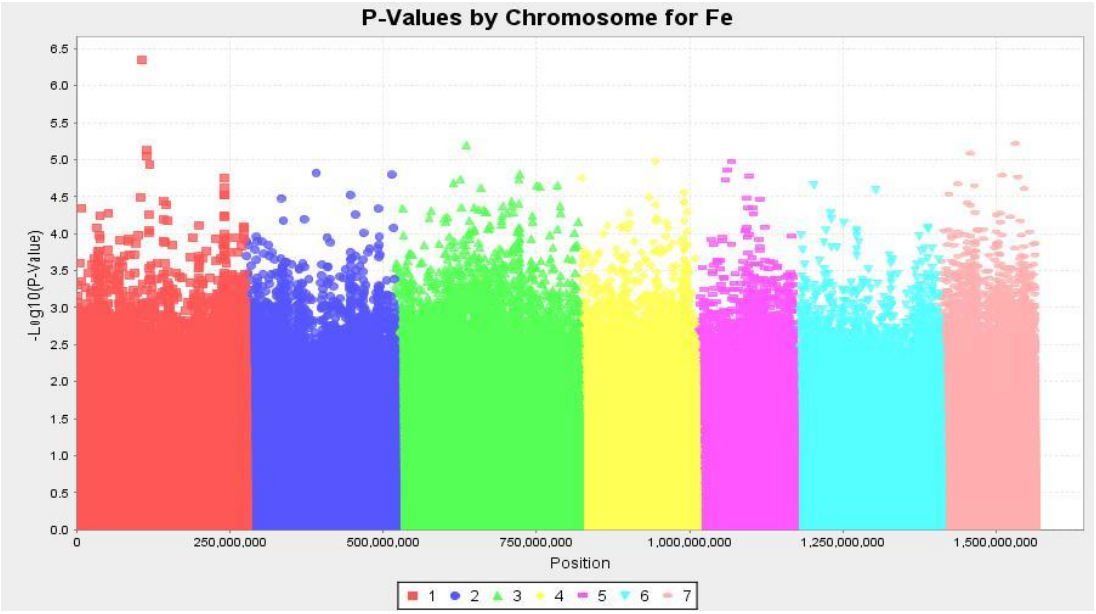


Figure 6.21, GWAS-based Manhattan plots built in the TASSEL v5.2.38 environment exhibiting the lowest P-values (measured by the MLM model) associated with Fe content using >3,000,000 genome-wide GBS SNPs in pearl millet. The x-axis illustrates the relative density of *Pennisetum glaucum* reference genome-based SNPs physically mapped on 7 chromosomes. The y-axis displays the -log10 (P)-value for the degree of association of SNP loci with Fe concentrations

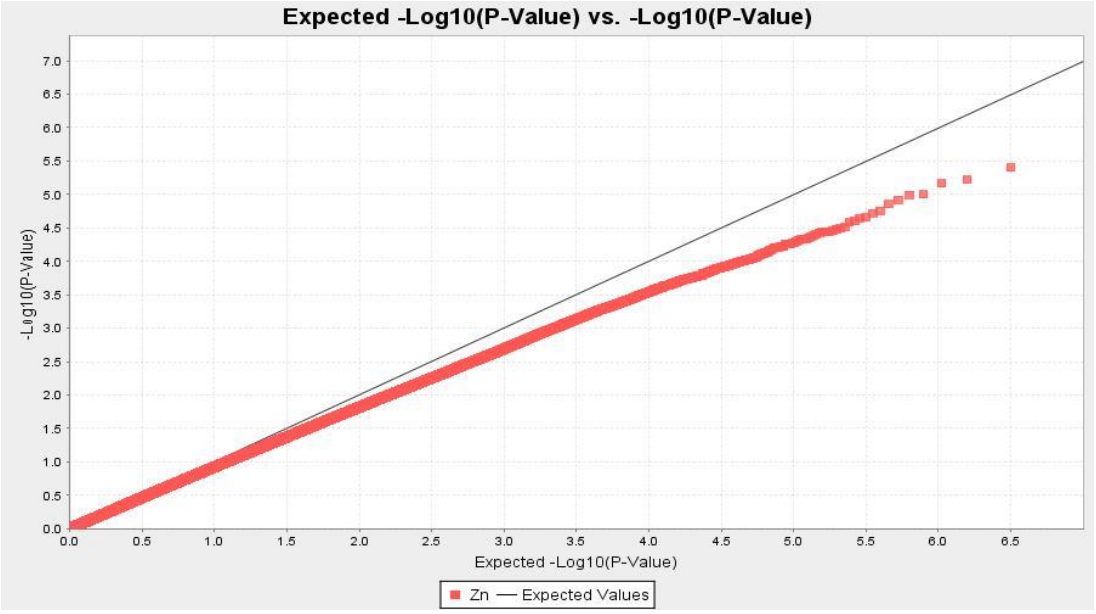


Figure 6.22, Quantile-quantile plot of Zn using the MLM model, built in the TASSEL v5.2.38 environment

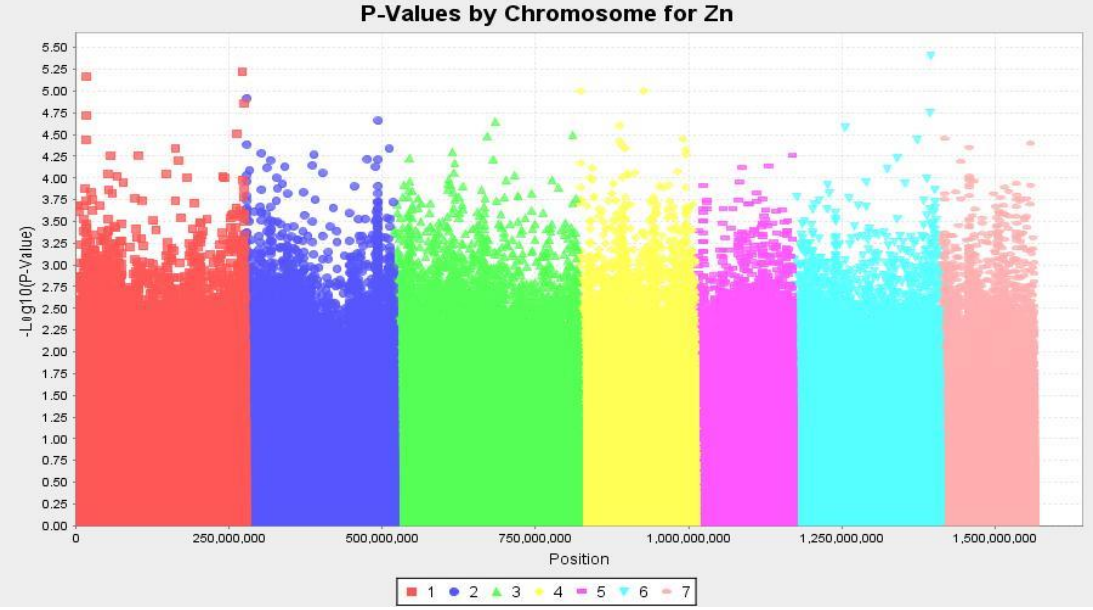


Figure 6.23, GWAS-based Manhattan plots built in the TASSEL v5.2.38 environment exhibiting the lowest P-values (measured by the MLM model) associated with Zn content using >3,000,000 genome-wide GBS SNPs in pearl millet. The x-axis illustrates the relative density of *Pennisetum glaucum* reference genome-based SNPs physically mapped on 7 chromosomes. The y-axis displays the -log10 (P)-value for the degree of association of SNP loci with Zn concentrations

Table 6.10: Top 4 SNPs associated with Fe and Zn uptake, P=<0.001. Bonferroni corrected threshold = 2.65E-09.

[a] Percentage of phenotypic variation explained.

Trait	Marker	Chromosome	Position	Df	F	P-Value	Marker r^2	$\%r^{2a}$	Cumulative $\%r^{2b}$
Fe	Chr1_105654630	1	105654630	2	15.8229	**4.56e-07**	0.17095	**17.10**	58.63
Fe	Chr7_121182972	7	121182972	2	12.8339	**6.06e-06**	0.12938	**12.94**	
Fe	Chr3_117017388	3	117017388	2	12.8008	**6.45e-06**	0.13629	**13.63**	
Fe	Chr1_113738268	1	113738268	2	12.6858	**7.54e-06**	0.14961	**14.96**	
Zn	Chr6_225937596	6	225937596	2	13.3155	**3.99e-06**	0.13930	**13.93**	56.22
Zn	Chr1_271435624	1	271435624	2	12.8531	**6.06e-06**	0.14347	**14.35**	
Zn	Chr1_16534195	1	16534195	2	12.7095	**6.84e-06**	0.14944	**14.94**	
Zn	Chr4_4700502	4	4700502	2	12.2976	**1.00e-07**	0.13004	**13.00**	

[b] Phenotypic variance (cumulative r^2) explained by the genetic effects of all associated SNPs.

When comparing the results to the previous data set where >37,000 markers were used, there were vast differences (Table 6.11). For example, increased marker density gave lower P-values by chance, when using >3,000,000 SNPs, as compared to >37,000 SNPs (Table 6.11). The MTA's explained a large proportion of observed phenotypic variation with individual marker $\%r^2$ values ranging from 12.94 – 17.10%, this was slightly higher than what was found previously (9.55 – 11.47%).

Cumulative %r^2 values ranged from 56.22 - 58.63% for all SNPs found associated with each mineral. Interestingly, when comparing the 6 significant Fe/Zn markers found from the >37,000 SNP data set directly to the p-value of the same marker in the >3,000,000 SNP data set, only two of the markers still gave a p-value below 0.001. These were 153484722 (found to be associated with Fe), which was now at 4.51E-04 as compared to a reading of 3.65E-05, previously and 199712868 (found to be associated with Zn), which was now at 6.05E-05 as compared to a reading of 3.96E-05. The remaining 4 markers previously found to be associated with Fe/Zn did not meet the 0.001 P-value threshold in the >3,000,000 SNP data set. Upon observation of the linkage groups where the associations were found, where LG4 was previously thought to be the dominant location, accounting for 50% of significant markers in the >37,000 data set, the dominant location now appears to be LG1, again accounting for 50% of the low P-value markers, according to the >3,000,000 SNP dataset (Table 6.11).

Table 6.11, A comparison of low P-value MTA's and linkage groups between >37,000 SNPs and >3,000,000 SNPs. MTA = Marker Trait Associations. 'Top'= Strongest by order of significance.

Trait	P-Values for top SNPs	
	>37,000 SNP dataset (Top 3 MTA's)	**>3,000,000 SNP dataset (Top 4 MTA's)**
Fe	3.65 E-05	4.56 E-07
Fe	1.40 E-04	7.54 E-07
Fe	1.69 E-04	6.45 E-06
Fe	-	6.06 E-06
Zn	6.42 E-05	1.00 E-07
Zn	3.96 E-05	6.84 E-06
Zn	2.74 E-05	6.06 E-06
Zn	-	3.99 E-06
	Linkage groups of MTA's	
Fe	4	1
Fe	1	7
Fe	5	3
Fe	-	1
Zn	4	6
Zn	2	1
Zn	4	1
Zn	-	4

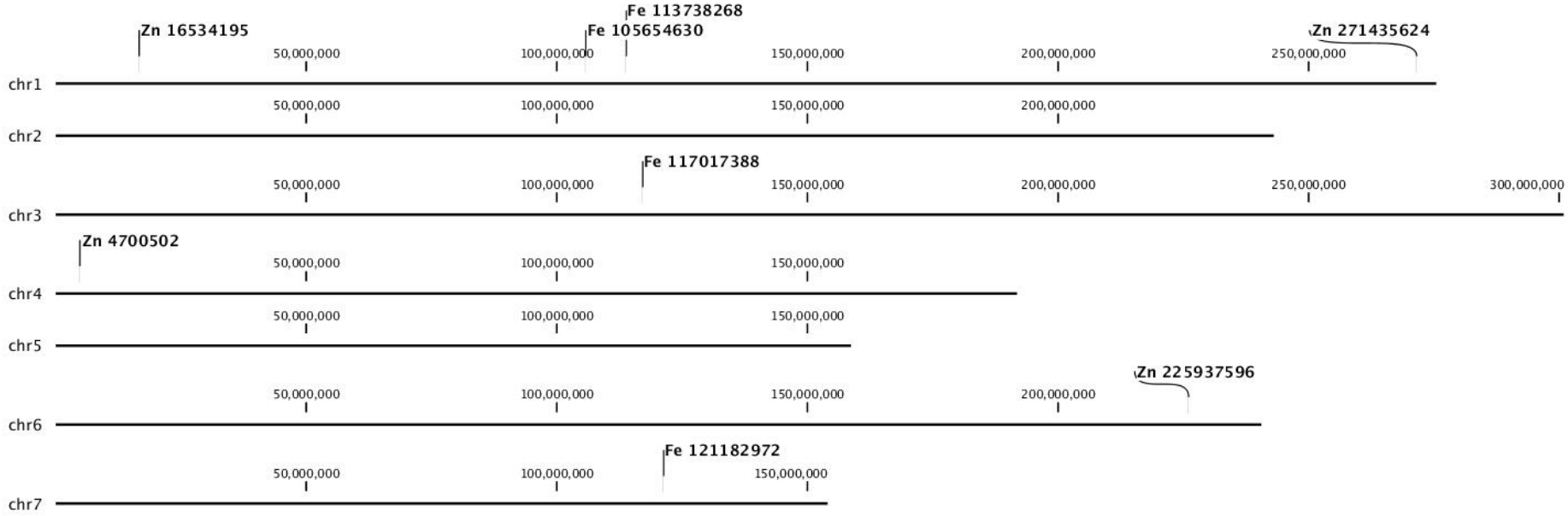


Figure 6.24, SNPs with the lowest P-values from GWAS mapped onto the *Pennisetum glaucum* genome. Chromosome numbers 1 – 7 correspond to chr1 – 7, respectively. Image built in the CLC Genomic Workbench environment.

Table 6.12, BLASTn alignments of low P-value SNPs mapped onto the *Pennisetum glaucum* reference genome assembly + 4kb.

SNP (Chromosome no._Position)	Score	E Value	Identities	Gaps	Range	Candidate Gene(s)
Fe 1 105654630	562 bits(304)	8e-158	581/707(82%)	50/707(7%)	34741313 to 34742005	4482 bp at 5' side: uncharacterized protein LOC101779298 isoform X2 59889 bp at 3' side: histone-lysine N-methyltransferase SETD2-like
	496 bits(268)	8e-138	531/652(81%)	42/652(6%)	32191624 to 32192263	25488 bp at 5' side: 4-hydroxyphenylacetaldehyde oxime monooxygenase-like 65466 bp at 3' side: uncharacterized protein LOC101771994
	479 bits(259)	8e-133	528/652(81%)	42/652(6%)	18744777 to 18745416	48361 bp at 5' side: protein NRT1/ PTR FAMILY 1.2 127804 bp at 3' side: uncharacterized protein LOC101757657
	473 bits(256)	4e-131	530/655(81%)	48/655(7%)	10288202 to 10288841	28285 bp at 5' side: aquaporin SIP1-1 4000 bp at 3' side: S-norcoclaurine synthase
	470 bits(254)	5e-130	531/657(81%)	49/657(7%)	30543271 to 30543913	19746 bp at 5' side: pentatricopeptide repeat-containing protein At4g14850 819 bp at 3' side: non-specific lipid-transfer protein C6
Fe 7 121182972	263 bits(142)	9e-68	279/344(81%)	13/344(3%)	12643199 to 12643539	31605 bp at 5' side: uncharacterized protein LOC101753094 44484 bp at 3' side: anthocyanidin 5,3-O-glucosyltransferase
Fe 3 117017388	158 bits(85)	4e-36	322/438(74%)	9/438(2%)	10400145 to 10400578	28384 bp at 5' side: uncharacterized protein LOC101753560 70932 bp at 3' side: probable E3 ubiquitin-protein ligase XBOS35
Fe 1 113738268	189	1e-45	592/827(72%)	39/827(4%)	9066176 to	35240 bp at 5' side: uncharacterized protein

	bits(102)				9066979	LOC101783463 78996 bp at 3' side: zinc finger protein 7
Zn 6 225937596	106 bits(57)	2e-20	111/137(81%)	4/137(2%)	33445909 to 33446043	5711 bp at 5' side: uncharacterized protein LOC101762365 569 bp at 3' side: F-box/FBD/LRR-repeat protein At5g22700 isoform X2
Zn 1 271435624	3273 bits(1772)	0.0	2100/2253(93 %)	44/2253(1%)	33215505 to 33217735	uncharacterized protein LOC101786921
Zn 1 16534195	486 bits(263)	5e-135	361/409(88%)	4/409(0%)	13393271 to 13393677	9295 bp at 5' side: probable indole-3-pyruvate monooxygenase YUCCA11 30178 bp at 3' side: transcription factor CSA
	468 bits(253)	2e-129	357/408(88%)	4/408(0%)	34824051 to 34824456	21639 bp at 5' side: histone-lysine N-methyltransferase SETD2-like 32170 bp at 3' side: uncharacterized protein LOC101781443
Zn 4 4700502	134 bits(72)	7e-29	205/266(77%)	21/266(7%)	1471782 to 1472030	6816 bp at 5' side: magnesium protoporphyrin IX methyltransferase, chloroplastic 2467 bp at 3' side: glucan endo-1,3-beta-glucosidase 11

Fe 1 105654630

4-hydroxyphenylacetaldehyde oxime monooxygenase-like was located 25488 bp at the 5' side. This protein is involved in step 2 of the sub-pathway that synthesises dhurrin (a cyanogenic glucoside defence compound) from L-tyrosine in *Sorghum bicolour* (Clausen *et al.,* 2015). The association with an Fe SNP may be due to the Fe (heme axial ligand) binding site. Interestingly, an Fe associated marker 153484722, in the region of plant cytochrome P450 71A1, from the >37,000 SNP data set was also found to be involved in dhurrin synthesis, although this was located on chromosome 4.

NRT1/ PTR FAMILY 1.2 was located 48361 bp at the 5' side. Family members were originally identified as nitrate or di/tri-peptide transporters. However, recent studies now suggest that this transporter family also transports the plant hormones auxin, ABA, and gibberellin, as well as secondary metabolites (glucosinolates) (Chiba *et al.,* 2015). Some family members are also associated with Fe deficiency responses. For example, a study by Liu *et al.,* (2015) revealed that NRT1.1 is down-regulated by Fe deficiency. 28285 bp at the 5' side, aquaporin SIP1-1 was detected, which facilitates the transport of water across cell membranes. Additionally, plant aquaporin family members may also transport various small molecules including glycerol, urea (Maurel *et al*. 2002), ammonia (Loqué *et al*. 2005) and CO_2 (Uehlein *et al.* 2003). These studies are vastly improving our knowledge of aquaporins and other transporter genes in plants. Although both of these candidate genes are associated with the transport of a wide variety of molecules, it is currently unknown whether they transport minerals. This may change with research in the future.

Zn 1 16534195

Probable indole-3-pyruvate monooxygenase, YUCCA11 was detected 9295 bp at the 5' side. YUCCA (YUC) flavin monooxygenase was first identified as a key auxin biosynthesis enzyme because overexpression of YUC in *Arabidopsis thaliana* was found to cause auxin overproduction (Cheng *et al.,* 2007). Auxin plays a key role in embryogenesis and seedling development. Furthermore, auxin synthesised by YUCCA flavin monooxygenase is essential for the establishment of the basal body region during embryogenesis and for the formation of embryonic and postembryonic organs, i.e., the part of a seed, consisting of precursor tissues for the leaves, stem and

roots (Chen *et al.*, 2007). The association with Zn could stem from the fact that it is well documented in the literature that Zn is required for auxin synthesis and may be part of the pathway. Studies suggest that disturbance in the metabolism of auxins, especially IAA results in stunted growth and "little leaf" syndrome- the two most distinct visible symptoms of Zn deficiency (Alloway, 2004). Additionally, a study by Begum *et al.*, (2016) reported that auxin signalling may trigger Zn uptake, transport and chelation in rice seedlings to withstand Zn-deficiency. The study also elucidates the involvement of auxin with Zn-efficiency in a Zn-efficient rice variety, *Pokkali*. This variety showed no significant decrease in physiological features, electrolyte leakage and total soluble proteins as a result of Zn deficiency as compared with Zn sufficient seedlings. However, an auxin inhibitor under Zn deficiency severely affected these characteristics, suggesting that Zn efficiency is associated with auxin signalling. Results also revealed a significant reduction in the expression of Zn transporter genes (OsIRT1, OsZIP4 and OsZIP1), OsDMAS1 (deoxymugeneic acid synthase) and phytochelatin in roots due to the auxin inhibitor. When the two key findings are linked, that i) YUCCA11 may regulate auxin biosynthesis and ii) auxin signalling triggers Zn efficiency, it can be elucidated that YUCCA11 may indirectly affect Zn efficiency in pearl millet.

6.5 Conclusions

In a study by Kumar *et al.*, (2016), using 305 loci, a linkage map was constructed to map QTLs for grain Fe and Zn content using replicated samples of 106 pearl millet RILs derived from two PMiGAP lines: ICMB 841-P3 × 863B-P2. On the basis of phenotypic mineral data from two environments, two co-localised QTLs for grain Fe and Zn content were identified on linkage group (LG) 3 by composite interval mapping. For OP seeds, the analysis also led to the identification of two QTLs for grain Fe content on LG3 and 5, and two QTLs for grain Zn content on LG3 and 7. The results differ to the findings generated in this study. For example, in Table 6.7, which shows the top 3 low P-value SNPs and chromosome positions detected for all minerals using >37,000 SNPs mapped to pearl millet, the SNPs for Fe and Zn fall instead on chromosomes 1, 2, 4 and 5. Similarly, for the >3,000,000 SNP data set, according to Table 6.10 none of the SNPs for Zn appeared on LG 3/7. However, 1 strongly associated Fe SNP was located on LG3. Instead, the majority of Fe/Zn associated markers fell on chromosome 1 and none were co-localised, in both cases.

Interestingly, the significance of chromosome 1 coincides with results from another study by Kumar *et al.,* (2018), where a genetic linkage map was constructed using 317 pearl millet RILs, derived from two Iniadi lines (ICMS 8511-S1-17-2-1-1-B-P03 × AIMP 92901-S1-183-2-2-B-08). Three large-effect QTLs for both minerals were located, one on LG1 and two on LG7. The differing results may be due to a number of factors, namely; (i) environment and pollen source play important roles in Fe/Zn concentrations in seeds, which consequently may explain the different chromosome positions of detected SNPs, (ii) the transport and accumulation of minerals in seeds is a complex trait that requires a combination of different genes on different chromosomes and (iii) the phenotype data in this study may be limited for the reasons discussed previously in Chapter 3, including the effect of the environment on selfed seed set, mineral distribution, grain size (small seeds were shown to be susceptible to concentration effects) and uptake.

As previously discussed in Chapter 1, the strong correlation between grain Fe and Zn content has been studied in several crops, with results, showing similar trends. Similar to this study, Kumar *et al.,* (2016) also reported that grain Fe and Zn contents were strongly and positively associated. This may be owing to common molecular mechanisms controlling the uptake and metabolism of these minerals in the seed or common transporters controlling the movement of these minerals within the plant (as previously discussed in Chapter 1). Kumar *et al.,* (2016) also hypothesised that the co-segregation of QTLs on LG3 for both Fe and Zn might be the reason for the strong association between grain Fe and Zn content. In this study, findings from the >37,000 SNP data set revealed that low p-value SNPs were located on chromosome 4 for both Fe and Zn, as seen in Table 6.7. Although the SNPs are located on the same chromosome, they are not as close together as demonstrated in other studies (Kumar *et al.,* 2016; Jin *et al.,* 2013). Similar findings from the >3,000,000 SNP data set also revealed that low P-value SNPs located on chromosome 1 (2 × Fe and 2 × Zn) were not close together. Given that there are other SNPs between the significant markers in both cases, it is not likely that they are of the same association. Despite the distance, this study still indicates that chromosomes 1 and 4 may hold SNPs for the uptake of both Fe and Zn. Interestingly, upon observation of the SNPs associated with grain Fe and Zn content as a result of 663 SNPs from the *Setaria italica* genome (Table 6.4), the SNPs are

much closer together, located at positions 4623914 and 3820759 on chromosome 9, respectively – thus indicative of some degree of co-localisation. Other studies have also reported this in other crops, including one by Jin *et al.*, (2013), who investigated the genetic architecture of Fe and Zn content in *Zea mays* grains as revealed by QTL mapping and meta-analysis. It was found that for Zn content, 4 QTLs were identified on chromosomes 2, 5 and 10, whereas for Fe content, only one QTL was located on chromosome 5. The QTLs identified for Fe and Zn on chromosome 5 were in the marker interval umc1429–umc1060, and therefore considered co-localised. The co-localisation of nutrient element QTLs may be due to tight linkage of distinct genes, pleiotropism or physiological association of micronutrient accumulation; this suggests that there is a relationship at the molecular level among these traits (Jin *et al.,* 2013). Additionally, the co-localisation of QTLs for the content of multiple elements has been reported in in wheat (Wu *et al.,* 2008), *Brassica oleracea* (Broadley *et al.,* 2008), *Arabidopsis thaliana* (Vreugdenhil *et al.,* 2004) and rice (Shimizu & Guerta, 2005). Co-localisation is also demonstrated to a more significant degree in two pairs of Ca associated SNPs from the 663 SNP data set, as indicated by overlapping of SNPs at positions 5882761 and 5882763 on chromosome 1 (2bp apart) and 37189739 and 37189769 on chromosome 6 (30bp apart). This has also been demonstrated in a study by Broadley *et al.*, (2008), who reported that QTLs for shoot Ca and Mg uptake potentially co-localise on chromosome 6, 8 and 9 in *Brassica oleracea*.

A surprising finding from this study was the lack of well-known Fe/Zn uptake genes found within the region of low P-value markers, such as the YSL proteins, the ZIPs or the IRT proteins (as described to a fuller extent in Chapter 1). This may be attributable to the 4kb sequence surrounding SNPs from the pearl millet genome assembly, in the case of the >37,000 and >3,000,000 SNP data sets being BLASTed against the *Setaria italica* reference genome. This is because the *Pennisetum glaucum* genome was not on the NCBI genome viewer at the time of this research. This may have affected the quality and accuracy of the downstream analysis, thus findings from the GWAS may be less valid than in if a pearl millet reference genome was used, simply because if there is too much sequence variation present between the two species, this may reduce the accuracy in finding genes when conducting a BLASTn search. In terms of comparative mapping between the two species, a study

by Devos *et al.,* (2000) investigated the homoeology between the genomes of pearl millet and foxtail millet and a comparison revealed that despite the close taxonomic relationship, their genomes were highly rearranged and most of these rearrangements may have taken place in pearl millet. The study involved mapping pearl millet RFLP probes onto a foxtail millet population. The pearl millet genetic map was compared with that of foxtail millet and it was found that large chromosomal rearrangements that took place in the millet genomes, relative to rice could be classified as either species-specific or as characteristic to the taxonomic group. When comparing pearl millet to foxtail millet, LG1 from pearl millet was shown to be homoeologous (of similar genetic constitution) with a segment of foxtail millet (FM) chromosome 8. Pearl millet LG2 is homoeologous with segments of FM9, FM4 and FM1. LG4 in pearl millet is homoeologous with a segment of FM3, most of FM6 and the duplicated regions of FM7 and FM8. LG6 is largely homoeologous to FM5 and LG7 is homoeologous to FM2. An additional comparison of the organisation of the genomes of rice, foxtail millet, sugar cane, sorghum, pearl millet, maize, wheat and oat revealed that most of these rearrangements were present in pearl millet only and therefore must be of recent origin. The pearl millet genome also carries at least one and two duplications between LG1 and 4, respectively, which are likely to be independent events. One corresponds to the duplication found between the short arms of rice chromosomes 11 and 12 and occurred before the divergence of the *Panicoideae* and *Oryzoideae* subfamilies. The other seems to be specific to pearl millet (Devos *et al.,* 2000).

Another reason for the lack of well-known Fe/Zn uptake genes may be due to few (3/35) p-values of MTAs exceeding the Bonferroni corrected threshold. Even though the Bonferroni correction is considered to be extremely stringent, it is a necessary step to ensure the prevention of genome-wide type 1 errors (false positive associations). It also reduces the probability of identifying SNPs with small effect size (Stringer *et al.,* 2011). Multiple testing must be taken seriously since false claims of associations that show only nominal significance are seldom replicated. Thus, it is important to avoid this, even if it means sacrificing positive results from our investigations. In Varshney *et al*.'s (2017) study, multiple testing was accounted for using FDR (False Discovery Rate) at a 0.001 threshold level and only p-values lower than 1E−10 were considered. Multiple testing was accounted for in this study

using the Bonferroni threshold correction method at the 5% significance level. When the corrected threshold was calculated at 1.26E-05 for the 663 SNP data set, 2.23E-07 for the >37,000 SNP data set and 2.65E-09 for the >3,000,000 SNP data set, it was found that just three SNPs from a total of 35 were above the thresholds. These were Na_ 3_221196763, Na_5_65686933 and Na_3_193997829 from the >37,000 SNP data set. The lack of significant p-values that exceeded the Bonferroni corrected threshold may be due to known errors, i.e. the relatively small size of the PMiGAP population combined with the lack of replicates. A larger population with at least 3 replicates would significantly increase the power of GWAS for future work. For example, Varshney *et al.,* (2017) did GWAS on 288 PMiGAP lines, as compared to 221 lines used in this study for the >37,000 and >3,000,000 SNP data sets and 171 lines used for the 663 SNP data set. In terms of replicates, two replications in three test environments were used in Varshney *et al.'s* (2017) study, which greatly increased the validity and reliability of results.

Chapter 7: Haplotype Analysis for the Verification of the YUCCA11 Gene

7.1 Summary

In this chapter, 9 haplotypes covering the YUCCA11 gene were idenitifed and their association with trait data (combined Fe and Zn levels) from 42 PMiGAP lines was assessed.

Since haplotype analysis provides more evidence for associations than single SNP analysis from GWAS, haplotypes around the most relevant candidate gene, YUCCA11, were identified and analysed to determine whether an association with Fe/Zn uptake was present. This was achieved by looking at combinations of SNPs rather than individual SNPs, to strengthen the conclusion that these SNPs are associated on the YUCCA11 gene. A total of 9 haplotypes in 42 PMiGAP lines were detected, and findings revealed no association between any haplotype and Fe/Zn uptake. This may have been attributable to a number of reasons, the most likely of which being the fact that the MTA was likely a chance outcome. Other points to consider are known issues with the phenotype data (i.e. lack of replicates) affecting the downstream analysis and the relatively small sample size. Additionally, data was examined to determine whether any haplotype was associated with a particular region, of which no association was found. This reflects the lack of population structure in the PMiGAP.

7.2 Introduction

A haplotype is defined as a group of genes within an organism that are inherited together from a single parent. In addition, the term 'haplotype' may also refer to the inheritance of a cluster of SNPs. The analysis of haplotypes with the grouping and interaction of several variants is sometimes superior to individual SNP analysis techniques and significantly improves the power and robustness of association studies (Wu *et al.,* 2014). However, if the causal connection between SNP and phenotype is truly driven by just one SNP, then the haplotype-based approach may perform worse than the one-SNP-at-a-time approach (Clark, 2004).

SNPs are considered choice markers for GWAS as they are the most abundant class of sequence variability in the genome and they provide the highest map resolution

(N'Daiye *et al.,* 2017). However, SNPs are typically bi-allelic so each provides less polymorphism information content than that of other markers, such as SSRs (which are multi-allelic), therefore marker density should be increased. This limitation is typically overcome by merging SNPs into haplotypes (Lu *et al.* 2012). Although haplotype-based analyses are more commonly used in human genetics studies, similar efforts are gaining ground in several staple crops, including maize (Lu *et al.,* 2012), rice (Lestari *et al.,* 2011) and soybean (Langewisch *et al.,* 2014).

7.2.1 Single-Point Analysis from GWAS vs. Haplotyping

GWAS almost invariably use single point analysis because the traditional methods used to analyse large-scale genetic data are lagging behind the rapid advances in industrial omics/NGS technology. Traditional genetic analyses from association mapping explore likely single markers associated with a trait of interest and as a result identify only a small proportion of genetic variants responsible. This contributes to an incomplete understanding of complex phenotypic traits. In addition, the current popular single-point analysis of GWAS data suffers from low validation and replication rates (Panoutsopoulou & Zeggini, 2009). There is a growing consensus that genetic factors attributing to complex pathways, such as mineral uptake in plants is contributed to by multiple genes/SNPs, rather than by the mutations of individual genes (Dunn *et al.,* 2007). For example, Dunn *et al.,* (2007) suggests that many proteins/genes may be involved in Fe uptake, metabolism and homeostasis in plants including Divalent Metal Transporter-1, Ferroportin-1 and Heme Carrier Protein-1, to name a few. Hence, to further interpret the underlying molecular mechanisms behind mineral uptake, systematic dissection of the interactions between multiple genes as well as their functionalities is essential.

A disappointing finding from many GWAS is the lack of association between relevant candidate genes and their cognate traits, couplied with the discovery of a wide range of genomic regions, some containing no genes at all, that have a small effect size on traits. This is common in the literature. For example, in a study by Anuradha *et al.,* (2017), a pearl millet association mapping panel of 130 diverse lines was phenotyped for Fe and Zn levels and MTAs were analysed with 267 markers (250 SSRs and 17 genic markers). 3 markers were found to be significantly associated with Fe/Zn uptake in pearl millet, across three different environments.

However only 1 matched a segment of the pearl millet reference genome, annotated as the gene *Asp1*, which does not appear to be directly linked to Fe/Zn uptake. The results from the GWAS, described in Chapter 6 indicated that none of the well-known genes for Fe/Zn uptake, including the YSL and ZIP family transporter genes, were identified. Additionally, there were no overlapping candidate genes between the three GWAS. This has prompted some suggestions that a two tier system is needed for the improvement of association mapping. Haplotype analysis is an effective technique that overcomes many limitations of single-point analysis and is a plausible method of investigating the lack of association between candidate genes and Fe/Zn uptake. Reasons for the lack of association could be due to several factors, namely: (i) Differences in the effect size of the SNPs, therefore those associated with Fe/Zn uptake were not detectable. Normal practise is to select a low significance threshold; this reduces the probability of identifying SNPs with small effect size (Stringer *et al.,* 2011). (ii) The QTL may not have been segregating at a sufficiently high frequency to be detected. (iii) Differences in the density of SNP coverage for the candidate genes relating to mineral uptake (especially in the case of the 663 SNP data set) (iv), Differences in the degree of LD across the region (Barendse, 2011) and (v), the fact that few SNPs exceeded the Bonferroni corrected threshold may point to the lack of true MTA's, i.e. false positives.

7.2.2 Experimental Aims

42 PMiGAP lines were selected for haplotype analysis based on high, medium and low grain Fe/Zn content. DNA was extracted from young leaf tissue and 6 candidate genes were chosen for haplotype analysis. This was subsequently reduced to 1 candidate gene (with the most lowest P-value) – YUCCA11, due to the weak MTA's detected in Chapter 6. A transposome-based Nextera XT kit was used to generate libraries for sequencing on a MiSeq (Illumina platform). Data analysis included the determination of haplotype trait associations (HTA's) and whether any haplotype was associated with a particular region.

7.3 Materials and Methods

7.3.1 Plant Material

See Chapter 2, Table 2.3.

PMiGAP leaf tissue samples were selected based on low (14), medium (18) and high (16) combined grain Fe/Zn content. The low Fe/Zn lines ranged between 59.83 – 72.9 mg/kg, the medium Fe/Zn lines ranged between 106.34 – 109.37 mg/kg and the high Fe/Zn lines ranged between 158.04 – 214.58 mg/kg.

7.3.2 DNA Extraction from Young Leaf Tissue

DNA was extracted using the Qiagen DNeasy® 96 well plant kit. Liquid nitrogen was used to transfer the frozen leaf tissue samples into 2cm collection tubes. A tungsten carbide bead was added to each tube and tissue was homogenised using a Qiagen TissueLyser for 2 × 40 second bursts, at 25 Hz. Samples were then pulse centrifuged at 3000 rpm to bring any excess tissue down from the caps. A working lysis solution was made up with 30mL API (preheated to 65°C, using a water-bath), 75mL RNAse A and 75mL reagent DX. 400µL of the working lysis solution was added to each collection tube and left for 2 minutes on the workbench. 130µL P3 precipitation buffer was added to each collection tube (to remove carbohydrates and proteins) and resealed using new caps. Samples were then shaken vigorously by hand for 1 minute, ensuring a 90° turn every 10 seconds, then pulse centrifuged at 3000 rpm to collect any remaining solution from the caps. The tubes were then stored at -20°C for 20 minutes. Samples were centrifuged for 5 minutes at 6000 rpm and 400µL of the supernatant was added to new collection tubes. 600µL AW1 buffer was added to each tube and closed with new caps. Samples were then shaken vigorously for 15 seconds, and pulse centrifuged at 1000 rpm to collect any solution from the caps. A DNeasy 96 well plate was placed on top of an S-block and 1mL of each sample was added to each well. The plate was sealed with an Airpore tape sheet and centrifuged for 4 minutes at 6000 rpm. The tape was removed and 800µL AW2 buffer was added to each sample. Samples were centrifuged for a further 15 minutes at 6000 rpm, without tape to evaporate any remaining EtOH. 100µL AE buffer was added and samples were incubated at room temperature for 1 minute, then centrifuged for 2 minutes at 6000 rpm. An additional 100µL AE buffer was added and samples were again incubated at room temperature for 1 minute, and then centrifuged for 2 minutes at 6000 rpm. Samples were mixed by vortex to ensure homogeneity and a final pulse spin at 6000 rpm ensured all the sample was removed from the lid.

A 1% agarose gel confirmed the presence of DNA by visualisation (Figure 7.1) and each sample was quantified using an Epoch spectrophotometer (BioTek Ltd) to measure light absorption at 260nM for individual readings. The DNA was then stored at -20°C.

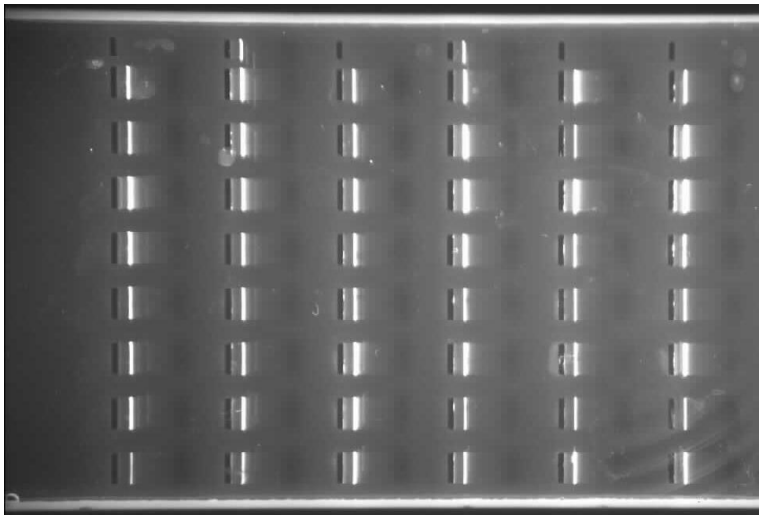


Figure 7.1, 1% agarose gel image, confirming the presence of DNA from PMiGAP DNA extractions

7.3.3 Candidate Genes

At first, 6 candidate genes were selected for haplotype analysis due to their association with Fe/Zn uptake, as reported in the literature (Table 7.1).

Table 7.1, Candidate genes selected for haplotype analysis

SNP (Chromosome no._Position)	SNP dataset	CG of interest	Location	Function
Ca 7_34304700	663	V-type proton ATPase subunit D	20506 bp at 3' side	There is some experimental data that highlights the effect of proton-pumping V-ATPase activity on metal transport. E.g. proton gradient-dependent transport of metals across the tonoplast and antiporter activity depends on the presence of a proton gradient across the vacuolar membrane and thus, indirectly on V-ATPase (Dietz *et al.,* 2001).
Fe 4_153484722	>37,000	Cytochrome P450 71A1-like	54260 bp at 5' side	C-P450's are a large group of enzymes that contain heme as a cofactor. Some C-P450 family members accumulate in response to Fe deficiency (Colangelo & Guerino, 2004).
Fe 5_58171748	>37,000	Cytochrome P450 76C2-like	35844 bp at 5' side	
Na 5_65686933	>37,000	Cytochrome P450 99A2-like	206267 bp at 5' side	
Fe 4_153484722	>37,000	LRR receptor-like serine/threonine-protein kinase ERECTA	11388 bp at 3' side	The ERECTA locus has been identified as a major QTL for mineral concentrations, including Fe and Zn in *Arabidopsis thaliana* (Ghandilyan *et al.,* 2009).
Zn 1_16534195	>3,000,000	Indole-3-pyruvate monooxygenase YUCCA11	9295 bp at 5' side	YUCCA11 is associated with auxin biosynthesis. In turn, auxin drives Zn efficiency by triggering uptake, transport and chelation in Zn-deficient conditions (Begum *et al.,* 2016).

For each candidate gene, the base position (start/end region), origin gene sequence and the FASTA sequence were located by BLASTn search of the NCBI database.

7.3.4 PCR Primer Design

For each candidate gene, the FASTA sequences were imported into the program PRIMER3 (http://bioinfo.ut.ee/primer3-0.4.0/). The product size range was set to be within 100bp of the start/end region. All the default settings were used, including optimum 50% CG content and primer size range was set to 18 – 27bp. 1 CG clamp was added to each primer. Primers for each candidate gene can be seen in Table 7.2 and were ordered from Sigma Aldrich, UK at 0.025µmole, desalted and in dry format.

Table 7.2 Forward and reverse primer sequences for CG's. T_m = Melting Temperature.

DNA Oligos	Sequence			
	FWD	T_m°C	REV	T_m°C
V-ATPase subunit D	CGCTGCTGAAGAAGAAGTCC	64.2	GATGATGTCGTCGTCCTTCTC	63.7
Cytochrome P450 99A2	TATGGTGATGCCACAAGGTG	64.5	CAGCGACGCTTATCTTCCTC	63.9
Cytochrome P450 71A1	CGACAAGGGTCAAGGGATAC	63.3	GTTAGACGCCTTCGATCAGG	63.7
LRR ERECTA	TACATCAGGCAGAGCGAGAC	63.2	GACCCGGTCGACTTCCTG	65.9
Cytochrome P450 76C2	AAGGCCTAGGATGGCTTGTC	64.3	CGACCCCACACTTCTCTTTC	63.6
YUCCA11	TCCCCTACCTTGTCGTTGAG	64	TGGCAATGTTGTTAGCATCC	63.5

7.3.5 PCR Optimisation

Primers were made up with 0.1 × TE buffer to 100µM, to create the stock solution, which was then diluted to 5µM with distilled water to make the working solution. 50µL of the forward and reverse primers, per candidate gene were combined.

Two PMiGAP lines with large amounts of DNA were selected, measuring at 25.4 ng/µL and 15.4 ng/µL, respectively. A 50µL PCR reaction was conducted using 25µL Phusion High-Fidelity DNA Polymerase, 2µL Primer, 2/3µL DNA (2µL for the 25.4 ng/µL DNA sample and 3µL for the 15.4 ng/µL DNA sample) and 21/20µL

distilled water (21µL for the 25.4 ng/µL DNA sample and 20µL for the 15.4 ng/µL DNA sample). The PCR cycle was set to:

95ºC 10 minutes

95ºC 30 seconds
58ºC 30 seconds
72ºC 30 seconds
72ºC 3 minutes
× 35 cycles

72ºC 10 minutes
4ºC hold

58°C was chosen for the initial annealing temperature, due to it being 5°C below the lowest primer sequence T_m. A 1% agarose gel confirmed the retention of PCR products (Figure 7.2). When the PCR products showed evidence of mispriming at 58°C due to a laddering pattern, higher annealing temperatures including 59°C, 60°C, 61°C, 62°C and 63°C were tested, keeping all conditions the same as above. However, results still showed the same laddering pattern in all cases (Figure 7.2).

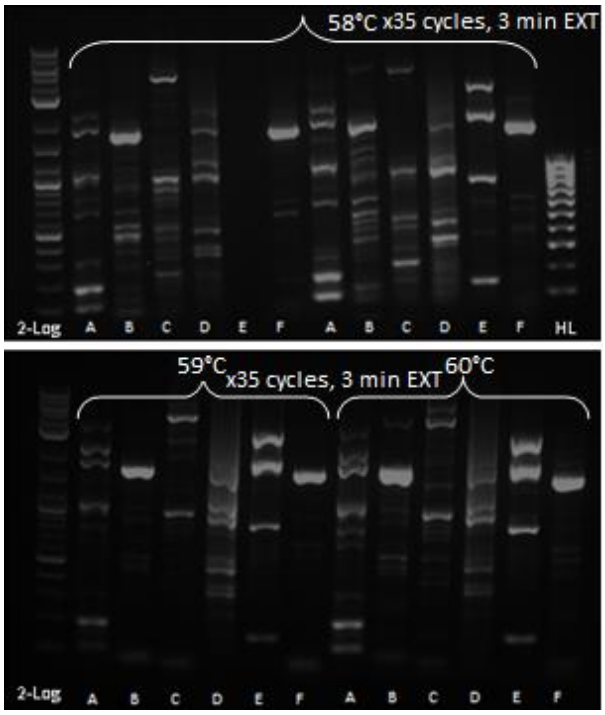


Figure 7.2, PCR products as confirmed by 1% agarose gel. Image shows mispriming in most cases at 58°C (Top image), 59°C (Bottom image, Wells 2-7) and 60°C (Bottom image, Wells 8-13). EXT = Extension time, 2-log = 2-Log DNA Ladder, HL = HyperLadder, A = V-ATPase subunit D, B = Cytochrome P450 99A2, C =

Cytochrome P450 71A1, D = LRR ERECTA, E = Cytochrome P450 76C2, F = YUCCA 11.

A touch-down PCR was conducted, where the initial annealing temperature was set to 63°C and gradually reduced by -5°C increments over 10 cycles until the T_m or "touch-down temperature" was reached, as below.

12.5µL Phusion
1µL DNA (24.5ng/µL)
1µL Primer
10.5µL Water

Heat lid 112°C
95°C 10 minutes
95°C 30 seconds
63°C 30 seconds (-0.5) } × 10
72°C 3 minutes
95°C 30 seconds
58°C 30 seconds } × 25
72°C 3 minutes

72°C 10 minutes
12°C Hold

A 1% agarose gel revealed that the touchdown PCR failed in the majority of cases, as confirmed by a laddering/smear pattern (Figure 7.3).

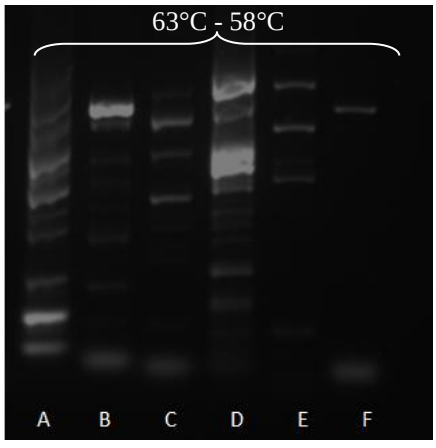


Figure 7.3: Touchdown PCR products as confirmed by 1% agarose gel, 63°C - 58°C. A = V-ATPase subunit D, B = Cytochrome P450 99A2, C = Cytochrome P450 71A1, D = LRR ERECTA, E = Cytochrome P450 76C2, F = YUCCA 11.

In an effort to address the issue of the laddering patterns, the number of PCR cycles was altered to ×20, ×25 and ×30, respectively. Although there was an improvement

to the PCR product, mispriming could still be seen for the majority of cases (Figure 7.4).

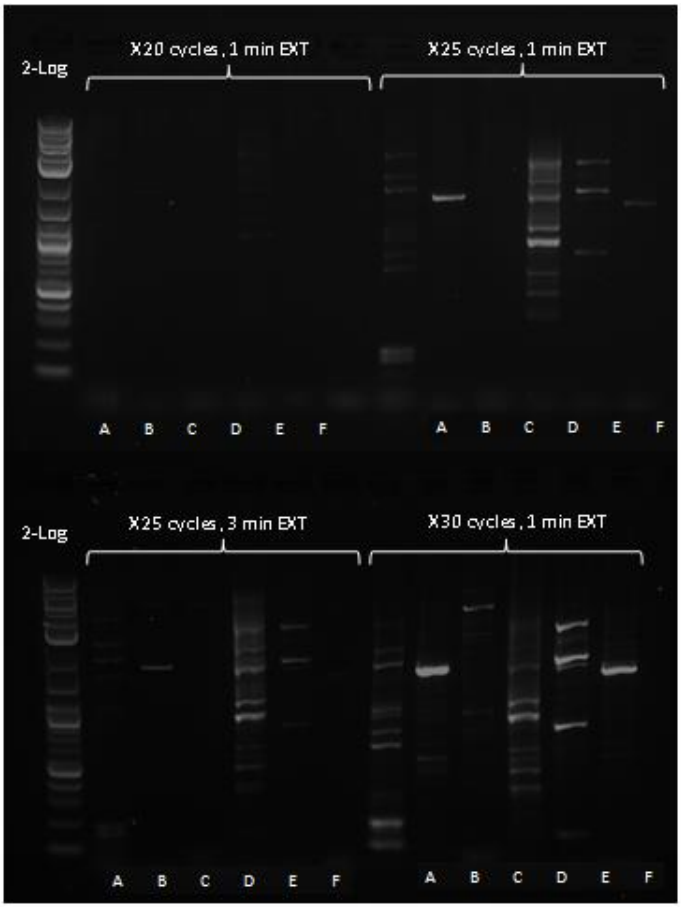


Figure 7.4, PCR products as confirmed by 1% agarose gel. Image shows mispriming in most cases at ×20, ×25 and ×30 cycles. EXT = Extension time, 2-Log = 2 Log DNA ladder. A = V-ATPase subunit D, B = Cytochrome P450 99A2, C = Cytochrome P450 71A1, D = LRR ERECTA, E = Cytochrome P450 76C2, F = YUCCA 11.

A different reaction mix, ImmoMix (based on IMMOLASE™ DNA polymerase) was used in place of Phusion buffer, under the following conditions:

7.5µL Immomix
2µL DNA (24.5ng/µL)
1µL Primer
4.5µL Water

95ºC 10 minutes

95ºC 30 seconds
59ºC 30 seconds
72ºC 30 seconds
72ºC 1 minute

× 35 cycles

72ºC 10 minutes
4ºC hold

The PCR product was AMPure cleaned (0.6 AMPure: 1 DNA) with 100µL 80% EtOH and resuspended in 12µL Illumina resuspension buffer.

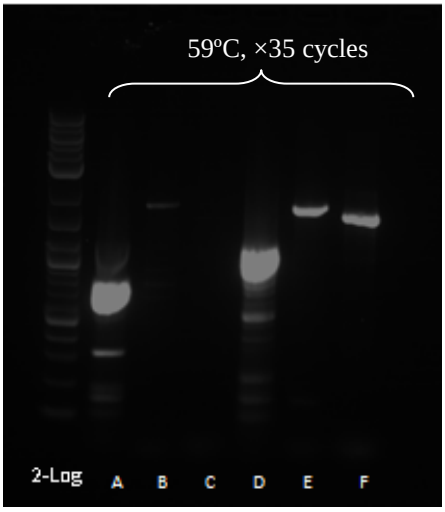


Figure 7.5, PCR products as confirmed by 1% agarose gel. Image shows successful PCR in lanes B, E and F. 2-Log = 2 Log DNA ladder. A = V-ATPase subunit D, B = Cytochrome P450 99A2, C = Cytochrome P450 71A1, D = LRR ERECTA, E = Cytochrome P450 76C2, F = YUCCA 11.

Whilst the PCR conditions were deemed successful for wells 3 6 and 7 (Figure 7.5), corresponding to the candidate genes; Cytochrome P450 99A2, Cytochrome P450 76C2 and YUCCA11, respectively, the laddered PCR products in wells 2 and 5, corresponding to the candidate genes; V-ATPase subunit D and LLR ERECTA were further improved by reducing the temperature from 59°C to 58°C and changing the number of cycles from ×35 to ×30 and the extension time from 1 minute to 1 minute 30 seconds. The PCR product was AMPure cleaned (0.6 AMPure: 1 DNA) with 100µL 80% EtOH and resuspended in 12µL Illumina resuspension buffer. This proved to be successful (Figure 7.6).

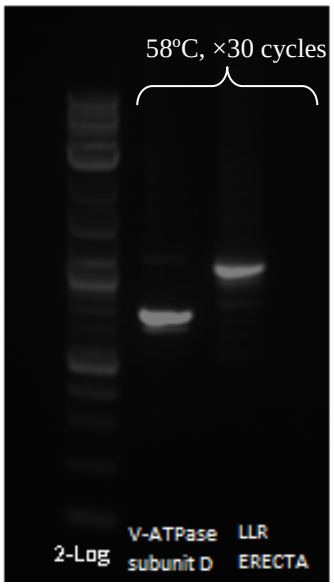


Figure 7.6, PCR products as confirmed by 1% agarose gel. Image shows successful PCR's corresponding to the CG's: V-ATPase subunit D and LLR ERECTA, respectively. 2-Log = 2 Log DNA ladder.

7.3.6 Sanger Sequencing

Sanger sequencing was used to confirm the correct amplification of the PCR product. 5µL of AMPure cleaned PCR product plus 2pmol of primer (forward and reverse, separate) were subjected to Sanger sequencing, using an accredited service by Ms Caron Evans, University of Aberystwyth, Division of Genomics. The raw data was analysed in the Chromas v.2.6.4 environment (https://technelysium.com.au/wp/chromas/) and low-quality sequence was trimmed (Figure 7.7). The remaining FASTA sequence for each sample was BLASTed against all assemblies available on the NCBI website to confirm the correct amplification of DNA sequence.

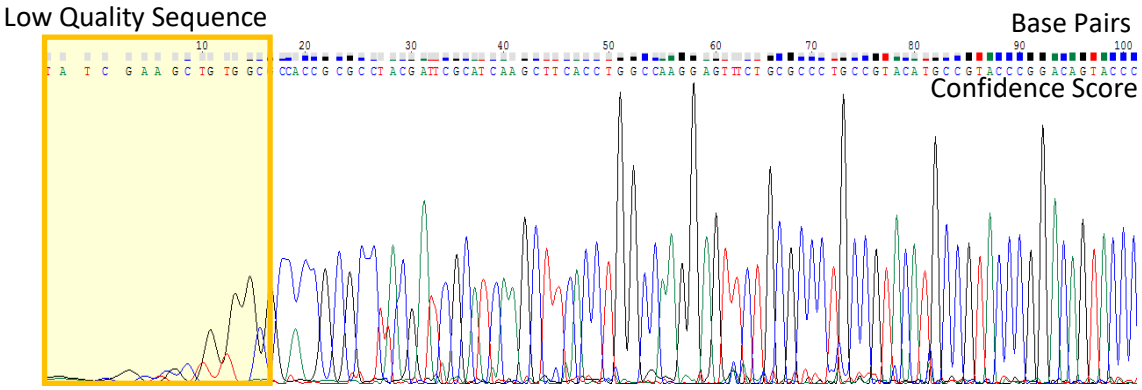


Figure 7.7, Chromatogram file built in the Chromas v.2.5.4 environment. Base pairs are decoded by fluorescence as per the 4 different colour peaks.

Table 7.3, BLASTn results from Sanger Sequencing

CG of interest	Primer (FWD/REV)	BLASTn hit	Bits/identities/E-Value
V-ATPase Subunit D	CGCTGCTGAAGAAGAAGTCC (FWD)	No Similar Sequence found	-
V-ATPase Subunit D	GATGATGTCGTCGTCCTTCTC (REV)	*Setaria italica* V-type proton ATPase subunit D (LOC101759408), mRNA	87.8 bits(96) 99/121(82%) 2e-13
CP450 99A2	TATGGTGATGCCACAAGGTG (FWD)	*Setaria italica* cytochrome P450 99A2 (LOC101776139), mRNA	657 bits(728) 450/507(89%) 0.0
CP450 99A2	CAGCGACGCTTATCTTCCTC (REV)	*Setaria italica* cytochrome P450 99A2 (LOC101776139), mRNA	1054 bits(1168) 704/784(90%) 0.0
LLR ERECTA	TACATCAGGCAGAGCGAGAC (FWD)	*Setaria italica* LRR receptor-like serine/threonine-protein kinase ERECTA (LOC101756593), mRNA	1519 bits(1684) 911/953(96%) 0.0
LLR ERECTA	GACCCGGTCGACTTCCTG (REV)	*Setaria italica* LRR receptor-like serine/threonine-protein kinase ERECTA (LOC101756593), mRNA	1443 bits(1600) 891/941(95%) 0.0
CP450 76C2	AAGGCCTAGGATGGCTTGTC (FWD)	*Setaria italica* cytochrome P450 76C2 (LOC101763431), mRNA	1247 bits(1382) 791/851(93%) 0.0
CP450 76C2	CGACCCCACACTTCTCTTTC (REV)	*Setaria italica* cytochrome P450 76C2 (LOC101763431), mRNA	749 bits(830) 475/513(93%) 0.0
YUCCA 11	TCCCCTACCTTGTCGTTGAG (FWD)	*Setaria italica* probable indole-3-pyruvate monooxygenase YUCCA11 (LOC101766463), mRNA	699 bits(774) 450/489(92%) 0.0
YUCCA 11	TGGCAATGTTGTTAGCATCC (REV)	*Setaria italica* probable indole-3-pyruvate monooxygenase YUCCA11 (LOC101766463), mRNA	356 bits(394) 221/237(93%) 8e-94

7.3.7 PCR and AMPure clean-up of YUCCA11

A decision was made amongst all members of the research group that since the weak MTA's, as per the results from the GWAS in Chapter 6 meant that there was a strong likelihood that this work may not yield any meaningful results, instead of following up the six candidate genes previously described, with varying levels of statistical significance, the association target with the lowest p-value would be taken forward for the haplotype analysis. Namely, YUCCA11 (found in association with the SNP Zn 1 16534195, P = 6.84E-06).

All 48 PMiGAP DNA extractions were subjected to PCR, under the following conditions:

7.5µL Immomix
2-5µL DNA
1µL Primer
1.5-4.5µL Water

The reaction volume was 15µL.

95ºC 10 minutes

95ºC 30 seconds
59ºC 30 seconds
72ºC 30 seconds
72ºC 1 minute
× 35 cycles

Extension time: 72ºC 10 minutes
4ºC hold

The PCR product was AMPure cleaned (0.6 AMPure: 1 DNA) with 100µL 80% EtOH and resuspended in 12µL Illumina resuspension buffer.

A 1% agarose gel confirmed the presence of PCR products in 45 out of 48 cases (Figure 7.8).

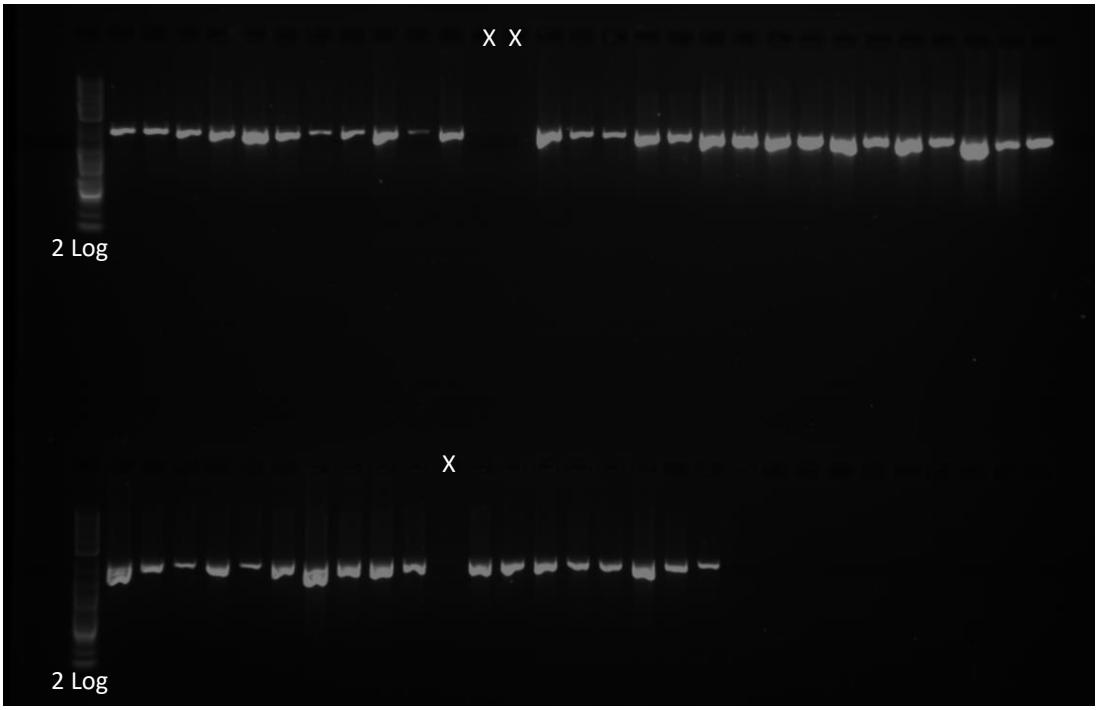


Figure 7.8, 1% agarose gel confirming the presence of PCR products after AMPure clean-up in 45 out of 48 of cases. × = unsuccessful PCR, 2-Log = 2 Log DNA ladder.

7.3.8 Transposome-based Nextera XT Amplicon Libraries

PCR products were Qubit DNA quantified and diluted to 1ng/µL in a 10µL volume with RNAase free water, for use in a transposome-based Nextera XT kit. Indexed paired-end libraries were created for each isolate using the Nextera XT DNA sample preparation kit (Version C protocol, Illumina,), for sequencing on Illumina platforms.

10µL Tagment DNA Buffer plus 5µL AMPure cleaned PCR product were added to a hard-shell skirted PCR plate. 5µL Amplicon Tagment Mix was added to each well. The plate was mixed by pipetting and centrifuged at 280 × g at 20°C for 1minute. The following tagmentation program was set up on a PCR machine:

Preheat lid
55°C 5 minutes
Hold 10°C

The program was run and when the temperature reached 10°C, 5µL Neutralize Tagment Buffer was added to each well and the plate was centrifuged at 280 × g at 20°C for 1 minute. The plate was incubated at room temperature for 5 minutes.

Index primers were arranged in the TruSeq Index Plate Fixture, as per the Illumina Nextera protocol. 5µL of each index adapter was added to each respective column/row and 15µL Nextera PCR Master Mix was added to each isolate containing index adapters. The plate was centrifuged at 280 × g at 20°C for 1 minute and the following programme was run on a PCR machine:

Preheat lid
72ºC 3 minutes
95ºC 30 seconds
95ºC 10 seconds
55ºC 30 seconds
72ºC 30 seconds
72ºC 5 minutes
× 12 cycles
10ºC hold

The PCR product were run on a 1% agarose gel. If a smear was present, the PCR was deemed successful (Figure 7.9).

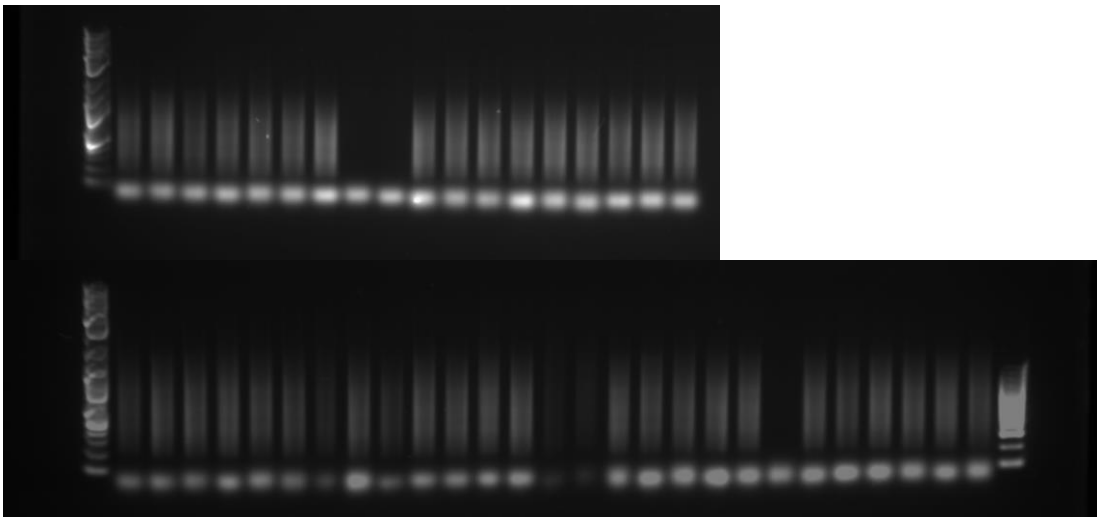


Figure 7.9, PCR products from Nextera library amplification, as confirmed by the appearance of "smearing" on a 1% agarose gel.

The successful PCR products were AMPure cleaned (0.5 AMPure: 1 DNA) with 100µL 80% EtOH and resuspended in 47µL Illumina resuspension buffer. DNA was Epoch quantified and equal quantities of each sample were pooled into a single microfuge tube. A 5µL aliquot of the pooled sample was run on a 1% agarose gel against 2-Log DNA ladder and Hyperladder to observe the average size, which was 0.7kb (Figure 7.10).

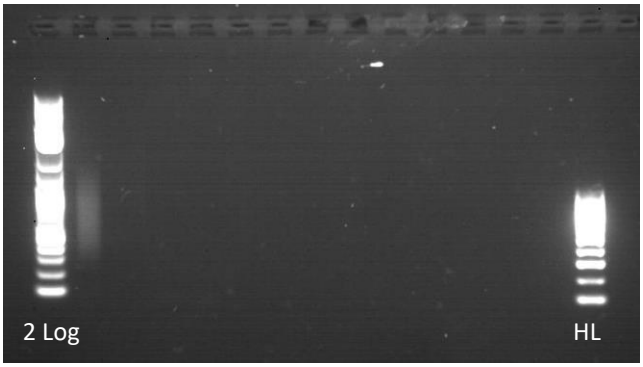


Figure 7.10, 1% agarose gel confirming the presence of PCR product after AMPure clean up (well 2). 2-Log = 2 Log DNA ladder, HL = HyperLadder.

The pooled sample was Qubit DNA quantified (reading at 1.73ng/µL). From these readings the molarity was calculated for future dilution calculations for sequencing using the following formula:

$$\left(\frac{x\ ng/\mu L}{660\ \times Average\ size\ of\ DNA}\right) \times 1{,}000{,}000 = y\ nM$$

$$\left(\frac{1.73\ ng/\mu L}{660\ \times 700}\right) \times 1{,}000{,}000 = 3.7nM$$

2% of the total library was spiked into an Illumina HiSeq run (not relevant to this project) to assess library balance. Samples returned approzimately equal numbers of reads within the 2% spike (2.3 million reads total, 1.15 paired reads), which was intended.

7.3.9 Next Generation Sequencing and Analysis

The Nextera XT libraries were sequenced on a MiSeq (Illumina platform) at 8pmol concentration. The MiSeq was chosen instead of the previously used HiSeq, for the longer available read-lengths (2×300bp). This resulted in 12.3 million pairs of high quality reads.

Bioinformatics analysis was performed using the CLC Genomics Workbench v.6.5 (CLC Bio, Aarhus, Denmark) software. The FASTQ files were imported as Illumina files (distance: 1-1000) in paired format. The reads were subsequently mapped to the YUCCA11 sequence using the following CLC Genomics Workbench settings: Masking mode = no masking, Mismatch cost = 2, Insertion cost = 3, Deletion cost = 3, Length fraction = 0.8, Similarity fraction = 0.85, Global alignment = No,

Auto-detect paired distances = Yes, Non-specific match handling = Ignore. In terms of read alignment, 80% of the read lengths were matched to the reference sequence with 85% similarity. Once mapped, each sample was subjected to resequencing analysis to identify polymorphisms using probabilistic variant detections. Parameters were set to: Ignore non-specific matches and Ignore broken pairs, minimum coverage=50, variant probability=95%, variant count=20.

The data generated for each genotype was imported into Microsoft Excel, according to the variant files created in the CLC Genomics Workbench environment and recoded according to the IUPAC nucleotide codes. These were then converted into PLINK format for loading into TASSEL. In TASSEL v5.2.38, a GLM genotype/phenotype association analysis was conducted, applying all default settings, which generated a table of p-values. Additionally, the haplotype file was used to generate a cladogram tree image.

7.4 Results and Discussion

Most haplotype based analyses have focused on the mapping of human disease alleles, using estimated haplotypes and either case-control or family-based designs (for a more comprehensive review see Liu *et al.,* 2008). In contrast, this study was designed to be relevant to the candidate gene YUCCA11, revealed as a result of GWAS of selfed PMiGAP lines. Because only one gene was used for the haplotype analysis, only limited conclusions could be drawn about the conditions under which haplotype markers might be preferable to single-SNP markers.

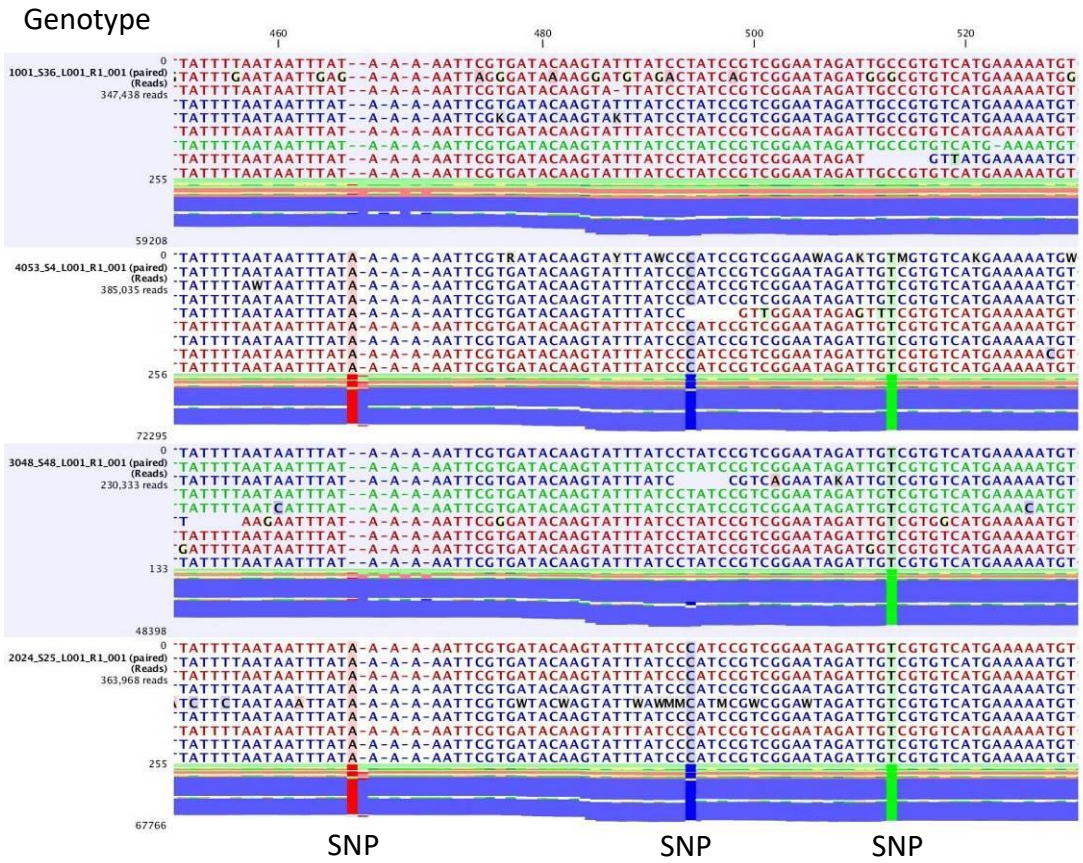
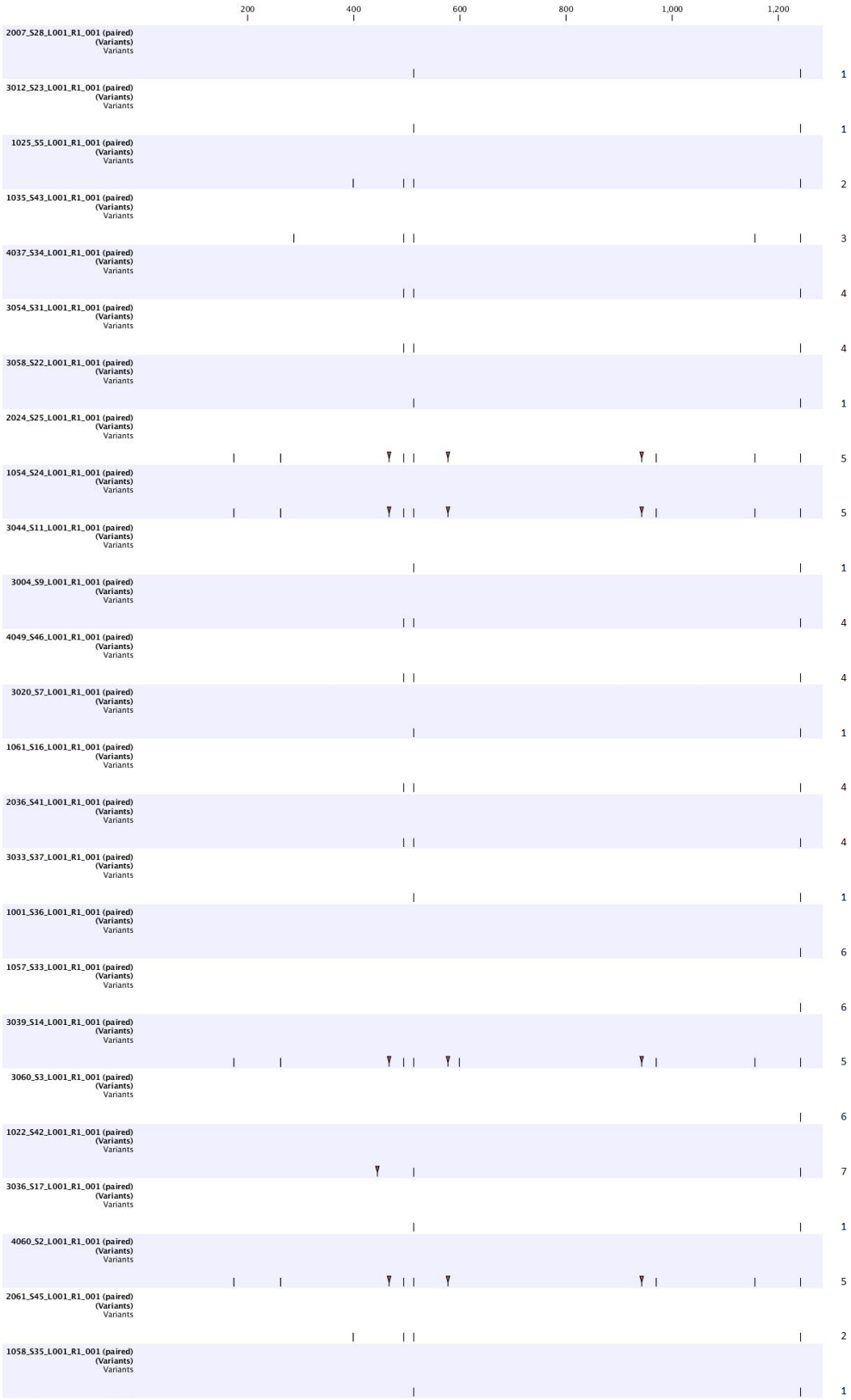


Figure 7.11, Haplotype example built in the CLC Genomics Workbench environment.


200
400
600
800
1,000
1,200
2007_S28_L001_R1_001 (paired) (Variants) Variants 1
3012_S23_L001_R1_001 (paired) (Variants) Variants 1
1025_S5_L001_R1_001 (paired) (Variants) Variants 2
1035_S43_L001_R1_001 (paired) (Variants) Variants 3
4037_S34_L001_R1_001 (paired) (Variants) Variants 4
3054_S31_L001_R1_001 (paired) (Variants) Variants 4
3058_S22_L001_R1_001 (paired) (Variants) Variants 1
2024_S25_L001_R1_001 (paired) (Variants) Variants 5
1054_S24_L001_R1_001 (paired) (Variants) Variants 5
3044_S11_L001_R1_001 (paired) (Variants) Variants 1
3004_S9_L001_R1_001 (paired) (Variants) Variants 4
4049_S46_L001_R1_001 (paired) (Variants) Variants 4
3020_S7_L001_R1_001 (paired) (Variants) Variants 1
1061_S16_L001_R1_001 (paired) (Variants) Variants 4
2036_S41_L001_R1_001 (paired) (Variants) Variants 4
3033_S37_L001_R1_001 (paired) (Variants) Variants 1
1001_S36_L001_R1_001 (paired) (Variants) Variants 6
1057_S33_L001_R1_001 (paired) (Variants) Variants 6
3039_S14_L001_R1_001 (paired) (Variants) Variants 5
3060_S3_L001_R1_001 (paired) (Variants) Variants 6
1022_S42_L001_R1_001 (paired) (Variants) Variants 7
3036_S17_L001_R1_001 (paired) (Variants) Variants 1
4060_S2_L001_R1_001 (paired) (Variants) Variants 5
2061_S45_L001_R1_001 (paired) (Variants) Variants 2
1058_S35_L001_R1_001 (paired) (Variants) Variants 1

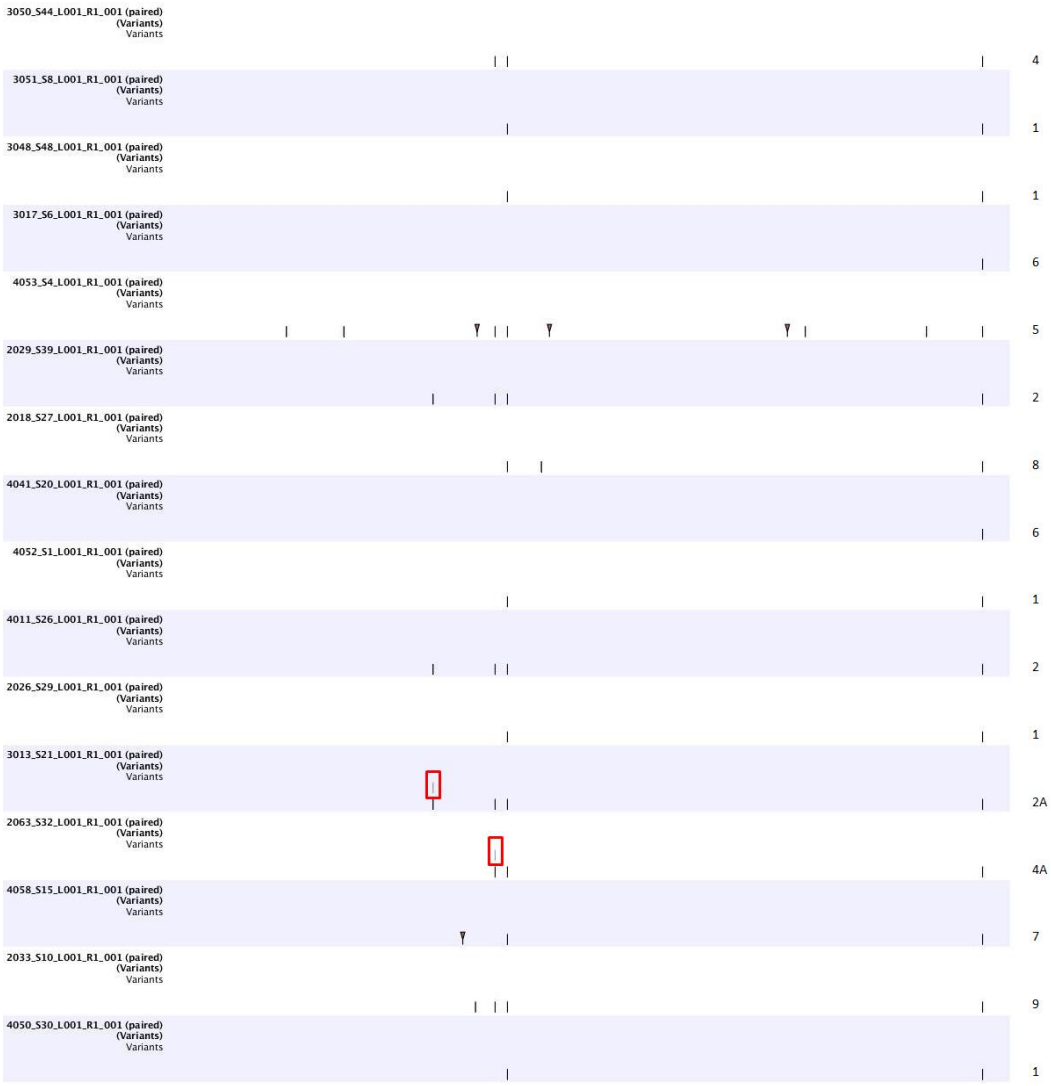


Figure 7.12, Haplotype score per genotype. Numbers 1-9 on the right hand side of the image indicate which haplotype each genotype belongs to. Haplotypes labelled 2/4A represent the presence of heterozygous calls (red box). Upside-down triangle = Insertion. Image built in the CLC Genomics Workbench environment.

A haplotype example can be seen in Figure 7.11. Haplotypes were manually scored (Figure 7.12) and a total of 9 different haplotypes were found to be present across the 42 PMiGAP lines. The calls were mostly homozygous, which was to be expected from inbred lines. This was the case with the exception of two lines – 3013 (IP11346) and 2063 (IP11311), which contained heterozygous calls. This can be explained by residual heterozygosity. Since the PMiGAP lines used in this study were an F6 population, the heterozygosity has been reduced by 50% each time the plant was selfed and therefore may still remain in small numbers.

Table 7.4, Percentage of haplotypes corresponding to 42 high, low and medium Fe/Zn PMiGAP lines

High Fe/Zn		Low Fe/Zn		Medium Fe/Zn	
Haplotype	**%**	**Haplotype**	**%**	**Haplotype**	**%**
1	28.57	1	38.46	1	35.71
2	14.29	2	15.38	2	7.14
3	7.14	3	0.00	3	0.00
4	28.57	4	15.38	4	14.29
5	0.00	5	15.38	5	21.43
6	14.29	6	7.69	6	14.29
7	7.14	7	0.00	7	7.14
8	0.00	8	7.69	8	0.00
9	0.00	9	0.00	9	7.14

According to the data in Table 7.4, none of the haplotypes that cover the YUCCA11 gene were particularly prominent in any group, suggesting that there was no haplotype potent enough to influence levels of Fe/Zn. The limited results seen here may also be attributable to the relatively small sample size (n=42) and the small effect size of the haplotypes. Some studies, albeit those using human population data, recommend sample sizes in excess of 725 subjects (Osabe *et al.*, 2007). Although the literature reports little guidance on the determination of the sample size needed to achieve the desired power for haplotype association studies in plants, it may be beneficial to measure HTA's in pilot studies in order to gain knowledge of the distribution of haplotypes in the target population. Therefore, it is recommended that this analysis should be repeated with at least 100 high/low Fe/Zn PMiGAP lines, and subsequent results should be analysed to determine whether more samples are needed, after that.

Table 7.5, Haplotype trait association within the gene YUCCA11: TASSEL v5.2.38 output.

Trait	Marker	Df	F	P-Value	Marker r^2	Trait	Marker	Df	F	P-Value	Marker r^2
Zn	565	1	2.04822	0.160151	0.048711	Zn	399	2	0.948191	0.396191	0.04637
Fe	174	1	1.776248	0.190153	0.042518	Fe	287	1	0.655247	0.423032	0.016117
Fe	262	1	1.776248	0.190153	0.042518	Fe	399	2	0.675917	0.514548	0.033501
Fe	466_467	1	1.776248	0.190153	0.042518	Zn	1156	1	0.305241	0.58369	0.007573
Fe	577_578	1	1.776248	0.190153	0.042518	Fe	494	2	0.520595	0.598238	0.026003
Fe	942_943	1	1.776248	0.190153	0.042518	Zn	494	2	0.470342	0.628286	0.023552
Fe	970	1	1.776248	0.190153	0.042518	Fe	444_445	1	0.168529	0.683613	0.004196
Zn	287	1	1.454579	0.23488	0.035089	Fe	464	1	0.08295	0.774826	0.002069
Zn	174	1	1.373734	0.248105	0.033203	Zn	464	1	2.72E-04	0.986921	6.80E-06
Zn	942_943	1	1.373734	0.248105	0.033203						
Zn	262	1	1.373734	0.248105	0.033203						
Zn	466_467	1	1.373734	0.248105	0.033203						
Zn	577_578	1	1.373734	0.248105	0.033203						
Zn	970	1	1.373734	0.248105	0.033203						
Fe	565	1	1.218557	0.276241	0.029563						
Zn	444_445	1	1.195941	0.280677	0.029031						
Fe	513	1	0.923354	0.342369	0.022563						
Zn	513	1	0.90635	0.346802	0.022157						
Fe	1156	1	0.74836	0.392155	0.018365						

The HTA data in Table 7.5 for individual variants revealed that there were no significant p-values ($P \leq 0.001$), indicative of the lack of correspondence with the trait value. Although disappointing, this result was predicted due to the weak MTA's from the GWAS (that did not exceed the Bonferroni corrected threshold), as indicated in Chapter 6. Thus, further demonstrating the importance of resolving known issues with phenotype data in the early stages of research, including lack of replicates and small sample size so that they do not affect the downstream analysis, as was the case in this study.

For YUCCA11, there was also no association in terms of haplotypes distributed per geographical location (Figure 7.13). This coincides with the lack of population structure in the PMiGAP, as per the results in Chapter 6.

Figure 7.13, Cladogram tree image built in the TASSEL v5.2.38 environment plus origin data. High/Med/Low = Group to which each genotype belongs in terms of combined Fe/Zn level.

7.5 Conclusions

Even though the literature clearly demonstrates that GWAS greatly benefit from haplotype-based analyses in terms of statistical power (N'Diaye *et al.*, 2017), the results from this study reveal no apparent advantage of haplotype-based analysis over individual SNP analysis. However, the results here are limited, attributable to a small population size and do not apply to a highly significant candidate gene, as a result of weak MTAs from GWAS. Although no HTAs were found, the research contributes to the development of methods that, with minor improvements are fully compatible with a larger data set, with more candidate genes. Reasons for the lack of HTAs point to known issues with the phenotype data, which in turn affected the GWAS confidence levels. Therefore, it is recommended that robust phenotype data with at least three replicates is generated for the improvement of downstream genotypic analysis.

This study represents the first attempt to verify a candidate gene associated with Zn uptake in pearl millet using haplotypes. Although we could not determine any favourable haplotypes using association analysis, it is possible to gain insights as to why this was the case from other haplotype verification studies in the literature. For example, in a study by Zhang *et al.*, (2017), candidate genes were verified by testing the significant differences among major haplotypes for important QTLs associated with Fe or Zn toxicity tolerance. Haplotype analysis was carried out for 22 candidate genes in 10 important QTL regions and results indicated that several haplotypes were found to be associated with phenotypic differences, which verified the candidate genes in many cases. The positive results are likely due to efficient experimental design, including a favourable sample size (10 plants per 211 accessions, with 2 replicates for both control and stress conditions) and a large abundance of markers used for GWAS (395,553 SNP markers). At the genotype level, other parameters were also in place. For example, haplotype analysis was only carried out if more than two significant SNPs were distributed in one gene.

Chapter 8: Discussion, Conclusions and Future Work

8.1 Aims and Background

Increased grain Fe and Zn content in pearl millet is an important breeding target for food and nutrition security of people living in poverty stricken areas within India and Africa.

The study detailed in this thesis was undertaken to decipher genomic regions associated with elevated grain Fe and Zn levels in pearl millet using genetic association studies. This involved assessment of grain from a germplasm diversity panel of pearl millet (the PMiGAP) for mineral content by ICPAES (Chapter 3), as well as the presence of anti-nutrients, including phytate and metal-chelating soluble phenolics (Chapters 4 and 5, respectively), that might hinder the bioavailability of Fe and Zn. GBS libraries were prepared for high-density genotyping and the resulting SNPs were used, in addition to SNPs from the recently published pearl millet genome, to conduct GWAS of the phenotype data for the identification of MTAs and in turn candidate genes (Chapter 6). In addition, haplotypes covering the most significant candidate gene were identified and their association with the trait data was assessed (Chapter 7).

The research detailed within this thesis lay down a foundation for future studies and discoveries, not only within the fields of molecular genetics, plant biotechnology and nutrition for human health but also with respect to analytical chemistry, for the consideration of anti-nutrients. Without this ground work, future endeavours may not be possible.

8.2 Overview and Outcomes

8.2.1 Plant Material

Chapter 2 describes the plant material used within this thesis and three seed multiplication trials that took place under glasshouse conditions for the production of selfed and OP seed.

Matured seed was harvested for phenotyping purposes (Table 2.3), as well as young leaf tissue for DNA extractions (Chapter 7, Section 7.3.2). Lessons learned from the

first glasshouse trial facilitated the optimisation of conditions for good growth and yield during subsequent trials.

8.2.2 Phenotyping the PMiGAP

In Chapters 3, 4 and 5, extensive phenotypic analysis by ICPAES, phytate assays and HPLC-MS was conducted to quantify levels of minerals, phytate and soluble phenolics in PMiGAP grain samples, respectively. The ICPAES data generated in Chapter 3 is considered to be reliable and precise as the analysis was conducted by an accredited service within IBERS, Aberystwyth University. The Megazyme® Phytic Acid (Phytate)/Total Phosphorus assay kit was used to quantify levels of phytate in Chapter 4 and has been extensively tested by British Geological Survey, Centre for Environmental Geochemistry. According to the performance characteristics attained, the kit was found to give accurate, reproducible and reliable data. For example, reproducibility for wheat grain samples was demonstrated with a precision of 10% between separate runs, meeting validation criteria (Reason *et al.,* 2015). Furthermore, the HPLC-MS data generated in Chapter 5 is extremely robust due to the extensive improvements to the compound extraction protocol, implemented during method development process, as discussed.

Upon analysis of the ICPAES mineral data, grain Fe and Zn content was highly variable and consistently positively correlated when grown in different environments. However, upon analysis of the mean mineral concentrations across multi-environments, there were major differences, which indicated that the environment had a stronger effect on mineral uptake than the genetics. Therefore, the data should be approached with caution due to an unstable phenotype. Furthermore, the mineral content of OP and selfed seed was also compared and it was found that mineral levels were significantly higher in selfed populations, which may be indicative of seed concentration effects. Levels of phytate also differed between two populations grown in different environments. This may be attributable to the effects of cultivars, environment and their interactions with phytate being highly significant. Interestingly, there was no significant correlation between grain Fe/Zn content and phytate, thus indicative of the possibility of breeding pearl millet lines with low phytate content and high grain Fe/Zn levels. Research by Al Hasan *et al.,* (2016) suggests two ‘critical values’ based on molar ratios of phytate: Fe/Zn of 1 and 15

respectively, where when exceeded may result in poor Fe/Zn bioavailability (Al Hasan *et al.*, 2016). Analysis of 235 PMiGAP lines revealed that 100% of the population exceeded the critical value for Fe and 69% exceeded the critical value for Zn. This suggests that the vast majority of the PMiGAP population is limited in bioavailability of Fe and Zn.

In terms of metal-chelating soluble phenolics, the research in Chapter 5 primarily focused on two flavonone glycosides; apigenin and luteolin (and their derivatives). These compounds were extracted using MeOH from raw pearl millet grain and peak areas from HPLC chromatograms (as a measure of relative abundance) were identified, quantified and compared. Findings revealed that the content of all compounds were highly variable between lines. The correlation between polyphenol content and micronutrients was also investigated and although no significant correlations were observed, upon application of a statistical filter, some clustering was revealed between apigenin and luteolin derivatives.

8.2.3 GWAS for the Identification of Markers/Candidate Genes Associated with Fe and Zn Uptake

In Chapter 6, three GWAS were conducted using different marker sets for the identification of MTAs and in turn candidate genes that may be associated with elevated Fe/Zn levels. Prior to GWAS, population structure and LD were accounted for to safeguard against confounding factors that may lead to a type 1 error, i.e. false positive MTAs. Population structure was insignificant in all cases, which was the pattern expected from inbred lines derived from a highly allogamous species (Varshney *et al.*, 2017). LD and LD decay differed between the marker sets. For example, where low numbers of markers were used, only a small amount of marker pairs were in LD based on r^2 values and LD decay was extremely rapid. Whereas when marker density was high, LD was more prominent and decayed at a slower rate, indicating an adequate number of markers for GWAS. The three GWAS facilitated the selection of 35 MTAs and when the regions surrounding each significant SNP were BLASTed against the most annotated *Setaria italica* reference genome (because a complete pearl millet genome was not available on the NCBI genome sequence viewer at the time of this research), 6 candidate genes were found to be directly associated with mineral uptake including: V-type proton ATPase subunit D, 3 × cytochrome P450 family members, ERECTA and YUCCA11. The

YUCCA11 gene had the lowest P-value (P = 6.84E-06). The reported MTAs/candidate genes may be useful in MAS programs, for genomic selection and for population improvement programs for the creation of elite lines.

8.2.4 Haplotype analysis of the YUCCA11 gene

In Chapter 7, a total of 9 haplotypes in 42 high, medium and low Fe/Zn PMiGAP lines were detected, and findings revealed no significant association between any haplotype and Fe/Zn uptake. This may be attributable to known issues with the phenotype data affecting the downstream analysis or the relatively small sample size. Furthermore, no particular haplotype was found to be associated with any geographical region, which reflects the lack of population structure in the PMiGAP.

8.3 General Discussion

Pearl millet is a climate change ready crop with excellent nutritional value, as compared to many other staples (Varshney *et al.,* 2017). It serves as a traditional staple grain for millions of people living in dry, semi-arid regions within Africa and India (areas where micronutrient malnutrition is prominent). To utilise the nutritional value of this crop, biofortification based research has been initiated to mitigate the challenges associated with micronutrient malnutrition (Kumar *et al.,* 2018).

Pearl millet biofortification has been most successful in India and it is thought that similar achievements could be realised for Western and Central Africa (Pfeiffer *et al.,* 2018). In particular, most of the research for the development of Fe/Zn-rich varieties has taken place at ICRISAT headquarters, Patencheru where similar to this study, a positive correlation between grain Fe and Zn content was observed, indicating good prospects for simultaneous selection for both micronutrients (Velu *et al.,* 2008). Furthermore, Fe and Zn are largely under additive genetic control, thus Fe/Zn hybrids require that both parental lines have high Fe/Zn density (Govindaraj *et al.,* 2016). The most recently biofortified OP pearl millet variety, Dhanashakti (71 mg/kg Fe density and 40 mg/kg Zn density) was released by ICRISAT in 2014 and was rapidly adopted by over 65,000 farmers in 2015. Currently, several pearl millet hybrids, with up to 75 mg/kg Fe density and 25-35% higher grain yield have been developed, and are at various stages of testing in national agricultural trials in India (ICRISAT, 2016). A breeding target of >77 mg/kg is now the goal of many

institutes. This target will be achieved when markers/genes associated with elevated grain Fe/Zn levels are identified and incorporated via breeding-based biofortification programmes. Therefore, the research detailed within this thesis, coupled with the recently published pearl millet genome will certainly help with such endeavours not only in terms of breeding for elevated grain Fe and Zn content, but also in accounting for mineral bioavailability in terms of metal-chelating antinutrient compounds.

8.3.1 Challenges and Future Research

No Seed, No Science

PMiGAP seed from Indian field trails was obtained directly from ICRISAT headquarters in Patencheru, although difficulty in obtaining the seed was experienced as there was several years delay due to shipping restrictions and quarantine issues. The seed multiplication trials described in Chapter 2 were vital for the investigations into phenotype stability, as described in Chapter 3. The biggest challenge for bulking seed was maintaining optimal conditions within the glasshouse environment. This was overcome by carefully observing plants on a daily basis and quickly responding to possible stress factors e.g. insect infestations and overcrowding during the early growth stages. As discussed, when a new trial was conducted, observations from the previous trial were taken into account for an improved success rate.

Despite the best intentions, limited seed availability meant that subsets of the PMiGAP had to be used in many cases for phenotyping purposes and findings were generalised to the whole population. Furthermore, there was not enough seed to replicate the majority of the analysis. Although care was taken to ensure that bias was minimal, some subsets were as small as 20 lines (Table 2.3). A sample size that is too small may reduce the statistical power of the study and increase the margin of error, which could render the study meaningless. Thus, many of the findings in this thesis should be approached with caution and analysis should be repeated with as many samples as possible.

Good, robust trait data is required for GWAS and the contribution of a gene to a trait may vary depending on the environmental conditions (Korte and Farlow, 2013). If

this study were to be improved the most important objective would be to resolve the known issues with the phenotype data and to obtain more stable phenotypes, where GEI effects would be minimal. Throughout this study, the lack of replicated phenotype data has been a recurrent problem, the causes of which have been discussed in previous chapters. This has undoubtedly affected the downstream analysis and GWAS.

The use of selfed PMiGAP lines are ideal for GWAS as they can be maintained as inbred lines via continued self-fertilisation, thus it is possible to repeatedly phenotype genetically identical individuals and assign MTAs to individual lines. Phenotyping the same lines many times, under different environmental conditions increases precision of the trait mean, but also allows the estimation of phenotypic variance. In the scope of this thesis, there was only a limited amount of time to conduct seed multiplication trials. In addition, the seed produced from the trials was shared among other projects. Ideally, to evaluate phenotypic variation in terms of mineral concentration in pearl millet, a minimum of 3 environments should be used, which would take a number of years to achieve. The environments should reflect all the possible conditions in which pearl millet is grown within field trials conducted in India and Africa. Repetitively phenotyping the same genotype in this many environments would allow a robust investigation into phenotypic stability, where stable phenotypes could hopefully be identified. This would improve the statistical power of GWAS. For example, in a study by Sun *et al.*, (2017) excellent phenotypic data was achieved as a result of cotton field trials conducted in 8 environments over 2 years. From the data, stable phenotypes could be selected with increased confidence; this in turn resulted in highly significant MTAs from GWAS.

Phenotypic Analysis

The extraction of trace elements and phytate from pearl millet grain, as described in Chapters 3 and 4 was targeted in that known compounds were extracted and quantified. However in Chapter 5, the extraction process was untargeted and as a result, a wide range of soluble phenolics were extracted and investigated (16 different compounds in total) thanks to an extensive and robust MeOH extraction protocol. It was considered that other solvents such as ethyl acetate, acetone or chloroform may facilitate the extraction of different classes of compounds or

different yields. For example, oil extracted with a non-polar solvent is reported to be different in fatty acid composition as compared with oil from extraction procedures using polar solvents (Jellum and Powell, 1971). Thus, future research should be conducted to compare the efficiency of different extraction solvents in extracting phytochemicals from matured pearl millet grain. Other studies suggest that MeOH was indeed the best choice of extraction solvent and can be used for the extraction of many phytochemicals from leaves/grain. This has been proven by Dhawan and Gupta (2017), where the effect of different extraction solvents including distilled water, MeOH, acetone, chloroform, ethyl acetate and hexane were compared for the extraction of bioactive components such as alkaloids, flavonoids, saponins, steroids and tannins from dried leaves of Devil's Trumpet (*Datura metel*). Phytochemical screening tests were performed to detect the presence of bioactive components in plant extracts and findings revealed that MeOH was the most efficient solvent in extracting phytochemicals as per percentage yield. In terms of flavonoid extraction, MeOH was again the most efficient solvent. However, maximum phenol concentration was observed when ethyl acetate was used as the solvent for extraction. Interestingly, when chloroform and distilled water were used as extraction solvents, phenol extraction was poor (Dhawan and Gupta, 2017).

Increasing the Power of GWAS

The aim of GWAS is to evaluate the association between each genotyped marker and a trait of interest that has been scored across a large number of individuals (Korte and Farlow, 2013). The PMiGAP was an ideal population to conduct GWAS on as it included many geographically distant accessions, thus maximising the genetic variance. However, many samples had to be excluded from the 663 SNP dataset (as described in Chapter 6) due to missing/low quality data. Excluding poorly genotyped variants often introduces an unequal sample size across sites, making the downstream statistical analysis more complex.

GWAS are largely suited to the identification of traits that are underpinned by simple genetic architecture, i.e. a small number of loci with large effect sizes. However, many traits may possess more complex architectures that may complicate GWAS, which could be the reason for the lack of well-known Fe/Zn uptake genes found within this study. In addition to the reasons described in Chapters 6 and 7, two

additional possibilities are that a trait of interest may be controlled by; i) many rare variants, with large effect size on the phenotype, or ii) many common variants with small phenotypic effect. In both scenarios the causative variants may be clustered in one, a small number of genes, or across many genes (Korte and Farlow, 2013). Future research studies should investigate the possibility of increasing the power of GWAS to locate true associations when variants are either at low frequency or have a small effect size. Several necessary considerations include; adequate sample size, incomplete genotyping, genetic heterogeneity and accounting for confounding genetic factors (Korte and Farlow, 2013).

According to the literature, GWAS is routinely conducted in many major, minor and orphan crop species for the identification of a wide variety of traits. It is hoped that increasing the population size will improve the power of GWAS to reveal meaningful associations. Although, some studies have found that sample sizes as little as 100 lines can produce significant MTAs (Atwell *et al.,* 2010), this suggests that the traits considered were underpinned by only a few loci that explain a large portion of the phenotypic variance. In contrast, human genetics based GWAS differ considerably with respect to sample size where typically a large number of small effect loci are revealed and most analyses require >2000 individuals for detection (Manolio *et al.,* 2009).

It is necessary to consider multiple statistical testing when conducting GWAS since a decision in whether a MTA is true is based on the result of more than one test. In a typical GWAS, thousands of tests are simultaneously conducted for each marker and each generates a false positive probability. This creates a cumulative likelihood of detecting multiple false associations over the whole analysis. The most routinely used correction for multiple testing is a 5% Bonferroni threshold, which was used in this study. However, this is considered aggressive by many researchers and as a result only 3 out of 35 markers remained significant after the threshold was applied. A more relaxed approach such as False Discovery Rate (FDR) may be used in some cases. When considering the aims of a study, a high FDR may be considered for some investigations (e.g. investigating the genetic architecture of a trait) and a low FDR for others (e.g. identifying candidate loci for follow-up studies) (Korte and Farlow, 2013).

Improved Haplotyping

For the reasons previously discussed in Chapter 7, single SNP analysis from GWAS is not always reliable and often suffers from low replication and validation rates. This has been recognised and is the reason why the haplotype-based analysis was conducted as a follow-up to GWAS. However, better experimental design is needed for future work. For example, phenotype data with replications and DNA extractions from a larger PMiGAP population would be required for the haplotype trait analysis and more genes should be considered. Thus, the original objective of following up combinations of genes rather than individual genes to find associations with Fe/Zn uptake should be resumed. This would strengthen the conclusions that these genes are associated with a certain trait. The research in Chapter 7 should therefore be viewed as method development that should motivate and pave the way for future research.

Beyond the Genes

Following the suggested improvements to the work conducted within this thesis, it is hoped that any resulting genes associated with increased grain Fe/Zn content would be incorporated into elite lines via biofortification-based breeding programmes to benefit women and children living in resource poor areas. As previously discussed, accessibility of such lines to farmers living in remote areas is paramount. Thus, whilst it is recognised that GM is an important approach in improving/introducing beneficial traits of interest, the cost of this would make the seed largely inaccessible to smallholder farmers (Horna *et al.*, 2008). Therefore traditional plant breeding, which relies on natural genetic variation, is highly attractive in this respect as prices are kept more affordable. However, since micronutrient malnutrition is an ongoing challenge that is largely exacerbated by an increasing population, climate change and increasing food prices, getting biofortified lines out to communities in need is seen as a matter of urgency. In light of this, more timely solutions should be considered since conventional breeding relies on time-consuming extensive back-crossing programs to introgress the selected traits into elite lines and vast amounts of land. Additionally, the availability of beneficial alleles in nature also limits what can be achieved using this approach (Hou *et al.*, 2014). The CRISPR/Cas9 system offers a more timely solution that is safe, simple, efficient, and highly specific. Furthermore,

multiple traits can be modified simultaneously and fewer off-target events are produced than in traditional plant breeding, which is largely based on chance. It is thus considered a promising tool for gene editing in plants and should provide a more efficient approach to accelerated plant breeding (Bortesi and Fischer, 2015). Gene editing allows precise, rapid and predictable genome modifications directly in elite lines, saving time-consuming backcrossing procedures required in traditional breeding programmes. The CRISPR system relies on the endonuclease enzyme Cas9, which uses a guide RNA molecule to target specific DNA sequences, and then edits, deletes or replaces the target DNA, leading to genome modifications during the repair process or the insertion of new sequences (Ledford, 2015). Due to the orphan status of pearl millet, little work has been performed so far on the nutritional enhancement of grain via gene editing – thus presenting a significant gap in the literature. However, work on major cereals and reliable techniques have demonstrated that genetic improvements are possible using gene editing approaches (O'Kennedy *et al*., 2006).

Factors that may influence micronutrient uptake in the soil should also be investigated in future research. In particular, plant-associated microorganisms present in the soil may stimulate growth and influence yield and quality of edible parts by affecting nutrient mobilisation and transport (Pii *et al.,* 2016). Because of this, the root rhizosphere microbiome is considered to be one of the key determinants of productivity and plant health (Manwaring *et al.,* 2016). Investigations into Arbuscular mycorrhizal fungi, soil inhabitants which are obligate symbionts that require plant host roots to complete their life cycle (Borde *et al.,* 2011), would be a good place to start. Arbuscular mycorrhizal fungi have been most recently characterised in pearl millet as alleviators of the adverse effects of salt on growth. A study by Borde *et al.,* (2011) revealed that total chlorophyll content in pearl millet was significantly higher in Arbuscular mycorrhizal inoculated plants, which lead to increased growth and nutrient uptake capacity. This could be investigated further with the aim of enhancing Fe/Zn status even more so when elite lines have been developed for efficient micronutrient uptake. Root exudates are also thought to play a significant role in efficient micronutrient uptake, including that of Fe and Zn. For example, a variety of low and high molecular weight organic compounds are triggered if plants are exposed to certain abiotic stressors (Dakora and Phillips, 2002)

and due to their solubilising, chelating, redox capacities, they play a fundamental role in enhancing nutrient bioavailability, uptake, translocation and allocation (Manwaring *et al.*, 2016).

Although increasing Fe and Zn content through biofortification is an important and necessary approach to mitigate the challenges associated with micronutrient malnutrition, it was considered that Fe and Zn levels could also be toxic when present at high concentrations. For example, unregulated high affinity binding of Zn to S-, N- and O- containing functional groups in biological molecules and uncontrolled displacement of cofactor metal cations, e.g. Mn^{2+} and Fe^{2+}, can cause serious damage (Palmgren *et al.*, 2008). Additionally, the physiological range between deficiency and toxicity of Fe/Zn is narrow (Maret and Sandstead, 2006). To guard against damage, plants have the ability to balance the uptake, utilisation and storage of these metals in an effort to maintain metal homeostasis (Le and An, 2009). For example, with respect to Fe, once taken up into the cells, Fe is subjected to tight control to avoid cellular toxicity and ferritin (an Fe storage unit) is believed to play a key role in this process (Briat and Lobreaux, 1997). However, metal homeostasis in plants still remains incompletely understood, as discussed in Chapter 1. In light of this, before the effect of biofortification strategies can be predicted, more research is needed regarding the biological processes that govern the uptake and distribution of Fe and Zn in pearl millet and future research should also work towards pin-pointing the exact concentration where Fe/Zn uptake becomes toxic in pearl millet.

Chapter 9: References

Abdurakhmonov, I.Y. and Abdukarimov, A., 2008. Application of association mapping to understanding the genetic diversity of plant germplasm resources. *International Journal of Plant Genomics*, *2008*.

Al Hasan, S.M., Hassan, M., Saha, S., Islam, M., Billah, M. and Islam, S., 2016. Dietary phytate intake inhibits the bioavailability of iron and calcium in the diets of pregnant women in rural Bangladesh: a cross-sectional study. *BMC Nutrition*, *2*(1), p.24.

Alloway, B.J., 2004. Zinc in soils and crop nutrition.

Al-Sa'aidi, J., The anti-hyperglycemic effect of Solenostemma argel compared with Glibenclamide.

Andersson, M.S., Saltzman, A., Virk, P.S. and Pfeiffer, W.H., 2017. Progress update: crop development of biofortified staple food crops under HarvestPlus. African Journal of Food, Agriculture, Nutrition and Development, 17(2), pp.11905-11935.

Andrews, D.J. and Kumar, K.A., 1996. Use of the West African pearl millet landrace Iniadi in cultivar development. *Plant Genetic Resources Newsletter*, *105*, pp.15-22.

Anon (2018). Millets make their presence felt on mid-day meal plates. Available: https://www.thehindubusinessline.com/economy/agri-business/millets-make-their-presence-felt-on-midday-meal-plates/article10030563.ece. Last accessed 07/03/2018.

Anuradha, N., Satyavathi, C.T., Bharadwaj, C., Nepolean, T., Sankar, S.M., Singh, S.P., Meena, M.C., Singhal, T. and Srivastava, R.K., 2017. Deciphering genomic regions for high grain iron and zinc content using association mapping in pearl millet. *Frontiers in plant science*, *8*, p.412.

Atwell, S., Huang, Y.S., Vilhjálmsson, B.J., Willems, G., Horton, M., Li, Y., Meng, D., Platt, A., Tarone, A.M., Hu, T.T. and Jiang, R., 2010. Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature, 465(7298), p.627.

Badau, M.H., Nkama, I. and Jideani, I.A., 2005. Phytic acid content and hydrochloric acid extractability of minerals in pearl millet as affected by germination time and cultivar. *Food Chemistry*, *92*(3), pp.425-435.

Bailey, R.L., West Jr, K.P. and Black, R.E., 2015. The epidemiology of global micronutrient deficiencies. *Annals of Nutrition and Metabolism*, *66*(Suppl. 2), pp.22-33.

Bais, H.P., Weir, T.L., Perry, L.G., Gilroy, S. and Vivanco, J.M., 2006. The role of root exudates in rhizosphere interactions with plants and other organisms. *Annu. Rev. Plant Biol.*, *57*, pp.233-266.

Bak, S., Tax, F.E., Feldmann, K.A., Galbraith, D.W. and Feyereisen, R., 2001. CYP83B1, a cytochrome P450 at the metabolic branch point in auxin and indole glucosinolate biosynthesis in Arabidopsis. *The Plant Cell*, *13*(1), pp.101-111.

Bantilan, M.C.S. and Joshi, P.K., 1998. *Assessing joint research impacts: proceedings of an International Workshop on Joint Impact Assessment of NARS/ICRISAT Technologies for the Semi-Arid Tropics, 2-4 Dec 1996, ICRISAT, Patancheru, India*. International Crops Research Institute for the Semi-Arid Tropics.

Bänziger, M. and Long, J., 2000. The potential for increasing the iron and zinc density of maize through plant-breeding. *Food and Nutrition Bulletin*, *21*(4), pp.397-400.

Barendse, W., 2011. Haplotype analysis improved evidence for candidate genes for intramuscular fat percentage from a genome wide association study of cattle. *PLoS One*, *6*(12), p.e29601.

Basford, K.E. and Cooper, M., 1998. Genotypexenvironment interactions and some considerations of their implications for wheat breeding in Australia. *Australian Journal of Agricultural Research*, *49*(2),

Bassoli, B.K., Cassolla, P., Borba-Murad, G.R., Constantin, J., Salgueiro-Pagadigorria, C.L., Bazotte, R.B., da Silva, R.S.D.S. and de Souza, H.M., 2008. Chlorogenic acid reduces the plasma glucose peak in the oral glucose tolerance test: effects on hepatic glucose release and glycaemia. Cell biochemistry and function, 26(3), pp.320-328.

Bedhomme, M., Hoffmann, M., McCarthy, E.A., Gambonnet, B., Moran, R.G., Rébeillé, F. and Ravanel, S., 2005. Folate metabolism in plants an arabidopsis homolog of the mammalian mitochondrial folate transporter mediates folate import into chloroplasts. *Journal of Biological Chemistry*, *280*(41), pp.34823-34831.

Begum, M.C., Islam, M., Sarkar, M.R., Azad, M.A.S., Huda, A.N. and Kabir, A.H., 2016. Auxin signaling is closely associated with Zn-efficiency in rice (Oryza sativa L.). *Journal of Plant Interactions*, *11*(1), pp.124-129.

Belozersky, M.A., Sarbakanova, S.T. and Dunaevsky, Y.E., 1989. Aspartic proteinase from wheat seeds: isolation, properties and action on gliadin. *Planta*, *177*(3), pp.321-326.

Bennett, M.D., Bhandol, P. and Leitch, I.J., 2000. Nuclear DNA amounts in angiosperms and their modern uses—807 new estimates. *Annals of botany*, *86*(4), pp.859-909.

Bertin, I., Zhu, J.H. and Gale, M.D., 2005. SSCP-SNP in pearl millet—a new marker system for comparative genetics. *Theoretical and Applied Genetics*, *110*(8), pp.1467-1472.

Berwal, M.K., Chugh, L.K., Goyal, P., Kumar, R. and Vart, D., 2017. Protein, Micronutrient, Antioxidant Potential and Phytate Content of Pearl Millet Hybrids and Composites Adopted for Cultivation by Farmers of Haryana, India. *Int. J. Curr. Microbiol. App. Sci*, *6*(3), pp.376-386.

Bhattacharjee, R., Bramel, P., Hash, C., Kolesnikova-Allen, M. and Khairwal, I., 2002. Assessment of genetic diversity within and between pearl millet landraces. *TAG Theoretical and Applied Genetics*, *105*(5), pp.666-673.

Bhattacharjee, R., Khairwal, I.S., Bramel, P.J. and Reddy, K.N., 2007. Establishment of a pearl millet [Pennisetum glaucum (L.) R. Br.] core collection based on geographical distribution and quantitative traits. *Euphytica*, *155*(1-2), pp.35-45.

Black, R.E., Victora, C.G., Walker, S.P., Bhutta, Z.A., Christian, P., De Onis, M., Ezzati, M., Grantham-McGregor, S., Katz, J., Martorell, R. and Uauy, R., 2013. Maternal and child undernutrition and overweight in low-income and middle-income countries. *The lancet*, *382*(9890), pp.427-451.

Blackmore, T., Thorogood, D., Skøt, L., McMahon, R., Powell, W. and Hegarty, M., 2016. Germplasm dynamics: the role of ecotypic diversity in shaping the patterns of genetic variation in Lolium perenne. *Scientific reports*, *6*.

Borde, M., Dudhane, M. and Jite, P., 2011. Growth photosynthetic activity and antioxidant responses of mycorrhizal and non-mycorrhizal bajra (Pennisetum glaucum) crop under salinity stress condition. *Crop Protection*, *30*(3), pp.265-271.

Bortesi, L. and Fischer, R., 2015. The CRISPR/Cas9 system for plant genome editing and beyond. Biotechnology advances, 33(1), pp.41-5

Bouis, H., Biofortification Progress Briefs August 2014. Washington DC.: Harvest Plus; 2014 August 2014. 82 p.

Bouis, H.E., Hotz, C., McClafferty, B., Meenakshi, J.V. and Pfeiffer, W.H., 2011. Biofortification: a new tool to reduce micronutrient malnutrition. *Food and nutrition bulletin*, *32*(1_suppl1), pp.S31-S40.

Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y. and Buckler, E.S., 2007. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, *23*(19), pp.2633-2635.

Briat, J.F. and Lobréaux, S., 1997. Iron transport and storage in plants. Trends in plant science, 2(5), pp.187-193.

Brnić, M., Hurrell, R.F., Songré-Ouattara, L.T., Diawara, B., Kalmogh;o-Zan, A., Tapsoba, C., Zeder, C. and Wegmüller, R., 2016. Effect of phytase on zinc absorption from a millet-based porridge fed to young Burkinabe children. *European Journal of Clinical Nutrition*.

Broadley, M.R., Hammond, J.P., King, G.J., Astley, D., Bowen, H.C., Meacham, M.C., Mead, A., Pink, D.A., Teakle, G.R., Hayden, R.M. and Spracklen, W.P., 2008. Shoot calcium and magnesium concentrations differ between subtaxa, are highly heritable, and associate with potentially pleiotropic loci in Brassica oleracea. *Plant Physiology*, *146*(4), pp.1707-1720.

Broadley, M.R., White, P.J., Hammond, J.P., Zelko, I. and Lux, A., 2007. Zinc in plants. *New Phytologist*, *173*(4), pp.677-702.

Brune, M., Rossander, L. and Hallberg, L., 1989. Iron absorption and phenolic compounds: importance of different phenolic structures. *European journal of clinical nutrition*, *43*(8), pp.547-557.

Buck, H.T., Nisi, J.E. and Salomón, N. eds., 2007. *Wheat Production in Stressed Environments: Proceedings of the 7th International Wheat Conference, 27 November-2 December 2005, Mar Del Plata, Argentina* (Vol. 12). Springer Science & Business Media.

Bundock, P., Christopher, J., Eggler, P., Ablett, G., Henry, R. and Holton, T., 2003. Single nucleotide polymorphisms in cytochrome P450 genes from barley. *TAG Theoretical and Applied Genetics*, *106*(4), pp.676-682.

Burton, G.W., 1952. The immediate effect of gametic relationship upon seed production in pearl millet, Pennisetum glaucum. *Agronomy Journal*, *44*(8), pp.424-427.

Bush, W.S. and Moore, J.H., 2012. Genome-wide association studies. *PLoS computational biology*, *8*(12), p.e1002822.

Busk, P.K. and Møller, B.L., 2002. Dhurrin synthesis in sorghum is regulated at the transcriptional level and induced by nitrogen fertilization in older plants. *Plant Physiology*, *129*(3), pp.1222-1231.

Butte, N.F., Wong, W.W., Treuth, M.S., Ellis, K.J. and Smith, E.O.B., 2004. Energy requirements during pregnancy based on total energy expenditure and energy deposition. *The American journal of clinical nutrition*, *79*(6), pp.1078-1087.

Cai, H., Hudson, E.A., Mann, P., Verschoyle, R.D., Greaves, P., Manson, M.M., Steward, W.P. and Gescher, A.J., 2004. Growth-inhibitory and cell cycle-arresting properties of the rice bran constituent tricin in human-derived breast cancer cells in vitro and in nude mice in vivo. British journal of cancer, 91(7), pp.1364-1371.

Carriaga, M.T., Skikne, B.S., Finley, B., Cutler, B. and Cook, J.D., 1991. Serum transferrin receptor for the detection of iron deficiency in pregnancy. *The American journal of clinical nutrition*, *54*(6), pp.1077-1081.

Casas, M.I., Duarte, S., Doseff, A.I. and Grotewold, E., 2014. Flavone-rich maize: an opportunity to improve the nutritional value of an important commodity crop. *Frontiers in plant science*, *5*, p.440.

Caulfield, L.E. and Black, R.E., 2004. Zinc deficiency. *Comparative quantification of health risks: global and regional burden of disease attributable to selected major risk factors*, *1*, pp.257-280.

Cercamondi, C.I., Egli, I.M., Mitchikpe, E., Tossou, F., Zeder, C., Hounhouigan, J.D. and Hurrell, R.F., 2013. Total iron absorption by young women from iron-biofortified pearl millet composite meals is double that from regular millet meals but less than that from post-harvest iron-fortified millet meals. *The Journal of nutrition*, *143*(9), pp.1376-1382.

Cha, B.Y., Shi, W.L., Yonezawa, T., Teruya, T., Nagai, K. and Woo, J.T., 2009. An inhibitory effect of chrysoeriol on platelet-derived growth factor (PDGF)-induced proliferation and PDGF receptor signaling in human aortic smooth muscle cells. Journal of pharmacological sciences, 110(1), pp.105-110.

Chandrasekara, A. and Shahidi, F., 2011. Determination of antioxidant activity in free and hydrolyzed fractions of millet grains and characterization of their phenolic profiles by HPLC-DAD-ESI-MS n. *Journal of Functional Foods*, *3*(3), pp.144-158.

Chapple, C., 1998. Molecular-genetic analysis of plant cytochrome P450-dependent monooxygenases. *Annual review of plant biology*, *49*(1), pp.311-343.

Chardonnens, A.N., Koevoets, P.L., van Zanten, A., Schat, H. and Verkleij, J.A., 1999. Properties of enhanced tonoplast zinc transport in naturally selected zinc-tolerant Silene vulgaris. *Plant Physiology*, *120*(3), pp.779-786.

Chasapis, C.T., Loutsidou, A.C., Spiliopoulou, C.A. and Stefanidou, M.E., 2012. Zinc and human health: an update. *Archives of toxicology*, *86*(4), pp.521-534.

Chen, J., Ding, J., Ouyang, Y., Du, H., Yang, J., Cheng, K., Zhao, J., Qiu, S., Zhang, X., Yao, J. and Liu, K., 2008. A triallelic system of S5 is a major regulator of the reproductive barrier and compatibility of indica–japonica hybrids in rice. *Proceedings of the National Academy of Sciences*, *105*(32), pp.11436-11441.

Chen, W., Wang, W., Peng, M., Gong, L., Gao, Y., Wan, J., Wang, S., Shi, L., Zhou, B., Li, Z. and Peng, X., 2016. Comparative and parallel genome-wide association studies for metabolic and agronomic traits in cereals. *Nature communications*, *7*.

Cheng, N.H., Liu, J.Z., Nelson, R.S. and Hirschi, K.D., 2004. Characterization of CXIP4, a novel Arabidopsis protein that activates the H+/Ca2+ antiporter, CAX1. *FEBS letters*, *559*(1-3), pp.99-106.

Chiba, Y., Shimizu, T., Miyakawa, S., Kanno, Y., Koshiba, T., Kamiya, Y. and Seo, M., 2015. Identification of Arabidopsis thaliana NRT1/PTR FAMILY (NPF) proteins capable of transporting plant hormones. *Journal of plant research*, *128*(4), pp.679-686.

Cichy, K.A., Caldas, G.V., Snapp, S.S. and Blair, M.W., 2009. QTL analysis of seed iron, zinc, and phosphorus levels in an Andean bean population. *Crop Science*, *49*(5), pp.1742-1750.

Clark, A.G., 2004. The role of haplotypes in candidate gene studies. *Genetic epidemiology*, *27*(4), pp.321-333.

Clausen, M., Kannangara, R.M., Olsen, C.E., Blomstedt, C.K., Gleadow, R.M., Jørgensen, K., Bak, S., Motawie, M.S. and Møller, B.L., 2015. The bifurcation of the cyanogenic glucoside and glucosinolate biosynthetic pathways. *The Plant Journal*, *84*(3), pp.558-573.

Colangelo, E.P. and Guerinot, M.L., 2004. The essential basic helix-loop-helix protein FIT1 is required for the iron deficiency response. *The Plant Cell*, *16*(12), pp.3400-3412.

Cook, N.C. and Samman, S., 1996. Flavonoids—chemistry, metabolism, cardioprotective effects, and dietary sources. *The Journal of nutritional biochemistry*, *7*(2), pp.66-76.

Costa, A., Barbaro, M.R., Sicilia, F., Preger, V., Krieger-Liszkay, A., Sparla, F., De Lorenzo, G. and Trost, P., 2015. AIR12, a b-type cytochrome of the plasma membrane of Arabidopsis thaliana is a negative regulator of resistance against Botrytis cinerea. *Plant Science*, *233*, pp.32-43.

Croft, D., Mundo, A.F., Haw, R., Milacic, M., Weiser, J., Wu, G., Caudy, M., Garapati, P., Gillespie, M., Kamdar, M.R. and Jassal, B., 2013. The Reactome pathway knowledgebase. *Nucleic acids research*, *42*(D1), pp.D472-D477.

Crozier, A., Burns, J., Aziz, A. A., Stewart, A. J., Rabiasz, H. S., Jenkins, G. I., ... & LEAN, M. E. (2000). Antioxidant flavonols from fruits, vegetables and beverages: measurements and bioavailability. *Biological Research*,*33*(2), 79-88.

Crozier, A., Clifford, M.N. and Ashihara, H. eds., 2008. *Plant secondary metabolites: occurrence, structure and role in the human diet*. John Wiley & Sons.

Curie, C., Cassin, G., Couch, D., Divol, F., Higuchi, K., Le Jean, M., Misson, J., Schikora, A., Czernic, P. and Mari, S., 2009. Metal movement within the plant: contribution of nicotianamine and yellow stripe 1-like transporters. *Annals of botany*, *103*(1), pp.1-11.

Cuyckens, F. and Claeys, M., 2004. Mass spectrometry in the structural analysis of flavonoids. *Journal of Mass Spectrometry*, *39*(1), pp.1-15.

Dakora, F.D. and Phillips, D.A., 2002. Root exudates as mediators of mineral acquisition in low-nutrient environments. In Food Security in Nutrient-Stressed Environments: Exploiting Plants' Genetic Capabilities (pp. 201-213). Springer, Dordrecht.

De Morais Cardoso, L., Pinheiro, S.S., Martino, H.S.D. and Pinheiro-Sant'Ana, H.M., 2017. Sorghum (Sorghum bicolor L.): Nutrients, bioactive compounds, and potential impact on human health. *Critical reviews in food science and nutrition*, *57*(2), pp.372-390.

Deschamps, S., Llaca, V. and May, G.D., 2012. Genotyping-by-sequencing in plants. *Biology*, *1*(3), pp.460-483.

Desjardins, A.E., McCormick, S.P. and Corsini, D.L., 1995. Diversity of sesquiterpenes in 46 potato cultivars and breeding selections. *Journal of agricultural and food chemistry*, *43*(8), pp.2267-2272.

Devos, K.M., 2005. Updating the 'crop circle'. *Current opinion in plant biology*, *8*(2), pp.155-162.

Devos, K.M., Pittaway, T.S., Reynolds, A. and Gale, M.D., 2000. Comparative mapping reveals a complex relationship between the pearl millet genome and those of foxtail millet and rice. *Theoretical and Applied Genetics*, *100*(2), pp.190-198.

Dhawan, D. and Gupta, J., 2017. Comparison of Different Solvents for Phytochemical Extraction Potential from Datura metel Plant Leaves. International Journal of Biological Chemistry, 11, pp.17-22.

Dickin, E., Steele, K., Frost, G., Edwards-Jones, G. and Wright, D., 2011. Effect of genotype, environment and agronomic management on β-glucan concentration of naked barley grain intended for health food use. *Journal of Cereal Science*, *54*(1), pp.44-52.

Dietz, K.J., Tavakoli, N., Kluge, C., Mimura, T., Sharma, S.S., Harris, G.C., Chardonnens, A.N. and Golldack, D., 2001. Significance of the V-type ATPase for the adaptation to stressful growth conditions and its regulation on the molecular and biochemical level. *Journal of experimental botany*, *52*(363), pp.1969-1980.

Dong, X., Gao, Y., Chen, W., Wang, W., Gong, L., Liu, X. and Luo, J., 2015. Spatiotemporal distribution of phenolamides and the genetics of natural variation of hydroxycinnamoyl spermidine in rice. *Molecular plant*, *8*(1), pp.111-121.

Douchkov, D., Gryczka, C., Stephan, U.W., Hell, R. and Bäumlein, H., 2005. Ectopic expression of nicotianamine synthase genes results in improved iron accumulation and increased nickel tolerance in transgenic tobacco. *Plant, Cell & Environment*, *28*(3), pp.365-374.

Du, J. and Du, W., 2009. Correlation of mineral elements between milled and brown rice and soils in Yunnan studied by ICP-AES. *Spectroscopy and Spectral Analysis*, *29*(5), pp.1413-1417.

Dunn, L.L., Rahmanto, Y.S. and Richardson, D.R., 2007. Iron uptake and metabolism in the new millennium. *Trends in cell biology*, *17*(2), pp.93-100.

Dunwell, J.M. and Gane, P.J., 1998. Microbial relatives of seed storage proteins: conservation of motifs in a functionally diverse superfamily of enzymes. *Journal of molecular evolution*, *46*(2), pp.147-154.

Dykes, L. and Rooney, L.W., 2007. Phenolic compounds in cereal grains and their health benefits. *Cereal foods world*, *52*(3), pp.105-111.

Egli, I., Davidsson, L., Zeder, C., Walczyk, T. and Hurrell, R., 2004. Dephytinization of a complementary food based on wheat and soy increases zinc, but not copper, apparent absorption in adults. *The Journal of nutrition*, *134*(5), pp.1077-1080.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S. and Mitchell, S.E., 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS one*, *6*(5), p.e19379.

Eltayeb, M.M., Hassn, A.B. and Babiker, E.E., 2016. Effect of processing followed by fermentation on antinutritional factors content of pearl millet (Pennisetum glaucum L.) cultivars. *University of Khartoum, staff publication.*

Engström, K., Widmark, A.K., Brishammar, S. and Helmersson, S., 1999. Antifungal activity toPhytophthora infestans of sesquiterpenoids from infected potato tubers. *Potato Research*, *42*(1), pp.43-50.

Ernst, W.H.O. and Nelissen, H.J.M., 2000. Life-cycle phases of a zinc-and cadmium-resistant ecotype of Silene vulgaris in risk assessment of polymetallic mine soils. *Environmental Pollution*, *107*(3), pp.329-338.

Etcheverry, P., Grusak, M.A. and Fleige, L.E., 2012. Application of in vitro bioaccessibility and bioavailability methods for calcium, carotenoids, folate, iron, magnesium, polyphenols, zinc, and vitamins B6, B12, D, and E. *Frontiers in physiology*, *3*.

Eyzaguirre, R.Z., Nienaltowska, K., De Jong, L.E., Hasenack, B.B. and Nout, M.J., 2006. Effect of food processing of pearl millet (Pennisetum glaucum) IKMP-5 on the level of phenolics, phytate, iron and zinc. *Journal of the Science of Food and Agriculture*, *86*(9), pp.1391-1398.

Feil, B. and Fossati, D., 1997. Phytic acid in triticale grains as affected by cultivar and environment. *Crop science*, *37*(3), pp.916-921.

Feil, B., Moser, S.B., Jampatong, S. and Stamp, P., 2005. Mineral composition of the grains of tropical maize varieties as affected by pre-anthesis drought and rate of nitrogen fertilization. *Crop Science*, *45*(2), pp.516-523.

Felgines, C., Texier, O., Morand, C., Manach, C., Scalbert, A., Régerat, F., & Rémésy, C. (2000). Bioavailability of the flavanone naringenin and its glycosides in rats. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, *279*(6), G1148-G1154.

Fernandez, M.T., Mira, M.L., Florencio, M.H. and Jennings, K.R., 2002. Iron and copper chelation by flavonoids: an electrospray mass spectrometry study. *Journal of Inorganic Biochemistry*, *92*(2), pp.105-111.

Ferreres, F., Gil-Izquierdo, A., Andrade, P. B., Valentão, P., & Tomás-Barberán, F. A. (2007). Characterization of C-glycosyl flavones O-glycosylated by liquid chromatography–tandem mass spectrometry. *Journal of Chromatography A*, *1161*(1), 214-223.

Ferreres, F., Ribeiro, V., Izquierdo, A. G., Rodrigues, M. Â., Seabra, R. M., Andrade, P. B., & Valentão, P. (2006). Rumex induratus leaves: interesting dietary source of potential bioactive compounds. *Journal of agricultural and food chemistry*, *54*(16), 5782-5789.

Fitzgerald, S.L., Gibson, R.S., de Serrano, J.Q., Portocarrero, L., Vasquez, A., De Zepeda, E., Lopez-Palacios, C.Y., Thompson, L.U., Stephen, A.M. and Solomons, N.W., 1993. Trace element intakes and dietary phytate/Zn and Ca x phytate/Zn millimolar ratios of periurban Guatemalan women during the third trimester of pregnancy. *The American journal of clinical nutrition*, *57*(2), pp.195-201.

Gale, M.D. and Devos, K.M., 1998. Comparative genetics in the grasses. *Proceedings of the National Academy of Sciences*, *95*(5), pp.1971-1974.

Gan, Y., Kumimoto, R., Liu, C., Ratcliffe, O., Yu, H. and Broun, P., 2006. GLABROUS INFLORESCENCE STEMS modulates the regulation by gibberellins of epidermal differentiation and shoot maturation in Arabidopsis. *The Plant Cell*, *18*(6), pp.1383-1395. Gershater, M.C. and Edwards, R., 2007. Regulating biological activity in plants with carboxylesterases. *Plant science*, *173*(6), pp.579-588.

Garris, A.J., McCOUCH, S.R. and Kresovich, S., 2003. Population structure and its effect on haplotype diversity and linkage disequilibrium surrounding the xa5 locus of rice (Oryza sativa L.). *Genetics*, *165*(2), pp.759-769.

Garvin, D.F., Welch, R.M. and Finley, J.W., 2006. Historical shifts in the seed mineral micronutrient concentration of US hard red winter wheat germplasm. *Journal of the Science of Food and Agriculture*, *86*(13), pp.2213-2220.

Ghandilyan, A., Ilk, N., Hanhart, C., Mbengue, M., Barboza, L., Schat, H., Koornneef, M., El-Lithy, M., Vreugdenhil, D., Reymond, M. and Aarts, M.G., 2009. A strong effect of growth medium and organ type on the identification of QTLs for phytate and mineral concentrations in three Arabidopsis thaliana RIL populations. *Journal of Experimental Botany*, *60*(5), pp.1409-1425.

Ghandilyan, A., Vreugdenhil, D. and Aarts, M.G., 2006. Progress in the genetic understanding of plant iron and zinc nutrition. *Physiologia Plantarum*, *126*(3), pp.407-417.

Ghanem, M.E., Albacete, A., Martínez-Andújar, C., Acosta, M., Romero-Aranda, R., Dodd, I.C., Lutts, S. and Pérez-Alfocea, F., 2008. Hormonal changes during salinity-induced leaf senescence in tomato (Solanum lycopersicum L.). *Journal of Experimental Botany*, *59*(11), pp.3039-3050.

Gibson, R.S., Bailey, K.B., Gibbs, M. and Ferguson, E.L., 2010. A review of phytate, iron, zinc, and calcium concentrations in plant-based complementary foods used in low-income countries and implications for bioavailability. *Food and nutrition bulletin*, *31*(2_suppl2), pp.S134-S146.

Gibson, R.S., Hotz, C., Temple, L., Yeudall, F., Mtitimuni, B. and Ferguson, E., 2000. Dietary strategies to combat deficiencies of iron, zinc, and vitamin A in developing countries: development, implementation, monitoring, and evaluation. *Food and Nutrition Bulletin*, *21*(2), pp.219-231.

Gibson, S.W. and Todd, C.D., 2015. Arabidopsis AIR12 influences root development. *Physiology and Molecular Biology of Plants*, *21*(4), pp.479-489.

Gingerich, D.J., Gagne, J.M., Salter, D.W., Hellmann, H., Estelle, M. and Vierstra, R.D., 2005. Cullin 3A and B assemble with members of the broad complex/tramtrack/bric-A-brac (BTB). *Journal of Biological Chemistry*.

Godiard, L., Sauviac, L., Dalbin, N., Liaubet, L., Callard, D., Czernic, P. and Marco, Y., 1998. CYP76C2, an Arabidopsis thaliana cytochrome P450 gene expressed during hypersensitive and developmental cell death. *FEBS letters*, *438*(3), pp.245-249.

Goldman, J.J., Hanna, W.W., Fleming, G. and Ozias-Akins, P., 2003. Fertile transgenic pearl millet [Pennisetum glaucum (L.) R. Br.] plants recovered through microprojectile bombardment and phosphinothricin selection of apical meristem-, inflorescence-, and immature embryo-derived embryogenic tissues. *Plant cell reports*, *21*(10), pp.999-1009.

Gomez-Becerra, H.F., Yazici, A., Ozturk, L., Budak, H., Peleg, Z., Morgounov, A., Fahima, T., Saranga, Y. and Cakmak, I., 2010. Genetic variation and environmental stability of grain mineral nutrient concentrations in Triticum dicoccoides under five environments. *Euphytica*, *171*(1), pp.39-52.

Goto, F., Yoshihara, T., Shigemoto, N., Toki, S. and Takaiwa, F., 1999. Iron fortification of rice seed by the soybean ferritin gene. *Nature biotechnology*, *17*(3), pp.282-286.

Govindaraj, M., Rai, K.N. and Shanmugasundaram, P., 2016. Intra-population genetic variance for grain iron and zinc contents and agronomic traits in pearl millet. The Crop Journal, 4(1), pp.48-54.

Govindaraj, M., Rai, K.N., Shanmugasundaram, P., Dwivedi, S.L., Sahrawat, K.L., Muthaiah, A.R. and Rao, A.S., 2013. Combining ability and heterosis for grain iron and zinc densities in pearl millet. *Crop Science*, *53*(2), pp.507-517.

Graham, R.D., Welch, R.M. and Bouis, H.E., 2001. Addressing micronutrient malnutrition through enhancing the nutritional quality of staple foods: principles, perspectives and knowledge gaps. *Advances in agronomy*, *70*, pp.77-142.

Grandbastien, M.A., 1998. Activation of plant retrotransposons under stress conditions. *Trends in plant science*, *3*(5), pp.181-187.

Gray, D. and STECKEL, J.R., 1986. Self-and open-pollination as factors influencing seed quality in leek (Allium porrum). *Annals of applied biology*, *108*(1), pp.167-170.

Green, D.I., Agu, R.C., Bringhurst, T.A., Brosnan, J.M., Jack, F.R. and Walker, G.M., 2015. Maximizing alcohol yields from wheat and maize and their co-products for distilling or bioethanol production. *Journal of the Institute of Brewing*, *121*(3), pp.332-337.

Grillet, L., Mari, S. and Schmidt, W., 2014. Iron in seeds–loading pathways and subcellular localization. *Frontiers in plant science*, *4*, p.535.

Grotz, N., Fox, T., Connolly, E., Park, W., Guerinot, M.L. and Eide, D., 1998. Identification of a family of zinc transporter genes from Arabidopsis that respond to zinc deficiency. *Proceedings of the National Academy of Sciences*, *95*(12), pp.7220-7224.

Guerinot, M.L., 2000. The ZIP family of metal transporters. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, *1465*(1), pp.190-198.

Guerrero, F. and Mullet, J.E., 1986. Increased abscisic acid biosynthesis during plant dehydration requires transcription. *Plant Physiology*, *80*(2), pp.588-591.

Gugler, R., Leschik, M. and Dengler, H.J., 1975. Disposition of quercetin in man after single oral and intravenous doses. *European journal of clinical pharmacology*, *9*(2), pp.229-234.

Gupta, S.K., Rai, A.K., Kanwar, S.S. and Sharma, T.R., 2012. Comparative analysis of zinc finger proteins involved in plant disease resistance. *PLoS One*, *7*(8), p.e42578.

Gupta, S.K., Velu, G., Rai, K.N. and Sumalini, K., 2009. Association of grain iron and zinc content with grain yield and other traits in pearl millet (Pennisetum glaucum (L.) R. Br.). *Crop Improvement*, *36*(2), pp.4-7.

Gurmu, F., Mohammed, H. and Alemaw, G., 2009. Genotype x environment interactions and stability of soybean for grain yield and nutrition quality. *African Crop Science Journal*, *17*(2).

Hallberg, L., Brune, M. and Rossander, L., 1989. Iron absorption in man: ascorbic acid and dose-dependent inhibition by phytate. *The American journal of clinical nutrition*, *49*(1), pp.140-144.

Hambidge, K.M., Huffer, J.W., Raboy, V., Grunwald, G.K., Westcott, J.L., Sian, L., Miller, L.V., Dorsch, J.A. and Krebs, N.F., 2004. Zinc absorption from low-phytate hybrids of maize and their wild-type isohybrids. *The American journal of clinical nutrition*, *79*(6), pp.1053-1059.

Hambidge, M., 2000. Human zinc deficiency. *The Journal of nutrition*, *130*(5), pp.1344S-1349S.

Han, S.K., Song, J.D., Noh, Y.S. and Noh, B., 2007. Role of plant CBP/p300-like genes in the regulation of flowering time. *The Plant Journal*, *49*(1), pp.103-114.

Hanna, W.W., 1986, April. Utilization of wild relatives of pearl millet. In *Proceedings of the international pearl millet workshop* (pp. 7-11).

Hardie, D.G., 1999. Plant protein serine/threonine kinases: classification and functions. *Annual review of plant biology*, *50*(1), pp.97-131.

Hash, C.T., Sharma, A., Kolesnikova-Allen, M.A., Singh, S.D., Thakur, R.P., Raj, A.B., Rao, M.R., Nijhawan, D.C., Beniwal, C.R., Sagar, P. and Yadav, H.P., 2006. Teamwork delivers biotechnology products to Indian small-holder crop-livestock producers: Pearl millet hybrid "HHB 67 Improved" enters seed delivery pipeline. *Journal of SAT Agricultural Research*, *2*(1), pp.1-3.

Hattan, J., Kanamoto, H., Takemura, M., YOKOTA, A. and KOHCHI, T., 2004. Molecular characterization of the cytoplasmic interacting protein of the receptor kinase IRK expressed in the inflorescence and root apices of Arabidopsis. *Bioscience, biotechnology, and biochemistry*, *68*(12), pp.2598-2606.

Hawksworth, G., Drasar, B.S., and Hill, M.J. (1971) Intestinal bacteria and the hydrolysis of glycosidic bonds. *Journal of Medical Microbiology.* 4: 451-459.

Heim, M.A., Jakoby, M., Werber, M., Martin, C., Weisshaar, B. and Bailey, P.C., 2003. The basic helix–loop–helix transcription factor family in plants: a genome-wide study of protein structure and functional diversity. *Molecular biology and evolution*, *20*(5), pp.735-747.

Hill, W.G. and Weir, B.S., 1988. Variances and covariances of squared linkage disequilibria in finite populations. *Theoretical population biology*, *33*(1), pp.54-78.

Hindu, V., Palacios-Rojas, N., Babu, R., Suwarno, W.B., Rashid, Z., Usha, R., Saykhedkar, G.R. and Nair, S.K., 2018. Identification and validation of genomic regions influencing kernel zinc and iron in maize. *Theoretical and Applied Genetics*, pp.1-15.

Hirschi, K.D., 1999. Expression of Arabidopsis CAX1 in tobacco: altered calcium homeostasis and increased stress sensitivity. *The Plant Cell*, *11*(11), pp.2113-2122.

Hoddinott, J., 2013. The economic cost of malnutrition. In *The Road to Good Nutrition* (pp. 64-73). Karger Publishers.

Horna, D., Smale, M., Al-Hassan, R., Falck-Zepeda, J. and Timpo, S.E., 2008. Insecticide use on vegetables in Ghana: Would GM seed benefit farmers?. Intl Food Policy Res Inst.

Hornick, S.B., 1992. Factors affecting the nutritional quality of crops. *American Journal of Alternative Agriculture*, *7*(1-2), pp.63-68.

Horton, J.S., Wakano, C.T., Speck, M. and Stokes, A.J., 2015. Two-pore channel 1 interacts with citron kinase, regulating completion of cytokinesis. *Channels*, *9*(1), pp.21-29.

Hou, H., Atlihan, N. and Lu, Z.X., 2014. New biotechnology enhances the application of cisgenesis in plant breeding. Frontiers in plant science, 5, p.389.

Hsieh, M.H. and Goodman, H.M., 2002. Molecular characterization of a novel gene family encoding ACT domain repeat proteins in Arabidopsis. *Plant physiology*, *130*(4), pp.1797-1806.

Hu, Z., Mbacké, B., Perumal, R., Guèye, M.C., Sy, O., Bouchet, S., Prasad, P.V. and Morris, G.P., 2015. Population genomics of pearl millet (Pennisetum glaucum (L.) R. Br.): Comparative analysis of global accessions and Senegalese landraces. *BMC genomics*, *16*(1), p.1048.

Huang, X. and Han, B., 2014. Natural variations and genome-wide association studies in crop plants. *Annual review of plant biology*, *65*, pp.531-551.

Huang, X., Kurata, N., Wang, Z.X., Wang, A., Zhao, Q., Zhao, Y., Liu, K., Lu, H., Li, W., Guo, Y. and Lu, Y., 2012. A map of rice genome variation reveals the origin of cultivated rice. Nature, 490(7421), p.497.

Huehn, M., 1990. Nonparametric measures of phenotypic stability. Part 1: Theory. *Euphytica*, *47*(3), pp.189-194.

Hurrell, R. and Egli, I., 2010. Iron bioavailability and dietary reference values. *The American journal of clinical nutrition*, *91*(5), pp.1461S-1467S.

Hurrell, R.F., 1997. Preventing iron deficiency through food fortification. *Nutrition reviews*, *55*(6), pp.210-222.

Hurrell, R.F., 2003. Influence of vegetable protein sources on trace element and mineral bioavailability. *The Journal of nutrition*, *133*(9), pp.2973S-2977S.

Hyder, S.Z., Persson, L.Å., Chowdhury, M., Lönnerdal, B.O. and Ekström, E.C., 2004. Anaemia and iron deficiency during pregnancy in rural Bangladesh. *Public health nutrition*, *7*(08), pp.1065-1070.

Hyten, D.L. and Lee, D.J., 2016. Plant Genetic Mapping Techniques. *eLS*.

ICRISAT (International Crops Research Institute for the Semi- arid Tropics). 2016. Pearl millet. [2016-01-20]. http:// exploreit.icrisat.org/page/pearl_millet/680/274 2016.

ICRISAT. (2016). ADDRESSING MALNUTRITION BY MAINSTREAMING PEARL MILLET BIOFORTIFICATION BREEDING. Available: http://www.icrisat.org/addressing-malnutrition-by-mainstreaming-pearl-millet-biofortification-breeding/. Last accessed 09/07/2018.

Inoue, H., Higuchi, K., Takahashi, M., Nakanishi, H., Mori, S. and Nishizawa, N.K., 2003. Three rice nicotianamine synthase genes, OsNAS1, OsNAS2, and OsNAS3 are expressed in cells involved in long-distance transport of iron and differentially regulated by iron. *The Plant Journal*, *36*(3), pp.366-381.

Iqbal, T.H., Lewis, K.O. and Cooper, B.T., 1994. Phytase activity in the human and rat small intestine. *Gut*, *35*(9), pp.1233-1236.

Ito, O., O'Toole, J.C. and Hardy, B. eds., 1999. *Genetic improvement of rice for water-limited environments*. Int. Rice Res. Inst..

Iwahashi, H., Akata, K., Sunaga, A., Tone, Y., Yamada, N. and Iijima, K., 2004. The inhibitory effect of luteolin-7-O-glucoside on the formation of pentyl and 7-carboxyheptyl radicals from 13-hydroperoxy-9, 11-octadecadienoic acid in the presence of iron (II) ions. *Free radical research*, *38*(8), pp.869-876.

Jalaja, N., Maheshwari, P., Naidu, K.R. and Kavi Kishor, P.B., 2016. In vitro regeneration and optimization of conditions for transformation methods in Pearl millet, Pennisetum glaucum (L.). *International Journal of Clinical and Biological Sciences*, *1*(1), pp.34-52.

James, C., 2015. Global status of commercialized biotech/GM crops: 2014. *ISAAA brief*, *49*.

Jaspers, P., Brosché, M., Overmyer, K. and Kangasjär, J., 2010. The transcription factor interacting protein RCD1 contains a novel conserved domain. *Plant signaling & behavior*, *5*(1), pp.78-80.

Jellum, M.D. and Powell, J.B., 1971. Fatty Acid Composition of Oil from Pearl Millet Seed 1. Agronomy Journal, 63(1), pp.29-33.

Jiao, Y., Zhao, H., Ren, L., Song, W., Zeng, B., Guo, J., Wang, B., Liu, Z., Chen, J., Li, W. and Zhang, M., 2012. Genome-wide genetic changes during modern breeding of maize. *Nature genetics*, *44*(7), pp.812-815.

Jin, L., Lu, Y., Xiao, P., Sun, M., Corke, H. and Bao, J., 2010. Genetic diversity and population structure of a diverse set of rice germplasm for association mapping. *Theoretical and Applied Genetics*, *121*(3), pp.475-487.

Jones, E.R.L., 1997. Brown rust of wheat. *Tellus*, *2*, p.1.

Jones, E.S., Breese, W.A., Liu, C.J., Singh, S.D., Shaw, D.S. and Witcombe, J.R., 2002. Mapping quantitative trait loci for resistance to downy mildew in pearl millet. *Crop Science*, *42*(4), pp.1316-1323.

Jun, H. K., Park, K. Y., & Jo, J. B. (1989). Inhibitory effects of Ginseng saponins on Aflatoxin production in culture. In *Chem. Abstr* (Vol. 106, pp. 116-199).

Kalinowski, S.T., 2011. The computer program STRUCTURE does not reliably identify the main genetic clusters within species: simulations and implications for human population structure. *Heredity*, *106*(4), pp.625-632.

Kanatti, A., Rai, K.N., Radhika, K., Govindaraj, M., Sahrawat, K.L. and Rao, A.S., 2014. Grain iron and zinc density in pearl millet: combining ability, heterosis and association with grain yield and grain size. *SpringerPlus*, *3*(1), p.763.

Kannan, B., Senapathy, S., Bhasker Raj, A.G., Chandra, S., Muthiah, A., Dhanapal, A.P. and Hash, C.T., 2014. Association analysis of SSR markers with phenology, grain, and stover-yield related traits in Pearl Millet (Pennisetum glaucum (L.) R. Br.). *The scientific world journal*, *2014*.

Kaplan, B., Davydov, O., Knight, H., Galon, Y., Knight, M.R., Fluhr, R. and Fromm, H., 2006. Rapid transcriptome changes induced by cytosolic Ca2+ transients reveal ABRE-related sequences as Ca2+-responsive cis elements in Arabidopsis. *The Plant Cell*, *18*(10), pp.2733-2748.

Kapur, M., Wang, W., Maloney, M.T., Millan, I., Lundin, V.F., Tran, T.A. and Yang, Y., 2012. Calcium tips the balance: a microtubule plus end to lattice binding switch operates in the carboxyl terminus of BPAG1n4. *EMBO reports*, *13*(11), pp.1021-1029.

Kassebaum, N.J., Jasrasaria, R., Naghavi, M., Wulf, S.K., Johns, N., Lozano, R., Regan, M., Weatherall, D., Chou, D.P., Eisele, T.P. and Flaxman, S.R., 2014. A systematic analysis of global anemia burden from 1990 to 2010. *Blood*, *123*(5), pp.615-624.

Kato, Y., Murakami, S., Yamamoto, Y., Chatani, H., Kondo, Y., Nakano, T., Yokota, A. and Sato, F., 2004. The DNA-binding protease, CND41, and the degradation of ribulose-1, 5-bisphosphate carboxylase/oxygenase in senescent leaves of tobacco. *Planta*, *220*(1), pp.97-104.

Kato-Noguchi, H., 2011. Convergent or parallel molecular evolution of momilactone A and B: Potent allelochemicals, momilactones have been found only in rice and the moss Hypnum plumaeforme. *Journal of plant physiology*, *168*(13), pp.1511-1516.

Kaur, R. and Soodan, A.S., 2017. Reproductive biology of Sorghum halepense (L.) Pers.(Poaceae; Panicoideae; Andropogoneae) in relation to invasibility. *Flora*, *229*, pp.32-49.

Khairwal, I.S. and Hash, C.T., 2007. HHB 67-improved–The first product of marker-assisted crop breeding in India. Asia-Pacific Consortium on Agricultural Biotechnology (APCoAB) e-News.

Khairwal, I.S., Rai, K.N., Diwakar, B., Sharma, Y.K., Rajpurohit, B.S., Nirwan, B. and Bhattacharjee, R., 2007. Pearl Millet Crop Management and Seed Production Manual.

Khairwal, I.S., Yadav, S.K., Rai, K.N., Upadhyaya, H.D., Kachhawa, D., Nirwan, B., Bhattacharjee, R., Rajpurohit, B.S. and Dangaria, C.J., 2007. Evaluation and identification of promising pearl millet germplasm for grain and fodder traits. *Journal of SAT Agricultural Research*, *5*(1), pp.1-6.

Khatkar, B.S., Bell, A.E. and Schofield, J.D., 1995. The dynamic rheological properties of glutens and gluten sub-fractions from wheats of good and poor bread making quality. *Journal of Cereal Science*, *22*(1), pp.29-44.

Kim, J.H., Cheon, Y.M., Kim, B.G. and Ahn, J.H., 2008. Analysis of flavonoids and characterization of theOsFNS gene involved in flavone biosynthesis in Rice. *Journal of Plant Biology*, *51*(2), p.97.

Klaassen, C.D. and Rozman, K. (1991) Absorption, distribution, and excretion of toxicants. In: Casarett and Doull's Toxicology - the Basic Science of Posions, 4th Edition (Amdur, M.O., Doull, J. and Klaassen C.D., eds), pp. 50-87. Pergamon Press, Elmsford, NY.

Klug, A. and Schwabe, J.W., 1995. Protein motifs 5. Zinc fingers. *The FASEB journal*, *9*(8), pp.597-604.

Knip, M., de Pater, S. and Hooykaas, P.J., 2012. The SLEEPER genes: a transposase-derived angiosperm-specific gene family. *BMC plant biology*, *12*(1), p.192.

Knox, J., Hess, T., Daccache, A. and Wheeler, T., 2012. Climate change impacts on crop productivity in Africa and South Asia. *Environmental Research Letters*, *7*(3), p.034032.

Knuckles, B.E., Kuzmicky, D.D. and Betschart, A.A., 1985. Effect of phytate and partially hydrolyzed phytate on in vitro protein digestibility. *Journal of Food Science*, *50*(4), pp.1080-1082.

Kobayashi, T. and Nishizawa, N.K., 2012. Iron uptake, translocation, and regulation in higher plants. *Annual review of plant biology*, *63*, pp.131-152.

Kodkany, B.S., Bellad, R.M., Mahantshetti, N.S., Westcott, J.E., Krebs, N.F., Kemp, J.F. and Hambidge, K.M., 2013. Biofortification of pearl millet with iron and zinc in a randomized controlled trial increases absorption of these minerals above physiologic requirements in young children. *The Journal of nutrition*, *143*(9), pp.1489-1493.

Kolbe, W., 1970. Further studies on the reduction of cereal yields by aphid infestation. *Pflanzenschutz-Nachrichten Bayer*, *23*(2), pp.144-162.

Komaki, S., Abe, T., Coutuer, S., Inzé, D., Russinova, E. and Hashimoto, T., 2010. Nuclear-localized subtype of end-binding 1 protein regulates spindle organization in Arabidopsis. *J Cell Sci*, *123*(3), pp.451-459.

Korte, A. and Farlow, A., 2013. The advantages and limitations of trait analysis with GWAS: a review. Plant methods, 9(1), p.29.

Kulp, K. ed., 2000. *Handbook of Cereal Science and Technology, revised and expanded*. CRC Press.

Kumar, A.L.P.A.N.A. and Chauhan, B.M., 1993. Effects of phytic acid on protein digestibility (in vitro) and HCl-extractability of minerals in pearl millet sprouts. *Cereal chemistry*, *70*, pp.504-504.

Kumar, S., 2011. Development of new mapping population and marker-assisted improvement of iron and zinc grain density in pearl millet [Pennisetum glaucum (L.) R. Br.]. *Bikaner: Dissertation, Swami Keshwanand Rajasthan Agricultural University*.

Kumar, S., Hash, C.T., Nepolean Thirunavukkarasu, G.S., Rajaram, V., Rathore, A., Senapathy, S., Mahendrakar, M.D., Yadav, R.S. and Srivastava, R.K., 2016. Mapping Quantitative Trait Loci Controlling High Iron and Zinc Content in Self and Open Pollinated Grains of Pearl Millet [Pennisetum glaucum (L.) R. Br.]. *Frontiers in plant science*, *7*.

Kumar, S., Hash, C.T., Nepolean, T., Mahendrakar, M.D., Satyavathi, C.T., Singh, G., Rathore, A., Yadav, R.S., Gupta, R. and Srivastava, R.K., 2018. Mapping Grain Iron and Zinc Content Quantitative Trait Loci in an Iniadi-Derived Immortal Population of Pearl Millet. Genes, 9(5), p.248.

Kumar, S., Hash, T., Nepolean, T., Mahendrakar, M., 2, Satyavathi, C.T., Singh, G., Rathore, A., Yadav, Y., Gupta, R.,. 2018. Mapping Grain Iron and Zinc Content Quantitative Trait Loci in an Iniadi-Derived Immortal Population of Pearl Millet. *Genes*. 9 (5), pp248

Lamberti, L.M., Fischer Walker, C.L. and Black, R.E., 2016. Zinc Deficiency in Childhood and Pregnancy: Evidence for Intervention Effects and Program Responses. In *Hidden Hunger* (Vol. 115, pp. 125-133). Karger Publishers.

Langewisch, T., Zhang, H., Vincent, R., Joshi, T., Xu, D. and Bilyeu, K., 2014. Major soybean maturity gene haplotypes revealed by SNPViz analysis of 72 sequenced soybean genomes. *PloS one*, *9*(4), p.e94150.

Lasat, M.M. and Kochian, L.V., 2000. Physiology of Zn hyperaccumulation in Thlaspi caerulescens. *Phytoremediation of contaminated soil and water*, pp.159-169.

Ledford, H., 2015. CRISPR, the disruptor. Nature News, 522(7554), p.20.

Lee, S. and An, G., 2009. Over-expression of OsIRT1 leads to increased iron and zinc accumulations in rice. Plant, cell & environment, 32(4), pp.408-416.

Lee, S., Jeon, U.S., Lee, S.J., Kim, Y.K., Persson, D.P., Husted, S., Schjørring, J.K., Kakei, Y., Masuda, H., Nishizawa, N.K. and An, G., 2009. Iron fortification of rice seeds through activation of the nicotianamine synthase gene. *Proceedings of the National Academy of Sciences*, *106*(51), pp.22014-22019.

Leff, B., Ramankutty, N. and Foley, J.A., 2004. Geographic distribution of major crops across the world. *Global Biogeochemical Cycles*, *18*(1).

Lestari, P., Lee, G., Ham, T.H., Woo, M.O., Piao, R., Jiang, W., Chu, S.H., Lee, J. and Koh, H.J., 2011. Single nucleotide polymorphisms and haplotype diversity in rice sucrose synthase 3. *Journal of Heredity*, *102*(6), pp.735-746.

Lestienne, I., Besançon, P., Caporiccio, B., Lullien-Péllerin, V. and Tréche, S., 2005. Iron and zinc in vitro availability in pearl millet flours (Pennisetum glaucum) with varying phytate, tannin, and fiber contents. *Journal of agricultural and food chemistry*, *53*(8), pp.3240-3247.

Lestienne, I., Icard-Vernière, C., Mouquet, C., Picq, C. and Trèche, S., 2005. Effects of soaking whole cereal and legume seeds on iron, zinc and phytate contents. *Food chemistry*, *89*(3), pp.421-425.

Limagrain Europe. (2015). REVELATION WINTER WHEAT. Available: http://www.lgseeds.co.uk/uploads/Revelation-Technical-Summary.pdf. Last accessed 05/11/2016

Lind, T., Lönnerdal, B., Stenlund, H., Gamayanti, I.L., Ismail, D., Seswandhana, R. and Persson, L.Å., 2004. A community-based randomized controlled trial of iron and zinc supplementation in Indonesian infants: effects on growth and development. *The American journal of clinical nutrition*, *80*(3), pp.729-736.

Linthorst, H.J., Melchers, L.S., Mayer, A., Van Roekel, J.S., Cornelissen, B.J. and Bol, J.F., 1990. Analysis of gene families encoding acidic and basic beta-1, 3-glucanases of tobacco. *Proceedings of the National Academy of Sciences*, *87*(22), pp.8756-8760.

Liotta, L.J. and James-Pederson, M., 2008. Identification of an Unknown Compound by Combined Use of IR, 1H NMR, 13C NMR, and Mass Spectrometry: A Real-Life Experience in Structure Determination. *Journal of chemical education*, *85*(6), p.832.

Liu, H., Zhou, H., Wu, Y., Li, X., Zhao, J., Zuo, T., Zhang, X., Zhang, Y., Liu, S., Shen, Y. and Lin, H., 2015. The impact of genetic relationship and linkage disequilibrium on genomic selection. *PloS one*, *10*(7), p.e0132379.

Liu, H.C., Liao, H.T. and Charng, Y.Y., 2011. The role of class A1 heat shock factors (HSFA1s) in response to heat and other stresses in Arabidopsis. *Plant, cell & environment*, *34*(5), pp.738-751.

Liu, N., Zhang, K. and Zhao, H., 2008. Haplotype-association analysis. *Advances in genetics*, *60*, pp.335-405.

Liu, X., Cui, H., Li, A., Zhang, M. and Teng, Y., 2015. The nitrate transporter NRT1. 1 is involved in iron deficiency responses in Arabidopsis. *Journal of Plant Nutrition and Soil Science*, *178*(4), pp.601-608.

Lo, S.F., Yang, S.Y., Chen, K.T., Hsing, Y.I., Zeevaart, J.A., Chen, L.J. and Yu, S.M., 2008. A novel class of gibberellin 2-oxidases control semidwarfism, tillering, and root development in rice. *The Plant Cell*, *20*(10), pp.2603-2618.

Lönnerdal, B.O., 2000. Dietary factors influencing zinc absorption. *The Journal of nutrition*, *130*(5), pp.1378S-1383S.

Lopez, A., Cacoub, P., Macdougall, I.C. and Peyrin-Biroulet, L., 2016. Iron deficiency anaemia. *The Lancet*, *387*(10021), pp.907-916.

Loqué, D., Ludewig, U., Yuan, L. and von Wirén, N., 2005. Tonoplast intrinsic proteins AtTIP2; 1 and AtTIP2; 3 facilitate NH3 transport into the vacuole. *Plant physiology*, *137*(2), pp.671-680.

Lu, Y., Xu, J., Yuan, Z., Hao, Z., Xie, C., Li, X., Shah, T., Lan, H., Zhang, S., Rong, T. and Xu, Y., 2012. Comparative LD mapping using single SNPs and haplotypes identifies QTL for plant height and biomass as secondary traits of drought tolerance in maize. *Molecular Breeding*, *30*(1), pp.407-418.

Lucht, J.M., 2015. Public acceptance of plant biotechnology and GM crops. *Viruses*, *7*(8), pp.4254-4281.

Ma, G., Jin, Y., Piao, J., Kok, F., Guusje, B. and Jacobsen, E., 2005. Phytate, calcium, iron, and zinc contents and their molar ratios in foods commonly consumed in China. *Journal of Agricultural and Food Chemistry*, *53*(26), pp.10285-10290.

Madaule, P., Furuyashiki, T., Reid, T., Ishizaki, T., Watanabe, G., Morii, N. and Narumiya, S., 1995. A novel partner for the GTP-bound forms of rho and rac. *FEBS letters*, *377*(2), pp.243-248.

Maestre, M., Poole, N. and Henson, S., 2017. Assessing food value chain pathways, linkages and impacts for better nutrition of vulnerable groups. *Food Policy*, *68*, pp.31-39.

Mahadevan, B., Sivakumar, S., Dinesh Kumar, D. and Ganeshram, K., 2013. Redesigning midday meal logistics for the Akshaya Patra Foundation: OR at work in feeding hungry school children. Interfaces, 43(6), pp.530-546.

Makanya, Z., 2004. Twelve reasons for Africa to reject GM crops. Seedling magazine, 17.

Mamo, B.E., Barber, B.L. and Steffenson, B.J., 2014. Genome-wide association mapping of zinc and iron concentration in barley landraces from Ethiopia and Eritrea. *Journal of Cereal Science*, *60*(3), pp.497-506.

Manach, C., Scalbert, A., Morand, C., Rémésy, C. and Jiménez, L., 2004. Polyphenols: food sources and bioavailability. *The American journal of clinical nutrition*, *79*(5), pp.727-747.

Manach, C., Williamson, G., Morand, C., Scalbert, A. and Rémésy, C., 2005. Bioavailability and bioefficacy of polyphenols in humans. I. Review of 97 bioavailability studies. *The American journal of clinical nutrition*, *81*(1), pp.230S-242S.

Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A. and Cho, J.H., 2009. Finding the missing heritability of complex diseases. Nature, 461(7265), p.747.

Manwaring, H.R., Bligh, H.F.J. and Yadav, R., 2016. The Challenges and Opportunities Associated with Biofortification of Pearl Millet (Pennisetum glaucum) with Elevated Levels of Grain Iron and Zinc. *Frontiers in Plant Science*, *7*.

Maret, W. and Sandstead, H.H., 2006. Zinc requirements and the risks and benefits of zinc supplementation. Journal of Trace Elements in Medicine and Biology, 20(1), pp.3-18.

Markham, K.R., Ryan, K.G., Bloor, S.J. and Mitchell, K.A., 1998. An increase in the luteolin: apigenin ratio in Marchantia polymorpha on UV-B enhancement. *Phytochemistry*, *48*(5), pp.791-794.

Marroni, F., Pinosio, S., Zaina, G., Fogolari, F., Felice, N., Cattonaro, F. and Morgante, M., 2011. Nucleotide diversity and linkage disequilibrium in Populus nigra cinnamyl alcohol dehydrogenase (CAD4) gene. *Tree genetics & genomes*, *7*(5), pp.1011-1023.

Marschner, H. and Rimmington, G., 1988. Mineral nutrition of higher plants. *Plant Cell Environ*, *11*, pp.147-148.

Marshall, S.D., Putterill, J.J., Plummer, K.M. and Newcomb, R.D., 2003. The carboxylesterase gene family from Arabidopsis thaliana. *Journal of molecular evolution*, *57*(5), pp.487-500.

Matus-Cadiz, M.A., Hucl, P., Perron, C.E. and Tyler, R.T., 2003. Genotype× environment interaction for grain color in hard white spring wheat. *Crop Science*, *43*(1), pp.219-226.

Maurel, C., Javot, H., Lauvergeat, V., Gerbeau, P., Tournaire, C., Santoni, V. and Heyes, J., 2002. Molecular physiology of aquaporins in plants. *International review of cytology*, *215*, pp.105-148.

McCabe, C.D. and Innis, J.W., 2005. A genomic approach to the identification and characterization of HOXA13 functional binding elements. *Nucleic acids research*, *33*(21), pp.6782-6794.

McCarty, M.F., 2005. A chlorogenic acid-induced increase in GLP-1 production may mediate the impact of heavy coffee consumption on diabetes risk. Medical hypotheses, 64(4), pp.848-853.

Messina, M. J. (1999). Legumes and soybeans: overview of their nutritional profiles and health effects. *The American journal of clinical nutrition*, *70*(3), 439s-450s.

Meza, F.J. and Silva, D., 2009. Dynamic adaptation of maize and wheat production to climate change. *Climatic change*, *94*(1), pp.143-156.

Miean, K.H. and Mohamed, S., 2001. Flavonoid (myricetin, quercetin, kaempferol, luteolin, and apigenin) content of edible tropical plants. *Journal of agricultural and food chemistry*, *49*(6), pp.3106-3112.

Mira, L., Tereza Fernandez, M., Santos, M., Rocha, R., Helena Florêncio, M., & Jennings, K. R. (2002). Interactions of flavonoids with iron and copper ions: a mechanism for their antioxidant activity. *Free radical research*,*36*(11), 1199-1208.

Miret, S., Tascioglu, S., van der Burg, M., Frenken, L. and Klaffke, W., 2009. In vitro bioavailability of iron from the heme analogue sodium iron chlorophyllin. *Journal of agricultural and food chemistry*, *58*(2), pp.1327-1332.

Miyagishima, S.Y., Nishida, K., Mori, T., Matsuzaki, M., Higashiyama, T., Kuroiwa, H. and Kuroiwa, T., 2003. A plant-specific dynamin-related protein forms a ring at the chloroplast division site. *The Plant Cell*, *15*(3), pp.655-665.

Moghaddam, M.J. and Pourdad, S.S., 2009. Comparison of parametric and non-parametric methods for analysing genotypexenvironment interactions in safflower (Carthamus tinctorius L.). *The Journal of Agricultural Science*, *147*(5), p.601.

Monteiro, M., Farah, A., Perrone, D., Trugo, L.C. and Donangelo, C., 2007. Chlorogenic acid compounds from coffee are differentially absorbed and metabolized in humans. The Journal of nutrition, 137(10), pp.2196-2201.

Morohashi, K., Casas, M.I., Ferreyra, M.L.F., Mejía-Guerra, M.K., Pourcel, L., Yilmaz, A., Feller, A., Carvalho, B., Emiliani, J., Rodriguez, E. and Pellegrinet, S., 2012. A genome-wide regulatory framework identifies maize pericarp color1 controlled genes. *The Plant Cell*, *24*(7), pp.2745-2764.

Morrissey, J. and Guerinot, M.L., 2009. Iron uptake and transport in plants: the good, the bad, and the ionome. Chemical reviews, 109(10), pp.4553-4567.

Moumouni, K.H., Kountche, B.A., Jean, M., Hash, C.T., Vigouroux, Y., Haussmann, B.I.G. and Belzile, F., 2015. Construction of a genetic map for pearl millet, Pennisetum glaucum (L.) R. Br., using a genotyping-by-sequencing (GBS) approach. *Molecular breeding*, *35*(1), p.5.

Muthamilarasan, M., Dhaka, A., Yadav, R. and Prasad, M., 2016. Exploration of millet models for developing nutrient rich graminaceous crops. *Plant Science*, *242*, pp.89-97.

N'Diaye, A., Haile, J.K., Cory, A.T., Clarke, F.R., Clarke, J.M., Knox, R.E. and Pozniak, C.J., 2017. Single marker and haplotype-based association analysis of semolina and pasta colour in elite durum wheat breeding lines using a high-density consensus map. *PloS one*, *12*(1), p.e0170941.

National Research Council, 2011. *Advancing the science of climate change*. National Academies Press.

Neggers, Y.H., Cutter, G.R., Acton, R.T., Alvarez, J.O., Bonner, J.L., Goldenberg, R.L., Go, R.C. and Roseman, J.M., 1990. A positive association between maternal serum zinc concentration and birth weight. *The American journal of clinical nutrition*, *51*(4), pp.678-684.

Nekohashi, M., Ogawa, M., Ogihara, T., Nakazawa, K., Kato, H., Misaka, T., Abe, K. and Kobayashi, S., 2014. Luteolin and quercetin affect the cholesterol absorption mediated by epithelial cholesterol transporter Niemann–Pick C1-Like 1 in caco-2 cells and rats. *PloS one*, *9*(5), p.e97901.

Nelson, R.J., Naylor, R.L. and Jahn, M.M., 2004. The role of genomics research in improvement of" orphan" crops. *Crop Science*, *44*(6), p.1901.

Nestel, P., Bouis, H. E., & Meenakshi, J. V. (2015). Biofortification of staple food crops: Six questions.

Neuffer, M.G., Coe, E.H. and Wessler, S.R., 1997. Mutants of maize Cold Spring Harbor Laboratory Press.

Nielsen, A.V., Tetens, I. and Meyer, A.S., 2013. Potential of phytase-mediated iron release from cereal-based foods: a quantitative view. *Nutrients*, *5*(8), pp.3074-3098.

Nodine, M.D., Yadegari, R. and Tax, F.E., 2007. RPK1 and TOAD2 are two receptor-like kinases redundantly required for Arabidopsis embryonic pattern formation. *Developmental cell*, *12*(6), pp.943-956.

O'Kennedy, M.M., Grootboom, A. and Shewry, P.R., 2006. Harnessing sorghum and millet biotechnology for food and health. *Journal of Cereal Science*, *44*(3), pp.224-235.

Okubo K, Kudou S, Uchida T, Yoshiki Y, Yoshikoshi M, Tonomura M (1994). Soybean saponin and isoflavonoids: Structure and antiviral activity against Human Immunodeficiency virus in-vitro. A Cs.Symp.Ser.1994, Food Phytochem. Cancer Prevent. I. 330-339

Oladosu, Y., Rafii, M.Y., Abdullah, N., Hussin, G., Ramli, A., Rahim, H.A., Miah, G. and Usman, M., 2016. Principle and application of plant mutagenesis in crop improvement: a review. *Biotechnology & Biotechnological Equipment*, *30*(1), pp.1-16.

Oldach, K., Morgenstern, A., Rother, S., Girgi, M., O'Kennedy, M. and Lörz, H., 2001. Efficient in vitro plant regeneration from immature zygotic embryos of pearl millet [Pennisetum glaucum (L.) R. Br.] and Sorghum bicolor (L.) Moench. *Plant cell reports*, *20*(5), pp.416-421.

Olenichenko, N.A., Ossipov, V.I. and Zagoskina, N.V., 2006. Effect of cold hardening on the phenolic complex of winter wheat leaves. *Russian Journal of Plant Physiology*, *53*(4), pp.495-500.

Olsen, L.I. and Palmgren, M.G., 2014. Many rivers to cross: the journey of zinc from soil to seed. *Front Plant Sci*, *5*, pp.79-84.

Ortiz-Monasterio, J.I., Palacios-Rojas, N., Meng, E., Pixley, K., Trethowan, R. and Pena, R.J., 2007. Enhancing the mineral and vitamin content of wheat and maize through plant breeding. *Journal of Cereal Science*, *46*(3), pp.293-307.

Osabe, D., Tanahashi, T., Nomura, K., Shinohara, S., Nakamura, N., Yoshikawa, T., Shiota, H., Keshavarz, P., Yamaguchi, Y., Kunika, K. and Moritani, M., 2007. Evaluation of sample size effect on the identification of haplotype blocks. *BMC bioinformatics*, *8*(1), p.200.

Oshodi, HN Ogungbenle, MO Oladimeji, A.A., 1999. Chemical composition, nutritionally valuable minerals and functional properties of benniseed (Sesamum radiatum), pearl millet (Pennisetum typhoides) and quinoa (Chenopodium quinoa) flours. *International journal of food sciences and nutrition*, *50*(5), pp.325-331.

Oumar, I., Mariac, C., Pham, J.L. and Vigouroux, Y., 2008. Phylogeny and origin of pearl millet (Pennisetum glaucum [L.] R. Br) as revealed by microsatellite loci. *Theoretical and Applied Genetics*, *117*(4), pp.489-497.

Pahlavani, M.H. and Abolhasani, K., 2006. Xenia effect on seed and embryo size in cotton (Gossypium hirsutum L.). *Journal of applied genetics*, *47*(4), pp.331-335.

Palmgren, M.G., Clemens, S., Williams, L.E., Krämer, U., Borg, S., Schjørring, J.K. and Sanders, D., 2008. Zinc biofortification of cereals: problems and solutions. *Trends in plant science*, *13*(9), pp.464-473.

Panoutsopoulou, K. and Zeggini, E., 2009. Finding common susceptibility variants for complex disease: past, present and future. *Briefings in Functional Genomics and Proteomics*, *8*(5), pp.345-352.

Patil, J.V. ed., 2016. *Millets and Sorghum: Biology and Genetic Improvement*. John Wiley & Sons

Patterson, C., 2002. A new gun in town: the U box is a ubiquitin ligase domain. *Science Signaling*, *2002*(116), pp.pe4-pe4.

Peiter, E., Maathuis, F.J., Mills, L.N. and Knight, H., 2005. The vacuolar Ca2+-activated channel TPC1 regulates germination and stomatal movement. *Nature*, *434*(7031), p.404.

Peleg, Z., Cakmak, I., Ozturk, L., Yazici, A., Jun, Y., Budak, H., Korol, A.B., Fahima, T. and Saranga, Y., 2009. Quantitative trait loci conferring grain mineral nutrient concentrations in durum wheat× wild emmer wheat RIL population. *Theoretical and Applied Genetics*, *119*(2), pp.353-369.

Persson, D.P., Hansen, T.H., Laursen, K.H., Schjoerring, J.K. and Husted, S., 2009. Simultaneous iron, zinc, sulfur and phosphorus speciation analysis of barley grain tissues using SEC-ICP-MS and IP-ICP-MS. *Metallomics*, *1*(5), pp.418-426.

Pfeiffer, W., Andersson, M., Govindaraj, M., Parminder, V., Cherian, B., Illona, P., Magezi, S. and Mulambu, J., 2018. Biofortification in Underutilized Staple Crops for Nutrition in Asia and Africa.

Philpott, C.C., 2014. Pumping iron. *Elife*, *3*, p.e03997.

Pii, Y., Borruso, L., Brusetti, L., Crecchio, C., Cesco, S. and Mimmo, T., 2016. The interaction between iron nutrition, plant species and soil type shapes the rhizosphere microbiome. *Plant Physiology and Biochemistry*, *99*, pp.39-48.

Pitt, S.J., Lam, A.K., Rietdorf, K., Galione, A. and Sitsapesan, R., 2014. Reconstituted human TPC1 is a proton-permeable ion channel and is activated by NAADP or Ca2+. *Science signaling*, *7*(326), pp.ra46-ra46.

Pletsch-Rivera, L.A. and Kaeppler, S.M., 2007. Phosphorus accumulation in maize grain is not influenced by xenia (Zea mays L.). *Maydica*, *52*(2), p.151.

Poland, J.A. and Rife, T.W., 2012. Genotyping-by-sequencing for plant breeding and genetics. *The Plant Genome*, *5*(3), pp.92-102.

Porras-Hurtado, L., Ruiz, Y., Santos, C., Phillips, C., Carracedo, Á. and Lareu, M.V., 2013. An overview of STRUCTURE: applications, parameter settings, and supporting software. *Frontiers in genetics*, *4*.

Preger, V., Tango, N., Marchand, C., Lemaire, S.D., Carbonera, D., Di Valentin, M., Costa, A., Pupillo, P. and Trost, P., 2009. Auxin-responsive genes AIR12 code for a new family of plasma membrane b-type cytochromes specific to flowering plants. *Plant physiology*, *150*(2), pp.606-620.

Prentice, A.M., Mendoza, Y.A., Pereira, D., Cerami, C., Wegmuller, R., Constable, A. and Spieldenner, J., 2017. Dietary strategies for improving iron status: balancing safety and efficacy. *Nutrition Reviews*, *75*(1), pp.49-60.

Primikyri, A., Mazzone, G., Lekka, C., Tzakos, A.G., Russo, N. and Gerothanassis, I.P., 2014. Understanding zinc (II) chelation with quercetin and luteolin: a combined NMR and theoretical study. *The Journal of Physical Chemistry B*, *119*(1), pp.83-95.

Primm, T.P. and Gilbert, H.F., 2001. Hormone binding by protein disulfide isomerase, a high capacity hormone reservoir of the endoplasmic reticulum. *Journal of Biological Chemistry*, *276*(1), pp.281-286.

Pritchard, J.K., Stephens, M. and Donnelly, P., 2000. Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), pp.945-959.

Qu, L.Q., Yoshihara, T., Ooyama, A., Goto, F. and Takaiwa, F., 2005. Iron accumulation does not parallel the high expression level of ferritin in transgenic rice seeds. *Planta*, *222*(2), pp.225-233.

Raboy, V., Below, F.E. and Dickinson, D.B., 1989. Recurrent selection for maize kernel protein and oil has altered phytic acid levels. *J Heredity*, *80*, pp.311-315.

Raboy, V., Dickinson, D.B. and Below, F.E., 1984. Variation in seed total phosphorus, phytic acid, zinc, calcium, magnesium, and protein among lines of Glycine max and G. soja. *Crop Science*, *24*(3), pp.431-434.

Ragaee, S., Abdel-Aal, E.S.M. and Noaman, M., 2006. Antioxidant activity and nutrient composition of selected cereals for food use. *Food Chemistry*, *98*(1), pp.32-38.

Rai, K.N., Govindaraj, M., Pfeiffer, W.H. and Rao, A.S., 2015. Seed set and xenia effects on grain iron and zinc density in pearl millet. *Crop Science*, *55*(2), pp.821-827.

Rai, K.N., Kumar, K.A., Andrews, D.J., Rao, A.S., Raj, A.G.B. and Witcombe, J.R., 1990. Registration of'ICTP 8203'pearl millet. *Crop Science*, *30*(4), pp.959-959.

Rai, K.N., Yadav, O.P., Rajpurohit, B.S., Patil, H.T., Govindaraj, M., Khairwal, I.S. and Rao, A.S., 2013. Breeding pearl millet cultivars for high iron density with zinc density as an associated trait. *Journal of SAT Agricultural Research*, *11*, pp.1-7.

Ramirez-Wong, B., Sweat, V.E., Torres, P.I. and Rooney, L.W., 1994. Cooking time, grinding, and moisture content effect on fresh corn masa texture. *Cereal chemistry (USA)*.

Rani, S., Singh, R., Sehrawat, R., Kaur, B.P. and Upadhyay, A., 2018. Pearl millet processing: a review. *Nutrition & Food Science*, (just-accepted), pp.00-00.

Rao, N.K. and Bramel, P.J., 2000. *Manual of genebank operations and procedures*. International Crops Research Institute for the Semi-Arid Tropics.

Rao, P.P., Birthal, P.S., Reddy, B.V., Rai, K.N. and Ramesh, S., 2006. Diagnostics of sorghum and pearl millet grains-based nutrition in India. *International Sorghum and Millets Newsletter*, *47*, pp.93-96.

Rasmussen, S.K. and Hatzack, F., 1998. Identification of two Low-Phytate Barley (Hordeum Vulgare l.) Grain Mutants by TLC and Genetic Analysis. *Hereditas*, *129*(2), pp.107-112.

Rawat, N., Neelam, K., Tiwari, V.K. and Dhaliwal, H.S., 2013. Biofortification of cereals to overcome hidden hunger. *Plant Breeding*, *132*(5), pp.437-445.

Ray, S., Mondal, W.A. and Choudhuri, M.A., 1983. Regulation of leaf senescence, grain-filling and yield of rice by kinetin and abscisic acid. *Physiologia Plantarum*, *59*(3), pp.343-346.

Reason, D.A., Watts, M.J. and Devez, A., 2015. Quantification of phytic acid in grains.

Reddy, A.S., Ali, G.S., Celesnik, H. and Day, I.S., 2011. Coping with stresses: roles of calcium-and calcium/calmodulin-regulated gene expression. *The Plant Cell*, *23*(6), pp.2010-2032.

Reddy, B.V., Ramesh, S. and Longvah, T., 2005. Prospects of breeding for micronutrients and b-carotene-dense sorghums. *International Sorghum and Millets Newsletter*, *46*, pp.10-14.

Reddy, N.R., Pierson, M.D., Sathe, S.K. and Salunkhe, D.K., 1989. *Phytates in cereals and legumes*. CRC Press.

Reddy, N.R., Sathe, S.K. and Salunkhe, D.K., 1982. Phytates in legumes and cereals. *Advances in food research*, *28*, pp.1-92.

Reusch, W. (2013). Nuclear Magnetic Resonance Spectroscopy. Available: https://www2.chemistry.msu.edu/faculty/reusch/virttxtjml/spectrpy/nmr/nmr1.htm. Last accessed 09/05/2018.

Ricachenevsky, F.K., Menguer, P.K. and Sperotto, R.A., 2013. kNACking on heaven's door: how important are NAC transcription factors for leaf senescence and Fe/Zn remobilization to seeds?. *Frontiers in plant science*, *4*, p.226.

Ricachenevsky, F.K., Menguer, P.K., Sperotto, R.A., Williams, L.E. and Fett, J.P., 2013. Roles of plant metal tolerance proteins (MTP) in metal storage and potential use in biofortification strategies. *Frontiers in plant science*, *4*, p.144.

Riveras, E., Alvarez, J.M., Vidal, E.A., Oses, C., Vega, A. and Gutiérrez, R.A., 2015. The calcium ion is a second messenger in the nitrate signaling pathway of Arabidopsis. *Plant physiology*, *169*(2), pp.1397-1404.

Ross, J. A., & Kasum, C. M. (2002). Dietary flavonoids: bioavailability, metabolic effects, and safety. *Annual review of Nutrition*, *22*(1), 19-34.

Ross, J., Li, Y., Lim, E.K. and Bowles, D.J., 2001. Higher plant glycosyltransferases. *Genome Biology*, *2*(2), pp.reviews3004-1.

Rossander-Hultén, L., Brune, M., Sandström, B., Lönnerdal, B. and Hallberg, L., 1991. Competitive inhibition of iron absorption by manganese and zinc in humans. *The American journal of clinical nutrition*, *54*(1), pp.152-156.

RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/.

Saltzman, A., Birol, E., Bouis, H.E., Boy, E., De Moura, F.F., Islam, Y. and Pfeiffer, W.H., 2013. Biofortification: progress toward a more nourishing future. *Global Food Security*, *2*(1), pp.9-17.

Salunkhe, D. K., Jadhav, S. J., Kadam, S. S., Chavan, J. K., & Luh, B. S. (1983). Chemical, biochemical, and biological significance of polyphenols in cereals and legumes. *Critical Reviews in Food Science & Nutrition*, *17*(3), 277-305.

Samia, M., AbdelRahaman, B. and Elfadil, E., 2005. Effect of malt pretreatment followed by fermentation on antinutritional factors and HCl extractability of minerals of pearl millet cultivars. *Journal of food technology*, *3*(4), pp.529-534.

Sampedro, J. and Cosgrove, D.J., 2005. The expansin superfamily. *Genome biology*, *6*(12), p.242.

Sandberg, A.S. and Andlid, T., 2002. Phytogenic and microbial phytases in human nutrition. *International journal of food science & technology*, *37*(7), pp.823-833.

Sanders, D., Pelloux, J., Brownlee, C. and Harper, J.F., 2002. Calcium at the crossroads of signaling. *The Plant Cell*, *14*(suppl 1), pp.S401-S417.

Sandström, B., Davidsson, L., Cederblad, Å. and Lönnerdal, B., 1985. Oral iron, dietary ligands and zinc absorption. *The Journal of nutrition*, *115*(3), pp.411-414.

Scalbert, A., Morand, C., Manach, C. and Rémésy, C., 2002. Absorption and metabolism of polyphenols in the gut and impact on health. *Biomedicine & Pharmacotherapy*, *56*(6), pp.276-282.

Schmidhuber, J. and Tubiello, F.N., 2007. Global food security under climate change. *Proceedings of the National Academy of Sciences*, *104*(50), pp.19703-19708.

Sehgal, D., Skot, L., Singh, R., Srivastava, R.K., Das, S.P., Taunk, J., Sharma, P.C., Pal, R., Raj, B., Hash, C.T. and Yadav, R.S., 2015. Exploring potential of pearl millet germplasm association panel for association mapping of drought tolerance traits. *PloS one*, *10*(5), p.e0122165.

Senthilvel, S., Jayashree, B., Mahalakshmi, V., Kumar, P.S., Nakka, S., Nepolean, T. and Hash, C.T., 2008. Development and mapping of simple sequence repeat markers for pearl millet from data mining of expressed sequence tags. *BMC Plant Biology*, *8*(1), p.119.

Seo, P.J., Lee, S.B., Suh, M.C., Park, M.J., Go, Y.S. and Park, C.M., 2011. The MYB96 transcription factor regulates cuticular wax biosynthesis under drought conditions in Arabidopsis. *The Plant Cell*, *23*(3), pp.1138-1152.

Shanmuganathan, M., Gopalan, A. and Mohanraj, K., 2006. Genetic analysis of pearl millet for phytic acid content. *Journal of Agricultural Sciences*, *2*(2).

Sharma, A. and Kapoor, A.C., 1996. Levels of antinutritional factors in pearl millet as affected by processing treatments and various types of fermentation. *Plant Foods for Human Nutrition*, *49*(3), pp.241-252.

Sharma, C.B., Goel, M. and Irshad, M., 1978. Myoinositol hexaphosphate as a potential inhibitor of α-amylases. *Phytochemistry*, *17*(2), pp.201-204.

Shekhar, Hossain Uddin, Howlader, Zakir Hossain, Kabir, Yearul (2016). Exploring the Nutrition and Health Benefits of Functional Foods. IGI Global. 278.

Sheu, J.R., Hsiao, G., Shen, M.Y., Chou, C.Y., Lin, C.H., Chen, T.F. and Chou, D.S., 2003. Inhibitory mechanisms of kinetin, a plant growth-promoting hormone, in platelet aggregation. *Platelets*, *14*(3), pp.189-196.

Shimizu, A., Guerta, C.Q., Gregorio, G.B., Kawasaki, S. and Ikehashi, H., 2005. QTLs for nutritional contents of rice seedlings (Oryza sativa L.) in solution cultures and its implication to tolerance to iron-toxicity. *Plant and Soil*, *275*(1), pp.57-66.

Shimura, K., Okada, A., Okada, K., Jikumaru, Y., Ko, K.W., Toyomasu, T., Sassa, T., Hasegawa, M., Kodama, O., Shibuya, N. and Koga, J., 2007. Identification of a biosynthetic gene cluster in rice for momilactones. *Journal of Biological Chemistry*, *282*(47), pp.34013-34018.

Simwemba, C.G., Hoseney, R.C., Varriano-Marston, E. and Zeleznak, K., 1984. Certain B vitamin and phytic acid contents of pearl millet [Pennisetum americanum (L.) Leeke]. *Journal of agricultural and food chemistry*, *32*(1), pp.31-34.

Singh, P. and Raghuvanshi, R.S., 2012. Finger millet for food and nutritional security. *African Journal of Food Science*, *6*(4), pp.77-84.

Singh, R., Singh, D.P. and Tyagi, P.K., 2003. Effect of Azotobacter, farmyard manure and nitrogen fertilization on productivity of pearl millet hybrids (Pennisetum glaucum (l) r. br) in semi-arid tropical environment. *Archives of Agronomy and Soil Science*, *49*(1), pp.21-24

Slatkin, M., 2008. Linkage disequilibrium—understanding the evolutionary past and mapping the medical future. *Nature reviews. Genetics*, *9*(6), p.477.

Snyder, L.R., Kirkland, J.J. and Glajch, J.L., 2012. *Practical HPLC method development*. John Wiley & Sons.

Soetan KO, Oyekunle MA, Aiyelaagbe OO, Fafunso MA (2006). Evaluation of the antimicrobial activity of saponins extract of Sorghum bicolor L. Moench. *African journal of Biotechnology*. 5(23): 2405-2407.

Soetan, K. O. (2008). Pharmacological and other beneficial effects of antinutritional factors in plants-A review. *African journal of Biotechnology*,*7*(25).

Solomons, N.W. and Jacob, R.A., 1981. Studies on the bioavailability of zinc in humans: effects of heme and nonheme iron on the absorption of zinc. *The American journal of clinical nutrition*, *34*(4), pp.475-482.

Sperotto, R.A., Ricachenevsky, F.K. and Fett, J.P., 2007. Iron deficiency in rice shoots: identification of novel induced genes using RDA and possible relation to leaf senescence. *Plant cell reports*, *26*(8), pp.1399-1411.

Sperotto, R.A., Ricachenevsky, F.K., Williams, L.E., Vasconcelos, M.W. and Menguer, P.K. eds., 2014. *From soil to seed: micronutrient movement into and within the plant*. Frontiers E-books.

Stangoulis, J.C., Huynh, B.L., Welch, R.M., Choi, E.Y. and Graham, R.D., 2007. Quantitative trait loci for phytate in rice grain and their relationship with grain micronutrient content. *Euphytica*, *154*(3), pp.289-294.

Steele, K.A., Price, A.H., Shashidhar, H.E. and Witcombe, J.R., 2006. Marker-assisted selection to introgress rice QTLs controlling root traits into an Indian upland rice variety. *Theoretical and Applied Genetics*, *112*(2), pp.208-221.

Stich, B., Haussmann, B.I., Pasam, R., Bhosale, S., Hash, C.T., Melchinger, A.E. and Parzies, H.K., 2010. Patterns of molecular and phenotypic diversity in pearl millet [Pennisetum glaucum (L.) R. Br.] from West and Central Africa and their relation to geographical and environmental parameters. *BMC plant biology*, *10*(1), p.216.

Stone, S.L., Hauksdóttir, H., Troy, A., Herschleb, J., Kraft, E. and Callis, J., 2005. Functional analysis of the RING-type ubiquitin ligase family of Arabidopsis. *Plant physiology*, *137*(1), pp.13-30.

Stringer, S., Wray, N.R., Kahn, R.S. and Derks, E.M., 2011. Underestimated effect sizes in GWAS: fundamental limitations of single SNP analysis for dichotomous phenotypes. *PLoS One*, *6*(11), p.e27964.

Stuart, B., 2005. *Infrared spectroscopy*. John Wiley & Sons, Inc..

Sun, Z., Wang, X., Liu, Z., Gu, Q., Zhang, Y., Li, Z., Ke, H., Yang, J., Wu, J., Wu, L. and Zhang, G., 2017. Genome-wide association study discovered genetic variation and candidate genes of fibre quality traits in Gossypium hirsutum L. Plant biotechnology journal, 15(8), pp.982-996.

Symonowicz, M. and Kolanek, M., 2012. Flavonoids and their properties to form chelate complexes.

Takahashi, S., Yeo, Y.S., Zhao, Y., O'Maille, P.E., Greenhagen, B.T., Noel, J.P., Coates, R.M. and Chappell, J., 2007. Functional characterization of premnaspirodiene oxygenase, a cytochrome P450 catalyzing regio-and stereo-specific hydroxylations of diverse sesquiterpene substrates. *Journal of Biological Chemistry*, *282*(43), pp.31744-31754.

Tako, E., Beebe, S. E., Reed, S., Hart, J. J., & Glahn, R. P. (2014). Polyphenolic compounds appear to limit the nutritional benefit of biofortified higher iron black bean (Phaseolus vulgaris L.). *Nutrition journal*, *13*(1), 1.

Tako, E., Reed, S. M., Budiman, J., Hart, J. J., & Glahn, R. P. (2015). Higher iron pearl millet (Pennisetum glaucum L.) provides more absorbable iron that is limited by increased polyphenolic content. *Nutrition journal*, *14*(1), 11.

Tedman-Jones, J.D., Lei, R., Jay, F., Fabro, G., Li, X., Reiter, W.D., Brearley, C. and Jones, J.D., 2008. Characterization of Arabidopsis mur3 mutations that result in constitutive activation of defence in petioles, but not leaves. *The Plant Journal*, *56*(5), pp.691-703.

Torheim, L.E., Ferguson, E.L., Penrose, K. and Arimond, M., 2010. Women in resource-poor settings are at risk of inadequate intakes of multiple micronutrients. *The Journal of nutrition*, *140*(11), pp.2051S-2058S.

Torii, K.U., Mitsukawa, N., Oosumi, T., Matsuura, Y., Yokoyama, R., Whittier, R.F. and Komeda, Y., 1996. The Arabidopsis ERECTA gene encodes a putative receptor protein kinase with extracellular leucine-rich repeats. *The Plant Cell*, *8*(4), pp.735-746.

Turnlund, J.R., King, J.C., Keyes, W.R., Gong, B. and Michel, M.C., 1984. A stable isotope study of zinc absorption in young men: effects of phytate and alpha-cellulose. *The American journal of clinical nutrition*, *40*(5), pp.1071-1077.

Uehlein, N., Lovisolo, C., Siefritz, F. and Kaldenhoff, R., 2003. The tobacco aquaporin NtAQP1 is a membrane CO2 pore with physiological functions. *Nature*, *425*(6959), pp.734-737.

Ullah, A., Ahmad, A., Khaliq, T. and Akhtar, J., 2016. Recognizing production options for pearl millet in Pakistan under changing climate scenarios.

Uno, Y., Furihata, T., Abe, H., Yoshida, R., Shinozaki, K. and Yamaguchi-Shinozaki, K., 2000. Arabidopsis basic leucine zipper transcription factors involved in an abscisic acid-dependent signal transduction pathway under drought and high-salinity conditions. *Proceedings of the National Academy of Sciences*, *97*(21), pp.11632-11637.

Upadhyaya, H.D., Bajaj, D., Das, S., Kumar, V., Gowda, C.L.L., Sharma, S., Tyagi, A.K. and Parida, S.K., 2016. Genetic dissection of seed-iron and zinc concentrations in chickpea. *Scientific reports*, *6*.

Upadhyaya, H.D., Pundir, R.P.S., Dwivedi, S.L. and Gowda, C.L.L., 2009. *Mini Core Collections for Efficient Utilization of Plant Genetic Resources in Crop Improvement Programs. Information Bulletin No. 78*. International Crops Research Institute for the Semi-Arid Tropics.

Urbano, G., Lopez-Jurado, M., Aranda, P., Vidal-Valverde, C., Tenorio, E. and Porres, J., 2000. The role of phytic acid in legumes: antinutrient or beneficial function?. *Journal of physiology and biochemistry*, *56*(3), pp.283-294.

Van, A.E. and Young, A.E., 2014. Prevalence and impacts of genetically engineered feedstuffs on livestock populations. *Journal of animal science*, *92*(10), pp.4255-4278.

Varshney, R.K., Hoisington, D.A. and Tyagi, A.K., 2006. Advances in cereal genomics and applications in crop breeding. *Trends in biotechnology*, *24*(11), pp.490-499.

Varshney, R.K., Ribaut, J.M., Buckler, E.S., Tuberosa, R., Rafalski, J.A. and Langridge, P., 2012. Can genomics boost productivity of orphan crops?. *Nature Biotechnology*, *30*(12), pp.1172-1176.

Varshney, R.K., Shi, C., Thudi, M., Mariac, C., Wallace, J., Qi, P., Zhang, H., Zhao, Y., Wang, X., Rathore, A. and Srivastava, R.K., 2017. Pearl millet genome sequence provides a resource to improve agronomic traits in arid environments. *Nature Biotechnology*, pp.1-13.

Vasil, V. and Vasil, I.K., 1981. Somatic embryogenesis and plant regeneration from tissue cultures of Pennisetum americanum, and P. americanum x P. purpureum hybrid. *American Journal of Botany*, pp.864-872.

Velu, G., Bhattacharjee, R., Rai, K.N., Sahrawat, K.L. and Longvah, T., 2008. A simple and rapid screening method for grain zinc content in pearl millet. Journal of SAT Agricultural Research, 6, pp.1-4.

Velu, G., Rai, K.N., Muralidharan, V., Kulkarni, V.N., Longvah, T. and Raveendran, T.S., 2007. Prospects of breeding biofortified pearl millet with high grain iron and zinc content. *Plant Breeding*, *126*(2), pp.182-185.

Velu, G., Rai, K.N., Muralidharan, V., Longvah, T. and Crossa, J., 2011. Gene effects and heterosis for grain iron and zinc density in pearl millet (Pennisetum glaucum (L.) R. Br). *Euphytica*, *180*(2), pp.251-259.

Viana, J.M.S., Mundim, G.B., Silva, F.F. and Garcia, A.A.F., 2016. Efficiency of genome-wide association study in open-pollinated populations. *bioRxiv*, p.050955.
Vom Brocke, K., Christinck, A., Weltzien, R.E., Presterl, T. and Geiger, H.H., 2003. Farmers' seed systems and management practices determine pearl millet genetic diversity patterns in semiarid regions of India. *Crop Science*, *43*(5), pp.1680-1689.

Voorrips, R.E., 2002. MapChart: software for the graphical presentation of linkage maps and QTLs. *Journal of heredity*, *93*(1), pp.77-78.

Vos, P.G., Paulo, M.J., Voorrips, R.E., Visser, R.G., van Eck, H.J. and van Eeuwijk, F.A., 2017. Evaluation of LD decay and various LD-decay estimators in simulated and SNP-array data of tetraploid potato. *Theoretical and Applied Genetics*, *130*(1), pp.123-135.

Vreugdenhil, D., Aarts, M.G.M., Koornneef, M., Nelissen, H. and Ernst, W.H.O., 2004. Natural variation and QTL analysis for cationic mineral content in seeds of Arabidopsis thaliana. *Plant, Cell & Environment*, *27*(7), pp.828-839.

Vukics, V., & Guttman, A. (2010). Structural characterization of flavonoid glycosides by multi-stage mass spectrometry. *Mass Spectrometry Reviews*,*29*(1), 1-16.

Walker, D.R., Scaboo, A.M., Pantalone, V.R., Wilcox, J.R. and Boerma, H.R., 2006. Genetic mapping of loci associated with seed phytic acid content in CX1834-1-2 soybean. *Crop science*, *46*(1), pp.390-397.

Wang, H., Qi, Q., Schorr, P., Cutler, A.J., Crosby, W.L. and Fowke, L.C., 1998. ICK1, a cyclin-dependent protein kinase inhibitor fromArabidopsis thalianainteracts with both Cdc2a and CycD3, and its expression is induced by abscisic acid. The Plant Journal, 15(4), pp.501-510.

Wang, H., Zhou, Y., Gilmer, S., Whitwill, S. and Fowke, L.C., 2000. Expression of the plant cyclin-dependent kinase inhibitor ICK1 affects cell division, plant growth and morphology. *The Plant Journal*, *24*(5), pp.613-623.

Wang, Z., Zhao, X., Wang, B., Liu, E., Chen, N., Zhang, W. and Liu, H., 2016. Overexpression of an Arabidopsis heterogeneous nuclear ribonucleoprotein gene, AtRNP1, affects plant growth and reduces plant tolerance to drought and salt stresses. *Biochemical and biophysical research communications*, *472*(2), pp.353-359.

Welch, R.M. and Graham, R.D., 2004. Breeding for micronutrients in staple food crops from a human nutrition perspective. *Journal of experimental botany*, *55*(396), pp.353-364.

White, P.J. and Broadley, M.R., 2009. Biofortification of crops with seven mineral elements often lacking in human diets–iron, zinc, copper, calcium, magnesium, selenium and iodine. *New Phytologist*, *182*(1), pp.49-84.

White, P.J. and Broadley, M.R., 2011. Physiological limits to zinc biofortification of edible crops. *Frontiers in plant science*, *2*, p.80.

Williamson, G., & Manach, C. (2005). Bioavailability and bioefficacy of polyphenols in humans. II. Review of 93 intervention studies. *The American journal of clinical nutrition*, *81*(1), 243S-255S.

Winkel-Shirley, B., 2001. Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant physiology*, *126*(2), pp.485-493.

World Health Organization, 2001, Geneva, 2001, Iron deficiency anaemia assessment, prevention, and control: a guide for programme managers

World Health Organization, 2008. Worldwide prevalence of anaemia 1993-2005: WHO global database on anaemia.

World Health Organization, Food and Agricultural Organization of the United Nations (2006). Guidelines on Food Fortification with Micronutrients. Available at: http://www.who.int/nutrition/publications/micronutrients/9241594012/en/ (accessed December 15, 2016).

Wu, J., Li, L.T., Li, M., Khan, M.A., Li, X.G., Chen, H., Yin, H. and Zhang, S.L., 2014. High-density genetic linkage map construction and identification of fruit-related QTLs in pear using SNP and SSR markers. *Journal of experimental botany*, *65*(20), pp.5771-5781.

Wu, J., Yuan, Y.X., Zhang, X.W., Zhao, J., Song, X., Li, Y., Li, X., Sun, R., Koornneef, M., Aarts, M.G. and Wang, X.W., 2008. Mapping QTLs for mineral accumulation and shoot dry biomass under different Zn nutritional conditions in Chinese cabbage (Brassica rapa L. ssp. pekinensis). *Plant and Soil*, *310*(1-2), pp.25-40.

Wu, Y., Fan, H., Wang, Y., Zhang, L., Gao, X., Chen, Y., Li, J., Ren, H. and Gao, H., 2014. Genome-wide association studies using haplotypes and individual SNPs in Simmental cattle. *PloS one*, *9*(10), p.e109330.

Xu X, Harris KS, Wang HJ, Murphy PA, Hendrich S. 1995. Bioavailability of soybean isoflavones depends upon gut microflora in women. *Journal of Nutrition*. 125:2307–15

Xu, X. (1995). Human bioavailability and health protective effects of soy isoflavones.

Yadav, O.P. and Rai, K.N., 2013. Genetic improvement of pearl millet in India. *Agricultural Research*, *2*(4), pp.275-292.

Yadav, R.S., Sehgal, D. and Vadez, V., 2010. Using genetic mapping and genomics approaches in understanding and improving drought tolerance in pearl millet. *Journal of experimental botany*, *62*(2), pp.397-408.

Yang, A.H., Shi, X.Y., Li, X., Li, F.F., Zhang, Q.Q., Jiang, S.X., Cui, J.Z. and Gao, H.L., 2014. Spectroscopic and electrochemical studies on the evaluation of the radical scavenging activities of luteolin by chelating iron. *RSC Advances*, *4*(48), pp.25227-25233.

You J, Wang X, Yan Y, Jin F, Huang B (1993). Effects of active constituents of Chinese herbal medicine on HMG-CA Reductase. *Chemical Abstracts*. 120(7): 70.

Zhang, J., Chen, K., Pang, Y., Naveed, S.A., Zhao, X., Wang, X., Wang, Y., Dingkuhn, M., Pasuquin, J., Li, Z. and Xu, J., 2017. QTL mapping and candidate gene analysis of ferrous iron and zinc toxicity tolerance at seedling stage in rice by genome-wide association study. *BMC genomics*, *18*(1), p.828.

Zhang, G., Liu, X., Quan, Z., Cheng, S., Xu, X., Pan, S., Xie, M., Zeng, P., Yue, Z., Wang, W. and Tao, Y., 2012. Genome sequence of foxtail millet (Setaria italica) provides insights into grass evolution and biofuel potential. Nature biotechnology, 30(6), p.549.

Živanović, T., Branković, G., Zorić, M., Momirović, G.Š., Janković, S., Vasiljević, S. and Pavlov, J., 2012. Effect of recombination in the maize breeding population with exotic germplasm on the yield stability. *Euphytica*, *185*(3), pp.407-417.

Zollman, S., Godt, D., Prive, G.G., Couderc, J.L. and Laski, F.A., 1994. The BTB domain, found primarily in zinc finger proteins, defines an evolutionarily conserved family that includes several developmentally regulated genes in Drosophila. *Proceedings of the National Academy of Sciences*, *91*(22), pp.10717-10721.

Chapter 10: Appendices

Table 10.1: PMiGAP passport data.

Entry	Genotype	Origin
1001	IP 10820	Sudan
1002	IP 10964	Kenya
1006	IP 9407	Ghana
1007	IP 20349	Yemen
1008	IP 11229	Zimbabwe
1009	IP 13370	Tanzania
1010	IP 11353	Burkina Faso
1012	IP 18157	Mali
1013	IP 13964	Zimbabwe
1014	IP 4965	Uganda
1015	ICMB 90111-P6	ICRISAT- Patencheru
1016	IP 17690	Togo
1017	IP 6101	Niger
1018	IP 10488	Zimbabwe
1019	IP 13149	Niger
1020	IP 15533	Burkina Faso
1021	IP 10140	Mali
1022	IP 12925	Unknown
1023	IP 3890	India
1024	IP 11677	Sudan
1025	IP 11984	Nigeria
1026	IP 3175	India
1027	W 504-1-P1	India
1028	IP 22455	ICRISAT- Patencheru
1029	IP 16096	India
1030	IP 6060	Central African Republic
1031	IP 7633	India
1032	IP 19386	Namibia
1033	IP 6112	Niger
1035	IP 3757	India
1036	IP 18147	Pakistan
1037	IP 19448	Namibia
1038	ICMS 7703	ICRISAT- Patencheru
1039	IP 11577	Burkina Faso
1040	GB 8735	ICRISAT- Patencheru
1042	OKASHANA (ICMV 88908)	ICRISAT- Patencheru
1043	IP 9446	Ghana
1044	IP 21155	USA
1045	IP 11211	India
1046	IP6745	Malawi
1048	IP 10705	Mali

1049	GICKV 93191 (=ICMP 93191)	India
1050	AIMP 92901	India
1051	IP 18132	Pakistan
1052	PRLT 2/89-33	ICRISAT- Patencheru
1053	IP4020	India
1054	IP 13971	Zimbabwe
1055	IP 8767	Botswana
1057	863B-P2	ICRISAT- Patencheru
1058	ICTP 8203	ICRISAT- Patencheru
1059	IP4542	India
1060	IP17720	Togo
1061	H 77/833-2-P5 (NT)	ICRISAT- Patencheru
1062	843B	ICRISAT- Patencheru
2001	IP 6102	Niger
2002	IP 3616	India
2003	IP 18389	Namibia
2004	IP 9824	Mozambique
2005	IP 18293-P152	Unknown
2006	IP 22419	ICRISAT- Patencheru
2007	IP 11765	South Africa
2008	IP 6869	Kenya
2010	IP 10394	India
2011	IP 21517	Niger
2012	ICMV-IS 92222	ICRISAT- Patencheru
2013	IP 12839	Botswana
2014	IP 3732	India
2015	IP 8210	ICRISAT- Patencheru
2016	IP 7108	India
2017	IP 5272	Niger
2018	IP 7910	Niger
2019	IP 9391	Ghana
2020	IP 3471	India
2021	IP 3636	India
2022	IP 9710	Nigeria
2023	IP 3132	India
2024	IP 6037	Central African Republic
2025	IP 5713	Nigeria
2026	IP 6146	Cameroon
2027	ICMB89111-P2	ICRISAT- Patencheru
2029	IP 6460	Mali
2030	ICMV 221= ICMV88904	ICRISAT- Patencheru
2031	IP 4979	Nigeria
2032	IP 9351	Ghana
2033	IP 10539	Senegal
2035	IP 11310	Burkina Faso

2036	IP 15344	India
2037	IP 12058	Nigeria
2038	IP 19626	Niger
2039	Tift 186	USA
2040	IP 16403	Zimbabwe
2041	IP 2058	Nigeria
2043	IP 12845	Burkina Faso
2044	IP 5695	Nigeria
2045	IP 4962	Uganda
2046	IP 15320	India
2048	IP 8280	ICRISAT- Patencheru
2050	IP 3557	India
2051	IP 10271	Nigeria
2052	IP 8761	Botswana
2053	IP 9532	Ghana
2054	IP 8955	Togo
2057	IP 8344	India
2058	IP 8275	ICRISAT- Patencheru
2059	IP 10379	Nigeria
2060	IP 7095	India
2061	IP 9406	Ghana
2062	IP 17554	Togo
2063	IP 11311	Burkina Faso
3001	IP8166	ICRISAT- Patencheru
3002	IP 9969	Zambia
3003	IP 10543	Mali
3004	IP 7930	ICRISAT- Patencheru
3005	WSIL-P8	ICRISAT- Patencheru
3006	IP 18090	Pakistan
3007	P 1449-2-P1	Unknown
3008	IP 4927	Senegal
3009	IP 6882	Kenya
3010	IP 5253	Niger
3011	IP 18500	Namibia
3012	IP 8863	Zambia
3013	IP 11346	Burkina Faso
3014	IP 8069	India
3015	IP 4828	India
3016	IP 19584	Niger
3017	IP 5207	Niger
3018	IP 11275	Burkina Faso
3019	SOSAT-C88	ICRISAT- Patencheru
3020	IP6417	Mali
3021	IP 14439	Cameroon
3022	IP 9651	Nigeria

3023	LGD1-B-10	Unknown
3024	IP 5131	Niger
3025	IP 8187	ICRISAT- Patencheru
3026	WC-C75	ICRISAT- Patencheru
3027	81-B-P6	ICRISAT- Patencheru
3028	IP 13927	Zimbabwe
3029	IP 12967	Malawi
3030	ICMP 451-P8	ICRISAT- Patencheru
3031	ICMP 85410-P7	ICRISAT- Patencheru
3032	IP 15872	Tanzania
3033	IP 9971	Zambia
3034	IP 22420	ICRISAT- Patencheru
3035	IP 8074	ICRISAT- Patencheru
3036	Tift 383	USA
3038	IP 13817	Burkina Faso
3039	IP 13180	Nigeria
3041	IP 10701	Mali
3042	IP 21169	ICRISAT- Patencheru
3043	IP 8129	ICRISAT- Patencheru
3044	ICMP 451-P6	ICRISAT- Patencheru
3045	IP 11593	Burkina Faso
3046	IP 10471	Zimbabwe
3047	IP 7952	ICRISAT- Patencheru
3048	IP 9282	Togo
3049	IP 17611	Togo
3050	IP 7886	India
3051	ICMV-IS 89305	ICRISAT- Patencheru
3052	Raj 171 (ICMV 85404=RBC-IC 9)	ICRISAT- Patencheru
3053	IP 13384	Uganda
3054	IP 13344	Sudan
3055	IP 16638	Zimbabwe
3056	IP 8972	Togo
3057	IP 7967	ICRISAT- Patencheru
3058	IP 5441	Niger
3059	IP 16289	Zimbabwe
3060	IP 6682	Malawi
3061	IP 3138	India
3062	IP 11584	Burkina Faso
3063	IP 8172	ICRISAT- Patencheru
4001	IP 11218	Zimbabwe
4002	IP 18412	Namibia
4003	IP 8174	ICRISAT- Patencheru
4004	IP 6110	Niger
4005	IP 5438	Niger
4006	IP 11378	Burkina Faso

4007	IP 13608	India
4008	IP 8647	Sudan
4009	IP 7364	Tanzania
4010	IP 10759	Sudan
4011	IP 15536	Burkina Faso
4012	IP 6125	Cameroon
4013	IP 14624	Cameroon
4014	IP 12020	Nigeria
4015	IP 13290	Senegal
4016	IP 11763	South Africa
4017	IP 13520	India
4019	IP 12128	Nigeria
4020	IP 14311	Cameroon
4021	IP 15857	Tanzania
4022	IP 12364	Nigeria
4023	IP 18292	ICRISAT- Patencheru
4024	IP 13840	Burkina Faso
4025	IP 5931	Senegal
4027	IP 11929	Sierra Leone
4028	IP 10945	Sudan
4029	IP 7941	ICRISAT- Patencheru
4030	IP 17150	Zimbabwe
4031	IP 13459	India
4032	IP 8182	ICRISAT- Patencheru
4033	IP 14849	Cameroon
4034	IP 5900	Senegal
4035	IP 19334	Namibia
4036	IP 3110	India
4037	IP 7536	India
4038	IP 12395	South Africa
4039	IP 14418	Cameroon
4040	IP 17493	Togo
4041	IP 6415	Mali
4042	IP 16120	India
4043	IP 5560	Niger
4044	IP 12138	Nigeria
4045	IP15512	Burkina Faso
4046	IP 10579	Mali
4048	IP 13324	Sudan
4049	IP 19405	Chad
4050	IP 17125	Zimbabwe
4051	IP 5031	Nigeria
4052	IP 11358	Burkina Faso
4053	IP 7660	India
4054	IP 15553	Burkina Faso

4055	IP 18168	Burkina Faso
4056	IP 13363	Tanzania
4057	IP 6310	Mali
4058	IP 20679	Nigeria
4059	IP 10339	Nigeria
4060	IP 3201	India
4061	P 310-17-B	ICRISAT- Patencheru
4062	IP 14148	Zimbabwe
4063	IP 4952	Uganda

Table 10.2: October 2014 seed multiplication trial, seed emergence per day (Section 2.4.1).

Genotype	IP. No.	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Day 9	Day 10
1001	IP 10820	1	4	6	6	7	7	7	7
1002	IP 10964	0	5	6	7	8	8	8	8
1006	IP 9407	0	9	9	9	9	9	9	9
1007	IP 20349	0	7	8	9	9	9	10	10
1008	IP 11229	3	8	9	9	9	9	9	9
1009	IP 13370	2	5	5	5	5	5	5	5
1010	IP 11353	1	2	7	9	9	9	9	9
1012	IP 18157	0	6	8	9	9	9	9	9
1013	IP 13964	0	1	1	1	1	1	1	1
1014	IP 4965	3	8	10	10	10	10	10	10
1015	ICMB 90111-P6	0	7	7	7	7	7	7	7
1016	IP 17690	4	4	9	9	9	9	9	9
1017	IP 6101	0	2	2	4	5	5	5	5
1018	IP 10488	0	3	4	4	4	5	5	5
1019	IP 13149	2	7	10	10	10	10	10	10
1020	IP 15533	3	9	9	9	9	9	9	9
1021	IP 10140	3	6	7	9	9	9	9	10
1022	IP 12925	3	5	5	5	5	5	5	7
1023	IP 3890	0	4	8	8	8	8	8	8
1024	IP 11667	0	1	1	1	1	1	1	1
1025	IP 11984	0	6	6	6	6	6	7	7
1026	IP 3175	0	6	7	8	8	8	8	8
1027	W 504-1-P1	0	6	7	7	7	7	7	7

1028	IP 22455	0	1	7	7	7	7	7	8
1029	IP 16096	0	6	8	10	10	10	10	10
1030	IP 6060	0	8	9	9	9	9	9	9
1031	IP 7633	0	3	3	3	5	6	6	6
1032	IP 19386	1	5	6	7	8	8	8	8
1033	IP 6112	1	2	4	5	6	6	6	7
1035	IP 3757	1	5	5	7	7	7	7	7
1036	IP 18147	1	4	5	7	7	7	8	9
1037	IP 19448	1	1	2	4	5	6	6	6
1038	ICMS 7703	0	9	10	10	10	10	10	10
1039	IP 11577	0	0	6	8	8	8	8	8
1040	GB 8735	3	5	8	9	9	9	9	9
1042	OKASHANA (ICMV 88908)	3	3	5	8	9	9	9	9
1043	IP 9446	3	3	5	6	6	6	6	6
1044	IP 21155	6	9	9	10	10	10	10	10
1045	IP 11211	0	2	4	4	4	4	4	4
1046	IP6745	0	2	4	5	8	9	9	9
1048	IP 10705	6	8	8	9	9	9	9	9
1049	GICKV 93191 (=ICMP 93191)	0	6	6	7	8	8	8	8
1050	AIMP 92901	3	9	9	9	9	9	9	9
1051	IP 18132	1	7	7	8	9	9	9	9
1052	PRLT 2/89-33	1	9	9	9	9	9	9	9
1053	IP4020	1	3	3	3	4	4	4	4
1054	IP 13971	1	5	6	6	6	6	6	6
1055	IP 8767	0	3	3	3	3	3	3	4
1057	863B-P2	0	3	3	3	3	3	3	3

1058	ICTP 8203	0	6	8	8	8	8	8	8
1059	IP4542	0	2	2	2	2	2	2	2
1060	IP17720	0	4	8	8	9	9	9	9
1061	H 77/833-2-P5 (NT)	0	8	10	10	10	10	10	10
1062	843B	0	7	7	7	8	8	8	8
2001	IP 6102	2	4	6	7	7	7	7	7
2002	IP 3616	1	1	4	5	5	7	7	7
2003	IP 18389	0	1	9	10	10	10	10	10
2004	IP 9824	0	4	9	9	9	9	9	9
2005	IP 18293-P152	0	4	4	4	5	5	5	5
2006	IP 22419	0	3	5	8	8	8	8	8
2007	IP 11765	3	6	10	10	10	10	10	10
2008	IP 6869	3	7	7	7	7	7	7	8
2010	IP 10394	0	2	2	2	2	2	3	3
2011	IP 21517	0	7	8	8	8	8	8	8
2012	ICMV-IS 92222	0	0	0	1	1	2	2	2
2013	IP 12839	4	8	8	8	8	8	8	8
2014	IP 3732	4	4	5	5	8	8	8	8
2015	IP 8210	0	3	6	7	8	9	9	9
2016	IP 7108	0	1	7	8	9	9	9	9
2017	IP 5272	4	6	7	7	9	9	9	9
2018	IP 7910	0	2	6	7	7	7	7	7
2019	IP 9391	0	1	1	1	5	6	6	6
2020	IP 3471	4	4	6	6	6	7	7	8
2021	IP 3636	1	3	4	5	5	5	5	6
2022	IP 9710	4	7	7	8	8	10	10	10

2023	IP 3132	0	0	3	7	7	7	8	8
2024	IP 6037	0	3	6	9	9	10	10	10
2025	IP 5713	0	1	2	2	2	2	2	2
2026	IP 6146	0	6	6	8	9	10	10	10
2027	ICMB89111-P2	1	5	8	9	8	9	9	9
2029	IP 6460	0	8	10	10	10	10	10	10
2030	ICMV 221= ICMV88904	2	8	10	10	10	10	10	10
2031	IP 4979	0	3	4	6	6	7	7	7
2032	IP 9351	2	2	6	7	7	7	7	8
2033	IP 10539	0	3	6	8	8	9	9	10
2035	IP 11310	1	7	10	10	10	10	10	10
2036	IP 15344	0	6	9	9	10	10	10	10
2037	IP 12058	5	7	8	8	9	9	9	10
2038	IP 19626	0	2	2	3	5	5	6	6
2039	Tift 186	0	5	6	6	7	7	7	7
2040	IP 16403	0	0	1	1	1	1	1	1
2041	IP 2058	0	7	7	8	8	8	8	8
2043	IP 12845	0	4	5	9	9	9	9	9
2044	IP 5695	2	9	10	10	10	10	10	10
2045	IP 4962	2	5	8	8	9	10	10	10
2046	IP 15320	0	0	0	0	0	0	0	0
2048	IP 8280	5	6	6	6	6	7	7	8
2050	IP 3557	7	10	10	10	10	10	10	10
2051	IP 10271	0	2	2	2	2	3	3	3
2052	IP 8761	0	1	2	3	3	3	3	3
2053	IP 9532	0	9	9	9	9	9	9	9

2054	IP 8955	0	4	4	8	8	8	8	8
2057	IP 8344	0	2	3	3	3	3	3	3
2058	IP 8275	0	4	4	5	6	6	6	6
2059	IP 10379	8	10	10	10	10	10	10	10
2060	IP 7095	0	1	1	7	7	7	7	7
2061	IP 9406	0	3	5	5	7	7	7	7
2062	IP 17554	0	9	10	10	10	10	10	10
2063	IP 11311	0	3	2	4	4	4	4	4
3001	IP8166	1	1	1	1	1	1	1	1
3002	IP 9969	7	7	9	9	9	9	9	9
3003	IP 10543	6	9	10	10	10	10	10	10
3004	IP 7930	0	3	4	4	4	4	4	4
3005	WSIL-P8	0	0	2	3	3	3	3	3
3006	IP 18090	0	0	0	0	0	0	0	0
3007	P 1449-2-P1	2	6	9	9	9	9	9	9
3008	IP 4927	1	3	5	6	6	6	6	6
3009	IP 6882	1	2	6	7	7	7	7	7
3010	IP 5253	1	5	9	9	9	9	9	9
3011	IP 18500	7	8	8	9	9	10	10	10
3012	IP 8863	0	2	3	3	4	4	4	4
3013	IP 11346	0	1	3	3	3	3	3	4
3014	IP 8069	0	4	8	8	9	9	9	9
3015	IP 4828	0	2	2	2	2	3	3	3
3016	IP 19584	0	5	6	6	6	7	7	7
3017	IP 5207	2	3	4	6	6	6	6	7
3018	IP 11275	2	5	6	6	7	7	7	8

3019	SOSAT-C88	0	8	10	10	10	10	10	10
3020	IP6417	0	4	8	9	10	10	10	10
3021	IP 14439	0	4	6	6	6	6	6	6
3022	IP 9651	0	2	6	6	8	8	8	8
3023	LGD1-B-10	0	3	7	8	8	8	8	8
3024	IP 5131	0	7	8	8	8	8	8	8
3025	IP 8187	2	9	10	10	10	10	10	10
3026	WC-C75	0	4	7	7	9	9	9	9
3027	81-B-P6	0	3	6	6	7	7	7	7
3028	IP 13927	0	5	9	9	10	10	10	10
3029	IP 12967	4	7	9	9	9	9	9	9
3030	ICMP 451-P8	0	3	4	4	5	5	5	5
3031	ICMP 85410-P7	0	1	6	6	6	6	6	6
3032	IP 15872	0	2	2	2	2	2	2	2
3033	IP 9971	0	0	3	5	5	5	5	5
3034	IP 22420	1	4	6	6	7	7	7	7
3035	IP 8074	1	7	8	8	9	9	9	9
3036	Tift 383	0	5	5	5	5	5	5	5
3038	IP 13817	0	5	6	7	7	7	7	7
3039	IP 13180	0	8	10	10	10	10	10	10
3041	IP 10701	3	7	10	10	10	10	10	10
3042	IP 21169	0	4	8	8	9	9	9	9
3043	IP 8129	0	6	8	8	8	8	8	8
3044	ICMP 451-P6	0	5	6	6	6	6	6	6
3045	IP 11593	0	3	6	6	6	6	6	6
3046	IP 10471	0	6	7	7	7	7	7	7

3047	IP 7952	0	5	6	6	7	7	7	7
3048	IP 9282	0	0	0	0	0	0	0	0
3049	IP 17611	0	5	7	7	8	8	8	8
3050	IP 7886	0	3	4	4	4	4	4	4
3051	ICMV-IS 89305	6	9	10	10	10	10	10	10
3052	Raj 171	0	3	4	4	9	9	9	9
3053	IP 13384	0	8	10	10	10	10	10	10
3054	IP 13344	0	1	1	1	1	1	1	1
3055	IP 16638	0	5	8	8	8	8	8	8
3056	IP 8972	0	9	10	10	10	10	10	10
3057	IP 7967	0	3	4	4	7	7	7	7
3058	IP 5441	0	3	4	6	7	7	7	7
3059	IP 16289	2	8	8	8	9	9	9	9
3060	IP 6682	2	6	10	10	10	10	10	10
3061	IP 3138	0	5	6	6	8	8	8	8
3062	IP 11584	0	6	7	7	9	9	9	9
3063	IP 8172	0	3	8	8	8	8	8	8
4001	IP 11218	0	1	3	3	4	4	4	4
4002	IP 18412	0	1	3	3	6	6	6	6
4003	IP 8174	0	2	6	6	7	7	7	7
4005	IP 5438	1	1	2	2	2	2	2	2
4006	IP 11378	3	7	8	8	8	8	8	8
4007	IP 13608	0	0	2	2	2	2	2	2
4008	IP 8647	1	5	7	7	7	7	7	7
4009	IP 7364	0	1	3	3	3	3	3	3
4010	IP 10759	0	5	9	9	9	9	9	9

4011	IP 15536	0	4	7	7	9	9	9	9
4012	IP 6125	0	1	5	5	6	6	6	6
4013	IP 14624	8	9	9	9	9	9	9	9
4014	IP 12020	0	4	6	6	9	9	9	9
4015	IP 13290	0	4	5	5	8	8	8	8
4016	IP 11763	3	8	8	8	8	8	8	8
4017	IP 13520	0	5	8	8	8	8	8	8
4019	IP 12128	0	3	6	6	8	8	8	8
4020	IP 14311	0	1	3	3	5	5	5	5
4021	IP 15857	0	5	9	9	9	9	9	9
4022	IP 12364	0	3	7	7	7	7	7	7
4023	IP 19292	1	4	7	7	7	7	7	7
4024	IP 13840	2	4	7	7	10	10	10	10
4025	IP 5931	0	4	8	8	8	8	8	8
4027	IP 11929	1	4	6	6	8	8	8	8
4028	IP 10945	0	0	0	0	0	0	0	0
4029	IP 7941	3	8	8	8	8	8	8	8
4030	IP 17150	2	5	7	7	8	8	8	8
4031	IP 13459	1	8	8	8	9	9	9	9
4032	IP 8182	0	3	6	6	8	8	8	8
4033	IP 14849	0	2	3	3	6	6	6	6
4034	IP 5900	0	0	3	3	9	9	9	9
4035	IP 19334	1	6	6	6	6	6	6	6
4036	IP 3110	0	2	4	4	6	6	6	6
4037	IP 7536	0	2	4	4	5	5	5	5
4038	IP 12395	0	4	8	8	8	8	8	8

4039	IP 14418	0	3	4	4	4	4	4	4
4040	IP 17493	1	4	9	9	10	10	10	10
4041	IP 6415	0	0	2	3	4	4	4	4
4042	IP 16120	0	3	7	7	7	7	7	7
4043	IP 5560	1	5	6	7	7	7	7	7
4044	IP 12138	0	1	2	2	5	5	5	5
4045	IP15512	0	3	7	8	8	8	8	8
4046	IP 10579	0	1	5	5	5	5	5	5
4048	IP 13324	1	3	3	3	3	3	3	3
4049	IP 19405	0	1	3	8	8	8	8	8
4050	IP 17125	0	2	6	8	8	8	8	8
4051	IP 5031	0	3	5	5	5	5	5	5
4052	IP 11358	0	3	6	6	6	6	6	6
4053	IP 7660	0	0	1	4	4	4	4	4
4054	IP 15553	0	6	9	9	9	9	9	9
4055	IP 18168	2	7	8	8	8	8	8	8
4056	IP 13363	0	7	8	8	8	8	8	8
4057	IP 6310	0	0	5	5	5	5	5	5
4058	IP 20679	0	1	6	7	7	7	7	7
4059	IP 10339	0	0	3	3	3	3	3	3
4060	IP 3201	0	1	3	3	3	3	3	3
4061	P 310-17-B	0	1	7	8	8	8	8	8
4062	IP 14148	0	4	6	7	7	7	7	7
4063	IP 4952	0	0	3	4	5	5	5	5

Table 10.3: Seed yield - October 2014, February 2015 and August 2015 growth trials (Section 2.4.2).

Genotype	SMT1 S	SMT1 OP	SMT2 S	SMT2 OP	SMT3 S	SMT3 OP	Total S	Total OP	Total S+OP
1001	0	0	0	0	11.09	0	11.09	0	11.09
1002	0	4.7	3.96	19.16	15.05	3.65	19.01	27.51	46.52
1006	0	7.56	23.44	11.5	13.03	1.09	36.47	20.15	56.62
1007	6.66	0.44	0	0	0	13.75	6.66	14.19	20.85
1008	0.13	3.76	6.63	5.48	6.8	2.64	13.56	11.88	25.44
1009	2.9	5.41	13.96	3.22	31.24	8.52	48.1	17.15	65.25
1010	0	0.9	7.14	23.59	2.84	12.24	9.98	36.73	46.71
1012	1.1	4.12	9.49	0	6.37	0	16.96	4.12	21.08
1013	0	0.38	8.8	0	8.61	2.59	17.41	2.97	20.38
1014	1.6	5.69	16.67	28.98	17.2	2.3	35.47	36.97	72.44
1015	0.86	8.48	10.45	18.64	14.85	4.15	26.16	31.27	57.43
1016	3.28	0	25.69	2.28	4.09	0	33.06	2.28	35.34
1017	0	0	3.6	0	3.94	2.04	7.54	2.04	9.58
1018	0.16	0	3.38	0	21.87	0	25.41	0	25.41
1019	3.43	20.92	26.4	5.67	4.59	0	34.42	26.59	61.01
1020	0	0	0	0	0	2.52	0	2.52	2.52
1021	2.5	13.39	3.06	0.59	1.03	0	6.59	13.98	20.57
1022	2.77	0	15.34	0	27.44	10.18	45.55	10.18	55.73
1023	0	0	0	0	15.4	12.63	15.4	12.63	28.03
1024	0	1.66	0	0	0	0	0	1.66	1.66
1025	1.05	5.08	24.33	8.17	0	0	25.38	13.25	38.63
1026	6.6	5.15	33.33	2.8	11.66	0	51.59	7.95	59.54
1027	0	1.45	4.37	3.54	10.76	10.54	15.13	15.53	30.66

1028	0	1.24	10.27	4.95	0	0	10.27	6.19	16.46
1029	0	1.47	37.36	24.58	25.05	5.22	62.41	31.27	93.68
1030	0.64	3.24	32.8	13.41	0	0	33.44	16.65	50.09
1031	0	0	19.24	2.41	0	3.15	19.24	5.56	24.8
1032	2.5	15.14	1.17	8.15	12.82	5.95	16.49	29.24	45.73
1033	0	0	0	4.39	7.79	5.13	7.79	9.52	17.31
1035	1.1	0	7.31	0.34	12.4	0	20.81	0.34	21.15
1036	0	0	0	0	17.83	0	17.83	0	17.83
1037	0	4.35	4.16	8.05	2.76	0	6.92	12.4	19.32
1038	0	0	2.1	0	24.1	4.18	26.2	4.18	30.38
1039	0	0	0	0	0	0	0	0	0
1040	0	0.42	1.27	12.74	6.66	17.24	7.93	30.4	38.33
1042	0	0	2.7	0	18.63	2.18	21.33	2.18	23.51
1043	3.48	12.58	9.7	0.28	5.47	2.77	18.65	15.63	34.28
1044	2.34	1.66	0	0	29	14.06	31.34	15.72	47.06
1045	2.31	0	0	0	11.93	0	14.24	0	14.24
1046	6.13	1.44	3.2	0	2.12	0	11.45	1.44	12.89
1048	13.37	10.78	27.6	18.93	16	22.57	56.97	52.28	109.25
1049	0.45	0	24.75	3.5	9.1	0	34.3	3.5	37.8
1050	3.16	3.82	9.95	0	22.23	0	35.34	3.82	39.16
1051	7.09	11.98	18.3	0	31.32	0	56.71	11.98	68.69
1052	2.15	4.93	12.8	0	0	0	14.95	4.93	19.88
1053	1.05	11.78	17.67	0	12.3	6.02	31.02	17.8	48.82
1054	4.75	11.1	1.78	4.9	2.07	0	8.6	16	24.6
1055	5.89	2.85	14.8	24.21	4.71	27.8	25.4	54.86	80.26
1057	4.09	17.17	15.07	0	18.95	4.69	38.11	21.86	59.97

1058	0	1.06	0	0	17.3	0	17.3	1.06	18.36
1059	6.99	0	1.95	0	7.34	0	16.28	0	16.28
1060	2.56	1.58	12.11	0	0.99	0	15.66	1.58	17.24
1061	1.12	2.45	7.08	2.62	17.03	2.42	25.23	7.49	32.72
1062	0.85	1.61	0	0	10.77	0	11.62	1.61	13.23
2001	0	0	0	0	5.15	0	5.15	0	5.15
2002	0.02	0.87	4.08	0	7.04	0	11.14	0.87	12.01
2003	0	0	2.89	0	0	0	2.89	0	2.89
2004	0	0	2.36	2.72	1.31	0	3.67	2.72	6.39
2005	0	0	1.79	10.57	4.25	0	6.04	10.57	16.61
2006	5.78	18.81	3.06	0	10.86	6.96	19.7	25.77	45.47
2007	0	0	10.35	1.95	0	0	10.35	1.95	12.3
2008	0	0.9	0	0	7.4	0	7.4	0.9	8.3
2010	0	2.79	0	0	4.18	0	4.18	2.79	6.97
2011	6.16	0	14.28	20.63	14.48	7.2	34.92	27.83	62.75
2012	0	0	1.67	10.76	3.02	0	4.69	10.76	15.45
2013	0	4.7	4.9	6.25	2.84	6.63	7.74	17.58	25.32
2014	0	18.3	12.44	9.95	7.14	0	19.58	28.25	47.83
2015	0	0	0	0	21.08	0	21.08	0	21.08
2016	0	0.82	0	0	0	0	0	0.82	0.82
2017	2.12	8.3	25.19	31.03	69.12	5.31	96.43	44.64	141.07
2018	0	0	13.4	0	8.26	3.37	21.66	3.37	25.03
2019	3.1	5.26	3.05	10.18	1.28	0	7.43	15.44	22.87
2020	0	0	2.19	0	0	2.94	2.19	2.94	5.13
2021	7.35	14.73	14.01	9.19	7.75	0	29.11	23.92	53.03
2022	5.74	25.46	9.32	0	12.32	22.48	27.38	47.94	75.32

2023	3.87	13.81	5.66	1.16	6.68	5.35	16.21	20.32	36.53
2024	3.86	12.16	14.18	6.34	13	15.93	31.04	34.43	65.47
2025	6.22	0	11.33	2.44	12.98	0	30.53	2.44	32.97
2026	0	0	0	0	0	5.48	0	5.48	5.48
2027	1.15	0.96	0.78	0	6.2	0	8.13	0.96	9.09
2029	1.05	21.88	0	0	9.64	8.37	10.69	30.25	40.94
2030	6.85	7.86	2.18	0	2.01	0	11.04	7.86	18.9
2031	0	0	0	0	0	0	0	0	0
2032	6.17	0	5.22	0	5.8	15.33	17.19	15.33	32.52
2033	4.23	11.91	32.23	3.39	18.47	0	54.93	15.3	70.23
2035	3.54	7.47	19.02	9.68	16.01	5.71	38.57	22.86	61.43
2036	0	2.64	0	0	0	0	0	2.64	2.64
2037	5.16	27.68	14.47	6.95	5.8	0	25.43	34.63	60.06
2038	2.89	3.64	9.51	19.15	27.26	14.17	39.66	36.96	76.62
2039	0	6.54	1.94	0.79	4.37	0	6.31	7.33	13.64
2040	0	1.13	0	19.84	46.1	8.87	46.1	29.84	75.94
2041	0.52	0	1.11	0	9.6	0	11.23	0	11.23
2043	0	0	0	8.22	9.07	0	9.07	8.22	17.29
2044	1.05	0.78	1.82	6.19	2.21	0	5.08	6.97	12.05
2045	0	0	0	0	0	0	0	0	0
2046	0	0	0	0	11.93	4.13	11.93	4.13	16.06
2048	1.45	7.15	2.35	0.87	1.14	6.74	4.94	14.76	19.7
2050	0	0	0	0	0	0	0	0	0
2051	0	0	0	10.23	0	0	0	10.23	10.23
2052	10.34	0	7.28	8.19	4.76	0	22.38	8.19	30.57
2053	0.65	0	3.53	0	4.13	3.25	8.31	3.25	11.56

2054	1.46	24.05	0.44	0	0	0	1.9	24.05	25.95
2057	0	2.5	0	0	0	0	0	2.5	2.5
2058	0	0	0	1.88	7.31	3.99	7.31	5.87	13.18
2059	0.53	0	7.96	5.81	11.39	3.85	19.88	9.66	29.54
2060	3.11	2.42	10.44	7.74	5.91	0	19.46	10.16	29.62
2061	0	0	1.87	0	0	0	1.87	0	1.87
2062	0.37	6.04	0	20.12	3.03	12.15	3.4	38.31	41.71
2063	1.23	4.1	3.4	14.55	6.09	0	10.72	18.65	29.37
3001	0	0	0	7.32	1.46	0	1.46	7.32	8.78
3002	0	0	0.9	7.47	5.65	0	6.55	7.47	14.02
3003	0	0.78	6.68	0	0	0	6.68	0.78	7.46
3004	0.53	4.32	5.46	11.19	5	0	10.99	15.51	26.5
3005	0	0	1.54	0.13	7.72	4.05	9.26	4.18	13.44
3006	0	0	0	0	0	0	0	0	0
3007	0	0	0	0	23.92	2.69	23.92	2.69	26.61
3008	0	0.48	9.59	0	6.11	0	15.7	0.48	16.18
3009	3.93	1.66	4.28	10.4	41.82	0	50.03	12.06	62.09
3010	0	0.29	0.92	16	26.99	5.35	27.91	21.64	49.55
3011	0.34	30.2	5.05	2.74	14.61	0	20	32.94	52.94
3012	0	0	0	0	0	0	0	0	0
3013	0	0	2.16	0	0	0	2.16	0	2.16
3014	0	0.48	0	0	0	0	0	0.48	0.48
3015	2.01	2.97	0	4.99	3.18	9.55	5.19	17.51	22.7
3016	0	0.39	0	0	10.1	0	10.1	0.39	10.49
3017	0.42	0	0	0	19.8	3.96	20.22	3.96	24.18
3018	0	0.32	0	0	0	0	0	0.32	0.32

3019	1.28	2.38	0	0	0	0	1.28	2.38	3.66
3020	0	0	20	17.83	16.71	4.65	36.71	22.48	59.19
3021	0	0	0	0	0	0	0	0	0
3022	0	0	37.56	5.05	9.23	2.58	46.79	7.63	54.42
3023	0	0	0	0	2.65	0	2.65	0	2.65
3024	0.33	1.5	12.76	9.95	2.16	0	15.25	11.45	26.7
3025	3.2	0.53	7.55	0	9.59	6.52	20.34	7.05	27.39
3026	0.33	5.33	4.63	2.08	5.3	0	10.26	7.41	17.67
3027	0	0	0	0	0.97	0	0.97	0	0.97
3028	0	0.49	0	0	7.8	0	7.8	0.49	8.29
3029	0.06	2.52	0	0	0	0	0.06	2.52	2.58
3030	0	0	15.88	0	10.54	7.17	26.42	7.17	33.59
3031	0	0	4.92	0	0	0	4.92	0	4.92
3032	0.01	0	0	0	1.59	0	1.6	0	1.6
3033	0	4.69	0	0	0	0	0	4.69	4.69
3034	0	4.91	13.85	0	22.69	5.01	36.54	9.92	46.46
3035	0	0.35	6.46	0	2.01	0	8.47	0.35	8.82
3036	0	0	12.34	0	14.8	0	27.14	0	27.14
3038	0	3.48	3.13	0	0	0	3.13	3.48	6.61
3039	0	8.19	29.98	0	10.14	5.49	40.12	13.68	53.8
3041	1.92	6.26	0	2.81	2.17	9.8	4.09	18.87	22.96
3042	0	0	1.66	0	4.05	10.8	5.71	10.8	16.51
3043	0	0	0	0	17.8	0	17.8	0	17.8
3044	0	2.4	3.47	3.44	0	10.53	3.47	16.37	19.84
3045	0	0	0	0	0	0	0	0	0
3046	0	0	2.24	12.6	27.3	0	29.54	12.6	42.14

3047	0	0	0	0	14.05	6.32	14.05	6.32	20.37
3048	0	0	0	0	0	0	0	0	0
3049	0	4.19	9.53	0	7.6	0	17.13	4.19	21.32
3050	0	10.63	0	0	15.73	6.65	15.73	17.28	33.01
3051	2.18	0	8.45	33.03	53.8	0	64.43	33.03	97.46
3052	0	0	0	0	0	0	0	0	0
3053	0	2.22	8.88	10.03	0	10.07	8.88	22.32	31.2
3054	0	0.71	4.11	7.9	2.85	0	6.96	8.61	15.57
3055	0.27	0	6.65	0	0	0	6.92	0	6.92
3056	8.28	27.29	12.23	0	17.8	0	38.31	27.29	65.6
3057	0	15.43	22.42	6.2	12.15	2.67	34.57	24.3	58.87
3058	0	0	0	13.97	5.06	0	5.06	13.97	19.03
3059	1.01	1.38	7.5	6.5	3.79	0	12.3	7.88	20.18
3060	0.73	2.14	14.2	0	17.57	0	32.5	2.14	34.64
3061	1.27	6.58	0	0	0	13.5	1.27	20.08	21.35
3062	0	0	0	0	0	0	0	0	0
3063	1.56	1.67	11.53	0	6.61	3.34	19.7	5.01	24.71
4001	0	0	0	0	0	0	0	0	0
4002	0	0	2.46	0	0	0	2.46	0	2.46
4003	0	26.3	14.3	6.58	4.6	0	18.9	32.88	51.78
4004	0	0	0	0	0	0	0	0	0
4005	0	0	0	0	0	0	0	0	0
4006	0	0	0	0	0	0	0	0	0
4007	0	0	0	0	0	4.3	0	4.3	4.3
4008	0	0	7.55	0	1.59	0	9.14	0	9.14
4009	0	0	0	0	6.56	0	6.56	0	6.56

4010	0	0	0	0	0	0	0	0	0
4011	0	0	0	0	0	0	0	0	0
4012	0	0	0	0	0	0	0	0	0
4013	0	28.61	8.89	0	3.55	0	12.44	28.61	41.05
4014	0	0	0	0	0	0	0	0	0
4015	3.97	6.88	5.22	10.68	7	2.07	16.19	19.63	35.82
4016	0.3	4	4.74	0	34.9	0	39.94	4	43.94
4017	0	0	1.44	0	0	0	1.44	0	1.44
4019	0	0	0	0	0	0	0	0	0
4020	0	0	0	0	23.8	0	23.8	0	23.8
4021	0	0	6.77	15.07	39.8	0	46.57	15.07	61.64
4022	0	0	10.83	3.95	24.39	0	35.22	3.95	39.17
4023	0	0	0	0	27.8	11.62	27.8	11.62	39.42
4024	0	8	0	0	15.8	3.22	15.8	11.22	27.02
4025	0	0	13.72	0	5.39	0	19.11	0	19.11
4027	0	0	3.49	0	0	0	3.49	0	3.49
4028	0	0	0	0	0	0	0	0	0
4029	0	0	4.57	8.45	0	1.11	4.57	9.56	14.13
4030	0.68	28.35	5.82	16.55	43.67	13.25	50.17	58.15	108.32
4031	0	0	0	0	8.8	0	8.8	0	8.8
4032	0	0	0	0	4.76	0	4.76	0	4.76
4033	0	0	9.91	0	0	0	9.91	0	9.91
4034	0	0	6.41	0	22.8	3.81	29.21	3.81	33.02
4035	0	5.92	10.67	23.28	2.36	38.8	13.03	68	81.03
4036	0	1.91	7.93	0	9.8	0	17.73	1.91	19.64
4037	0	0	0	0	9.42	0	9.42	0	9.42

4038	0.48	2.56	0	0	9.73	10.96	10.21	13.52	23.73
4039	0.88	3.8	0	0	20.72	2.59	21.6	6.39	27.99
4040	0	0	0	0	0	0	0	0	0
4041	0	0.7	0	0	0	0	0	0.7	0.7
4042	0	0	0	0	12.76	5.94	12.76	5.94	18.7
4043	0	0	0	0	5.8	0	5.8	0	5.8
4044	0	0	0	0	11.31	0	11.31	0	11.31
4045	0	0	5.55	2.73	7.8	22.98	13.35	25.71	39.06
4046	0	0	0	0	3.81	0	3.81	0	3.81
4048	2.43	9.74	14.13	20.58	9.4	8.32	25.96	38.64	64.6
4049	0	6.49	0	29.51	16.55	14.17	16.55	50.17	66.72
4050	0	1.51	3.26	10.26	0	2.9	3.26	14.67	17.93
4051	5.66	9.07	0	0	8.8	0	14.46	9.07	23.53
4052	0	0	0	0	0	0	0	0	0
4053	2.34	2.54	23.27	33.42	15.64	4.09	41.25	40.05	81.3
4054	1.66	11.34	3.79	15.89	0	0	5.45	27.23	32.68
4055	5.85	0	0	0	25.16	0	31.01	0	31.01
4056	2.75	4.86	21.32	5.41	0	0	24.07	10.27	34.34
4057	0.84	0	0	0	14.01	0	14.85	0	14.85
4058	0	0	0	0	0	0	0	0	0
4059	0	0	0	0	0	0	0	0	0
4060	0	0	0	0	2.37	0	2.37	0	2.37
4061	4.4	0	17.6	17.62	37.4	0	59.4	17.62	77.02
4062	0	0	12.73	11.44	0	0	12.73	11.44	24.17
4063	0	0	3.99	0	3.9	0	7.89	0	7.89

SMT 1, 2 and 3 = Seed Multiplication Trials October 2014, February 2015 and August 2015; S = Selfed; OP = Open Pollinated

Table 10.4: ICPAES data, 229 PMiGAP lines grown in field conditions at ICRISAT, 2013 (Section 3.4.2).

Analyte	**Ca**	**K**	**Mg**	**Na**	**Fe**	**Zn**
Units	mg/100g	mg/100g	mg/100g	mg/100g	mg/kg	mg/kg
Genotype						
1001	31.00	369.5	116.3	7.33	106.91	54.47
1002	30.91	374.3	124.7	4.32	59.82	29.53
1006	38.65	417.8	113.0	4.36	95.55	55.70
1007	32.21	362.0	125.4	4.23	92.23	45.04
1008	40.49	319.8	110.5	4.97	64.94	42.51
1009	34.89	348.7	143.4	3.42	66.28	53.90
1010	41.89	378.2	126.2	8.75	55.76	38.75
1012	48.81	370.0	141.3	14.50	75.71	49.38
1013	19.89	329.2	129.5	6.66	93.37	59.18
1014	23.57	392.4	127.8	4.61	65.06	59.18
1015	45.00	488.0	111.3	13.21	61.48	34.86
1016	-	-	-	-	-	-
1017	42.38	418.0	147.0	5.31	78.25	59.56
1018	47.67	384.9	113.0	7.17	83.60	70.74
1019	25.23	452.9	115.3	13.09	69.32	44.79
1020	29.09	438.5	123.5	10.62	45.52	31.32
1021	33.49	466.8	147.5	9.84	95.99	58.38
1022	26.98	342.9	121.7	5.02	98.96	69.76
1023	30.06	391.2	126.5	4.20	66.90	49.53
1024	39.73	457.1	166.7	5.05	61.26	43.46
1025	22.92	364.5	122.8	6.65	54.39	51.95
1026	23.04	316.4	119.0	5.59	76.30	56.30
1027	39.94	415.2	113.2	4.32	54.68	38.28
1028	50.16	467.0	130.3	10.41	50.61	32.83
1029	28.98	372.0	127.5	3.73	60.13	34.97
1030	23.99	337.2	131.2	5.07	73.23	48.73
1031	34.53	360.6	129.0	6.28	86.25	39.45
1032	35.03	326.2	110.6	4.66	73.70	47.16
1033	47.17	289.2	137.1	4.55	72.62	47.31
1035	35.87	323.9	104.3	4.13	90.24	71.22
1036	31.81	456.7	125.8	7.51	65.70	44.28
1037	19.87	382.6	121.2	6.83	52.03	41.00
1038	28.48	387.6	114.5	5.12	43.69	33.79
1039	35.69	519.6	116.7	6.37	45.18	34.12
1040	23.16	383.3	132.4	4.22	102.88	69.40
1042	23.73	425.0	109.5	3.84	57.07	41.69
1043	49.65	488.7	157.3	4.00	111.40	44.41
1044	24.51	379.7	98.5	4.25	69.63	52.89
1045	22.36	345.6	94.5	4.29	61.76	48.83

1046	19.52	444.9	135.0	9.12	54.73	36.66
1048	36.75	445.5	127.4	5.96	76.21	63.52
1049	21.86	490.8	107.3	3.21	53.97	32.41
1050	42.33	343.1	121.5	3.81	64.50	53.84
1051	17.05	398.7	115.6	5.90	44.09	38.09
1052	20.81	413.9	148.1	3.46	60.80	52.18
1053	21.47	381.9	108.9	3.53	60.86	54.53
1054	33.55	359.6	135.9	3.60	36.87	34.62
1055	34.75	473.1	129.7	4.87	43.50	36.89
1057	32.62	374.2	104.3	6.43	117.84	54.15
1058	70.50	468.0	119.1	5.72	135.27	79.31
1059	22.62	346.2	132.4	5.33	86.67	52.71
1060	23.42	412.6	150.3	3.78	96.67	65.51
1061	35.02	379.6	146.7	5.19	58.66	47.00
1062	35.54	519.2	153.5	5.34	58.46	51.99
2001	26.01	423.6	136.2	4.62	75.13	59.24
2002	23.57	458.7	129.5	3.44	37.11	23.46
2003	-	-	-	-	-	-
2004	-	-	-	-	-	-
2005	25.23	395.1	150.0	3.54	85.44	79.39
2006	40.48	416.0	141.0	5.66	40.77	34.94
2007	37.95	378.8	131.8	3.33	37.24	32.19
2008	17.85	297.1	123.2	3.00	60.64	49.40
2010	10.94	272.5	121.6	3.26	52.94	33.90
2011	27.57	386.5	114.3	3.92	55.18	38.71
2012	26.26	412.3	133.4	8.83	65.47	36.01
2013	15.63	313.6	141.9	3.97	65.44	59.42
2014	38.41	425.8	146.9	4.89	56.97	54.56
2015	21.98	466.9	120.3	4.10	79.96	53.29
2016	32.13	458.0	160.4	4.25	81.58	70.29
2017	20.69	318.4	124.5	2.97	66.27	53.08
2018	32.14	424.9	135.4	2.61	34.55	25.74
2019	54.52	428.1	148.4	2.76	86.79	65.85
2020	29.12	386.1	149.2	4.52	48.28	36.89
2021	22.70	351.5	127.3	2.95	43.05	54.16
2022	29.36	366.3	128.2	7.65	77.95	74.12
2023	26.47	413.9	138.4	4.62	68.19	43.45
2024	34.27	386.4	138.2	3.79	35.94	26.57
2025	25.97	324.5	108.0	3.42	71.77	57.50
2026	37.72	539.7	142.0	4.24	38.59	25.81
2027	31.21	469.9	140.5	4.63	48.51	44.10
2029	40.75	483.1	145.3	3.65	100.23	70.76
2030	23.98	436.3	124.3	2.84	51.58	40.05
2031	35.25	315.2	136.5	8.61	67.69	77.37
2032	31.02	463.5	119.4	4.63	79.35	62.54

2033	27.71	435.0	124.2	6.98	58.11	50.05
2035	26.62	526.5	122.9	4.49	62.72	50.66
2036	42.05	347.5	129.3	2.85	111.41	82.76
2037	28.76	424.8	135.0	5.43	51.60	37.65
2038	20.50	386.2	126.9	3.41	47.22	45.44
2039	22.80	399.2	140.5	3.07	72.27	60.02
2040	20.52	373.4	141.2	3.17	70.45	57.30
2041	33.14	418.2	138.0	3.39	56.26	57.04
2043	32.45	346.7	135.1	3.88	56.93	33.90
2044	35.71	489.6	128.8	5.57	59.86	47.27
2045	-	-	-	-	-	-
2046	35.39	344.4	116.9	5.26	41.95	32.51
2048	16.59	353.9	123.3	3.18	56.63	45.60
2050	32.63	401.3	134.9	3.19	56.56	43.23
2051	18.50	375.5	129.4	2.47	45.84	50.74
2052	17.72	341.0	125.2	3.90	57.96	61.73
2053	21.06	356.2	132.3	1.66	77.34	63.78
2054	31.97	378.8	135.5	2.14	124.59	93.28
2057	23.45	422.0	132.6	2.27	87.05	62.64
2058	21.47	371.1	126.1	5.56	57.31	46.71
2059	24.35	375.4	122.2	3.37	49.49	43.67
2060	25.25	418.4	118.3	3.21	44.35	36.62
2061	52.00	372.1	129.3	2.76	113.95	62.72
2062	37.50	397.2	139.4	2.14	68.74	59.00
2063	21.99	352.0	125.7	2.23	36.78	36.10
3001	24.32	470.6	153.4	2.32	39.12	39.03
3002	32.18	454.5	141.8	2.20	40.43	30.49
3003	27.21	441.1	133.1	2.16	51.22	51.09
3004	24.45	354.6	141.9	4.51	50.91	55.25
3005	16.02	529.3	126.5	5.18	49.24	50.20
3006	23.84	389.5	127.3	2.28	41.97	42.50
3007	18.50	466.5	124.3	2.56	50.68	46.49
3008	21.99	414.1	130.5	2.12	53.92	49.09
3009	18.80	359.4	136.9	2.04	69.94	56.68
3010	31.85	376.8	151.8	1.77	69.64	57.09
3011	23.28	378.8	152.7	1.26	51.73	52.74
3012	24.38	429.9	136.8	2.67	37.06	35.84
3013	16.99	429.0	128.5	4.62	39.25	26.42
3014	25.94	400.0	141.9	2.61	70.88	46.59
3015	30.79	392.8	136.8	3.00	70.08	57.29
3016	14.37	332.9	162.8	2.24	71.43	68.07
3017	35.73	473.4	147.7	2.68	60.95	46.19
3018	14.44	382.9	111.6	2.55	45.64	43.40
3019	23.67	389.2	126.6	2.40	49.61	37.72
3020	16.40	421.4	133.5	4.61	56.29	48.51

3021	29.48	438.6	135.9	2.47	47.99	27.82
3022	33.93	392.3	112.9	4.32	34.33	39.55
3023	24.35	462.4	121.2	2.96	59.97	56.37
3024	49.33	457.7	122.4	2.61	56.22	43.03
3025	24.66	397.5	135.5	2.70	42.79	44.34
3026	28.84	410.7	158.4	3.44	67.51	80.80
3027	62.81	510.6	145.5	3.49	50.07	54.01
3028	29.24	375.0	142.4	2.26	54.86	47.57
3029	19.18	393.9	144.8	2.52	53.38	58.58
3030	40.12	384.6	128.7	4.52	72.34	55.24
3031	30.19	461.5	134.0	2.93	36.69	30.71
3032	20.62	388.6	157.8	2.82	56.89	56.60
3033	38.98	328.3	143.2	2.31	94.00	82.90
3034	73.71	448.0	150.6	4.84	52.47	43.59
3035	43.44	466.5	150.1	3.08	64.53	64.46
3036	18.80	373.6	129.2	2.37	53.34	56.21
3038	13.86	379.8	119.5	1.76	59.74	54.69
3039	31.94	445.1	126.3	2.14	62.20	46.08
3041	26.60	376.1	132.5	4.71	69.63	64.39
3042	43.29	418.1	148.7	2.95	62.37	45.53
3043	18.76	390.9	130.8	2.43	43.19	43.97
3044	31.59	406.8	130.0	1.64	64.30	44.63
3045	35.59	469.7	109.2	2.76	54.48	39.54
3046	35.04	370.2	151.4	1.46	73.39	64.91
3047	21.02	360.0	138.7	2.33	38.58	43.41
3048	13.50	319.4	111.9	2.59	111.01	66.22
3049	25.90	466.2	135.0	2.86	51.48	51.02
3050	19.98	364.0	140.8	3.65	93.74	66.95
3051	27.05	466.6	122.8	6.23	55.35	51.54
3052	68.88	359.4	141.9	4.51	89.45	79.16
3053	39.91	444.8	139.8	2.30	70.48	74.47
3054	22.52	372.9	115.9	2.00	33.63	34.83
3055	32.30	415.6	112.9	1.78	83.29	56.56
3056	18.54	345.7	119.3	2.08	49.21	38.49
3057	16.97	376.0	104.3	3.61	50.76	42.29
3058	30.71	327.4	122.6	2.08	33.27	34.71
3059	17.08	329.9	125.3	5.97	44.03	40.03
3060	33.81	366.0	138.3	3.93	71.66	36.84
3061	35.20	360.3	139.5	4.26	82.16	76.80
3062	19.85	437.0	107.0	3.40	43.55	34.32
3063	35.03	325.2	113.0	4.16	79.79	50.94
4001	22.65	438.1	154.9	3.42	64.25	45.96
4002	16.71	320.6	138.3	3.75	69.51	56.74
4003	16.52	382.1	145.3	4.02	67.13	48.80
4005	27.80	419.3	143.8	6.01	84.16	44.60

4006	34.72	497.0	122.9	7.11	55.01	47.68
4007	20.18	490.8	157.9	3.92	65.01	46.80
4008	27.99	336.8	120.9	4.27	96.16	46.72
4009	31.84	403.8	133.7	3.93	76.27	68.14
4010	-	-	-	-	-	-
4011	33.68	371.9	119.9	3.83	39.65	22.07
4012	17.07	413.9	133.2	3.64	74.93	66.05
4013	24.96	372.7	143.1	4.58	58.75	45.91
4014	26.69	395.4	120.0	8.54	86.08	44.80
4015	29.48	370.5	117.8	6.81	75.95	56.16
4016	23.06	340.0	110.4	4.57	64.56	47.22
4017	26.01	451.0	158.2	4.44	55.73	55.83
4019	33.69	568.1	151.0	4.76	54.36	28.66
4020	46.38	476.4	152.7	4.98	97.81	47.44
4021	26.12	466.9	133.1	5.08	58.60	50.92
4022	18.49	409.5	123.9	3.58	55.01	57.03
4023	18.78	348.3	126.8	4.13	36.61	42.54
4024	26.27	501.7	123.9	8.15	53.71	43.54
4025	-	-	-	-	-	-
4027	25.90	356.8	124.5	4.75	57.90	62.97
4028	15.48	348.2	109.4	3.94	49.10	33.97
4029	32.81	486.1	150.5	4.08	44.94	54.73
4030	13.20	384.0	131.7	4.11	75.57	49.00
4031	23.91	346.5	125.9	3.43	41.53	29.51
4032	24.51	333.9	164.7	4.22	51.99	44.18
4033	21.71	374.8	139.5	6.68	72.16	50.91
4034	41.31	429.6	153.5	3.56	64.50	47.11
4035	26.02	407.0	132.3	2.39	41.06	32.67
4036	19.60	363.3	161.2	3.87	48.62	40.18
4037	25.12	371.9	141.4	2.29	93.40	64.64
4038	28.08	418.6	130.0	2.30	56.94	36.09
4039	44.16	380.7	139.3	1.74	68.59	55.14
4040	33.01	474.6	151.7	1.91	44.01	40.89
4041	41.28	393.5	123.3	2.56	35.02	24.81
4042	32.62	335.8	128.3	6.56	43.40	28.85
4043	34.06	395.4	135.5	2.74	54.57	32.71
4044	34.87	431.5	141.0	2.42	41.74	33.90
4045	35.13	450.7	133.3	2.13	50.66	39.63
4046	27.05	430.1	136.3	1.64	60.00	56.22
4048	18.34	475.1	119.2	1.95	43.11	38.37
4049	33.90	369.8	126.2	2.74	102.28	78.52
4050	32.66	383.3	111.7	2.22	29.18	33.87
4051	28.87	359.4	118.3	2.08	49.11	47.65
4052	41.16	429.2	152.0	5.00	61.36	43.97
4053	25.47	373.8	151.9	3.29	53.60	54.63

4054	51.37	451.4	157.7	2.71	84.04	64.22
4055	27.26	376.5	127.7	3.70	37.00	28.74
4056	26.43	369.3	127.6	3.57	34.96	27.61
4057	23.94	499.8	154.3	2.44	50.61	43.51
4058	22.88	469.6	141.5	2.21	51.07	57.45
4059	32.98	414.9	149.6	3.02	45.13	40.88
4060	13.77	354.7	165.3	2.27	62.71	46.66
4061	48.78	460.6	124.2	4.12	62.91	40.91
4062	24.11	365.5	131.3	4.14	46.23	45.05
4063	27.08	397.1	144.3	4.62	103.82	70.09

Table 10.5: ICPAES data used to test GEI effects (Section 3.4.3).

Analyte	**Ca**	**K**	**Mg**	**Na**	**Fe**	**Zn**
Units	mg/100g	mg/100g	mg/100g	mg/100g	mg/kg	mg/kg
Genotype						
4056 SELF OCT '14	71.65	387.7	124.5	18.29	35.52	27.23
4015 SELF OCT '14	58.71	583.3	135.2	15.95	62.24	47.55
1019 SELF OCT '14	47.92	501.8	106.4	18.88	62.69	41.23
1014 SELF OCT '14	44.66	457.5	124.6	16.73	60.43	39.79
1032 SELF OCT '14	51.39	485	145.3	16.9	64.88	51.67
2037 SELF OCT '14	67.18	367.9	135.6	17.51	41.3	29.16
1055 SELF OCT '14	60.15	582	156.7	20.32	79.54	37.64
2021 SELF OCT '14	56.9	378.2	131.4	19.86	56.61	33.84
2038 SELF OCT '14	49.32	365.6	145.4	16.03	137.16	42
1048 SELF OCT '14	55.06	526.1	147.6	21.69	77.99	73.34
2023 SELF OCT '14	35.79	276.3	97.4	15.25	42.04	31.47
1009 SELF OCT '14	55.03	399.5	152.2	16.1	58.97	50.04
2063 SELF OCT '14	65.74	468.4	116.2	19.48	39.09	25.37
2060 SELF OCT '14	34.64	220.5	65.4	14.41	17.04	14.36
1021 SELF OCT '14	44.64	455.9	141.5	16.6	54.93	29.14
1061 SELF OCT '14	38.52	533.6	137.4	17.6	107.27	78.12
2035 SELF OCT '14	40.59	554.7	147.8	19.44	21.09	15.78
2019 SELF OCT '14	47.1	361.4	123.2	21.12	87.75	64.17
2024 SELF OCT '14	40.24	242.5	106.5	17.49	25.81	18.61
1054 SELF OCT '14	59.62	363.9	147.6	20.31	44.26	39.6
4056 OP OCT '14	52.71	295.6	126.7	18.69	31.67	26.95
4015 OP OCT '14	44.88	358.9	108.9	20.15	52.86	36.17
1019 OP OCT '14	50.4	382	112.3	15.28	35.56	39.84
1014 OP OCT '14	51.02	425.7	124.5	18.32	35.75	25.18
1032 OP OCT '14	48.94	369.6	128.1	19.75	50.39	30.74
2037 OP OCT '14	46.63	375	145.4	16.53	39	17.29
1055 OP OCT '14	47.31	376.8	151.1	16.27	38.78	17.32
2021 OP OCT '14	56.86	467	148.7	17.45	38.89	26.69

2038 OP OCT '14	56.55	324.5	130.2	20.6	49.43	32.85
1048 OP OCT '14	43.78	314.1	136.8	14.56	50.46	45.64
2023 OP OCT '14	47.62	402.3	125.4	19.25	56.42	31.25
1009 OP OCT '14	52.85	341.3	138.4	16.54	54.05	36.2
2063 OP OCT '14	48.83	357.4	140.9	16.27	47.53	39.44
2060 OP OCT '14	64.14	474.7	144.9	15.81	45.47	30.88
1021 OP OCT '14	48.46	363.5	120.2	16.71	38.66	20.89
1061 OP OCT '14	60.8	427.3	142.6	17.08	58.98	53.44
2035 OP OCT '14	50.96	577.2	155.7	20.22	73.15	65.14
2019 OP OCT '14	53.31	504.5	126.6	17.12	52.61	44.86
2024 OP OCT '14	46.55	379.9	119.3	15.98	23.4	17.58
1054 OP OCT '14	52.15	303.9	131.1	16.04	35.6	37.98
4056 SELF FEB '15	52.7	294	126.3	11.42	39.63	25.91
4015 SELF FEB '15	60.17	377.6	131.7	14.3	101.34	61.7
1019 SELF FEB '15	30.38	355.5	114	10.02	41.85	26.64
1014 SELF FEB '15	50.31	337.9	134.3	17.41	45.8	34.68
1032 SELF FEB '15	65.9	377.2	152.1	13.4	102.2	63.27
2037 SELF FEB '15	55.69	347.5	128.3	15.33	38.07	22.81
1055 SELF FEB '15	58.1	445.3	141.1	15.33	35.98	28.16
2021 SELF FEB '15	44.57	308.4	114.6	11.68	34.14	31.33
2038 SELF FEB '15	56.34	320.5	129.7	15.28	56.27	50.29
1048 SELF FEB '15	44.92	365.1	137.7	12.23	95.01	53.16
2023 SELF FEB '15	183.51	237.5	104.3	25.2	80.93	38.04
1009 SELF FEB '15	62.13	326	163.5	13.93	45.8	39.07
2063 SELF FEB '15	58.55	379.9	130	14.25	43.82	31.98
2060 SELF FEB '15	53.01	307.8	108.2	16.98	49.21	30.51
1021 SELF FEB '15	63.97	406.3	132.9	14.64	65.28	44.25
1061 SELF FEB '15	59.14	326.4	134.9	15.5	56.85	42.16
2035 SELF FEB '15	60.01	630.4	144.8	20.59	96.7	81.98
2019 SELF FEB '15	72.48	340.3	137.3	16.39	77.87	40.47
2024 SELF FEB '15	66.28	353.8	147.3	15.14	30.88	23.38
1054 SELF FEB '15	70.76	253.5	150.2	14.61	35.65	41.59
4056 OP FEB '15	55.37	287.7	123.1	13.97	46.43	32.48
4015 OP FEB '15	50.98	251.3	115.2	14.76	69.82	30.75
1019 OP FEB '15	47.41	293.8	122.3	15.19	52.9	37.88
1014 OP FEB '15	63.56	355.8	120.2	18.49	46.16	57.77
1032 OP FEB '15	52.64	309.3	128.9	14.57	54.32	30.75
2037 OP FEB '15	52.35	313.4	133.6	13.75	58.18	32.38
1055 OP FEB '15	58.73	368.3	130.3	20.11	31.47	23.99
2021 OP FEB '15	56.63	296.1	116.5	17.83	53.68	38.93
2038 OP FEB '15	44.82	319	126.9	13.94	78.15	46.93
1048 OP FEB '15	49.98	350.5	120.5	16.43	70.02	44.55
2023 OP FEB '15	57.66	371.8	135.6	18.63	70.55	42.23
1009 OP FEB '15	53.47	319.7	144.3	15.1	65.01	54.19
2063 OP FEB '15	58.4	320.2	129.7	21.49	50.29	36.13

2060 OP FEB '15	64.05	338	125.1	18.47	58.53	41.42
1021 OP FEB '15	51.45	345.4	126.7	15.22	55.12	45.58
1061 OP FEB '15	64.54	290.1	142.2	17.31	61.2	38.02
2035 OP FEB '15	56.52	406.3	138.7	16.93	123.11	47.24
2019 OP FEB '15	64.5	348.7	127.6	17.55	82.29	51.09
2024 OP FEB '15	53.58	301.9	131.8	14.95	48.05	30.61
1054 OP FEB '15	65.71	288.8	142.9	15.98	61.8	39.88

Table 10.6: ICPAES data from the August 2015 seed multiplication trial used to compare glasshouse and field grown seed (Section 3.4.4)

Analyte	**Ca**	**K**	**Mg**	**Na**	**Fe**	**Zn**
Units	mg/100g	mg/100g	mg/100g	mg/100g	mg/kg	mg/kg
Genotype						
1001	16.82	296.71	118.07	5.01	92.42	81.53
1002	14.42	384.52	100.45	5.65	34.43	24.05
1006	30.83	440.45	154.39	5.01	133.36	114.54
1009	21.4	435	160.63	5.81	53.3	49.5
1013	23.89	364.5	127.16	5.36	52.9	42.29
1014	19.84	329.2	124.29	5.32	53.47	56.08
1015	28.19	414.5	119.93	5.19	49.14	36.35
1018	30.36	380.54	121.1	4.81	84.66	86.12
1022	17.94	339.88	126.61	5.49	99.27	75.11
1023	50.19	438.88	131.3	5.04	43.26	78.39
1026	41.11	310.56	121	5.07	66.44	83.31
1027	27.21	361.76	125	5.06	71.01	59.36
1029	19.37	305.41	126.38	4.88	79.66	54.34
1032	23.19	340.09	121.16	5.33	68.08	55.05
1033	24.86	303.15	120.13	4.84	65.04	62.9
1035	36.42	332.81	128.8	5.3	145.5	161.06
1036	21.64	376.18	130.58	7.46	70.58	64.6
1038	31.69	463.99	148.37	6.93	76.15	54.06
1042	21.92	378.23	114.88	4.71	88.03	68.13
1044	23.27	351.1	122.83	5.78	126.25	89.57
1045	14.36	325.6	114.19	6.15	63.85	55.64
1048	24.83	361.35	129.51	5.25	107.95	85.17
1049	17.99	576.8	109.63	10.39	50.34	29.67
1050	22.07	363.26	107.37	5.05	99.27	92.47
1051	18.14	435.88	115.06	6.12	69.89	56.01
1053	I.S	400.69	137.94	7.36	134.75	IS
1057	29.97	372.61	105.58	5.75	84.66	98.19
1058	16.66	428.37	115.54	5.03	88.53	60.02
1059	20.41	343.17	121.09	4.57	80.95	75.51
1061	27.73	365.02	153.36	5.58	81.63	93.11
1062	27.99	613.23	157.63	7.28	121.6	149.63

2002	37.97	558.85	155.04	5.75	37.41	26.31
2006	28.47	316.18	116.66	5.07	37.79	30.35
2008	15.66	320.97	111.59	5.6	56.31	49.21
2011	35.14	352.59	129.87	4.68	52.74	60.52
2014	18.25	415.21	138.32	6.26	51.5	65.93
2015	10.11	429.19	120.63	6.55	58.03	67.48
2017	20.72	339.05	128.09	5.1	61.49	58.44
2018	17.06	363.91	112.09	5.28	86.75	76.26
2021	19.59	370.88	120.88	4.87	68.83	70.77
2022	17.84	332.66	125.26	5.38	108.63	88.22
2024	19.07	365.95	144.6	5	32.55	26.09
2025	20.75	316.15	118.25	5.05	113.01	97.22
2029	26.98	365.7	118.99	5.67	67.18	46.56
2033	16.69	349.09	109.07	5.35	97.32	87.29
2035	21.8	463.38	127.48	6.47	138.08	131.9
2038	17.9	314.06	108.51	4.85	84.6	69.67
2040	16.3	373.95	144.81	5.55	85.53	91.31
2041	32.32	353.27	115.97	4.51	121.95	111.27
2043	29.47	372.01	125.45	5.64	71.87	54.32
2046	35.18	369.16	150.88	5.85	68.15	50.2
2058	11.98	378.52	110.66	7.33	61.32	40.22
2059	21.23	428.71	132.58	6.06	106.9	95.42
3005	13.39	408.8	106.62	5.49	69.43	47.77
3007	7.78	388.32	123.27	5.96	65.81	66.48
3009	12.38	290.46	121.96	5.66	94.26	82.52
3010	11.65	427.96	134.57	6.03	75.54	86.12
3011	23.91	360.63	136.75	5.08	76.98	73.78
3016	10.31	368.57	175.43	5.61	94.99	120.12
3017	27.49	459.14	134.11	5.19	56.78	52.51
3020	16.69	371.77	129.51	7.01	70.54	56.11
3022	15.8	327.25	104.3	5.59	37.7	33.93
3025	21.17	339.3	137.67	5.91	97.05	82.03
3028	32.33	399.79	167.64	5.98	69.09	80.43
3030	29.51	366.66	143.55	4.94	69.95	47.65
3034	36.89	360.59	121	5.23	41.7	40.3
3036	33.81	323.52	141.76	5.8	97.02	79.55
3039	38.66	455.13	139.81	4.97	72.37	53.18
3043	26.8	366.37	126.46	5.56	49.64	56.44
3046	28.19	321.13	129	6.2	96.07	76.63
3047	28.77	322.61	128.74	7.19	48.17	50.51
3049	32.52	473.7	118.09	6.44	70.96	55.93
3050	35.42	404.28	145.55	6.48	68.36	78.95
3051	26.6	360.13	105.65	5.46	64.85	73.89
3056	34.7	577.67	137.01	8.91	79.58	61.94
3057	32.9	395.19	115.51	6.36	90.66	81.66

3060	47.03	361.64	136.35	9.3	78.09	57.59
4016	27.92	257.92	116.98	10.05	92.44	85.74
4020	29.01	389.19	134.67	5.75	74.31	61.56
4021	32.97	405.27	136.76	5.86	60.53	75.95
4022	29.22	367.42	121.95	8.5	67.68	81.46
4023	31.82	361.5	133.21	4.67	94.96	110.22
4024	22.85	390.81	113.75	5.02	83.93	90.11
4030	27.36	407.05	119.04	6.51	83.91	69.07
4031	47.43	446.01	148.31	7.16	78.71	63.39
4034	24.54	342.46	126.57	6.72	73.3	63.48
4036	25.54	410.97	170.94	8.08	82.99	80.81
4037	28.34	358.67	158.58	7.31	182	159.18
4038	43.75	526.96	146.33	9.52	101.24	91.98
4039	42.94	444.92	140.68	6.65	84.05	61.52
4042	35.32	371.22	127.37	6.99	56.66	36.77
4044	44.65	440.09	103.18	8.7	70.16	67.15
4045	40.05	453.38	115.44	8.76	66.96	57.56
4048	34.27	420.8	112.81	13.11	69.31	71.11
4049	32.83	360.73	120.39	5.17	141.09	135.75
4051	33.12	281.57	108.68	9.4	100.06	92.09
4053	34.24	296.87	145.18	5.7	85.68	107.79
4055	29.61	366.01	136.26	8.53	66.65	52.29
4057	37.02	367.92	157.18	8.78	90.74	94.27
4061	53.08	422.38	143.57	9.14	71.2	61.66

Table 10.7: ICPAES data of OP PMiGAP lines grown at ICRISAT, 2013 (Section 3.4.5)

Analyte	**Ca**	**K**	**Mg**	**Na**	**Fe**	**Zn**
Units	mg/100g	mg/100g	mg/100g	mg/100g	mg/kg	mg/kg
IP No.						
22295	33.31	410.2	172.2	12.4	55.66	65.37
4545	27.26	325	144.5	5	72.8	65.3
16403	23.13	338	130	7.29	43.43	40.06
7846	19.19	311.7	112.5	4.41	49.29	50.99
17720	-	-	-	-	-	-
15917	23.58	313.1	145	5.63	42.22	36.11
8767	27.42	334.5	122.9	8.52	38.45	30.7
12298	23.93	286.9	114.6	6.78	34.08	35.25
2058	30.7	336.7	129.9	4.62	45.6	47.6
3125	21.49	335.7	129.6	5.52	41.96	42.99
6584	18.24	295.1	128.4	4.91	51.66	43.9
4020	27.38	403.1	138	7.13	62.33	55.52
18132	22.69	367.6	128.4	6.78	46.21	40.5
16082	23.63	410.6	105.6	3.57	35.5	31.28

18062	19.74	345.3	120.1	5.15	49.24	36.95
22549	17.73	350.9	106.5	5.76	42.08	42.33
6745	18.17	371.9	134.1	9.49	45.12	30.29
9595	24.07	388.1	123.3	5.34	50.44	36.24
11211	21.81	358.3	110	4.79	56.07	41.66
10705	21.16	339.8	110.5	3.79	74.87	46.36
22274	18.35	392.5	113.2	5.12	43.76	36.69
22272	18.36	257.3	95.1	2.97	42.11	33.11
9426	24.77	363.9	131.3	4.58	57.86	46.39
7762	23.28	386	143.7	5.8	54.49	42.86
15551	28.42	387.7	132.9	5.46	64.9	44.46
9446	19.91	333.4	119	4.5	77.9	40.36
21155	21.13	290.3	97.1	3.48	54.88	38.5
10811	28.09	370	117.3	4.09	57.58	34.97
17632	19.65	275	105.7	3.26	36.46	34.98
11577	26.48	428.9	124.7	6.01	41.12	36.2
22454	15.04	334.3	115.3	4.09	66.43	45.65
22279	20.81	332.9	111.8	4.66	40.57	33.64
9351	-	-	-	-	-	-
15947	37.39	375.6	132.8	10.41	62.9	47.07
7970	26.11	310.7	126.1	10.14	40.47	35.73
3735	31.78	362.3	112.9	5.34	62.15	67.1
6112	28.88	302.7	127	3.23	47.57	41.84
19448	16.9	321.8	116.4	4.24	41.32	40.29
9242	19.9	365.6	118.4	4.37	28.57	32.92
22569	18.13	342	97	3.99	57.18	42.82
16096	17.72	289.3	114.8	3.24	58.29	33.53
7633	20.49	303.2	127	5.62	53.89	33.77
8198	-	-	-	-	-	-
6060	19.8	292.1	122.6	2.84	40.27	39.24
8276	16.63	360.6	118.8	3.27	41.8	43.34
22423	21.21	324.3	114.1	5.2	45.73	52.85
7848	30.98	298.4	131.2	2.69	36.71	36.75
9347	24.98	370.7	118.5	4.7	49.52	40.38
22455	23.02	342.4	117.2	4.65	23.78	28.64
3175	20.15	324.3	122.5	9.38	44.59	47.55
12840	20.2	371.5	120.4	7.11	42.21	43.2
19386	30.73	339.9	120.8	4.5	52.2	49.9
9710	18.78	310.5	123.2	2.65	69.32	63.61
5389	14	325.8	109.1	3.87	30.59	32.43
3890	21.69	384.7	129.1	5.81	48.41	40.59
11677	46.26	448.3	179.6	4.8	45.09	44.17
11984	23.47	329.3	126.4	3.9	42.39	48.58
17690	24.1	337	118.4	3.33	74.15	57.87

12925	17.39	302.2	114.8	4.3	65.63	50.45
10140	17.92	313	113.3	3.03	41.39	39.85
6037	-	-	-	-	-	-
3122	19.3	327.3	115.2	2.86	37.22	31.17
6146	29.27	462.4	133	6.56	34.4	25.62
15533	20.97	371.4	120.5	4.85	31.97	27.33
7942	24.36	351.7	140.8	4.52	50.15	42.61
9301	15.1	300.8	94.6	3.85	80.31	64.35
11311	20.23	318.2	119.4	4.26	36.38	37.33
9407	23.99	321	126.2	4.47	71.68	60.04
13154	14.13	353.7	107.7	4.57	33.06	33.9
19388	19.11	329.7	114.1	4.02	43.13	42.52
17554	24.88	329.9	128.8	4.59	46.26	46.98
9840	18.44	337	118.8	3.39	30.27	35.09
3098	21.27	369.4	132.6	2.31	41.87	46.4
22383	16.7	380.9	143.2	3.66	43.05	37.41
11353	24.03	335	113.5	3.47	38.49	34.48
10964	17.91	330.1	118.7	3.58	28.62	22.07
18157	26.05	278.1	128.6	2.41	50.4	40.16
4965	20.71	330.1	120	3.8	43.11	40.53
9496	22.03	309	97.5	3.77	38.74	44.12
10820	18.1	330.6	115.4	3.12	51.37	42.2
8949	16.89	344.6	135.7	4	49.12	52.54
8949	-	-	-	-	-	-
8181	21.44	334.5	141.9	5.27	44.84	43.05
11229	29.25	320.7	124.1	3.17	41.77	39.42
10343	23.48	389	121.9	4.94	35.49	30
15946	25.94	384.6	132.1	5.46	32.96	33.76
20349	18.46	335.3	121.8	2.81	51.28	35.19
13016	15.28	349.3	128.5	2.56	44.58	43.18
3865	19.56	329.6	134.3	3.45	65.21	62.28
10486	19.45	349.3	140.5	7.3	42.93	44.71
3557	29.72	405.2	136.5	4.29	43.62	35.24
4962	26.35	379.4	137.7	3.19	49.42	40.14
8344	23.78	368	119	2.62	39.59	37.45
5713	25.68	360.1	119.5	4.11	56.77	53.84
5695	28.68	456.6	121.3	8.12	60.48	51.65
12768	30.33	434.5	118.1	5.13	57.89	45.11
6179	21.06	315.5	103.3	3.95	37.78	31.52
7095	33.42	446.2	120.1	12.43	44.79	40.04
8761	15.47	290.8	120	2.67	42.56	48.86
8955	20.27	337.5	119.2	3.74	84.49	71.11
9406	46.36	310.4	125.8	2.18	112.8	64.04
9532	18.49	301	116.6	6.06	57.01	57.38

8275	18.94	356.4	126.9	5.18	44.34	47.65
8280	16.94	401	127.6	4.86	45.48	48.47
10379	21.98	287.7	116.1	2.51	44.37	49.21
10271	14.87	335.2	128.2	2.09	44.15	48.83
12322	18.18	362.1	124.4	3.13	42.51	46.68
19626	20.94	323.5	129.5	3.57	50.25	56.26
3564	19.28	281.7	116.7	3.02	37.75	45.48
6111	19.63	364.6	126.4	2.59	73.1	47.53
15320	-	-	-	-	-	-
12845	22.53	329.4	127.1	2.22	42.16	49.74
10953	24.51	388.4	120	1.94	36.1	34.36
22276	23.82	421.6	126.2	2.21	32.1	33.19
21206	12.88	355.5	107.3	3.33	29.15	32.2
3636	19.39	321.5	113.4	2.85	29.31	52.12
6099	19.38	378.6	129.3	3.67	36.47	38.78
10613	33.66	534.6	154.4	4.42	32.62	37.53
11310	23.83	445.6	122.2	4.64	49.63	47.93
10085	14.99	379.9	112.7	3.87	46.44	43.83
12839	15.79	285.4	139.3	2.81	50.82	50.62
10394	13.25	325.2	113	3.48	34.06	36.71
3732	18.34	356.4	133.2	3.17	47.03	52.07
9391	25.06	340.7	113	3.48	63.59	56.21
21020	16.87	358	119.4	2.12	41.26	45.86
3138	28.08	362.7	153.6	2.12	68.89	70.11
7952	16.97	315.8	135.1	2.23	34.77	38.84
16638	25.18	369.7	109.9	2.37	85.27	56.52
4974	15.86	323.1	124.8	2.18	38.85	43.06
8972	-	-	-	-	-	-
8426	25.2	352.3	125.1	2.98	46.4	47.35
7470	21.64	362.9	138.3	5.47	44.39	45.07
18389	36.23	370.1	137.9	3.86	39.11	46.45
7886	17.62	298.6	130.4	4.8	54.62	57.44
16289	28.77	356.7	133.5	15.51	76.31	45.67
6682	28.72	312.3	130.7	5.48	41.47	33.46
13384	22.38	323.9	117.5	3.53	55.35	54.42
22527	18.25	389.5	119.9	4.52	31.28	32.92
22424	19.93	334.6	115.3	4.83	30.31	30.65
22281	22.08	300.1	116.2	3.99	50.01	50.88
9282	17.27	283.6	122.1	2.75	110.85	66.05
22494	16.41	352.2	130.2	2.8	45.61	47.34
9824	-	-	-	-	-	-
3616	17.53	373.3	122.3	2.68	27.67	22.62
6102	19.64	368.7	123.4	3.81	63.21	47.84
18434	18.82	355	116.2	2.87	37.3	36.95

17611	14.67	351	112.5	2.81	43.81	39.78
7967	19.98	376.7	121.9	4.76	44.48	43.14
8172	21.55	350.1	111.5	3.59	44.32	40.25
13344	20.05	337.9	106.1	2.46	26.09	28.14
22419	25.07	336.7	132.9	2.75	34.97	37.18
13180	21.55	374.1	124.2	3.31	42.35	37.95
16402	18.69	253.9	120.6	2.94	50.64	43.07
3389	20.35	330	125.5	2.37	50.92	47.37
8074	24.41	417.7	147.5	3.66	39.37	41.58
21169	30.29	372.4	135.4	2.52	37.41	33.64
22420	42.78	385.2	131.6	4.18	30.68	30.07
7953	19.6	255.2	84.5	1.61	53.99	38.45
10701	18.23	323.5	132.8	2.41	54.18	51.95
10632	18.01	371.4	121.8	2.14	30.25	31.98
13817	15.65	336.1	108.7	2.8	46.68	44.57
8129	-	-	-	-	-	-
10471	19.91	348.7	114	2.97	44.5	39.31
11593	42.72	530.6	122.5	4.41	50.7	33.33
5131	17.54	304.6	101.7	2.6	31.48	29.89
22283	14.63	342.6	141.2	2.91	50.54	53.03
5923	22.68	342.4	127.2	1.97	38.54	35.93
8187	15.91	319.2	120.5	2.85	46.23	43.06
13927	15.43	287.3	115.2	2.48	39.87	38.63
15872	29.74	336.2	113.4	2.96	67.5	56.41
9692	29.06	341.9	117.9	2.71	55.96	50.03
9651	24.05	321.2	103.6	2.83	31.02	31.41
14439	21.18	412.6	131.2	3.63	35.59	25.08
7922	19.21	295.6	119.2	2.98	41.83	42.41
10761	28.19	450.2	144.1	3.48	40.62	41.48
4828	21.1	347.4	137.4	4.37	55	56.57
11275	14.49	356.7	105.7	2.12	40.93	46.6
22458	17.42	298.5	99.9	2.15	31.96	33.42
6417	13.49	351.9	125.4	2.24	42.52	46.8
4952	13.72	332.8	114.5	2.17	52.14	49.11
10543	-	-	-	-	-	-
9969	25.95	389.8	133.7	5.73	36.72	34.11
6882	16.09	307.1	115.6	4.64	37.18	38.66
14148	18.23	297.2	129.8	3.14	50.18	50.41
5816	19.63	324.1	124.4	3.16	41.84	44.09
8069	18.3	287.3	114.1	3.98	41.31	41.73
14497	21.05	382.5	139.7	3.06	36.08	34.34
18090	21.6	306	132.5	3.87	47.62	51.04
6329	33.62	382.9	119	2.8	47.07	43.06
7930	17.47	287	124.3	2.69	55.32	55.62

4927	19.23	308.3	113.7	1.91	43.55	43.71
5253	-	-	-	-	-	-
11346	20.32	287.2	100.3	2.55	40.73	34.59
6310	20.29	327.2	133.5	2.55	57.82	57.5
5031	23.03	301.2	107.2	2.04	41.16	46.25
11358	26.24	338.1	134.3	2.56	55.22	43.46
17125	20.27	318.3	101.4	2.26	25.38	31.75
17099	25.29	398.2	122.7	4.31	44.25	38.46
5121	37.15	351.7	139.4	4.81	52.45	44.25
18168	17.18	333.5	128.3	3.78	27.7	28.87
18246	13.91	338.7	112.5	2.97	26.17	30.62
15553	22.95	346.8	137	1.88	68.03	54.3
7660	19.8	297.9	139.1	3.18	52.93	53.78
12370	20.4	410.5	133	6.93	34.46	43.16
20679	25.03	375.5	126.9	2.1	49.61	54.86
8294	19.46	313.7	117.8	3.04	33.51	36.14
13324	24.1	358.8	114.8	2.08	56.88	43
3201	14.1	291.7	145.9	1.97	39.14	42.74
13363	-	-	-	-	-	-
12116	29.12	358.2	139.9	3.45	53.85	52.44
10339	28.69	374.3	146.9	2.27	43.43	38.75
18293	23.77	348.5	131.9	1.88	42.94	40.72
6415	31.57	327.4	96.9	2.94	36.6	27.91
17493	24.7	366.8	124.9	1.64	35.52	37.93
16120	24.6	331.9	138.9	1.69	42.39	36.99
12138	17.35	392.1	122.6	4.32	34.38	32.32
5560	26.98	436.7	136	3.61	38.79	30.94
10446	15.19	313.6	121.6	1.74	37.39	43.41
14418	26.2	348	129.2	3.23	47.24	44.75
7536	18.01	321.6	134.2	1.98	103.17	65.96
12395	23.17	341.9	120.7	4.98	51.32	45.27
13459	18.1	301.5	118.7	5.14	34.69	32.1
7941	23.92	469.7	143.3	3.01	44.55	55.25
17150	-	-	-	-	-	-
11929	23.66	295.4	120.9	1.5	50.5	57.16
5900	24.04	385.6	126.9	3.88	40.37	36.56
13840	16.84	469.6	119.7	3.38	32.73	41.04
5931	31.94	353.2	128	1.72	44.45	39.58
4378	24.97	371	132.7	2.94	57.97	49.35
8786	24.94	357.2	123.9	2.48	46.62	47.25
19361	28.48	321.1	102.8	2.83	42.99	40.49
6892	29.05	434.5	130.9	4.49	59.54	45.11
18292	19.52	354	124.8	2.93	46.79	47.55
14311	21.53	387	130.6	3.55	47.26	38.87

10456	21.46	437.3	135.1	5.18	43.17	31.99
13520	19.23	349.3	143.8	2.72	40.55	44.94
6098	25.29	345.4	111.9	1.84	43.08	38.73
13290	22.48	367	105.7	1.93	21.9	20.72
8647	26.74	326.5	110.6	1.51	60.3	43.4
11763	-	-	-	-	-	-
6125	32.55	335.8	116.4	2.16	39.7	33.07
14624	18.14	324.9	134.3	2.41	52.69	45.37
17028	18.45	317	111.1	2.57	39.59	39.6
14210	22.95	387.1	125.9	2.82	35.12	31.47
9981	33.59	486.9	111.1	3.25	40.05	30.82
11379	26.03	454.7	111.2	2.61	40.67	41.17
14398	20.98	466.3	111.9	1.91	28.27	27.48
10759	25.96	336.3	129.4	3.05	43.52	38.31
15536	27.29	311.7	117.8	1.79	57.93	47.55

Table 10.8: Phytate data used for comparison between grown trials conducted in October 2014 and August 2015 (Section 4.4)

		Phytic Acid Content (g/100g)	
Genotype	**IP No.**	**Oct-14**	**Aug-15**
1006	IP 9407	1.125455864	0.561938706
1009	IP 13370	1.167267795	1.102690628
1016	IP 17690	1.020384431	0.771198944
1019	IP 13149	0.68155614	1.024625478
1021	IP 10140	1.370044829	0.907609239
1022	IP 12925	0.907080597	0.833292519
1032	IP 19386	1.159698623	1.494403812
1044	IP 21155	1.26939685	1.319794751
1045	IP 11211	1.007229169	1.293312736
1046	IP6745	1.535110331	1.63844014
1048	IP 10705	1.623090525	1.18415516
1050	AIMP 92901	0.891769016	0.972083736
1051	IP 18132	0.721912118	1.09243696
1052	PRLT 2/89-33	1.049837573	0.990814918
1053	IP4020	0.984768276	1.304938986
1054	IP 13971	1.152996216	1.430674734
1055	IP 8767	1.524565713	1.357041813
1057	863B-P2	1.43628572	1.144755088
1059	IP4542	1.193664527	1.056342934
1060	IP17720	1.35911512	0.999455199
2011	H 77/833-2-P5 (NT)	1.312693731	0.862562803
2019	IP 9391	1.003416184	1.117452021
2021	IP 3636	1.248616343	1.082073076
2022	IP 9710	1.047456989	1.02156704

2023	IP 3132	1.045810938	1.196049564
2024	IP 6037	1.219714811	1.121673373
2025	IP 5713	1.134155697	0.979956576
2030	ICMV 221= ICMV88904	1.170568065	0.983172843
2032	IP 9351	1.370338647	1.21233192
2033	IP 10539	1.027505267	0.915832238
2052	IP 8761	1.379690457	0.913799895
2060	IP 7095	0.96562414	0.838824221
3009	IP 6882	1.006612925	1.338722266
3015	IP 4828	1.715360749	0.513447865
3025	IP 8187	1.495294748	0.934034293
3056	IP 8972	0.840668039	1.191300219
4015	IP 13290	1.378894558	1.039722893
4034	IP 5900	1.173154735	0.630518314
4035	IP 19334	1.450127495	0.901695388
4048	IP 13324	1.206184533	0.365664489
4051	IP 5031	1.222102508	0.85020607
4054	IP 15553	1.152262394	0.861949248

Table 10.9: Phytate data from 249 PMiGAP lines (Section 4.4).

IP No.	**Phytic Acid (g/100g)**	**IP No.**	**Phytic Acid (g/100g)**
22295	1.0411	22424	0.6118
4545	0.8659	22281	0.4811
16403	0.6753	9282	0.3779
7846	0.6784	22494	0.6528
17720	0.7552	9824	0.5943
15917	0.7019	3616	0.7495
8767	0.6639	6102	0.7485
12298	0.7049	18434	0.6297
2058	0.7629	17611	0.6009
3125	0.7621	7967	0.6667
6584	0.7786	8172	0.5594
4020	0.6478	13344	0.6595
18132	0.8107	22419	0.285
16082	0.7892	13180	0.5023
18062	0.7077	16402	0.4237
22549	0.7701	3389	0.6786
6745	0.9296	8074	0.931
9595	0.9382	21169	0.4924
11211	0.7054	22420	0.6262
10705	0.84	7953	0.5275
22274	0.9244	10701	0.6159
22272	0.7253	10632	0.6033

9426	1.1315	13817	0.2614
7762	1.0826	8129	0.3154
15551	0.8577	10471	0.337
9446	0.7214	11593	0.5281
21155	0.6841	5131	0.4952
10811	0.6954	22283	0.8118
17632	0.8267	5923	0.8756
11577	0.7751	8187	0.3438
22454	0.5558	13927	0.7045
22279	0.6637	15872	0.7846
9351	0.8343	9692	0.8009
15947	0.4722	9651	0.6936
7970	0.6769	14439	0.9083
3735	0.9135	7922	0.9661
6112	1.0248	10761	0.9109
19448	0.6447	4828	0.7547
9242	0.3828	11275	0.6854
22569	0.585	22458	0.7012
16096	0.775	6417	0.987
7633	0.4854	4952	0.9009
8198	0.6556	10543	1.0054
6060	0.7139	9969	1.0784
8276	0.3107	6882	1.0359
22423	0.6079	14148	1.0451
7848	0.7807	5816	1.0972
9347	0.7107	8069	1.0361
22455	0.6731	14497	1.1406
3175	0.7043	18090	1.0948
12840	0.7758	6329	1.0675
19386	0.6743	7930	1.3032
9710	0.75	4927	1.0361
5389	0.6857	5253	1.0194
3890	0.8973	11346	0.9319
11677	0.9245	6310	1.1988
11984	0.7893	5031	1.0545
17690	0.7266	11358	0.7235
12925	0.7223	17125	0.7027
10140	0.7338	17099	0.7488
6037	0.7864	5121	0.7055
3122	0.8617	18168	0.7889
6146	1.1344	18246	0.6772
15533	0.8964	15553	0.7895
7942	0.9425	7660	0.8382
9301	0.7943	12370	0.9969
11311	0.8098	20679	0.8994

9407	0.9023	8294	0.7089
13154	0.786	13324	0.8824
19388	0.7068	3201	0.8964
17554	0.5014	13363	0.7921
9840	0.7771	12116	1.0071
3098	0.8518	10339	1.2028
22383	0.6811	18293	0.7968
11353	0.2493	6415	0.6629
10964	0.6322	17493	0.6881
18157	0.2691	16120	0.793
4965	0.7139	12138	0.8273
9496	0.3907	5560	0.9782
10820	0.769	10446	0.6486
8949	0.7804	14418	0.7715
8949	0.58	7536	0.7609
8181	0.5377	12395	0.623
11229	0.689	13459	0.7105
10343	0.7049	7941	0.836
15946	0.5563	17150	0.7134
20349	0.6888	11929	0.7934
13016	0.6206	5900	0.672
3865	0.7055	13840	0.6928
10486	0.6018	5931	0.6668
3557	0.6365	4378	0.7723
4962	0.7098	8786	0.7318
8344	0.6448	19361	0.8395
5713	0.801	6892	0.8277
5695	0.6161	18292	0.8206
12768	0.629	14311	0.8879
6179	0.4994	10456	1.1076
7095	0.6249	13520	0.8596
8761	0.6415	6098	0.6685
8955	0.6205	13290	0.6857
9406	0.5235	8647	0.7423
9532	0.6413	11763	0.782
8275	0.7992	6125	0.3609
8280	0.531	14624	1.0221
10379	0.3951	17028	0.9929
10271	1.0714	14210	0.7079
12322	0.7414	9981	0.7202
19626	0.7074	11379	0.9028
3564	0.5423	14398	0.6917
6111	0.8772	10759	0.7715
15320	0.6848	15536	0.6543
12845	0.8113	21020	0.8289

10953	0.7507	3138	1.0357
22276	0.7189	7952	0.8473
21206	0.7363	16638	0.7197
3636	0.6773	4974	0.7304
6099	0.7731	8972	0.7786
10613	0.8163	8426	0.7985
11310	0.6874	7470	0.8686
10085	0.8907	18389	0.8072
12839	0.7205	7886	0.9599
10394	0.6293	16289	0.9802
3732	0.7484	6682	0.4105
9391	0.6662	13384	0.6855
22527	0.6957		

Table 10.10 Key for Table 10.11.

Genotype number	G .No
Apigenin7-glucoside	A
Ethyl Ferulate	B
P-Coumaric Acid	C
Apigenin	D
Luteolin	E
Vitexin	F
Orientin	G
Lut-7-*O*-β-D-glucoside	H
Hydroxy Apigenin	I
Dicaffeoyl Spermidine	J
2"-O-Hex-C-Hex-Ap	K
C-O-Dihexosyl-Lut	L
C-Hex-C-Pent-Ap	M
Ap-GlucGluc	N

Table 10.11: HPLCMS peak area data for 55 PMiGAP lines (Section 5.6.3).

G.no	A	B	C	D	E	F	G	H	I	J	K	L	M	N
2046	424.0	229.1	17201.2	-	60.4	33907.6	6679.5	158.9	2255.5	34235.7	27334.8	31588.8	513.4	59.5
1027	251.5	592.5	74729.8	-	215.9	6975.5	13715.4	407.4	3184.1	44520.3	117852.1	54398.7	62.4	90.8
3022	2301.3	792.0	98035.0	-	518.7	54726.9	3804.9	461.8	5350.8	30520.5	391483.3	53931.4	425.0	1112.5
1023	48.7	677.7	63335.5	-		18209.6	15296.0	138.0	4281.3	40422.9	74397.8	43416.5	294.8	286.9
1045	3211.7	350.9	135688.5	-	114.2	12013.0	6624.6	100.4	2094.0	27923.5	169570.0	53359.8	314.1	76.7
3043		563.4	71039.2	-	86.2	5117.0	6673.2	344.8	4256.7	20447.9	77732.4	49202.4	71.3	41.3
3030	20479.7	138.8	98336.2	33.7	50.9	593035.4	3824.8	372.5	726.2	49584.9	337953.7	40782.8	5299.9	301.0
3047	504.3	54.3	25534.5	-	53.6	31433.5	88648.6	326.9	2122.0	78139.4	136176.9	52690.5	339.9	168.4
1013	83.9	103.7	189613.5	-	36.2	10202.4	3638.1	346.7	1500.7	39095.3	74112.7	44186.1	86.0	8.1
2011	24154.7	219.2	39785.2	-	261.1	363544.8	3766.1	1010.3	1341.1	15526.8	434175.6	52094.8	1906.5	654.0
2015	7473.0	271.7	19061.8	65.8	123.0	375957.1	3139.7	507.9	4339.9	17052.6	476228.8	45521.5	3247.0	609.0
1048	1005.9	115.3	68020.6	-	219.5	25935.2	101350.0	783.8	5104.0	32376.4	186428.0	57102.6	64.7	578.6
3020	4117.0	197.1	56187.0	-	185.1	133691.7	4615.8	544.2	3647.8	24225.0	230870.1	49825.7	931.2	104.1
3056	621.2	92.1	12426.9	-	24.8	73093.6	2788.0	126.2	383.5	38944.3	289777.9	34417.4	2188.2	134.0
1061	184.5	130.0	41673.3	-	79.1	38092.3	52606.4	510.4	824.2	28986.5	127786.4	54358.0	310.6	109.0
2018	165.1	185.2	15809.2	-	102.5	5640.7	5067.4	314.6	2503.8	52004.2	31148.2	49095.7	123.5	75.1
3007	709.9	183.8	23083.4	-	147.7	40294.3	6423.1	285.7	4577.8	13884.2	112275.6	43474.9	751.5	360.0
1014	4759.1	153.6	56426.0	-	128.3	176989.9	8535.4	811.5	2716.6	56574.7	443996.2	56790.9	864.2	676.5
3017	596.2	119.5	21150.7	-	69.8	72869.9	6244.3	612.9	4785.6	34790.7	372253.9	54004.5	1255.1	920.4
2043	1270.1	123.4	48786.9	-	36.2	162073.4	3070.7	440.6	3739.7	49056.1	327935.3	49020.0	1863.9	585.2
3046	666.1	169.2	28968.5	-	95.6	10220.4	4385.7	457.2	1882.5	40796.9	65550.6	49532.2	185.6	90.7
1022	7204.7	233.3	33530.9	-	285.2	480180.1	5778.7	751.4	9371.1	21539.7	330773.9	49494.3	2379.7	623.5
3010	231.6	156.8	13302.0	-	214.3	40986.5	13379.9	565.3	34653.6	22057.1	225290.1	59949.8	208.5	133.6

3036	760.3	125.9	22069.5	-	29.8	10270.0	4947.5	234.4	2225.7	37185.4	148666.8	50514.5	187.0	68.2
2038	5723.2	121.9	37875.3	-	26.9	227228.4	2926.8	288.8	776.4	39522.0	420142.0	46971.9	1790.4	437.5
1026	891.0	84.8	37053.2	-	31.6	68200.4	2823.8	373.0	3031.7	12336.9	355665.2	52692.5	1369.0	281.3
3025	1667.6	236.6	118517.9	-	158.7	50524.1	4079.1	656.5	1550.8	58588.5	326567.1	45142.5	881.4	195.4
2029	408.7	156.2	28493.9	-	204.1	49755.3	40570.4	398.6	3869.9	33439.5	47130.1	40594.4	442.9	119.1
1001	694.0	170.8	53073.4	-	2819.6	24966.9	67630.4	2195.4	2079.1	27337.5	210104.3	59362.8	11.1	708.0
1018	858.5	74.3	27967.2	-	357.2	26290.7	10863.4	935.6	3017.3	10326.5	225794.3	59025.3	56.1	275.6
2025	1270.9	167.2	54492.5	-	138.7	51458.4	24758.8	610.1	610.2	15721.8	115410.8	45831.5	422.5	149.0
1042	5748.0	179.9	30148.9	-	107.7	262104.0	2522.4	97.3	1185.8	26395.8	411582.7	49106.8	2880.6	1293.3
2014	200.7	211.5	90112.9	-	21.1	11134.0	10408.6	776.2	2578.9	28390.8	25949.2	49668.9	237.6	105.0
2041	16876.3	100.9	36667.2	52.0	318.8	409002.8	5747.0	615.1	2291.7	16999.6	357546.0	29031.7	4384.3	190.3
1044	1254.1	187.2	25330.6	-	60.8	88426.0	3006.0	222.7	1967.5	15795.9	354650.8	41770.0	2104.2	368.5
2017	1630.4	139.4	51380.3	-	67.3	32220.7	8251.3	452.7	3640.5	53539.0	239085.4	58602.1	293.9	217.5
2040	2176.9	170.4	27177.6	-	176.2	177382.9	5534.4	401.8	3323.3	40498.1	398872.3	46180.3	2763.0	471.7
1002	2657.2	409.8	14850.6	-	41.5	251793.3	4957.2	264.4	4739.5	23326.3	382808.4	49366.0	2619.4	390.7
3057	562.6	88.4	104313.3	-	900.2	11057.6	31495.7	566.7	3994.1	58585.8	164514.2	62154.0	37.2	210.1
3011	24493.5	266.0	46392.1	-	242.3	635497.1	1778.5	859.9	1098.5	10865.7	428213.2	51136.8	4398.0	650.4
3050	99.1	170.8	84967.7	-	123.2	51665.6	4806.7	622.5	1644.7	60203.4	284934.9	46201.5	1491.0	250.2
2059	1112.4	132.0	71768.2	-	15.2	18660.5	4990.0	224.8	678.5	25689.3	251153.8	50200.9	384.7	68.4
2002	1075.0	184.1	20766.9	-	49.5	27964.7	5266.7	181.7	2034.5	70506.0	385321.5	54400.0	640.2	376.1
3051	2284.9	129.8	49356.0	25.8	37.0	52742.8	2431.3	351.3	10147.2	21425.0	399192.5	55330.1	527.5	645.7
2022	3579.1	133.8	35203.7	-	409.1	312035.1	49794.2	580.5	2899.2	16429.4	344693.7	56481.3	932.2	499.2
3016	175.4	119.1	38468.8	-	38.1	16821.8	7481.0	769.0	3917.7	46893.1	87764.9	57927.4	96.0	225.2
1049	1229.6	110.1	17614.0	-	17.6	41080.3	3393.5	262.9	2912.0	30937.3	300090.7	36618.8	1489.2	1040.0
1006	1270.4	256.0	186661.5	20.9	87.0	18906.0	4750.1	347.3	1989.1	46417.7	74773.6	52609.6	193.7	47.3
1029	545.8	157.8	34863.9	-	34.8	81563.6	6290.3	297.3	1316.2	21729.7	335314.3	46188.1	941.8	340.5

3060	950.0	206.9	41193.1	-	36.4	118706.9	5575.2	492.5	1617.7	20013.8	355946.3	53586.5	408.2	295.6
1009	1568.2	149.1	46917.8	-	64.3	206769.7	12884.7	1570.1	1125.5	25979.1	307489.8	22460.1	2477.8	118.5
3005	2257.7	183.4	67571.2	-	73.3	111157.6	2606.7	260.2	2860.4	29346.1	388121.4	49292.3	653.4	409.4
2058	56.9	150.5	166853.1	-	46.0	14786.5	6665.6	173.8	4130.5	19222.3	57951.8	49953.1	83.4	111.5
3009	2287.0	166.8	178494.8	-	39.3	118931.3	15145.6	522.3	1238.6	29075.8	184425.4	46196.1	828.1	275.5
1062	5226.1	166.5	31565.5	-	47.8	468331.0	1120.9	55.0	1922.2	56879.1	392365.5	34001.8	4852.3	126.8

Table 10.12: Raw data used to generate calibration curves (Section 5.6.5)

Apigenin 7 glucoside		**Luteolin 7 Glucoside**	
Concentration (µg/ml)	Peak Area	Concentration (µg/ml)	Peak Area
0.1	22615.721	0.1	268.435
0.5	45453.418	0.5	687.745
1	74052.766	1	685.819
5	151270.547	5	1902.898
10	700406.188	10	5032.957

Biochanin A		**Vitexin**	
Concentration (µg/ml)	Peak Area	Concentration (µg/ml)	Peak Area
0.1	57158.617	0.1	5903.256
0.5	107928.664	0.5	25767.549
1	148859.406	1	22072.801
5	349052.188	5	129518.633
10	1172363.375	10	388052.906

***P*-Coumaric Acid**		**Luteolin**	
Concentration (µg/ml)	Peak Area	Concentration (µg/ml)	Peak Area
0.1	5172.164	0.1	992.997
0.5	8398.861	0.5	2110.707
1	13695.444	1	2885.103
5	39441.992	5	7642.998
10	147107.75	10	23313.98

Table 10.13: HPLCMS peak area data for 185 PMiGAP lines (Section 5.6.5).

IP No.	Apigenin7-glucoside	Biochanin A	*P*-Coumaric Acid	Luteolin	Vitexin	Orientin	Lut-7-*O*-β-D-glucoside	Hydroxy Apigenin	Dicaffeoyl Spermidine	2"-O-Hex-C-Hex-Ap	C-O-Dihexosyl-Lut	C-Hex-C-Pent-Ap	Ap-Gluc-Gluc
IP6329	1645.145	157835.4	80575.09	227.784	81204.06	14246.7	101.572	837.689	43016.82	115979.2	33002.26	113.47	0
IP6417	5003.049	180632.8	52390.47	380.267	37705.36	13155.73	238.003	1368.97	31581.55	128904.2	41103.62	244.47	32.612
IP6099	6319.898	195225.9	38296.41	290.8	211105.7	8346.151	277.402	495.338	35593.92	307642.8	27857.49	1923.208	68.662
IP6098	18699.53	164301.6	70459.22	242.518	105526.5	8039.368	531.81	985.112	34185.68	386792.7	39442.2	841.276	97.186
IP6415	10517.37	144185.5	82468.57	303.029	22516.75	26278.87	450.511	5419.644	30208.53	75251.21	55315.27	96.135	38.172
IP6310	2962.382	183759.9	116200.5	215.273	83539.8	10299.25	290.875	1065.033	62601.68	171982.1	37188.86	284.057	41.604
IP2058	83538.67	154166.2	61716.51	1061.264	178430.8	5252.119	918.736	560.195	49998.56	188792.1	23361.01	1077.077	0
IP6060	17652.2	166254.7	72094.98	281.424	111282.5	5686.311	523.042	392.357	39120.18	329220.7	35358.03	882.144	68.465
IP5560	7414.483	177198.8	60767.47	225.357	35685.72	14541	87.126	852.112	39059.55	144002.4	39146.93	290.749	2.917
IP5441	1429.784	157520.5	98492.66	152.256	470494.8	14467.12	168.174	608.006	86162.92	220413.8	25982.7	2821.273	37.179
IP6179	1937.378	142076	95936.05	142.243	54561.88	18845.76	138.178	351.287	54468.01	68098.36	25965.08	391.615	22.193
IP6112	1922.672	169617.6	37400.56	156.683	154238.2	9263.37	87.86	774.592	41885.19	232628.6	24436.59	1095.635	43.298
IP6111	9414.957	174671.5	74724.66	58.935	55559.27	4693.494	103.52	372.925	65375.63	90544.05	22915.59	608.229	41.197
IP5438	2503.355	157014.5	127092.5	680.501	101439.2	254014.8	461.474	821.594	69375.4	55026.99	49615.77	585.34	89.868
IP5900	11681.23	144395.3	94021.38	299.459	146959.6	30573.75	412.308	531.527	84027.82	249400	24042.62	1577.78	165.545
IP6101	8332.442	161952.4	64916.23	153.041	610366.2	15972.42	139.542	368.9	48273.28	357224.6	19267.64	5464.148	51.061
IP10471	3159.563	200025.9	51492.6	218.717	218296.4	127203.1	201.817	1100.392	39607.24	247083.9	37961.07	1442.23	35.51
IP6125	305.261	154074.4	82436.88	91.151	24142.95	28299.45	327.533	543.317	49405.17	37788.33	30837.79	411.2	0
IP6146	6329.711	166812	69435.83	225.522	358316.3	12219.37	148.597	539.397	64896.75	418920.2	31046.57	2314.129	149.059
IP6110	7278.802	158723.3	62485.78	168.559	175137.7	4074.576	352.401	627.121	38111.15	410008	34488.52	3138.28	219.071
IP5931	13142.13	183924.9	42555.67	307.367	77508.35	8462.107	269.378	1197.863	68366.31	143312.7	28208.56	804.865	9.468
IP5121	8962.738	186978.5	45058.47	171.855	55111.47	7993.11	234.673	1055.339	36000.42	165755	31334.49	1159.449	104.675

IP5931	9154.813	176933.5	51005.25	522.366	527420.1	54179.22	408.889	921.297	50236.05	317209.6	40390.81	2110.928	165.973
IP5695	17212.48	166207	38665.91	149.935	332293.4	26302.79	241.45	636.688	53629.44	328095.2	23069.68	3046.258	23.094
IP10632	4220.88	163230.8	96142.75	187.074	49861.09	32843.8	98.611	1172.417	40093.51	78465.29	27289.41	298.484	86.892
IP6102	2024.556	189894.3	32809.36	185.828	48247.41	40044.45	223.434	856.776	22631.23	49794.09	36510.07	238.641	27.503
IP5272	3522.62	180193.9	80197.72	357.376	18470.49	19733.67	323.424	856.682	49681.08	50038.54	37883.15	291.647	38.265
IP10701	6028.233	176214	66180.09	244.988	14073.67	11823.02	250.424	894.936	45937.37	57677.06	32645.35	270.734	31.711
IP10456	1658.552	202576.9	72247.01	145.735	25714.18	23967.81	247.396	1378.979	83844.38	105992.3	39161.32	75.754	33.477
IP10579	3240.084	160137.9	146026.8	317.046	36707.36	34373.91	332.112	4347.632	40882.7	177015.9	53812.72	129.604	27.476
IP5031	1589.509	184008.8	44476.44	342.022	12635.93	12752.55	278.24	1140.284	16087.93	64459.7	36940.59	247.67	29.588
IP5207	24683.44	138550.4	98364.06	331.879	379192.2	24688.59	221.8	496.108	49155.16	374409.5	25879.76	2517.239	48.692
IP10271	5240.07	180420.8	59891.88	148.002	137462.1	7124.015	132.299	357.54	99951.66	266065.9	25190.09	1614.781	17.885
IP10343	6580.595	184698.2	66599.44	117.695	265956.4	4272.171	257.89	643.465	49942.59	296404.8	21788.73	3023.188	66.789
IP10705	7672.576	218242.8	48229.1	742.059	67824.49	19584.77	452.804	1221.1	28909.66	262327.9	49570.87	746.94	69.471
IP10394	5102.194	202426.8	65810.73	269.922	46482.8	4146.73	188.093	1044.67	78096.74	201403.4	26330.64	453.663	49.35
IP5816	3825.25	194662.9	68115.88	258.924	335316.9	68644.41	792.652	5112.968	37408.51	212687	49711.11	414.052	58.517
IP10486	14479.87	202773.4	59900.99	174.552	118888.5	10210.3	295.696	640.262	46819.61	254247.6	27282.53	1931.852	95.775
IP10446	5092.985	231691.6	74955.03	206.209	44980.73	6608.597	233.722	894.001	62050.22	186100.6	27344.26	723.613	54.275
IP10539	25434.83	194342	46551.51	445.175	34545.34	60651.14	323.227	2908.119	39603.07	90706.09	48802.96	280.83	174.399
IP10379	5770.098	200155.7	80463.01	213.584	16068.18	11741.74	169.668	538.313	49818.25	87990.83	36078.14	225.31	31.19
IP5131	2968.269	195494.1	36523.47	226.791	184436.7	13970.7	254.119	866.86	42376.33	393847.8	39052.11	2125.942	274.962
IP4020	6763.792	235745.3	38115.51	350.718	69207.23	17671.11	509.337	750.172	49311.77	243689	48659.13	511.317	72.827
IP4979	6555.727	198044.2	149250	270.454	28993.44	10342.2	355.659	669.283	109233.2	138830.2	40321.74	276.232	62.862
IP4962	1818.82	226741.3	116181.8	74.059	18747.1	4440.772	74.368	461.416	93921.26	214498.3	28661.66	233.24	0
IP3557	13035.93	242731.4	67008.02	92.875	269113.8	3189.452	122.987	636.323	8395.161	247743.3	22498.58	37.194	42.005
IP4378	2309.795	210288.2	101739.5	453.242	39052.4	65527.15	435.51	666.353	58361.32	59691.85	43369.96	333.999	47.87
IP4952	2902.195	199265.4	67830.87	275.142	65069.49	73412.25	306.034	3158.046	42488.17	124028.4	51316.06	516.098	59.424

IP4927	8634.646	242025	77868.4	432.978	67965.9	16810.51	488.57	2351.406	65668.77	311157.5	48436.36	621.278	97.49
IP3757	1406.107	241572	55122.97	57.286	23807.77	64225.14	111.11	823.929	110597.4	62430.66	36075.69	355.32	12.506
IP3389	4216.125	205222.4	116047.4	122.985	93529.57	8720.015	339.898	424.683	112039.4	239525.3	40040.99	881.97	29.592
IP4542	1489.926	208935.1	74153.56	240.274	28369.13	25499.54	292.507	919.697	88801.06	115582.8	50722.47	168.71	50.858
IP3636	1891.067	221914	85279.6	70.418	12513.39	10915.88	221.77	1207.079	94941.89	61271.66	30752.4	247.885	7.757
IP3175	2391.021	211009.5	48693.02	309.55	44308.1	15328.24	321.944	1072.545	76217.99	150532.5	46537	386.01	39.077
IP3616	30592.7	198404.2	108241.9	246.607	82946.39	6717.936	217.404	737.564	100501.9	433902.3	41566.39	1219.107	81.484
IP3732	3959.671	200804.7	69726.8	100.404	16509.85	7110.624	230.362	430.836	43175.19	122126.2	21499.73	463.732	72.153
IP3890	3104.463	206089.3	116210.5	181.615	278136.6	87828.87	194.487	860.763	72832.2	121058	31321.99	1771.813	50.175
IP3138	2704.086	192090.2	44901.69	160.763	43414.72	15292.51	198.861	422.567	28770.41	154503.5	37385.88	387.73	0
IP3471	7596.9	201629.1	62277.58	112.547	197956.4	7621.892	274.632	716.083	116938.5	295924.6	24917.59	706.563	17.321
IP3201	1331.492	174096.2	75327.56	153.179	28241.73	7365.746	214.501	518.542	33946.7	94827.86	32604.63	118.414	63.429
IP3122	4759.418	184740.8	63854.88	178.71	48578.07	4745.147	327.101	513.187	69965.85	328217.3	39703.58	551.866	70.098
IP3865	4875.609	196755.7	102049.5	239.54	111470.6	24464.21	289.905	1004.099	62925.32	243489	33382.1	679.876	84.75
IP3125	11789.8	185772.2	77347.49	313.867	67073.93	26546.48	270.1	873.101	74282.7	227187.6	36961.51	540.426	95.775
IP3110	1472.853	186655.9	86915.06	168.042	25098.42	19801.27	300.504	1268.399	44553.38	155697	44981.24	117.14	34.108
IP4828	1333.767	188241	108014.3	75.863	13877.52	10911.65	195.577	360.702	72665.74	40174.24	31832.52	196.338	25.324
IP4965	8935.301	208429.3	61932.77	145.794	91895.59	5654.749	236.167	529.928	75129.31	342293.2	27516.23	1075.402	124.591
IP3564	14664.53	213669	75051.34	300.363	181473	9395.413	448.569	3079.156	64477.77	379704.8	41523.94	1141.087	131.164
IP3098	13738.54	170378.7	43904.79	104.184	127338.1	6882.664	172.958	2042.03	35367.29	328974.7	34053.52	1654.949	47.897
IP7967	3067.153	196306.9	66686.13	602.729	56614.33	36135.78	301.588	1262.059	114682	161765.1	54874.51	222.62	57.947
IP11218	9414.066	188899	149893.4	264.177	50599.88	32683.85	383.429	1028.087	99983.73	144306.6	43273.72	375.752	18.229
IP7952	4642.054	190491.4	101248	344.602	50407.27	76653.11	138.75	548.926	86535.46	55264.06	35916.94	937.371	25.255
IP7886	2761.229	183598.4	59562.24	177.265	129186.5	9212.189	565.302	881.209	151069.8	304094.2	32782.57	1858.68	71.548
IP7941	1208.789	194369.9	66518.84	0	19955.55	4108.629	88.729	191.066	52099.52	70571.59	14418.64	431.24	0
IP7470	2261.882	213100.8	80880.16	149.906	44423.64	8344.684	305.099	1485.34	108047.6	158737.5	33796.82	692.171	0

IP7762	1959.472	188772.3	61469.56	147.494	151002.1	9610.339	126.991	590.308	27931.14	232998.5	28824.51	668.438	39.416
IP7910	2579.752	200417.2	93539.62	4211.053	82363.34	97104.62	878.845	1292.382	158454.7	152825.7	61105.19	79.473	190.069
IP7660	4234.715	208202.8	132600.9	727.77	34255.38	41612.5	811.219	783.811	93605.56	184821.2	60558.66	119.65	84.126
IP7922	1658.554	220279.3	119070.7	174.2	25526.98	14230.17	183.887	884.95	98451.95	111037.4	42929.74	148.169	13.857
IP7108	1438.132	177759.7	119447.3	128.761	15143.84	6783.913	181.195	451.077	129104.9	55879.77	26957.9	85.111	14.007
IP7095	472.656	169267.8	50467.61	132.756	258522.5	11450.66	99.039	1194.1	83766.51	278413.3	31012.63	2111.994	67.398
IP7970	9221.104	186347	81493.26	341.068	559916.6	13984.19	13834.47	1134.339	66362.34	288650.7	34341.77	3356.496	105.845
IP7942	1831.092	194109.2	101499.8	173.616	84555.04	28358.73	1294.527	927.53	54134.91	54998.33	27865.52	653.986	45.314
IP7633	6239.901	237719.5	50770.57	110.088	192858.5	4354.096	386.375	453.774	150981.7	336852.9	23296.25	1832.689	83.734
IP7536	1673.346	206163.4	45719.04	95.4	33819.34	7317.267	228.771	598.908	70487.06	90540.96	36255.53	273.221	20.634
IP7846	9472.941	186152.4	28619.01	196.026	385305	4507.557	248.399	807.555	77464.97	313119.7	33132.43	3506.309	209.788
IP6882	1632.131	202910.8	28813.44	285.239	154235.3	47020.48	3253.322	2418.844	64125.75	115198.4	39171.25	349.819	52.336
IP6892	1812.398	201951.6	24140.67	211.997	232376	55733.48	418.414	1048.26	65171.47	234115.5	36461.4	1080.672	54.575
IP6682	2302.592	226591.3	24962.39	122.047	90771.77	10237.97	611.331	458.731	90617.84	312427.2	38854.82	558.376	151.441
IP9242	26628.76	220648.2	68605.55	409.544	370906.1	26931.02	279.964	1063.227	125859.9	382719	43394.44	1880.106	258.557
IP7930	3434.901	224252.7	54236.03	120.056	163437	3665.486	250.556	1521.395	88801.6	245868.1	27500.24	1400.402	46.155
IP8767	6765.158	197239.8	66877.07	215.477	354125.8	2348.639	215.515	611.88	131182.8	252595.7	18459.19	3350.708	0
IP6869	1274.189	170860.9	34599.73	172.56	19557.29	8298.262	530.429	1279.195	57337.27	59870.83	42760.87	152.176	24.296
IP9407	21346.38	196631.1	32964.81	129.881	523561.2	2837.802	196.078	798.757	50490.76	393691.4	23533.83	6683.259	111.835
IP9406	4189.62	167346.4	38831.04	249.512	96017.33	48629.04	307.224	1302.026	6580.998	193676	38786.45	842.606	46.955
IP8955	3499.345	182273.1	24353.07	80.951	137935.8	6890.48	107.282	905.203	23742.86	248733.8	20644.06	4303.639	81.932
IP6584	6001.411	185576.1	38391.89	172.359	163736.9	6522.635	207.901	357.341	90749.59	286362.1	26055.87	1135.446	59.359
IP8426	2997.566	199488.3	48747.26	117.266	59629.74	7722.847	362.061	648.355	80998.68	199289.3	23017.35	1481.272	24.84
IP7848	4566.865	183483.1	24379.01	81.905	33334.31	4742.471	94.588	557.024	80535.48	102314.7	27048.44	442.971	0
IP9391	35400.95	241999.1	29668.46	105.139	728385.8	4492.913	291.847	696.262	44961.8	433519.5	23365.78	5293.864	123.378
IP9347	3189.438	232218.1	27465.55	148.88	306243.9	26488.69	210.827	796.184	48411.63	288094.8	29439.23	1654.068	18.629

IP8786	12637.96	202214	75687.91	385.826	340993.9	13324.76	458.799	976.427	78399.84	354420.8	32626.87	3510.823	113.343
IP6460	3086.071	187967.5	36546.64	285.963	107798.5	110874.9	317.93	695.085	45434.31	181315.9	37462.82	802.912	10.955
IP9301	7432.429	189826	112354.1	103.37	485938.6	4405.161	203.794	502.769	91207.99	363311.2	22495.9	3961.554	80.037
IP6745	1398.786	160362.6	41524.06	111.3	21301.63	6709.332	350.208	757.052	113440	64842.59	31908.13	502.302	24.87
IP8210	6751.364	200970	58844.58	267.787	40827.48	42774.24	219.271	1156.151	82217.91	126182.3	49118.67	316.865	18.163
IP8182	843.277	230383.5	41807.32	245.979	10652.96	9288.616	192.039	879.723	2445.384	58604.86	32238.01	237.942	46.569
IP8172	3861.242	196022.9	74396.66	283.453	550735.1	20116.17	413.842	902.252	90842.71	248364.8	44581.78	2132.938	0
IP8276	4719.157	194628.1	58672.63	88.352	53662.7	4863.119	531.334	647.799	116944.2	288315.9	26205.42	1072.821	65.812
IP9282	6671.009	177849	54660.61	115.652	154256.7	6143.858	895.676	594.787	59770.56	271724.7	29400.23	1007.301	138.666
IP8181	2231.829	213209	59645.32	119.296	155606	4104.254	107.796	389.551	87493.42	227922	19088.75	1243.574	51.439
IP8166	51041.3	186368.5	85937.64	726.594	782306.1	64986.25	13598.67	733.145	134512.9	362103.3	27998.35	5669.229	221.053
IP8174	5553.342	182990.5	64107.31	112.358	73927.49	5777.571	1122.564	431.69	34649.01	237174.3	29405.02	823.732	15.906
IP8294	4739.466	226915.9	67571.01	345.997	112743.9	10668.54	358.338	729.736	82270.76	306994.8	36165.02	683.22	33.501
IP8275	3331.176	211978.8	143944.5	274.121	25863.02	7046.626	233.387	711.086	101064.3	96801.65	29439.07	357.861	38.793
IP8863	28735.09	192181.2	83654.13	75.802	77107.99	2919.968	239.122	396.564	118975	346302.2	19449.17	2149.644	0
IP8187	10249.52	185658.5	49022.4	303.653	109366.6	6791.334	235.944	570.697	115222.3	299961.8	28667.12	1212.176	65.272
IP8647	103979.8	203502.6	115485.8	387.879	693728.5	5094.607	201.141	243.355	141148.2	348549.1	18932.79	8279.042	80.349
IP8761	9800.999	191311.7	104994.9	380.039	49864.81	12096.6	165.594	681.598	46937.21	94470.35	32611.1	583.386	48.03
IP8280	5179.019	195349.2	62222.26	414.899	21762.92	4978.492	238.833	830.682	136981.2	60208.88	25543.31	480.186	17.585
IP10085	2354.525	177314.9	48697.71	170.521	10726.29	4256.678	127.611	528.624	38398.47	28649.86	17437.28	268.956	4.544
IP9840	1789.349	181530.1	121458.9	103.978	31058.58	14558.77	110.969	1038.891	112018	31788.18	25957.87	271.478	56.697
IP9496	10949.87	189833.8	57867.23	247.037	252573.2	12243.92	394.71	1059.627	45929.84	311731.2	34907.65	1299.606	177.113
IP9446	792.224	174418.5	33143.36	91.546	51893.94	19595.61	250.021	349.063	22633.05	97754.35	24019.53	534.385	6.131
IP9692	4441.453	192807.5	59414.38	79.002	43675.4	9362.751	195.744	233.026	28335.14	110252.2	21370.91	480.508	44.28
IP9426	4935.589	195035.4	79230.78	194.383	248815.4	5937.634	355.699	386.034	71254.14	281393.5	22744.15	1345.197	83.376
IP10339	7758.763	184334.5	67540.87	72.864	102590.6	2629.785	144.526	191.361	50156.47	314732.3	15047.49	734.795	93.746

IP9710	2380.661	172936.3	101898.5	226.735	29834.67	13634.13	420.921	1079.963	61253.08	113863.3	37167.31	182.196	23.842
IP9651	15736.07	237451.3	94743.71	356.577	156953.2	14476.73	1102.511	974.697	67935.72	368978.7	37452.22	523.204	329.389
IP9595	12618.8	195510.5	89666.84	406	153792.7	156563.8	393.391	1936.707	98537.05	113570.5	36679.53	621.465	0
IP8074	4109.07	227685.6	119249	204.758	291993.5	35758.02	250.639	377.393	141557.2	198011.2	26376.09	1465.881	163.883
IP10820	4283.882	222974.9	84170.27	895.528	284763.8	74645.14	488.618	339.213	77053.31	146271.8	48066.84	139.673	226.406
IP11229	3775.3	205896.2	60186.59	223.611	56094.73	17630.41	217.749	435.669	69446.38	184968	28322.34	323.667	23.889
IP9981	2655.77	213922.6	33798.06	652.331	145400.8	34396.52	414.124	3014.835	77035.47	227434.2	43642.35	423.898	44.235
IP9532	574.529	192326.3	61613.93	131.978	32992.95	10038.45	300.516	1438.39	43501.66	93657.2	38790.9	128.127	26.812
IP10964	9472.102	221082.3	44268.81	290.994	231052.3	4893.832	98.031	487.695	34596.81	305028.7	26620.54	1627.862	43.281
IP10759	36979.68	183366.8	44931.48	306.698	783605.8	3685.002	189.096	123.28	69661.13	379283	16716.95	4130.643	48.36
IP9426	2752.439	214735.6	113266.1	547.909	85372.12	48556.59	632.457	2484.505	83745.93	174017.2	47004.49	260.939	102.9
IP8069	1420.877	201099.5	61272.18	311.089	71515.74	32734.73	2017.218	942.889	48089.68	72835.2	28392.54	432.075	32.635
IP11275	728.304	214714.3	99607.57	267.779	41092.36	86036.35	542.279	4160.947	6621.512	53992.93	50844.52	176.558	19.847
IP10953	5084.763	199084	46221.02	839.496	86168.88	168183	557.316	3827.709	61948.9	144204.2	53980.22	279.417	0
IP10945	13194.94	224451.2	88566.31	213.909	56048.67	26886.2	256.908	649.083	59218.98	312892.5	34906.54	460.547	73.156
IP12128	3327.876	222160.6	95096.54	79.178	408059.2	28319.34	221.7	396.812	108031.6	317986.8	30623.01	57.884	18.15
IP10140	3639.201	213357.9	68938.31	207.871	82601.2	24747.34	302.114	1899.35	57917.35	187045.1	44722.27	438.092	83.307
IP9971	6223.98	240027.9	63128.95	320.231	66385.62	36516.22	464.655	703.687	131655.7	173678.5	48470.96	406.136	53.113
IP10811	3308.762	224824.3	63001.27	211.59	91653.74	16198.42	337.759	1230.107	113019.1	259382.7	42512.35	775.292	33.453
IP10761	1983.282	228933.3	115466.3	721.047	91657.91	103565.3	533.304	673.541	133836.6	204956.9	66022.91	313.32	27.903
IP11929	10449.23	219381.9	95320.26	387.801	373227.7	12477.65	229.843	521.895	84867.45	297699.2	32752.36	1721.718	0
IP9969	6496.104	250140.4	83500.53	171.791	85763.95	69582.29	503.264	1401.634	124557.6	120658.8	48590.19	375.981	0
IP7953	9282.616	185885.8	143761.2	329.067	50125.29	71753.96	419.393	584.97	70357.43	54510.54	33271.51	457.635	76.932
IP11311	2702.631	200186.4	99020.56	361.68	18453.64	17478.71	6505.937	551.811	85390.07	70283.46	26666	167.974	62.646
IP11984	3419.987	205695.5	43180.56	116.017	56562.62	13283.93	407.835	419.65	84499.84	154955.5	23940.03	801.017	36.797
IP11577	1930.148	187508.8	37213.26	202.106	94370.98	13160.14	431.901	567.142	90773.57	193901.8	31625.94	842.824	46.713

IP12116	15156.91	220305.6	44155.81	217.013	325577	9925.148	328.55	862.026	33909.1	442000.1	35598.24	806.356	389.794
IP11310	4671.771	205659.9	37027.16	387.106	71469.64	18298.92	216.293	701.108	41902.32	205400.5	33689.3	391.293	152.523
IP11353	8919.885	216727.6	121154.3	199.495	209644.3	7831.412	363.827	391.645	92975.5	345330.7	30902.24	1278.269	67.502
IP11211	2805.648	210484.1	60270.91	570.714	67208.98	10027.91	202.618	1150.232	57171.01	285800	37246.47	505.777	79.739
IP11584	10133.52	237274.6	142063.2	159.715	237153.2	3384.778	305.003	1311.692	136153.7	364798.8	26612.2	2237.844	0
IP12138	32418.54	189165.1	56051.2	325.154	261219.5	5881.82	325.12	896.855	9566.629	401693	30639.7	1261.403	196.015
IP11593	1812.938	213134.6	42289.31	386.751	52590.23	45703.52	321.39	4179.971	82281.19	171074.1	47582.49	307.619	19.371
IP11358	2939.789	203690.4	64360.85	181.617	56486.65	6713.52	221.013	490.662	90246.86	87200.41	18260.43	411.02	14.11
IP11765	4585.919	202204.5	48097.07	223.436	31234.37	32088.42	2309.592	1053.936	78667.22	104975.9	34820.31	211.465	0
IP11378	1883.605	220253.9	87418.06	83.83	60522.6	35820.27	246.068	544.926	98310.23	182711.3	19305.54	931.387	20.504
IP11346	9843.745	212752	48551.06	345.423	51433.54	85489.76	861.237	531.51	37853.45	49574.1	31059.89	902.217	88.519
IP11584	5876.855	219233	38946.49	223.593	306299.8	13524.05	368.863	245.3	8269.553	335398.3	21365.66	1679.395	91.595

Illumina Sequencing FASTQ Quality Control / Quality Assurance Report (663 SNP data set from foxtail millet):

Number of Samples: 343

Total Reads in Project: 664058567

Mean Reads per Sample: 1936030.8

Median Reads per Sample: 1718170

Read Goal per Sample: 1000000

Std. Dev of Reads per Sample: 1663440.1

Number of Samples Meeting Read Goals: 248

Sequencing Coefficient of Variation: 0.86

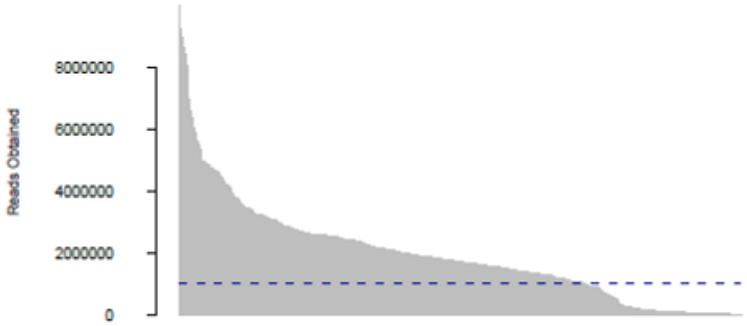
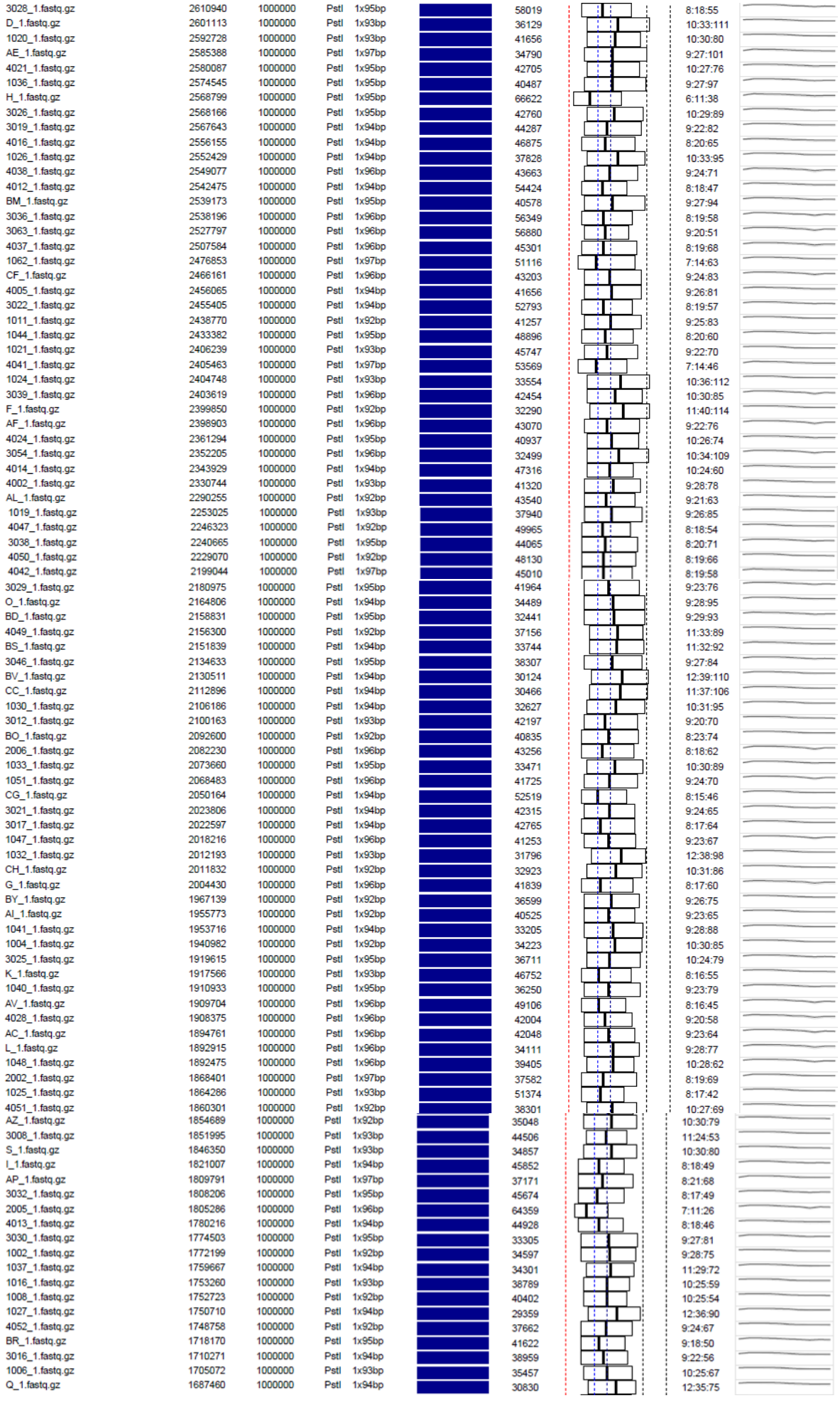
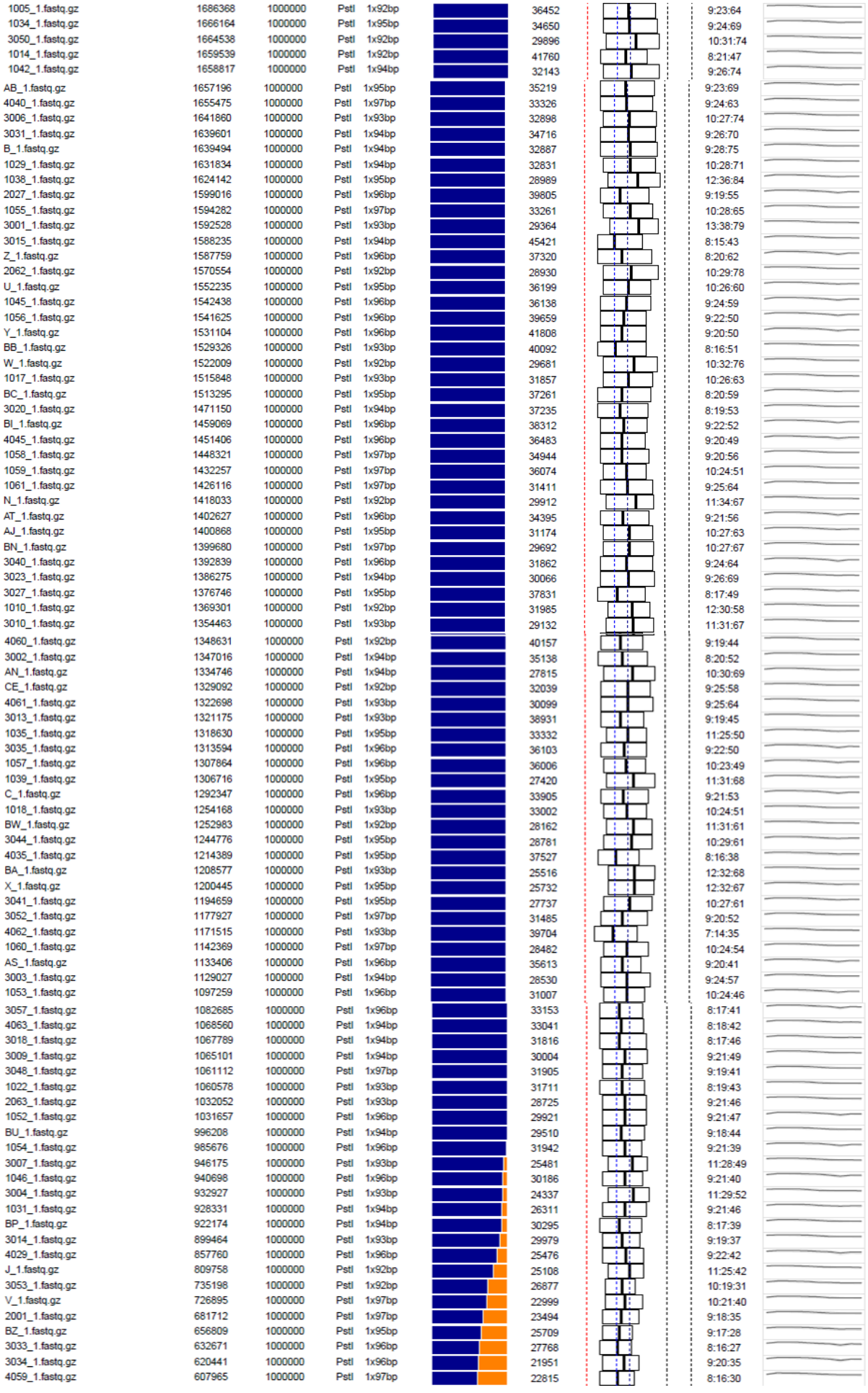


Figure 10.1: Sequencing distribution of 343 PMiGAP lines.

E) Individual Sample Sequencing Performance

Sample Name	NGS Reads	Read Goal	RE	NGS Type	Sequencing Distribution (0 % … 100 % of goal)	RAD Clusters (5x to 1000x)	Coverage Distribution (5x … 15x … 25x … 100x … 250x)	Coverages: Low. Quartile: Median: Upp. Quartile	Median Phred QScore vs Read Length (0bp … 92bp)
3047_1.fastq.gz	9985549	1000000	PstI	1x92bp		106215		8:15:49	
2003_1.fastq.gz	9249944	1000000	PstI	1x92bp		64075		9:29:178	
2011_1.fastq.gz	8963114	1000000	PstI	1x92bp		85898		9:27:93	
3055_1.fastq.gz	8665766	1000000	PstI	1x92bp		62599		9:31:178	
3060_1.fastq.gz	8417623	1000000	PstI	1x92bp		66251		9:26:160	
4032_1.fastq.gz	8024416	1000000	PstI	1x96bp		75867		8:20:107	
4036_1.fastq.gz	6957230	1000000	PstI	1x96bp		62070		9:27:133	
3056_1.fastq.gz	6591226	1000000	PstI	1x92bp		49145		8:33:185	
2020_1.fastq.gz	6356510	1000000	PstI	1x93bp		68795		9:25:104	
3059_1.fastq.gz	6060514	1000000	PstI	1x92bp		55742		9:30:139	
AH_1.fastq.gz	5878328	1000000	PstI	1x95bp		65635		8:15:94	
BK_1.fastq.gz	5615081	1000000	PstI	1x95bp		60599		8:22:106	
3062_1.fastq.gz	5504925	1000000	PstI	1x92bp		42011		11:46:201	
4039_1.fastq.gz	5337966	1000000	PstI	1x96bp		57298		9:24:112	
4009_1.fastq.gz	4977092	1000000	PstI	1x94bp		65787		9:21:63	
4018_1.fastq.gz	4975098	1000000	SbfI	1x94bp		47958		9:28:146	
4004_1.fastq.gz	4917051	1000000	PstI	1x93bp		55102		8:21:118	
AG_1.fastq.gz	4840067	1000000	PstI	1x94bp		65908		8:18:76	
4019_1.fastq.gz	4839787	1000000	PstI	1x95bp		71813		9:21:70	
4044_1.fastq.gz	4768927	1000000	PstI	1x96bp		64803		8:20:77	
AR_1.fastq.gz	4731624	1000000	PstI	1x97bp		46455		8:24:126	
4010_1.fastq.gz	4682784	1000000	PstI	1x94bp		46756		9:30:144	
4048_1.fastq.gz	4659956	1000000	PstI	1x92bp		64055		9:21:99	
4023_1.fastq.gz	4646882	1000000	PstI	1x95bp		62720		8:19:76	
4006_1.fastq.gz	4568678	1000000	PstI	1x94bp		49752		9:27:128	
4026_1.fastq.gz	4498284	1000000	PstI	1x95bp		58992		10:30:97	
AY_1.fastq.gz	4418398	1000000	PstI	1x92bp		50562		9:24:117	
4008_1.fastq.gz	4310116	1000000	PstI	1x94bp		62827		8:20:65	
4001_1.fastq.gz	4236484	1000000	PstI	1x93bp		57074		10:29:96	
R_1.fastq.gz	4232171	1000000	PstI	1x93bp		54904		9:23:94	
CJ_1.fastq.gz	4171066	1000000	PstI	1x94bp		49183		11:31:122	
1003_1.fastq.gz	4086650	1000000	PstI	1x92bp		52907		9:23:102	
4043_1.fastq.gz	3942080	1000000	PstI	1x96bp		44094		9:28:120	
4030_1.fastq.gz	3825652	1000000	PstI	1x96bp		51205		10:25:98	
4055_1.fastq.gz	3781928	1000000	PstI	1x92bp		66675		9:21:71	
4034_1.fastq.gz	3754585	1000000	PstI	1x96bp		59716		9:20:66	
1023_1.fastq.gz	3734946	1000000	PstI	1x93bp		39749		11:42:137	
4056_1.fastq.gz	3609172	1000000	PstI	1x97bp		55279		8:15:65	
4025_1.fastq.gz	3549831	1000000	PstI	1x95bp		47839		9:25:103	
1043_1.fastq.gz	3509642	1000000	PstI	1x95bp		45126		10:31:117	
BF_1.fastq.gz	3468259	1000000	PstI	1x93bp		44937		10:32:108	
AA_1.fastq.gz	3466136	1000000	PstI	1x94bp		51742		8:20:89	
M_1.fastq.gz	3435748	1000000	PstI	1x97bp		65779		8:16:52	
3058_1.fastq.gz	3426831	1000000	PstI	1x92bp		40554		9:31:123	
4033_1.fastq.gz	3382264	1000000	PstI	1x96bp		45136		9:29:97	
4057_1.fastq.gz	3272111	1000000	PstI	1x92bp		62915		8:18:66	
1009_1.fastq.gz	3245658	1000000	PstI	1x93bp		56721		8:18:65	
AQ_1.fastq.gz	3231452	1000000	PstI	1x95bp		61454		8:16:56	
1012_1.fastq.gz	3225735	1000000	PstI	1x92bp		41807		10:36:111	
AW_1.fastq.gz	3223869	1000000	PstI	1x92bp		49271		9:23:82	
4031_1.fastq.gz	3215145	1000000	PstI	1x96bp		47204		8:20:79	
3005_1.fastq.gz	3204519	1000000	PstI	1x93bp		57137		8:20:73	
1050_1.fastq.gz	3167065	1000000	PstI	1x95bp		40398		10:34:117	
4054_1.fastq.gz	3162977	1000000	PstI	1x92bp		41786		11:38:118	
AO_1.fastq.gz	3156258	1000000	PstI	1x97bp		40466		9:26:121	
4053_1.fastq.gz	3139243	1000000	PstI	1x92bp		42629		10:32:119	
4027_1.fastq.gz	3105609	1000000	PstI	1x96bp		53077		7:16:70	
3011_1.fastq.gz	3093361	1000000	PstI	1x93bp		53375		8:19:76	
2030_1.fastq.gz	3071468	1000000	PstI	1x94bp		39157		12:40:108	
BL_1.fastq.gz	3054902	1000000	PstI	1x92bp		37513		10:35:122	
4022_1.fastq.gz	3032341	1000000	PstI	1x95bp		46133		8:25:84	
4011_1.fastq.gz	2975654	1000000	PstI	1x94bp		33005		12:47:136	
1001_1.fastq.gz	2949753	1000000	PstI	1x92bp		42013		9:27:106	
3045_1.fastq.gz	2923551	1000000	PstI	1x95bp		44233		8:25:101	
3049_1.fastq.gz	2884656	1000000	PstI	1x96bp		51350		8:20:70	
3061_1.fastq.gz	2866460	1000000	PstI	1x92bp		57014		8:19:60	
3043_1.fastq.gz	2856154	1000000	PstI	1x95bp		48014		9:23:85	
3037_1.fastq.gz	2840339	1000000	PstI	1x95bp		51150		9:24:75	
1028_1.fastq.gz	2813413	1000000	PstI	1x94bp		33890		9:36:129	
4003_1.fastq.gz	2803640	1000000	PstI	1x93bp		50677		8:19:69	
3051_1.fastq.gz	2766936	1000000	PstI	1x97bp		38613		9:30:108	
4007_1.fastq.gz	2744154	1000000	PstI	1x94bp		45658		9:25:76	
CD_1.fastq.gz	2734841	1000000	PstI	1x95bp		37135		11:36:113	
CL_1.fastq.gz	2727002	1000000	PstI	1x94bp		43207		9:22:87	
4015_1.fastq.gz	2708797	1000000	PstI	1x95bp		46929		9:26:71	
4017_1.fastq.gz	2684420	1000000	PstI	1x95bp		52716		9:19:57	
AD_1.fastq.gz	2674292	1000000	PstI	1x92bp		41730		10:33:99	
1049_1.fastq.gz	2672093	1000000	PstI	1x95bp		50225		8:19:69	
4020_1.fastq.gz	2661724	1000000	PstI	1x95bp		54966		8:20:58	
4058_1.fastq.gz	2633245	1000000	PstI	1x92bp		55172		8:20:64	
3024_1.fastq.gz	2611623	1000000	PstI	1x95bp		40687		10:32:96	


3028_1.fastq.gz 2610940 1000000 PstI 1x95bp 58019 8:18:55
D_1.fastq.gz 2601113 1000000 PstI 1x93bp 36129 10:33:111
1020_1.fastq.gz 2592728 1000000 PstI 1x93bp 41656 10:30:80
AE_1.fastq.gz 2585388 1000000 PstI 1x97bp 34790 9:27:101
4021_1.fastq.gz 2580087 1000000 PstI 1x95bp 42705 10:27:76
1036_1.fastq.gz 2574545 1000000 PstI 1x95bp 40487 9:27:97
H_1.fastq.gz 2568799 1000000 PstI 1x95bp 66622 6:11:38
3026_1.fastq.gz 2568166 1000000 PstI 1x95bp 42760 10:29:89
3019_1.fastq.gz 2567643 1000000 PstI 1x94bp 44287 9:22:82
4016_1.fastq.gz 2556155 1000000 PstI 1x94bp 46875 8:20:65
1026_1.fastq.gz 2552429 1000000 PstI 1x94bp 37828 10:33:95
4038_1.fastq.gz 2549077 1000000 PstI 1x96bp 43663 9:24:71
4012_1.fastq.gz 2542475 1000000 PstI 1x94bp 54424 8:18:47
BM_1.fastq.gz 2539173 1000000 PstI 1x95bp 40578 9:27:94
3036_1.fastq.gz 2538196 1000000 PstI 1x96bp 56349 8:19:58
3063_1.fastq.gz 2527797 1000000 PstI 1x96bp 56880 9:20:51
4037_1.fastq.gz 2507584 1000000 PstI 1x96bp 45301 8:19:68
1062_1.fastq.gz 2476853 1000000 PstI 1x97bp 51116 7:14:63
CF_1.fastq.gz 2466161 1000000 PstI 1x96bp 43203 9:24:83
4005_1.fastq.gz 2456065 1000000 PstI 1x94bp 41656 9:26:81
3022_1.fastq.gz 2455405 1000000 PstI 1x94bp 52793 8:19:57
1011_1.fastq.gz 2438770 1000000 PstI 1x92bp 41257 9:25:83
1044_1.fastq.gz 2433382 1000000 PstI 1x95bp 48896 8:20:60
1021_1.fastq.gz 2406239 1000000 PstI 1x93bp 45747 9:22:70
4041_1.fastq.gz 2405463 1000000 PstI 1x97bp 53569 7:14:46
1024_1.fastq.gz 2404748 1000000 PstI 1x93bp 33554 10:36:112
3039_1.fastq.gz 2403619 1000000 PstI 1x96bp 42454 10:30:85
F_1.fastq.gz 2399850 1000000 PstI 1x92bp 32290 11:40:114
AF_1.fastq.gz 2398903 1000000 PstI 1x96bp 43070 9:22:76
4024_1.fastq.gz 2361294 1000000 PstI 1x95bp 40937 10:26:74
3054_1.fastq.gz 2352205 1000000 PstI 1x96bp 32499 10:34:109
4014_1.fastq.gz 2343929 1000000 PstI 1x94bp 47316 10:24:60
4002_1.fastq.gz 2330744 1000000 PstI 1x93bp 41320 9:28:78
AL_1.fastq.gz 2290255 1000000 PstI 1x92bp 43540 9:21:63
1019_1.fastq.gz 2253025 1000000 PstI 1x93bp 37940 9:26:85
4047_1.fastq.gz 2246323 1000000 PstI 1x92bp 49965 8:18:54
3038_1.fastq.gz 2240665 1000000 PstI 1x95bp 44065 8:20:71
4050_1.fastq.gz 2229070 1000000 PstI 1x92bp 48130 8:19:66
4042_1.fastq.gz 2199044 1000000 PstI 1x97bp 45010 8:19:58
3029_1.fastq.gz 2180975 1000000 PstI 1x95bp 41964 9:23:76
O_1.fastq.gz 2164806 1000000 PstI 1x94bp 34489 9:28:95
BD_1.fastq.gz 2158831 1000000 PstI 1x95bp 32441 9:29:93
4049_1.fastq.gz 2156300 1000000 PstI 1x92bp 37156 11:33:89
BS_1.fastq.gz 2151839 1000000 PstI 1x94bp 33744 11:32:92
3046_1.fastq.gz 2134633 1000000 PstI 1x95bp 38307 9:27:84
BV_1.fastq.gz 2130511 1000000 PstI 1x94bp 30124 12:39:110
CC_1.fastq.gz 2112896 1000000 PstI 1x94bp 30466 11:37:106
1030_1.fastq.gz 2106186 1000000 PstI 1x94bp 32627 10:31:95
3012_1.fastq.gz 2100163 1000000 PstI 1x93bp 42197 9:20:70
BO_1.fastq.gz 2092600 1000000 PstI 1x92bp 40835 8:23:74
2006_1.fastq.gz 2082230 1000000 PstI 1x96bp 43256 8:18:62
1033_1.fastq.gz 2073660 1000000 PstI 1x95bp 33471 10:30:89
1051_1.fastq.gz 2068483 1000000 PstI 1x96bp 41725 9:24:70
CG_1.fastq.gz 2050164 1000000 PstI 1x94bp 52519 8:15:46
3021_1.fastq.gz 2023806 1000000 PstI 1x94bp 42315 9:24:65
3017_1.fastq.gz 2022597 1000000 PstI 1x94bp 42765 8:17:64
1047_1.fastq.gz 2018216 1000000 PstI 1x96bp 41253 9:23:67
1032_1.fastq.gz 2012193 1000000 PstI 1x93bp 31796 12:38:98
CH_1.fastq.gz 2011832 1000000 PstI 1x92bp 32923 10:31:86
G_1.fastq.gz 2004430 1000000 PstI 1x96bp 41839 8:17:60
BY_1.fastq.gz 1967139 1000000 PstI 1x92bp 36599 9:26:75
AI_1.fastq.gz 1955773 1000000 PstI 1x92bp 40525 9:23:65
1041_1.fastq.gz 1953716 1000000 PstI 1x94bp 33205 9:28:88
1004_1.fastq.gz 1940982 1000000 PstI 1x92bp 34223 10:30:85
3025_1.fastq.gz 1919615 1000000 PstI 1x95bp 36711 10:24:79
K_1.fastq.gz 1917566 1000000 PstI 1x93bp 46752 8:16:55
1040_1.fastq.gz 1910933 1000000 PstI 1x95bp 36250 9:23:79
AV_1.fastq.gz 1909704 1000000 PstI 1x96bp 49106 8:16:45
4028_1.fastq.gz 1908375 1000000 PstI 1x96bp 42004 9:20:58
AC_1.fastq.gz 1894761 1000000 PstI 1x96bp 42048 9:23:64
L_1.fastq.gz 1892915 1000000 PstI 1x96bp 34111 9:28:77
1048_1.fastq.gz 1892475 1000000 PstI 1x96bp 39405 10:28:62
2002_1.fastq.gz 1868401 1000000 PstI 1x97bp 37582 8:19:69
1025_1.fastq.gz 1864286 1000000 PstI 1x93bp 51374 8:17:42
4051_1.fastq.gz 1860301 1000000 PstI 1x92bp 38301 10:27:69
AZ_1.fastq.gz 1854689 1000000 PstI 1x92bp 35048 10:30:79
3008_1.fastq.gz 1851995 1000000 PstI 1x93bp 44506 11:24:53
S_1.fastq.gz 1846350 1000000 PstI 1x93bp 34857 10:30:80
I_1.fastq.gz 1821007 1000000 PstI 1x94bp 45852 8:18:49
AP_1.fastq.gz 1809791 1000000 PstI 1x97bp 37171 8:21:68
3032_1.fastq.gz 1808206 1000000 PstI 1x95bp 45674 8:17:49
2005_1.fastq.gz 1805286 1000000 PstI 1x96bp 64359 7:11:26
4013_1.fastq.gz 1780216 1000000 PstI 1x94bp 44928 8:18:46
3030_1.fastq.gz 1774503 1000000 PstI 1x95bp 33305 9:27:81
1002_1.fastq.gz 1772199 1000000 PstI 1x92bp 34597 9:28:75
1037_1.fastq.gz 1759667 1000000 PstI 1x94bp 34301 11:29:72
1016_1.fastq.gz 1753260 1000000 PstI 1x93bp 38789 10:25:59
1008_1.fastq.gz 1752723 1000000 PstI 1x92bp 40402 10:25:54
1027_1.fastq.gz 1750710 1000000 PstI 1x94bp 29359 12:36:90
4052_1.fastq.gz 1748758 1000000 PstI 1x92bp 37662 9:24:67
BR_1.fastq.gz 1718170 1000000 PstI 1x95bp 41622 9:18:50
3016_1.fastq.gz 1710271 1000000 PstI 1x94bp 38959 9:22:56
1006_1.fastq.gz 1705072 1000000 PstI 1x93bp 35457 10:25:67
Q_1.fastq.gz 1687460 1000000 PstI 1x94bp 30830 12:35:75


1005_1.fastq.gz 1686368 1000000 PstI 1x92bp 36452 9:23:64
1034_1.fastq.gz 1666164 1000000 PstI 1x95bp 34650 9:24:69
3050_1.fastq.gz 1664538 1000000 PstI 1x92bp 29896 10:31:74
1014_1.fastq.gz 1659539 1000000 PstI 1x92bp 41760 8:21:47
1042_1.fastq.gz 1658817 1000000 PstI 1x94bp 32143 9:26:74
AB_1.fastq.gz 1657196 1000000 PstI 1x95bp 35219 9:23:69
4040_1.fastq.gz 1655475 1000000 PstI 1x97bp 33326 9:24:63
3006_1.fastq.gz 1641860 1000000 PstI 1x93bp 32898 10:27:74
3031_1.fastq.gz 1639601 1000000 PstI 1x94bp 34716 9:26:70
B_1.fastq.gz 1639494 1000000 PstI 1x94bp 32887 9:28:75
1029_1.fastq.gz 1631834 1000000 PstI 1x94bp 32831 10:28:71
1038_1.fastq.gz 1624142 1000000 PstI 1x95bp 28989 12:36:84
2027_1.fastq.gz 1599016 1000000 PstI 1x96bp 39805 9:19:55
1055_1.fastq.gz 1594282 1000000 PstI 1x97bp 33261 10:28:65
3001_1.fastq.gz 1592528 1000000 PstI 1x93bp 29364 13:38:79
3015_1.fastq.gz 1588235 1000000 PstI 1x94bp 45421 8:15:43
Z_1.fastq.gz 1587759 1000000 PstI 1x96bp 37320 8:20:62
2062_1.fastq.gz 1570554 1000000 PstI 1x92bp 28930 10:29:78
U_1.fastq.gz 1552235 1000000 PstI 1x95bp 36199 10:26:60
1045_1.fastq.gz 1542438 1000000 PstI 1x96bp 36138 9:24:59
1056_1.fastq.gz 1541625 1000000 PstI 1x96bp 39659 9:22:50
Y_1.fastq.gz 1531104 1000000 PstI 1x96bp 41808 9:20:50
BB_1.fastq.gz 1529326 1000000 PstI 1x93bp 40092 8:16:51
W_1.fastq.gz 1522009 1000000 PstI 1x92bp 29681 10:32:76
1017_1.fastq.gz 1515848 1000000 PstI 1x93bp 31857 10:26:63
BC_1.fastq.gz 1513295 1000000 PstI 1x95bp 37261 8:20:59
3020_1.fastq.gz 1471150 1000000 PstI 1x94bp 37235 8:19:53
BI_1.fastq.gz 1459069 1000000 PstI 1x96bp 38312 9:22:52
4045_1.fastq.gz 1451406 1000000 PstI 1x96bp 36483 9:20:49
1058_1.fastq.gz 1448321 1000000 PstI 1x97bp 34944 9:20:56
1059_1.fastq.gz 1432257 1000000 PstI 1x97bp 36074 10:24:51
1061_1.fastq.gz 1426116 1000000 PstI 1x97bp 31411 9:25:64
N_1.fastq.gz 1418033 1000000 PstI 1x92bp 29912 11:34:67
AT_1.fastq.gz 1402627 1000000 PstI 1x96bp 34395 9:21:56
AJ_1.fastq.gz 1400868 1000000 PstI 1x95bp 31174 10:27:63
BN_1.fastq.gz 1399680 1000000 PstI 1x97bp 29692 10:27:67
3040_1.fastq.gz 1392839 1000000 PstI 1x96bp 31862 9:24:64
3023_1.fastq.gz 1386275 1000000 PstI 1x94bp 30066 9:26:69
3027_1.fastq.gz 1376746 1000000 PstI 1x95bp 37831 8:17:49
1010_1.fastq.gz 1369301 1000000 PstI 1x92bp 31985 12:30:58
3010_1.fastq.gz 1354463 1000000 PstI 1x93bp 29132 11:31:67
4060_1.fastq.gz 1348631 1000000 PstI 1x92bp 40157 9:19:44
3002_1.fastq.gz 1347016 1000000 PstI 1x94bp 35138 8:20:52
AN_1.fastq.gz 1334746 1000000 PstI 1x94bp 27815 10:30:69
CE_1.fastq.gz 1329092 1000000 PstI 1x92bp 32039 9:25:58
4061_1.fastq.gz 1322698 1000000 PstI 1x93bp 30099 9:25:64
3013_1.fastq.gz 1321175 1000000 PstI 1x93bp 38931 9:19:45
1035_1.fastq.gz 1318630 1000000 PstI 1x95bp 33332 11:25:50
3035_1.fastq.gz 1313594 1000000 PstI 1x96bp 36103 9:22:50
1057_1.fastq.gz 1307864 1000000 PstI 1x96bp 36006 10:23:49
1039_1.fastq.gz 1306716 1000000 PstI 1x95bp 27420 11:31:68
C_1.fastq.gz 1292347 1000000 PstI 1x96bp 33905 9:21:53
1018_1.fastq.gz 1254168 1000000 PstI 1x93bp 33002 10:24:51
BW_1.fastq.gz 1252983 1000000 PstI 1x92bp 28162 11:31:61
3044_1.fastq.gz 1244776 1000000 PstI 1x95bp 28781 10:29:61
4035_1.fastq.gz 1214389 1000000 PstI 1x95bp 37527 8:16:38
BA_1.fastq.gz 1208577 1000000 PstI 1x93bp 25516 12:32:68
X_1.fastq.gz 1200445 1000000 PstI 1x95bp 25732 12:32:67
3041_1.fastq.gz 1194659 1000000 PstI 1x95bp 27737 10:27:61
3052_1.fastq.gz 1177927 1000000 PstI 1x97bp 31485 9:20:52
4062_1.fastq.gz 1171515 1000000 PstI 1x93bp 39704 7:14:35
1060_1.fastq.gz 1142369 1000000 PstI 1x97bp 28482 10:24:54
AS_1.fastq.gz 1133406 1000000 PstI 1x96bp 35613 9:20:41
3003_1.fastq.gz 1129027 1000000 PstI 1x94bp 28530 9:24:57
1053_1.fastq.gz 1097259 1000000 PstI 1x96bp 31007 10:24:46
3057_1.fastq.gz 1082685 1000000 PstI 1x96bp 33153 8:17:41
4063_1.fastq.gz 1068560 1000000 PstI 1x94bp 33041 8:18:42
3018_1.fastq.gz 1067789 1000000 PstI 1x94bp 31816 8:17:46
3009_1.fastq.gz 1065101 1000000 PstI 1x94bp 30004 9:21:49
3048_1.fastq.gz 1061112 1000000 PstI 1x97bp 31905 9:19:41
1022_1.fastq.gz 1060578 1000000 PstI 1x93bp 31711 8:19:43
2063_1.fastq.gz 1032052 1000000 PstI 1x93bp 28725 9:21:46
1052_1.fastq.gz 1031657 1000000 PstI 1x96bp 29921 9:21:47
BU_1.fastq.gz 996208 1000000 PstI 1x94bp 29510 9:18:44
1054_1.fastq.gz 985676 1000000 PstI 1x96bp 31942 9:21:39
3007_1.fastq.gz 946175 1000000 PstI 1x93bp 25481 11:28:49
1046_1.fastq.gz 940698 1000000 PstI 1x96bp 30186 9:21:40
3004_1.fastq.gz 932927 1000000 PstI 1x93bp 24337 11:29:52
1031_1.fastq.gz 928331 1000000 PstI 1x94bp 26311 9:21:46
BP_1.fastq.gz 922174 1000000 PstI 1x94bp 30295 8:17:39
3014_1.fastq.gz 899464 1000000 PstI 1x93bp 29979 9:19:37
4029_1.fastq.gz 857760 1000000 PstI 1x96bp 25476 9:22:42
J_1.fastq.gz 809758 1000000 PstI 1x92bp 25108 11:25:42
3053_1.fastq.gz 735198 1000000 PstI 1x92bp 26877 10:19:31
V_1.fastq.gz 726895 1000000 PstI 1x97bp 22999 10:21:40
2001_1.fastq.gz 681712 1000000 PstI 1x97bp 23494 9:18:35
BZ_1.fastq.gz 656809 1000000 PstI 1x95bp 25709 9:17:28
3033_1.fastq.gz 632671 1000000 PstI 1x96bp 27768 8:16:27
3034_1.fastq.gz 620441 1000000 PstI 1x96bp 21951 9:20:35
4059_1.fastq.gz 607965 1000000 PstI 1x97bp 22815 8:16:30

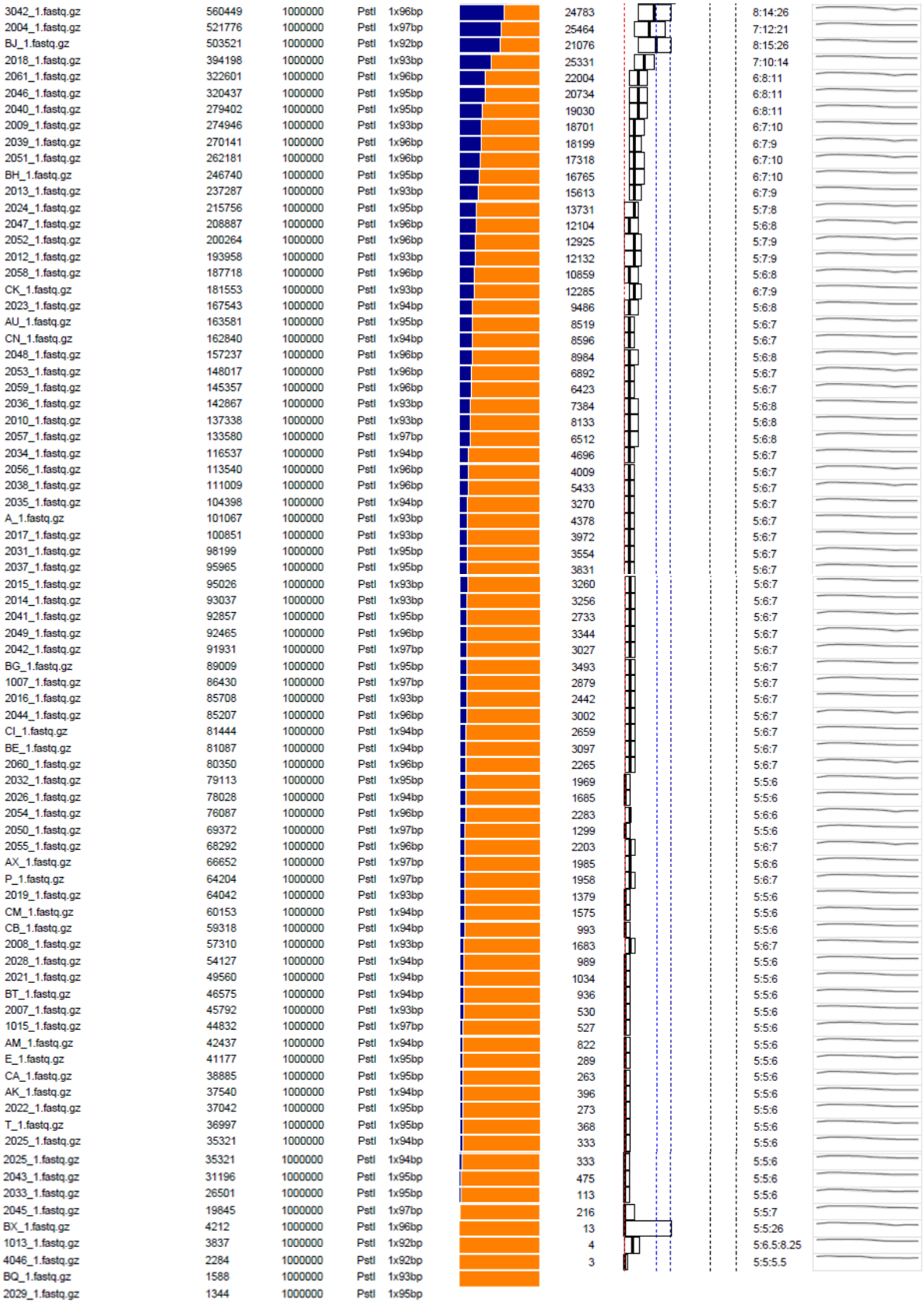


Figure 10.2 Individual sample sequencing performance of 343 PMiGAP lines sequenced at Floragenex, CA.

Tassel commands used in Section 6.3.8 to filter 32,901,665 SNPs downloaded from http://cegsb.icrisat.org/ipmgsc/:

Filtering:

0.05 minor allele frequency filter and a “minimum count’ threshold, where a call must be present in 80% of individuals at a given loci.

```
> filterAlignMinFreq 0.05
```

and

```
> filterAlignMinCount 276
```

The grep command was used to grab each line containing a specific word from a larger file e.g. "Fe" and then send these to a new file.

```
> grep "Fe" total_results_file.txt > Fe_results
```

The cat command was used to attach the column headings to the line grabbed previously using the grep command and then send this to another new file e.g.

```
> cat  headings Fe_results > Fe_results_with_headings
```

The R script used to generate the LD decay plots in Sections 6.4.1 and 6.4.2 (example uses the LD data generated in TASSEL 5.2.38 for >37,000 SNPs) and has been adapted from Marroni *et al.*, (2011):

```
> library(readr)
> LD_37k_table <- read_delim("~/Aberystwyth University-
PhD/GBS/SNPs/LD_37k_table.txt",
+     "\t", escape_double = FALSE, trim_ws = TRUE)
Parsed with column specification:
cols(
  Locus1 = col_integer(),
  Position1 = col_integer(),
  Site1 = col_integer(),
  NumberOfStates1 = col_integer(),
  States1 = col_character(),
  Frequency1 = col_character(),
  Locus2 = col_integer(),
  Position2 = col_integer(),
  Site2 = col_integer(),
  NumberOfStates2 = col_integer(),
  States2 = col_character(),
  Frequency2 = col_character(),
  Dist_bp = col_integer(),
  RSQ = col_double(),
  DPrime = col_double(),
  pDiseq = col_double(),
  N = col_integer()
)
|=================================================================================
=======| 100%  204 MB

> View(LD_37k_table)
> attach(subset(LD_37k_table, Locus1=='1'))
> n=7
> LD.data=(RSQ)
> distance=(Dist_bp)
> #n: sample size, i.e. number of sampled chromosomes, LD.data: estimates
of LD as r2 between pair of markers, distance: the distance between pair
of markers in bp.
> HW.st<-c(C=0.1)
> HW.nonlinear<-
nls(LD.data~((10+C*distance)/((2+C*distance)*(11+C*distance)))*(1+((3+C*di
stance)*(12+12*C*distance+(C*distance)^2))/(n*(2+C*distance)*(11+C*distanc
e))),start=HW.st,control=nls.control(maxiter=100))
> tt<-summary(HW.nonlinear)
> new.rho<-tt$parameters[1]
> fpoints<-
((10+new.rho*distance)/((2+new.rho*distance)*(11+new.rho*distance)))*(1+((
3+new.rho*distance)*(12+12*new.rho*distance+(new.rho*distance)^2))/(n*(2+n
ew.rho*distance)*(11+new.rho*distance)))
> #HW.nonlinear: object obtained after fitting the non-linear model,
new.rho: estimate of population recombination parameter, which is actually
C/distance, fpoints: vector of LD obtained fitting the linear model.
> ld.df<-data.frame(distance,fpoints)
> ld.df<-ld.df[order(ld.df$distance),]
> plot(distance,LD.data,pch=19,cex=0.9)
> lines(ld.df$distance,ld.df$fpoints,lty=3,lwd=1.2)
```

Table 10.14: Phenotype data used for haplotype trait analysis (Section 7.4).

Genotype	Fe/Zn Score	Fe (mg/kg)	Zn (mg/kg)	Haplotype Score
1001	High	106.91	54.47	6
1022	High	98.96	69.76	7
1035	High	90.24	71.22	3
1057	High	117.84	54.15	6
1058	High	135.27	79.31	1
2029	High	100.23	70.76	2
2036	High	111.41	82.76	4
2061	High	113.95	62.72	2
3033	High	94	82.9	1
3048	High	111.01	66.22	1
3050	High	93.74	66.95	4
4037	High	93.4	64.64	4
2005	High	85.44	79.39	1
4049	High	102.28	78.52	4
1054	Low	36.97	34.62	5
2007	Low	37.24	32.19	1
2018	Low	34.55	25.74	8
2024	Low	35.94	26.57	5
2026	Low	38.59	25.81	1
2063	Low	36.78	36.1	4A
3012	Low	37.06	35.84	1
3013	Low	39.25	26.42	2A
3054	Low	33.63	34.83	4
3058	Low	33.27	34.71	1
4011	Low	39.65	22.07	2
4041	Low	35.02	24.81	6
4050	Low	29.18	33.87	1
1025	Medium	54.39	51.95	2
1061	Medium	58.66	47	4
2033	Medium	58.11	50.05	9
3004	Medium	50.91	55.25	4
3017	Medium	60.95	46.19	6
3020	Medium	56.29	48.51	1
3036	Medium	53.34	56.21	1
3039	Medium	62.2	46.08	5
3044	Medium	64.3	44.63	1
3051	Medium	55.35	51.54	1
3060	Medium	71.66	36.84	6
4052	Medium	61.36	43.97	1
4053	Medium	53.69	54.63	5
4058	Medium	51.07	57.48	7
4060	Medium	62.71	46.66	5