Genetic analysis of carotenoid accumulation in tropical maize inbred lines using molecular tools and field tests for accelerating the biofortification of maize with provitamin A

By

Girum Azmach Mekonnen (BSc, MSc)

A thesis submitted to National University of Ireland Galway

College of Science, School of Natural Sciences

For fulfillment of the degree of Doctor of Philosophy

Academic Supervisor:

Prof. Charles Spillane, Discipline of Botany and Plant Science, School of Natural Sciences, National University of Ireland Galway

Research Supervisors:

Dr. Abebe Menkir, Maize Plant Breeder/Geneticist at International Institute for Tropical Agriculture (IITA)

Dr. Melaku Gedil, Cassava Molecular Geneticist at IITA

September 2014

NUI Galway
OÉ Gaillimh

Irish Aid
Department of Foreign Affairs
An Roinn Gnóthaí Eachtracha

IITA
Research to Nourish Africa

Table of Contents

List of Tables

List of Figures

List of Abbreviations and Symbols

Abbreviation/Symbol	Meaning
5'TE, 3'TE	transposable elements in the 5' and 3' regions of a gene
AD	number of days to anthesis
ANOVA	analysis of variance
ASI	anthesis silking interval = days to 50% silking minus days to 50% male flowering
BIC	Bayesian information criterion
BLUEs	best linear unbiased estimates
bpH	is best parent heterosis
CCD, *CCD*	carotenoid cleavage dioxygenase (enzyme, gene)
CGIAR's	Consultative Group for International Agricultrual Research
Chr	chromosome
CIMMYT	International Maize and Wheat Improvement Center
CMLM	compressed MLM
CRTISO, *crtISO*	carotene isomerase (enzyme, gene)
CRTRB, *crtRB*	β-carotene hydroxylase 1
CTAB	cetyl trimethyl ammonium bromide
CV%	coefficient of variation in %
cyp97a	cytochrome P450 13
cyp97c	cytochrome P450 14
DF	degree of freedom
DMAPP	dimethylallyl diphospate
DMSO	dimethyl Sulfoxide
DRC	Democratic Republic of Congo
DW	dry weight.
DXP	1-deoxy-D-xylulose 5-phosphate
DXS	DXP synthase;
DXR	DXP reductase
e	number of environments
EA	ear aspect
EH	Ear height
EMMA	efficient mixed model association
f	inbreeding coefficient
F1	first filial generation (hybrid)
FAO	Food and Agricultural Organization
FAOSTAT	FAO corporate Statistical Database
FDR	false discovery rate
G X E	Genotype by Environment
GA3P	glyceraldehyde 3-phosphate
GAPIT	Genetic Association and Prediction Integrated Tools
GBS	genotyping by sequencing
GCA	general combining ability
GCAf %	GCA female proportion
GCAf %	GCA female proportion
GCAm %	GCA male proportion
GGPP	geranylgeranyl pyrophosphate
GGPPS	GGPP Synthase
GLM	generalized linear model
GM	grand mean
GWA	genome-wide association
GWAS	GWAS Studies
GY	grain yield in t/ha
H	heterozygousity
HYD	β-carotene hydroxylase
HDR	4-hydroxy-3-methylbut-2-enyl diphosphate reductase
HPLC	High Pressure Liquid Chromatography

IGD	Institute for Genomic Diversity
IITA	International Institute for Tropical Agriculture
InDel	insertion deletion
IPP	Isopentyl diphospate
IPPI	IPP isomerase;
K	kinship
LCYB	lycopene β-cyclase
LCYE, *lcyE*	lycopene ε-cyclase (enzyme, gene)
LD	linkage disequilibrium
LOESS	localized regression
lut	lutein
MAF	minor allele frequency
MaizeGDB	maize Genome Database
Max	maximum
MEP	methylerythritol 4-phosphate
Min	minimum
MLM	mixed linear model
Mp	an average value of the corresponding parents for the hybrid
mpH	is mid-parent heterosis
NCBI	National Center for Biotechnology Information
NGS	next generation sequencing
NPK	Nitrogen-Phosphors-Potassium
P3D	population parameters previously determined
PA	plant aspect
PC	principal component
PCA	PC analysis
PDS	phytoene desaturase
PH	plant height in cm
PIC	polymorphic information content
PSY	phytoene synthase
PVA	provitamin A
Q	population structure
QTL	quantitative trait loci
r	number of repeats; number of replications per environment
R2 %	percentage of variation explained
R2	squared pairwise correlation coefficient
RAE	retinol activity equivalent
RAR	retinoic acid receptors
RXR	retinoid X receptors
SCA	specific combining ability
SCA%	SCA proportion
SD	number of days to silking
SE	Standard error
SFR™	super fine resolution
SNP	single nucleotide polymorophism
TASSEL	Trait Analysis by aSSociation Evolution and Linkage
Tbp	Calculated t value for F1 to bp comparison
tcar	total carotenoid
Tmp	calculated t value for F1 to mp comparison
UNSCN	United Nations System. Standing Committee on Nutrition
UPGMA	the unweighted pair-group method with arithmetical averages
VE	variance due to random errors
VG	is variance due to genetic effect
VGE	is variance due to genotype by environment interaction
Zea	zeaxanthin
ZEP	zeaxanthin epoxidase
Z-ISO	15 cis zeta-carotene isomerase
β br/αbr	ratio of carotenoids on β to α branch
βcar	β-carotene
βcryp	β-cryptoxanthin

Declaration

I certify that this thesis is my own work, and that I have not used this work in the course of another degree, either at National University of Ireland Galway, or elsewhere.

Signed: _____

(Girum Azmach Mekonnen)

Acknowledgment

My greatest thanks go to my supervisors Prof. Charles Spillane (NUIG), Dr. Abebe Menkir (IITA), and Dr. Melaku Gedil (IITA) for their supports and guidances.

This PhD study was realized because of the financial support from the Irish Aid. I am grateful to the IrishAid for its generosity in covering both living and research expenses throughout my study.

I am thankful to all the staff members of Maize Improvement Program, Bioscience Center and other units of IITA whose direct or indirect contributions were indespensable for the success of my study.

Thank you to IITA as a whole for hosting me as a research fellow for almost four years by providing excellent research and accomodation facilities.

I thank Dr. Sahron Michel of Cornell Univerisy for her answers and explanations to all my questions via email regarding the GBS data and related issues.

I am grateful to Dr. Alex Lipka of Cornell University for his explanations and answers to all my questions via email regarding his software GAPIT.

List of Publications

Azmach, G., Worku, M., Wolde, L., Abera, W., Tadesse, B., Keno, T., Chibsa, T., Spillane, C., and Menkir, A. (2011). Development of improved yellow maize germplasm in Ethiopia. In Meeting the Challenges of Global Climate Change and Food Security Through Innovative Maize Research. Proceedings of the National Maize Workshop of Ethiopia, B.M. Worku, M.; Twumasi Afriyie, S.; Wolde, L.; Tadesse, B.; Demisie, G.; Bogale, G.; Wegary, D.; Prasanna, ed (CIMMYT: Addis Ababa), pp. 58–65.

Worku, M., W. Legesse, T. Berhanu, D. Girma, Girum A., et al. Status and future directi on of maize research and production in Ethiopia. (2012). In Meeting the Challenges of Global Climate Change and Food Security through Innovative Maize Research. Proceedings of the Third National Maize Workshop of Ethiopia. B.M. Worku, M.; Twumasi Afriyie, S.; Wolde, L.; Tadesse, B.; Demisie, G.; Bogale, G.; Wegary, D.; Prasanna, ed (CIMMYT: Addis Ababa), pp. 17–23

Twumasi-Afriyie, S., A.K. Demisew, B. Gezahegn, A. Wende, Gudeta Nepir, N. Demoz, D. Friesen, Y. Kassa, A. Bayisa, Azmach G., and F. Wondimu. (2012). A decade of quality protein maize research progress in Ethiopia (2001–2011). InMeeting the Challenges of Global Climate Change and Food Security through Innovative Maize Research. Proceedings of the Third National Maize Workshop of Ethiopia. B.M. Worku, M.; Twumasi Afriyie, S.; Wolde, L.; Tadesse, B.; Demisie, G.; Bogale, G.; Wegary, D.; Prasanna, ed (CIMMYT: Addis Ababa), pp. 47–57

Azmach, G., Gedil, M., Menkir, a, and Spillane, C. (2013). Marker-trait association analysis of functional gene markers for provitamin A levels across diverse tropical yellow maize inbred lines. BMC Plant Biol. **13**: 227.

Abstract

Improving the nutritional content of staple crops through breeding (biofortification) is promoted as cheap and sustainable approach for alleviating vitamin A and other micronutrient deficiencies across the world. Maize is one of the target crops for provitamin A biofortification, since it is a major staple for millions of people in developing countries, and has a great genetic potential for accumulating provitamin A in its grain. This PhD study addressed three research objectives aimed at accelerating the biofortification of maize with provitamin A. To achieve the objectives, a diverse set of inbred lines developed from different crosses having mixed genetic background of temperate and tropical germplasm were employed. In the first research objective, the efficiency of PCR based functional markers proposed for high provitamin A maize breeding was investigated. Out of eight polymporphisms for three key carotenoid biosynthesis genes tested, the 3'TE and 5'TE polymorphisms of the gene *crtRB1* showed the strongest association. About 18 percent of the inbred lines studied had the favourable alleles for these two polymorphisms. The second research objective assessed the association of genome-wide SNP markers with provitamin A and non-provitamin A carotenoids using GWAS and pathway level association analysis. This study detected known major and minor effect carotenoid genes (*lcyE*, *crtRB1*, *ZEP1* and others), plus novel loci, such as *ARF20*, that may be involved in regulation of carotenoid biosynthesis in maize endsosperm. In the last research objective 24 yellow maize inbred lines were inter-crossed using factorial mating scheme for studying their combining abilities and heterotic effects for carotenoids and key agronomic traits. The result demonstrated the predominance of additive genetic effects in the inheritance of endosperm carotenoid content. Small percentage of non-additive genetic effect was also detected for provitamin A which is important for expression of heterosis in the trait. Hybrids combining good agronomic performance and high provitamin A content were identified. This study will serve as a basis for establishing heterotic pattern in IITA's yellow maize germplasm which is crucial for development of hybrid and synthetic varieties that can meet existing and emerging demands, while incorporating high level of provitamin A in the grain.

The adage 'health comes from the farm, not the pharmacy' is at the heart of ongoing international biofortification research and breeding programs (Mayer et al., 2008)**.**

Chapter 1: General Introduction

1.1 Micronutrient undernutrition

Micronutrient undernutrition, also known as hidden hunger (to imply that its effects have no immediate visible symptoms), is a poor diet and health condition caused by lack of essential vitamins and minerals that are required by the human body in small quantities (Tulchinsky, 2010; Nguyen et al., 2014). Half of the world's population is estimated to be affected by micronutrient undernutrition (Miller and Welch, 2013). Micronutrient undernutrition is a much bigger global problem than energy-protein or macronutrient malnutrition, imposing enormous short and long term adversities on societies in terms of ill-health, mortality and retarded physical and mental growth, all of which are in the long run translated to low quality of life and diminished economic productivity (Tulchinsky, 2010; Shetty, 2011). Among deficiencies of micronutrient, those of vitamin A and zinc are estimated to impose a significant disease burden on mothers and children (Black et al., 2008).

1.2 Vitamin A, its diatry sources, metabolism and functions

The knowledge of micronutrient's importance gained over the past two decades has lead to considerable research in order to better understand the physiological role and health risks associated with micronutrient-deficient diets, to establish criteria for defining the degree of public health severity of micronutrient malnutrition and to develop prevention and control strategies (FAO/WHO, 2004). One of the most important micronutrient targeted for scientific research is vitamin A. Vitamin A generally refers to a family of chemical compounds with biological activity of all-trans retinol (R-OH) which include retinaldehyde (retinal) (R-CHO), various retinyl esters (R-OO), and retinoic acids (R-OOH), where R represents the β-ionone ring plus chain moiety of the molecule (Figure 1.1) (West JR, Keith, Darnton-Hill, 2008). Higher animals and humans cannot synthesize vitamin A de novo, and thus need to obtain the nutrient from dietary sources. This can be either as preformed vitamin A from animal-based foods (e.g. liver, whole milk, and egg), or as precursors of vitamin A from coloured vegetables and fruits (e.g. carrots, dark green leaves and papaya) in the form of provitamin A carotenoids (West JR, Keith, Darnton-Hill, 2008). In animal-derived foods vitamin A occurs as retinyl esters of fatty acids associated with membrane bound lipids and fatty storage tissues. Whereas, in plant-derived foods precursors of vitamin A occur as provitamin A which is also

associated with cellular lipids but embedded in the more complex cellulose-containing matrix of chloroplasts or the pigment-containing portion of chromoplasts (FAO/WHO, 2004).

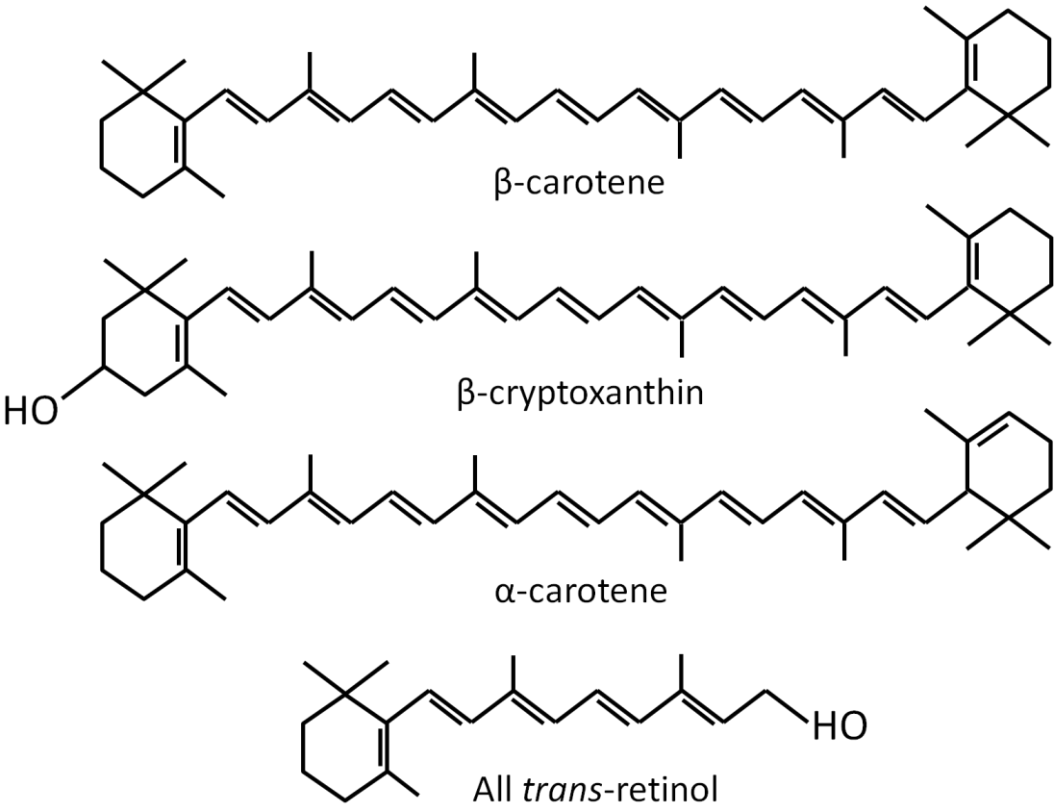


Figure 1.1 Chemical structure of all trans retinol, and its most common precursors: β-carotene, β-cryptoxanthin and α-carotene
(West JR, Keith, Darnton-Hill, 2008; Cuttriss et al., 2011)

Vitamin A is an essential micronutrient involved in regulating key biological processes in the human body throughout life including visual cycle, morphogenesis, organogenesis, reproduction, immunity, and cellular differentiation and proliferation (FAO/WHO, 2004; West JR, Keith, Darnton-Hill, 2008). Vitamin A in the form of 11-cis retinal serves as an indispensable light capturing chromophore in the eye assisting vision under dim light condition (West JR, Keith, Darnton-Hill, 2008; Saari, 2012). The visiual cycle is initiated with the photoisomerization of 11-cis retinal to all trans retinal and its release from rhodopsin, a photoreceptive protein in the retina. Upon release of all trans retinal, rhodopsin changes its conformation activating a signal transduction pathway that transmits visual information via the

optical nerv to the brain. The visual cycle completes when the rodopsin reattaches to a new 11-cis retinal restoring the pigment's photoreceptivity for the next visual cycle (Blomhoff and Blomhoff, 2006; West JR, Keith, Darnton-Hill, 2008; Saari, 2012). Lack of vitamin A can impair the visual cycle leading to night blindness, a well known visible disorder and clinical indicator of vitamin A deficiency (West JR, Keith, Darnton-Hill, 2008).

Most functions of vitamin A in relation to tissue/organ growth and development, cell differentiation and immunity stem from its ability to regulate gene transcription mediated by its all-trans and 9-cis retinoic acid forms (Mamoon et al., 2014; West JR, Keith, Darnton-Hill, 2008). These iso-forms of retinoic acid serve as ligands of transcription factors known as retinoic acid receptors (RAR) and retinoid X receptors (RXR) (Blomhoff and Blomhoff, 2006; West JR, Keith, Darnton-Hill, 2008). Activated transcription factors bind to short sequences of DNA near or within the target genes and initiate their transcription leading to translation of regulatory proteins involved in signaling, cell differentiation and apoptosis (Blomhoff and Blomhoff, 2006; West JR, Keith, Darnton-Hill, 2008). One of the visible consequences of perturbed regulation of transcription caused by vitamin A deficiency can be xerophthalmia, drying (xerosis) of the conjuctiva and cornea (West JR, Keith, Darnton-Hill, 2008).

Beta-carotene is the most potent vitamin A active carotenoid, since it can be broken down to two molecules of retinol. Other provitamin A carotenoids such as β-carotene, α-carotene and β-cryptoxanthin can provide only one molecule of vitamin A (US Institute of Medicine, 2001; Grune et al., 2010). Retinol active equivalent (RAE) is a notation used to describe the quantitative relationship between retinol and other dietary sources of vitamin A, where 1 μg RAE being equal to 1 μg all-trans-retinol, 12 μg β-carotene, and 24 μg α-carotene or β-cryptoxanthin (Otten et al., 2006). Beta-carotene provides 80% of the vitamin A consumption in developing countries (Biesalski, 2013). The efficiency of absorption of β-carotene from plant sources ranges from 5% to 65% in humans (Haskell, 2012). Depending on the type of dietry source, vitamin A equivalency ratios for β-carotene to retinol varies from 3.8:1 to 28:1 by weight (Tang, 2010; Haskell, 2012). RAE ratio for β-carotene from

biofortified maize (6.5:1) is lower than RAE ratio for vegetables (10:1 to 28:1) that have more complex food matrices (Haskell, 2012).

Vitamin A in humans and other higher animals is produced through metabolism of provitamin A carotenoids mainly inside the small intestine (Olson, 1989; D'Ambrosio et al., 2011; Blaner, 2013; Eroglu and Harrison, 2013; Jlali et al., 2014). Metabolism of vitamin A and provitamin A begins with gastro-intestinal digestion, in which the dietary forms of the nutrients are physically released from the food matrix and dissolved/emulsified in lipids forming mixed micelles (a mixture of bile salt, phospholipids, cholesterol, and lipid digestion products) (Reboul, 2013). Mixed micelles formation is presumed to facilitate absorption of vitamin A and carotenoids by the small intestine. Retinyl esters in the micelles are hydrolyzed into retinol and absorbed by the enterocytes along with carotenoids. Part of the provitamin A carotenoids that entered the cytoplasm of the mucosal cell can be converted to retinol (D'Ambrosio et al., 2011; Harrison, 2012; Reboul, 2013). The enzyme that converts dietary provitamin A carotenoids to vitamin A is a cleavage-oxygenase enzyme, β-carotene 15–15'-oxygenase (BCO1), formerly thought to be a monoxygenase but recently demonstrated as a dioxygenase (dela Seña et al., 2014). The retinol and provitamin A carotenoids inside the enterocytes are re-esterified and packaged with other lipid esters into chylomicrons which are then secreted to the lymphatic and circulatory system. The largest percentage of dietary vitamin A in the circulatory system is absorbed by the liver and stored in the hepatic stellate cells. The remaining vitamin A is transported to other peripheral tissues. In fasting condition, the stored vitamin A is mobilized from the liver and transported to the circulatory system assisted by retinoid binding proteins (D'Ambrosio et al., 2011).

1.3 Vitamin A deficiency

Retinol concentration in the blood plasma is under tight homeostatic regulation (US Institute of Medicine, 2001; Blomhoff and Blomhoff, 2006). Under normal condition, the concentration ranges between 1.0 and 1.5 μM among children and between 2 and 3 μM among adults (Biesalski, 2013). Plasma retinol concentration generally declines when liver vitamin A reserves fall below 20 μg/g liver (US Institute of Medicine, 2001). Plasma retinol concentration of below 0.7mol/l is considered to relfelct vitamin A deficiency in a population, although deficiency of

other nutrients (e.g. protein) can also contribute to the low retinol concentration (US Institute of Medicine, 2001; FAO/WHO, 2004).

Prolonged insufficient vitamin A in the diet can deplete vitamin A reserves in the liver leading to low plasma retinol concentration. This in turn can lead to possible malfunctions in any of the biological processes that are regulated with the involvement of vitamin A. Vitamin A inadequacy can cause non-specific systemic complications that can contribute to infection-related mortality and retarded growth. Visual disorders characterized by night blindness eye inflammation (xerophthalmia) and ulcer-related loss of vision (keratomalacia) are symptomes that are typical to vitamin A deficiency (Biesalski, 2013). To maintain adequate vitamin A levels in the body, the daily dietary intake should meet the recommended diatary allowance (RDA); which is defined as the average daily dietary vitamin A intake level sufficient to meet the nutrient requirement of nearly all (97 to 98 percent) healthy individuals in a particular life stage and gender group. On the other hand, this daily vitamin A intake shouldn't pass a certain limit which is known as tolerable upper intake level (UL), the highest average daily nutrient intake level that is likely to pose no risk of adverse health effects to almost all individuals in the general population (Otten et al., 2006). The RDA and UL for different age groups of healthy individuals are presented in Table 1.1 as adapted from US Institute of Medicine (2001).

Table 1.1 Recommended daily allowance (RDA) and daily tolerable upper intake levels of vitamin A by group

Group	Age (Years)	RDA (µg RAE/day)	UL (µg/day)
Infants	0–1	450	600
Children	1–3	300	600
	4–8	400	900
Boys	9–13	600	1,700
	14–18	900	2,800
Girls	9–13	600	1,700
	14–18	700	2,800
Adult Men	≥19	900	3,000
Adult Women	≥19	700	3,000
Lactiating	14–18	1,200	2,800
	≥19	1,300	3,000
Pregnant	14–18	750	2,800
	≥19	770	3,000

Source: US Institute of Medicine, 2001

Children, pregnant and lactating women have a high demand for vitamin (WHO, 2009). For this reason, vitamin A deficiency is considered a disabling and potentially fatal public health problem for children of less than 6 years of age and women during pregnancy and lactation (FAO/WHO, 2004). According to the WHO's 1995 – 2005 survey, which investigated low serum retinol content (< 0.7 μmole/litter), about 190 million pre-school children and 19 million pregnant women worldwide were vitamin A deficient with a prevalence of 33% and 15%, respectively (WHO, 2009). In Africa, it affected 54 million children and 4 million women with a prevalence of 44% and 13%, respectively (WHO, 2009). Improvements have been reported worldwide over time but with very slow progress in Africa and Southeast Asia, regions which are still lingering behind with severe and wide-spread effects (UNSCN, 2010). Up to half a million children lose their sight every year due to xerophtalmia casued by vitamin A deficiency rendering it the leading cause of preventable blindness (Sherwin et al., 2012). Lack of vitamin A together with other undernutrition problems, is responsible for the death of 3.1 million children annually (Biesalski, 2013).

The root cause of vitamin A deficiency in Sub-Saharan Africa and other less developed regions include poverty and lack of access to micronutrient rich foods aggravated by economic crises, price hiking and climate change (WHO, 2003; Bloem et al., 2010; West and Mehra, 2010). Traditional misconceptions, poor food habits, and diseases also contribute to vitamin A deficiency (FAO/WHO, 2004; Biesalski, 2013). Staple foods in these regions are dominated by starch-rich cereals, tuber and/or root crops (WHO, 2003) which supply little or no vitamin A to meet the recommended daily allowance for this nutrient (Ruel, 2003; FAO/WHO, 2004). Animal products are usually not affordable to the broader socio-economically disadvantaged families in the developing world. Besides, fruits and vegetables are inaccessible, and their availability depends on season and location (WHO, 2003; Biesalski, 2013). Vitamin A deficiency can be in a complex cause and effect relationship with chronic diseases aggravating the rate of child mortality and disability (Black et al., 2008).

1.4 Traditional intervention alternatives against vitamin A deficiency

The three commonly applied approaches for alleviating vitamin A deficiency are periodic deliveries of vitamin A supplements, industrial fortification of staple foods with vitamin A, and promote dietary diversification (West JR, Keith, Darnton-Hill, 2008; WHO, 2009; West and Mehra, 2010). Lasting and sustainable solutions to micronutrient undernutrition in developing countries can be achieved with economic development and education, as these are key to improving access to dietary quality and diversity, and by correcting preferences for less nutritious diets. This however is unlikely to be attained in the foreseeable future, as success is hampered by the prevailing malnutrition problem which adversely affects productivity, effectiveness and economic development of the society from generation to generation in a vicious circle (The World Bank, 2006; Biesalski, 2013).

Although fortification and supplementation are considered as fast and effective measures to curb chronic vitamin A deficiencies, these approaches are usually considered unsustainable and inaccessible to all people for economic, political, and/or logistical reasons (Welch and Graham, 1999). Also, unmonitored use of vitamin A supplements and fortified foods has been reported to cause toxicity and extra health problems (FAO/WHO, 2004; Biesalski, 2013).

1.5 Biofortification: agriculture's response to micronutrient undernutrition

Agriculture and agricultural biotechnology can play a crucial role in combating vitamin A and other micronutrient deficiencies through improvement of crop production and productivity in terms of both quantity and quality (Underwood, 2000; WHO, 2003; Newell-McGloughlin, 2008; UNSCN, 2010). One way of empowering agriculture towards its use as a tool for fighting vitamin A and other micronutrient deficiencies is through development of nutritious staple crops using the approach of crop biofortification. Biofortification can be defined as the application of conventional breeding and modern biotechnology to develop staple food crops with increased concentrations of bioavailable micronutrient in the edible portion without detrimental side effects (Nestel et al., 2006; Miller and Welch, 2013).

One challenge to the application of this approach has been lack of interest in nutritional attributes in the market, which has lead to underinvestment in biofortification. However, the biofortification approach has drawn significant attention and support from the public sector as awareness of the scourge of micronutrient undernutrition worldwide increases over time (Unnevehr et al., 2007).

Crop biofortification has emerged as a complementary intervention approach designed to mitigate micronutrient deficiencies. It has been proposed as a viable method to deliver safer, cost-effective and sustainable sources of vitamin A and other micronutrient via agricultural interventions (Underwood, 2000; Qaim et al., 2007; Bouis and Welch, 2010; Meenakshi et al., 2010a; Msangi et al., 2010). Ideally, through upfront investment in research and development, once the biofortified crop is put under production and consumption, farmers can grow and consume/sell the crop season after the season with cheaper running costs and minimal external intervention. In addition, biofortification could be less affected by political and economic instabilities, or poor infrastructural facilities (Poletti et al., 2004; Unnevehr et al., 2007; Meenakshi et al., 2010b; Bouis et al., 2011).

Crop biofortification endeavours have been in progress for more than a decade, particularly under the auspices of CGIAR's (Consultative Group for International Agricultural Research) Biofortification Challenge Programme known as HarvestPlus (Bouis and Welch, 2010). HarvestPlus has already released a range of vitamin A biofortified varieties fo major crops: maize for Zambia and Nigeria; cassava for Nigeria and Democratic Republic of Congo; and sweet potatoes for Uganda (HarvestPlus, 2012). However, except for the biofortified sweet potato, the target provitamin A concentrations for biofortified crops are yet to be attained. The initial goal of the HarvestPlus maize and cassava provitamin A biofortification programme was to generate fully fortified maize and cassava that can provide 60% and 50% of the daily vitamin A needs, yet the currently biofortified varieties can only meet 42% and 50 % of the target levels, respectively. Recent efficacy and impact analyses conducted on biofortified maize in Zambia did not show improvement in the vitamin A status of the target population based on serum retinol levels (Gannon et al., 2014; Palmer et al., 2014), while it did show positive results based on total body reserves of vitamin A (Gannon et al., 2014). This emphasizes the need for further

enhancement of the provitamin A concentration (and bioavailability) in maize endosperm to attain the targets set under the HarvestPlus program.

1.6 Maize as a vehicle for provitamin A biofortification

Maize is cultivated on over 175 million hectares of land with a total production of about 870 million tonnes around the world (FAOSTAT, 2012). This makes the crop world's second most widely grown cereal crop next to wheat, and the most important grain in terms of total production. Although it is largely used as animal feed globally (Smale et al., 2013), maize feeds one billion people as staple food in Africa and Latin America (http://www.HarvestPlus.org/content/vitamin-maize, accessed in December 2014). In Africa more than 300 million people subsist on maize (M'mboyi et al., 2010; Smale et al., 2013). The annual production of maize in the continent has shown a 40 % increase between 2005 and 2012, increasing from about 50 million metric tons produced on 30 hectares of land, to 70 million metric tons produced on 34 million hectares of land (FAOSTAT, 2012). This production increase occurred mainly as a result of yield improvement demonstrating the productivity potential of the crop in the region.

Maize is rich in energy with a calorie density of 365 kcal/100g, derived largely from the carbohydrates stored in the endosperm tissue in the form of starch. The grain also has appreciable levels of total protein and lipids but lacks important micronutrient such as vitamin A (especially white maize) and essential amino acids (Nuss and Tanumihardjo, 2010; Ranum et al., 2014). The crop represents a significant proportion of the total calorie intake of individuals in many countries in Africa, accounting for about 30% of the per-capita calorie consumption in Eastern and Southern Africa (FAOSTAT, 2011). For some Southern African countries like Zimbabwe, Malawi, Zambia, and Lesotho, the caloric contribution of maize is around 40 – 56% (FAOSTAT, 2011). Provitamin A biofortification of maize is therefore an ideal delivery route to improve the health status of vitamin A deficiency to vulnerable people in this region.

Maize is one of the most diverse crop species both genetically and phenotypically (Wang and Dooner, 2006; Xu et al., 2009). The diversity of the crop is well reflected in its endosperm carotenoid profile imparting the grain from pale yellow to dark

orange colours. Co-inheritance of the yellow pigment of maize endosperm and vitamin A has long been established (Hauge and Trost, 1928, 1930; Hauge, 1930). Mangelsdorf and Fraps (1931) did not only corroborate the association between the yellow pigment and Vitamin A, but also established the direct relationship between the quantity of vitamin A in the endosperm, the dosage effects of the genes for yellow pigmentation and the intensity of the yellow endosperm colour.

White maize, the most popular variant of maize in Africa, due to the stigma that associates yellow maize with food-aid and an industrial demand that had encouraged white maize cultivation in Africa the 20th century (Azmach et al., 2011), contains little or no carotenoids in its kernels. Yellow maize, on the other hand, is the only grain crop that can naturally accumulate a significant amount of diverse carotenoids in its seed. The major carotenoids in maize endosperm are lutein and zeaxanthin, which have no vitamin A activity, but are known for their antioxidant activities and correlation with eye/macular health (Burt et al., 2011b; Meyers et al., 2014). Provitamin A carotenoids (β-carotene, β-cryptoxanthin and α-carotene) are also found in maize kernel (Muzhingi et al., 2008b; Pixley et al., 2013). Maize germplasm that can accumulate up to 26 µg/g β-carotene and 30 µg/g of provitamin A in the endosperm are now available (Pixley et al., 2013). However, commonly cultivated varieties across the world contain low levels of provitamin A carotenoid ranging from 0.5 to 1.5 µg/g (Harjes et al., 2008). Hence, improvement of provitamin A carotenoids levels in commonly cultivated and consumed maize cultivars, using both conventional and genomic assisted breeding, is necessary.

Maize provitamin A biofortification is technically feasible owing to the amenability of maize to breeding and genetic analysis, readily available rich genetic and genomic resources, as well as the presence of maize germplasm with contrasting carotenoid composition and content (Wurtzel, 2007; Schnable et al., 2009; Xu et al., 2009; Farre et al., 2010; Lu et al., 2011; Wurtzel et al., 2012). This provides productive grounds for the application of conventional breeding in conjunction with metabolic engineering and genomic and marker-assisted breeding to enhance the provitamin A content of maize. In addition, the vitamin A requirement of millions of poor people in Sub-Saharan Africa, who are dependent on maize for their daily calorie needs, can

be sustainably and cost effectively met with biofortified maize, especially those living in remote rural areas.

1.7 Prospect of success and efficacy in maize biofortification

Since biofortification is a multi-disciplinary innovative approach, its success relies on the output of the concerted effort of breeders, nutritionists, health experts, agronomists, socio-economists and development workers (Bouis and Welch, 2010; Bouis et al., 2011). To date, specific target countries/populations with micronutrient deficiency (Iron, Zinc and vitamin A) have been selected and optimum target nutrient concentrations have been set for breeding micronutrient in several crops. These have been accomplished through studies on retention of the target nutrient in the food following food processing and cooking, bioavailability, and the nutrient requirements of the target population (Pfeiffer and McClafferty, 2007). In the case of maize, the initial target concentration of provitamin A for biofortification has been set at 15 µg/g. This is considered as an optimal concentration that can effect significant improvement on the health status of vitamin A deficiency-inflicted communities in target regions (HarvestPlus Maize Strategy: www.HarvestPlus.org/sites/default/files/HarvestPlus_Maize_Strategy.pdf, accessed in June 2014). The initial target concentration was determined by taking adult women of the target population as reference, and assuming 400 g/day consumption, 50 % retention after cooking, and 12:1 RAE ratio for β-carotene obtained from maize based food.

HarvestPlus and its partners have been conducting ex- and post-ante studies to ensure the approach will have sustainable and cost effective impact on the targeted societies. One of the areas of investigation in this respect involved farmer and consumer acceptance of the biofortification product (Meenakshi et al., 2010a; Bouis and Welch, 2010). Provitamin A biofortification in Eastern and Southern Africa illustrates such challenges of consumer acceptance, as maize consumers in many countries in these regions traditionally consider yellow maize inferior to white maize (De Groote and Kimenju, 2008; Stevens and Winter-Nelson, 2008). However, a number of studies carried out in those countries demonstrated that the public opinion on yellow maize may not be an obstacle for adoption of provitamin A biofortified maize (Stevens and Winter-Nelson, 2008; Muzhingi et al., 2008a; De Groote et al., 2011). In addition, since provitamin A biofortified maize kernels actually tend to be

orange coloured rather than yellow, a study by Meenakshi et al. (2010a) has shown that the negative perception of yellow maize might not affect the acceptance of the orange coloured provitamin A maize in Eastern and Southern Africa, by creating awareness among the target communities on the nutritional value of the biofortified variety. Pillay et al., (2011) also showed acceptance of provitamin A biofortified among preschool children but suggested the necessity for intensive nutrition education to convince adults to prefer biofortified yellow maize. On the other hand, there is no selective preference for yellow and white maize among Western and Central African farmers and consumers (Menkir et al., 2008).

Food habits of people are likely to change over time through influence of a number of interacting factors (WHO, 2003). However, a recent socioeconomic study by Msangi et al. (2010), speculated that the diet of the rural poor is far from shifting from the monotonous carbohydrate-dominated nutrition. This observation highlights the potential and long term relevance of biofortification to the rural communities in poor countries (Msangi et al., 2010), where biofortified staple crops like maize will be meaningful vehicle by which cheaper and sustainable sources of vitamin A can be delivered.

Another relevant question pertinent to provitamin A, and other micronutrient biofortification, is whether the biofortified micronutrient will be adversely affected by grain storage and processing, and to what extent the micronutrient will be biologically accessible and/or available to effect the intended impact. Studies have been carried out to answer this question in relation to maize. For example, Li et al. (2007) demonstrated the relatively good retention capacity of of *β*-carotene in maize following traditional food processing. Kean et al. (2008) reported maize-based food products as a good dietary source of bio-accessible carotenoids, noting the possible influence of specific food preparation methods on the relative bio-accessibility of individual carotenoid species. Their study showed that bio-accessibility of *β*-carotene was higher for maize porridge (40 to 63%) than for extruded puffs and bread (10 to 23%), suggesting wet cooking might improve the bioaccessibility of *β*-carotene. A number of studies have confirmed the bioavailability of vitamin A in maize using model animals (Howe and Tanumihardjo, 2006). A recent study using human subjects has demonstrated a good level of bioavailability of β-carotene in

biofortified maize (Li et al., 2010). This study verified that only 6.48 µg of the provitamin A carotenoids in biofortified maize is required to provide RAE of 1µg retinol, which is higher than the conversion rate for β-carotene from conventional vegetable sources (10:1 to 28:1) because of their complex food matrices (Haskell, 2012).

1.8 Carotenoids and the carotenoid biosynthesis pathway in maize endosperm

Exploitation of the diverse tropical maize inbred lines available at IITA and elsewhere for development of high provitamin A maize cultivars entails understanding and application of knowledge derived from molecular, genetic and biochemical studies of carotenoids and their biosynthesis in plants.

Carotenoids are naturally occurring organic pigments that are produced by plants and a few other photosynthetic organisms (Moran and Jarvik, 2010; Cazzonelli, 2011) They play essential structural and physiological roles in both plants and animals. Carotenoids are characterized by their extensive conjugated double bond along their carbon backbone giving them the capability to absorb lights in the blue to green range of the visible spectrum (Cuttriss et al., 2011; Moise et al., 2014). In plants, carotenoids are present mainly as indispensable integral components of the chloroplast, providing multiple services to the photosynthetic machinery participating in the light harvesting process and guarding the photosystems from possible damages by quenching reactive singlet oxygens and radicals created during photooxidation (Tracewell et al., 2001; Ballottari and Alcocer, 2014; Shinopoulos et al., 2014; Stamatakis et al., 2014).

Carotenoids are important in human nutrition and health as antioxidants to fight risks of chronic diseases such as cardiovascular, cancer, and age-related sight problems that arise from deficiencies of lutein and zeaxanthin (Fraser and Bramley, 2004; Krinsky and Johnson, 2005). Moreover, the β-ionone ring containing carotenoids, provitamin A carotenoids, including α-carotene, β-carotene and β-cryptoxanthin serve as precursors of vitamin A (West JR, Keith, Darnton-Hill, 2008; Cuttriss et al., 2011).

Carotenoids in plants are synthesized via the carotenoid biosynthetic pathway (Figure 1.2), which is an intermediate component of the isoprenoid biosynthetic pathway, a pathway responsible for the synthesis of thousands of primary and secondary metabolites (Young et al., 2012). The carotenoid biosynthesis pathway is one of the most extensively elucidated metabolic pathways in plants (Hirschberg et al., 1997; Hirschberg, 2001; Cazzonelli and Pogson, 2010; Farre et al., 2010; Moise et al., 2014; Giuliano, 2014), and is responsible for the biogenesis of an array of about 600 40-carbon isoprenoid compounds broadly classified as xanthophylls and carotenes. Xanthophylls (e.g. β-cryptoxanthin, lutein and zeaxanthin) are oxygenated versions of carotenes (e.g. α-carotene, β-carotene and lycopene).

The first reaction dedicated to siphoning substrates to the carotenoid biosynthesis pathway in plants is catalyzed by an enzyme known as phytoene synthase (PSY). In this reaction, two geranylgeranyl pyrophosphate (GGPP) molecules are condensed to produce the first colourless linear carotenoid compound, phytoene. The C_{20} intermediate GGPP is predominantly synthesized by the upstream plastidal biosynthesis pathway, methylerythritol 4-phosphate (MEP) pathway within the isoprenoid biosynthetic pathway, through condensation of 3 molecules of Isopentyl diphospate (IPP) and Dimethylallyl diphospate (DMAPP). Phytoene is then modified through a series of desaturation and isomerization reactions catalyzed by enzymes including phytoene desaturase (PDS) and carotenoid isomerase (CRTISO) yielding the red coloured carotenoid, lycopene. Lycopene is the forking point in the pathway that leads to two separate downstream branches called α and β branches. In the α branch, carotenoids like α-carotene and lutein are synthesized, while in the β branch carotenoids such as β-carotene, β-cryptoxanthin and zeaxanthin are generated following cyclization of the terminals of the linear structured lycopene. Key enzymes involved in the branched part of the pathway (Figure 1.2) include lycopene epsilon α-cyclase (LCYE) and lycopene epsilon β-cyclase (LCYB) and β-carotene hydroxylase (CRTRB). LCYB can add β-ionone rings in both ends of lycopene to give β-carotene; while LCYE can add ε-ring in one end only to give α-carotene. Among all carotenoid compounds, only β-carotene has full vitamin A activity due to its doubly ended β-ionone rings, while carotenoids that have single β ring, like α-carotene and β-cryptoxanthin, have half vitamin A activity of β-carotene (Wurtzel,

2007; Cuttriss et al., 2011; Young et al., 2012; Shumskaya and Wurtzel, 2013; Giuliano, 2014).

Understanding the regulatory steps of the carotenoid biosynthesis pathway in maize endosperm is one of the important steps towards successful exploitation of the pathway for provitamin A biosynthesis via metabolic engineering or breeding. The mechanism of regulation of the carotenoid biosynthesis is still not fully understood (Shumskaya and Wurtzel, 2013). However, tremendous progress has been made in this regard; which included elucidation of bottlenecks from the upstream methyl erythritol phosphate (MEP) pathway to phytoene synthases and the branch point enzymes (Cunningham, 2002; Cazzonelli and Pogson, 2010; Messias et al., 2014).

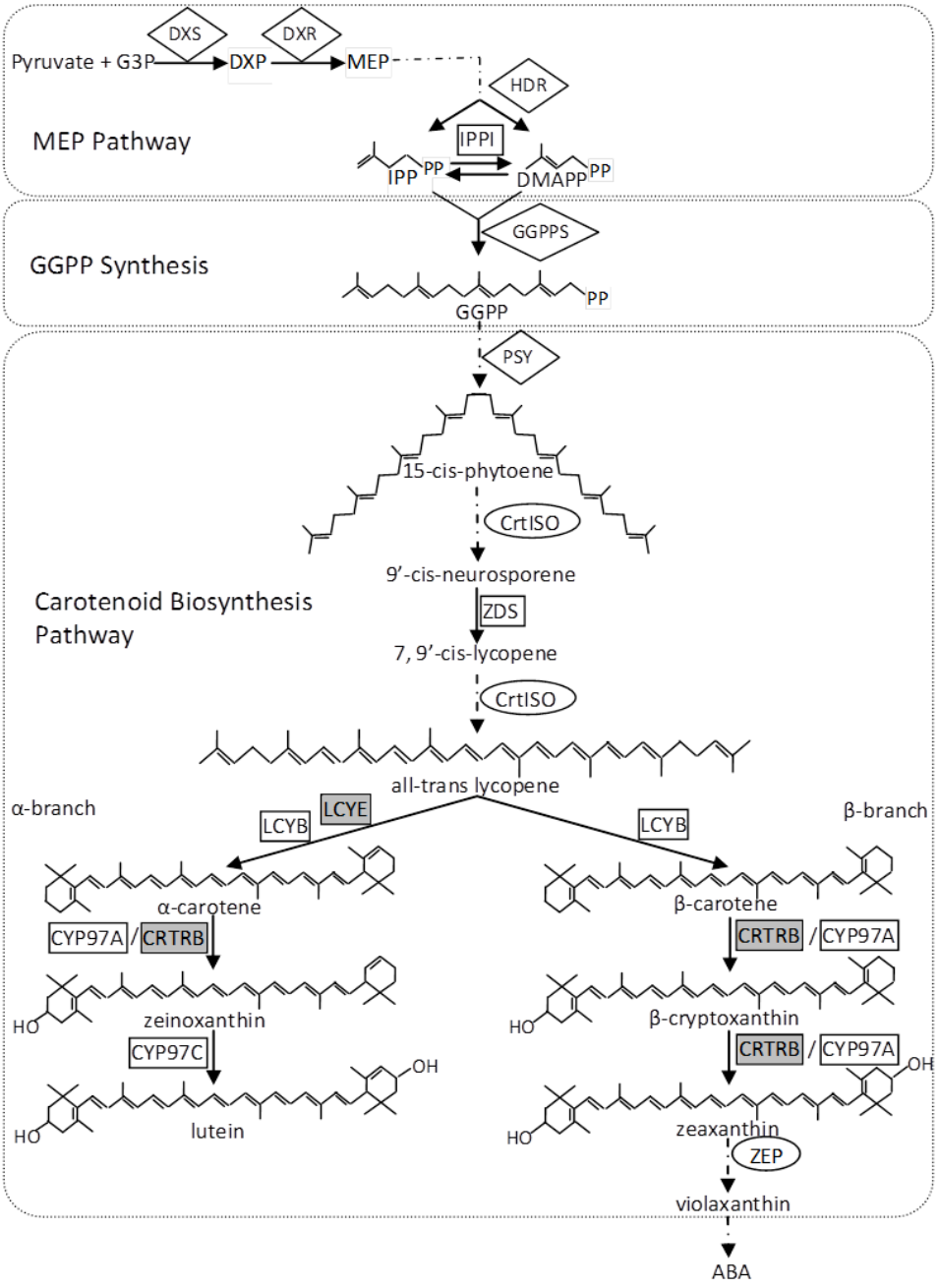


Figure 1.2 Schematic diagram of the carotenoid biosynthesis pathway and its upstream carbon supplying isoprenoid pathways in maize endosperm

Adapted from Vallabhaneni and Wurtzel (2009); Farre et al. (2010); Wurtzel et al. (2012). Arrows and boxed acronyms represent steps of enzymatic reactions. Only selected steps and enzymes indicated. Broken arrows are to suggest some of the skipped steps. Enzymes depicted in diamond and oval boxes were shown to have positive and negative correlation with maize kernel carotenoid content, respectively by Vallabhaneni and Wurtzel (2009). Enzymes indicated in shaded rectangular boxes were shown to have significant influence on β-caroten content in maize kernel by Harjes et al. (2008) and Yan et al. (2010). Acronyms: GA3P, glyceraldehyde 3-phosphate; DXP, 1-deoxy-D-xylulose 5-phosphate; DXS, DXP synthase; DXR, 1-deoxy-D-xylulose 5-phosphate reductoisomerase; MEP, methylerythritol 4-phosphate; HDR, 4-hydroxy-3-methylbut-2-enyl diphosphate reductase; IPP, isopentenyl pyrophosphate; IPPI, IPP isomerase; DMAPP, dimethylallyl pyrophosphate; GGPP, geranylgeranyl pyrophosphate; GGPPS, GGPP Synthase; PSY, phytoene synthase; CrtISO, carotene isomerase; LCYB, lycopene β-cyclase; LCYE, lycopene ε-cyclase; CRTRB, β-caroten hydroxylase family; and ZEP, zeaxanthin epoxidase.

1.9 Yellow maize germplasm under IITA's maize improvement program

The International Institute for Tropical Agriculture (IITA) is a CGIAR research institute with a mandate for developing and distributing maize germplasm adapted to Sub-Saharan Africa (http://www.iita.org/maize, accessed September 2014). In response to the need for agricultural intervention to the vitamin A deficiency problem, IITA initiated a breeding program in 1998 to develop maize varieties with high levels of provitamin A carotenoids (Maziya-Dixon et al., 2000). In the initial steps towards development of vitamin A enriched maize, different yellow maize materials within the breeding program of IITA were screened for their carotenoid content. The earliest work was reported by Maziya-Dixon et al. (2000) which demonstrated the existence of variation in total carotenoid concentration in selected maize germplasm. Another study also observed variability in the β-carotene levels of 17 varieties developed at IITA (Menkir and Mazia-Dixon, 2004). A more extensive carotenoid diversity study by Menkir et al. (2008) on tropical inbred lines clearly provided an account of the diversity of tropically adapted maize inbred lines developed at IITA both in terms of the composition and content of carotenoids. The non-provitamin A carotenoids lutein and zeaxanthin were the predominant carotenoids in these maize inbred lines, Promising inbred lines with relatively higher content of provitamin A carotenoids were also identified. The highest concentration of provitamin A recorded was 7.8 µg/g. The content and composition of carotenoid showed variation consistent with the two branches of the carotenoid biosynthesis pathway (Figure 1.2). Carotenoid content and composition of the inbred lines were demonstrated to be stable and less affected by replication, location and genotype by environment interaction indicating the significance of genetic control in the trait. This has paved the way for further studies to unravel the genetic basis of carotenoid content in adapted maize inbred lines and exploit this resource to facilitate the development of high provitamin A maize varieties for IITA's mandate area. Moreover, the study highlighted the potential for further germplasm enhancement and introgression of high provitamin A traits from the temperate maize germplasm (Menkir et al., 2008)

1.10 Target research areas of the PhD thesis project:

The PhD project aims to generate genetic/molecular tools and knowledge that can be used to advance the efficiency and speed of provitamin A maize breeding at IITA and in other breeding programs. The PhD research focused on three major research areas, which are presented in the following separate chapters, namely:

Chapter 2: Allele specific marker validation. In provitamin A breeding, selection is usually done by extracting and quantifying the various types of carotenoids of maize kernels using high performance liquid chromatography (HPLC) (Pfeiffer and McClafferty, 2007) since it is the most accurate currently available method. A visual selection based on kernel colour is ineffective because of the poor correlation between kernel colour and provitamin A (Harjes et al., 2008). HPLC, on the other hand, is highly expensive costing 50 to 70 US dollars per sample with very low throughput. The cost of HPLC makes it unattractive for routine selection decisions in provitamin A breeding (Pfeiffer and McClafferty, 2007). Marker assisted breeding can be a useful tool to overcome this bottleneck in provitamin A breeding. Scientists have identified and characterized three major rate limiting provitamin A carotenoid biosynthesis genes namely *PSY1*, *lcyE* and *crtRB1*, that were shown to significantly control the concentration of total carotenoids and β-carotene in maize endosperm. The researchers identified polymorphic sites in each of the key genes that were strongly associated with beta-carotene concentration in maize endosperm, and further developed breeder friendly allele specific markers that can be used for screening maize genotypes for favourable and unfavourable variants of the genes. The efficiency of these markers in different genetic backgrounds needed to be tested prior to their full scale use for provitamin A breeding of tropically adapted maize germplasm as there had been inconsistent previous results. This chapter presents the result of investigation of the allele specific markers using a set of diverse maize inbred lines developed at IITA from populations derived from various crosses of tropical and temperate germplasm.

Chapter 3: Genome-wide marker discovery using a high density SNP-carotenoid association scan. Current advances in next generation sequencing and the associated decline in price per unit datapoint has been introducing a paradigm shift in

population genetics and molecular breeding approaches for it is now possible to subject hundreds, if not thousands, of individuals to high density genotyping at genome-wide scale at increasingly lower prices. This is important as it can facilitate the dissection of the genetic basis for many agriculturally and nutritionally important complex traits and discover many useful alleles to design genomics assisted cultivars. Chapter 3 employs genotyping by sequencing (GBS) generated high density genome-wide SNPs to discover loci associated with the composition and quantity of carotenoids in maize endosperm. The rationale behind this research was the complexity of the carotenoid biosynthesis pathway and the involvement of many genes in carotenoid biosynthesis which suggests the potential small to medium effect influence of many genes involved in the pathway. Efficient marker assisted selection can be achieved if all loci contributing to the variation in carotenoid profile within a diverse The germplasm can be utilized through identification of sequence variation across the whole genome by making use of linkage disequilibrium based genome-wide association mapping. The result of this study can be used to develop cultivars with desirable traits through marker assisted breeding and genomic selection.

Chapter 4: Combining ability and heterosis for provitamin A. The PhD research also assessed the combining abilities of a set of inbred lines selected based on their carotenoid profiles. The genetic information obtained from this study can help understand the mode of inheritance and interaction of different profiles of provitamin A carotenoids in maize endosperm. The pattern of inheritance and interaction of the carotenoid levels will further be assessed in relation to the molecular markers causally linked to the traits to determine whether the markers can be efficiently used to predict provitamin A contents in hybrids based on molecular information generated from parental inbred lines.

1.11 Research Objectives

The objectives of this PhD research were:

1. To investigate the carotenoid content and composition of diverse yellow maize inbred lines developed for Sub-Saharan Africa from mixed genetic background of temperate and tropical germplasm.

2. To genotype the inbred lines using allele specific markers designed for *PSY1*, *lcyE* and *crtRB1*.
3. To conduct association analysis between carotenoid content and allele specific markers of *PSY1*, *lcyE* and *crtRB1*, in order to investigate their usefulness in provitamin A biofortification of inbred lines containing tropical and temperate germplasm in their genetic background.
4. To identify SNP polymorphisms associated with carotenoid content and profile across the maize genome using diverse but partially related maize inbred lines.
5. To determine the mode of inheritance of provitamin A carotenoid content and composition.
6. To assess the effect of combinations of different allelic variants on the provitamin A content in hybrids.

Chapter 2: Evaluation of allele specific DNA markers for key genes involved in regulation of provitamin A accumulation in maize endosperm

2.0 Abstract

Allele specific DNA markers have recently been reported for three key carotenoid biosynthesis genes (*PSY1*, *crtRB1* and *lcyE*) for possible application in marker assisted biofortification of maize with provitamin A. However, previous validation studies of these markers showed some inconsistencies. In addition, the populations employed for developing and validating the functional markers were largely of temperate origin where the favourable allele of the most significant marker was rare. In the present study the markers were further examined using a panel of diverse inbred lines developed by the International Institute of Tropical Agriculture (IITA) for Sub-Saharan Africa. The inbred lines were assessed for their carotenoid content and profile for two seasons, and lines that accumulated up to 16.38 µg/g dry weight (DW) β-carotene and 17.25 µg/g DW provitamin A were identified. The *PSY1* markers were found to be monomorphic across all the inbred lines for the favourable alleles. *CrtRB1* 5'TE and *crtRB1* 3'TE markers had the strongest effect on β-carotene content. Genotypes with the favourable alleles of these two markers (N = 20) achieved 3.22 fold average increases in β-carotene content over those genotypes without the favourable alleles (N = 106). Markers of *lcyE* were associated with lutein and ratio of alpha to beta branches carotenoids but not with provitamin A. However, the joint effects of functional markers of *lcyE* and *crtRB1* were larger than their individual effects for all carotenoids tested. The inbred line (KU1409/DE3/KU1409) S2-18-2-B-B-B-B was recommended for use as a donor parent of the favourable alleles of *crtRB1* polymorphisms, considering its relatively desirable agronomic traits and high levels of β-carotene and total provitamin A estimated at 11.30 and 12.11 µg/g DW, respectively. This study has demonstrated that the allele specific markers, especially, of *crtRB1* can be deployed to facilitate the biofortification of tropical maize with provitamin A which will contribute to the alleviation of vitamin A deficiency in Sub-Saharan Africa.

2.1 Introduction

Vitamin A deficiency is a global public health problem inflicting morbidity, stunted growth, night blindness as well as loss of sight and lives among millions of preschool children and pregnant and/or lactating women (Sherwin et al., 2012; WHO, 2009). Tackling this problem is the most important and benefitial priority among the challenges of the current world, which entails deploying sustainable and cost effective problem solving strategies (Copenhagen Consensus, 2012). Along with diet diversification, nutrient supplementation and industrial food fortification, agricultural approaches for enhancing the provitamin A content of staple crops such as maize has long been considered as sustainable and cheap solution to the problem of vitamin A deficiency (Underwood, 2000).

Maize is an important staple crop that is known for its natural capability of accumulating both provitamin A and non-provitamin A carotenoids in its endosperm (Burt et al., 2011b; Harjes et al., 2008; Menkir et al., 2008). However, provitamin A usually constitutes only 10 to 20% of the total carotenoids in maize seed, and the widely produced and consumed yellow maize cultivars in the world have less than 2 µg/g provitamin A (Pixley et al., 2013). Exploitation of the natural genetic diversity of maize in carotenoids through conventional and molecular breeding for Provitamin A biofortification (Bouis et al., 2011; Wurtzel et al., 2012) is proposed as a cost effective, safe and sustainable solution that can contribute towards the alleviation of vitamin A undernutrition and its consequences across the globe (Bouis and Welch, 2010; Meenakshi et al., 2010b; Msangi et al., 2010; Pixley et al., 2013; Qaim et al., 2007; Underwood, 2000).

One of the major challenges to the effort of maize provitamin A biofortification is the quantification of carotenoids in maize endosperm. High performance liquid chromatography (HPLC) is the commonly used method for this purpose. Although this procedure is accurate, it is highly expensive, time consuming and low throughput, limiting its use for routine breeding (Pfeiffer and McClafferty, 2007). Moreover, even though the various tints of yellow colour in maize endosperm are attributable to variations in carotenoid contents and profile, selection for high provitamin A maize based only on kernel colour is not reliable due to poor correlations between the two traits (Harjes et al., 2008; Mishra and Singh, 2010).

Marker assisted selection (Moose and Mumm, 2008) using functional markers (Andersen and Lübberstedt, 2003) can be used as a complementary breeding tool to screen a large number of breeding materials for carotenoid profile and content accurately, cost effectively and quickly, thus circumventing the challenge of complex time taking and expensive phenotyping.

Functional markers are gene targeted molecular markers devised based on polymorphic sequences which can be situated inside and around the gene of interest, and which have causative effects on the variation of the phenotype controlled by the gene (Andersen and Lübberstedt, 2003; Gupta and Rustgi, 2004; Varshney et al., 2007). Since they represent the nucleotide sequence variation underlying an observed trait, functional markers do not suffer from the disadvantage of uncoupling of marker-to-traits linkages as can occur with genetically linked markers, which are developed only on the basis of genetic linkage between markers and traits disregarding whether the markers are located on the nucleotides/locus controlling the trait. Several investigations have reported the development and successful application of functional markers for various traits of important crop plants (Bagge et al., 2007; Perumalsamy et al., 2010).

The genes involved in the carotenoid biosynthesis pathway (Figure 1.2) are well studied (Cunningham and Gantt, 1998; Vallabhaneni and Wurtzel, 2009; Wurtzel et al., 2012). Natural nucleotide sequence variants within the key carotenogenic genes have also been characterized and shown to have significant contribution to variation in accumulation of provitamin A and total carotenoids in maize endosperm (Fu et al., 2013; Harjes et al., 2008; Yan et al., 2010). Fu and colleagues (2013) identfied two polymorphisms in phytoene synthase (*PSY1)*, the gene acting at the gateway of the pathway controlling substrate influx, which explained 7 to 8% of the variation in total carotenoids. Harjes and colleagues (2008), on the other hand, identified four significant functional polymorphic sites in the gene encoding lycopene epsilon cyclase (*lcyE*) that altered the ratio of carotenoids in α to β branches of the carotenoid synthesis pathway, leading to a threefold increase in provitamin A. Yan and colleagues (2010) also demonstrated three significant functional polymorphic sites in another critical downstream gene encoding a β-carotene hydroxylase (*crtRB1*) accounted for 40% of the variation in β-carotene concentration in maize

endosperm. These researchers not only identified the functional polymorphic sites but developed breeder-friendly polymerase chain reaction (PCR) based functional markers for detecting the allelic variants of the significant polymorphic sites in the three genes.

These functional markers of provitamin A genes are important for faster and efficient biofortification of maize with high level of provitamin A. However, consistencies of the markers need to be investigated before their widespread utilization. Some validation studies examined the individual and combined effects of the functional markers of *lcyE* and *crtRB1* on carotenoids using different genetic backgrounds (Babu et al., 2013; Burt et al., 2011b; Fu et al., 2013; Vignesh et al., 2012). Burt and colleagues (2011b) observed that the diagnostic polymorphisms of *lcyE* could not distinguish between the high lutein and high zeaxanthin inbred lines representing carotenoids in the α- and β-branches of the pathway. The authors also observed inbred lines with high level of β-carotene despite they were harboring the unfavourable alleles of *crtRB1* at both 5'TE and 3'TE (Burt et al., 2011b). Similar discrepancies were reported by Vignesh and colleagues (2012) in which they observed lines carrying the favourable allele for 3'TE *crtRB1* exhibiting low levels of β-carotene. Babu and colleagues (2013) validated two of the three significant polymorphic sites of *lcyE* (5'TE and 3' indel) and one of the three functional polymorphisms of *crtRB1* (3'TE) using 26 different tropical segregating populations. Their result showed that the effects of *lcyE* on both ratio of α to β branch carotenoids and total provitamin A content were inconsistent across the populations. In contrast, they showed that *crtRB1* polymorphic site had large effect on β-carotene and provitamin A concentrations. On the other hand, Fu et al. (2013) detected significant effects of all the functional polymorphisms for individual and haplotypes of selected polymorphisms of *lcyE*, *crtRB1* and *PSY1* using inbred lines with tropical, subtropical and temperate backgrounds. These results demonstrate the inconsistency in the diagnostic markers of the favourable alleles of *lcyE* and *crtRB1* and thus necessitate further validation of the markers to identify the ones that are robust and effective enough for screening inbred lines with different diverse and mixed genetic background. Yellow inbred lines developed at IITA with both tropical and temperate genetic background are considered suitable for investigation of these markers.

The objectives of the study were (1) to determine the carotenoid content of a diverse set of yellow endosperm maize inbred lines (2) to screen the inbred lines using PCR-based allele specific markers of *lcyE, crtRB1 and PSY1* and (3) to conduct marker-trait association and investigate the usefulness of the markers under mixed genetic background of both tropical and temperate germplasm (4) to identify inbred lines that can serve as donor parents of the favourable alleles of the significant functional markers and at the same time displaying acceptable level of resistance to the major foliar diseases for the study area, and having good ear and plant aspect scores.

2.2 Material and Methods

2.2.1 Plant materials

One hundred and thirty diverse yellow maize inbred lines were used in the marker-trait association validation study. These inbred lines had been developed at the International Institute for Tropical Agriculture (IITA) from eight bi-parental crosses of tropical inbred lines, four broad based populations, and 28 backcrosses, all involving temperate lines as donors of high β-carotene alleles (Table 2.1).

Table 2.1 Origins of the 130 inbred lines used in the present study

Origin	Number of lines
4001	1
9450	1
KU1409	1
9450/KI28	2
9450/KI21	6
9450/CM116/9450	2
9450/KI21-1-4-1-1-1-B/DE3/9450/KI21-1-4-1-1-1-B	4
9450/KI28-1-2-1-1-B/DE3/9450/KI28-1-2-1-1-B	3
9450/KI21-1-4-1-1-1-B/DE3/9450/KI21-1-4-1-1-1-B	1
9450/KI21-1-5-3-2-1-B/DE3/9450/KI21-1-5-3-2-1-B	1
9450/KI21-1-5-3-2-2-B/DE3/9450/KI21-1-5-3-2-2-B	2
9450/KI21-3-2-2-1-3/KU1409/MO17LPA/KU1409	1
DE3/KU1414-SR/KU1414-SR	3
KU1409/NC358/KU1409	3
KU1414-SR	1
KU1414-SR/KVI11	2
KU1409/DE3/KU1414-SR	1
KU1409/DE3/KU1409	23
KU1409/KU1414-SR/A619	9
KU1409/KU1414-SR/KVI11	1
KU1409/KU1414-SR/KVI3	12
KU1409/KU1414-SR/M162W	3
KU1409/KU1414-SR/NC298	8
KU1409/KU1414-SR/NC350	10
KU1409/KU1414-SR/SC55	2
KU1409/SC55/KU1409	4
KU1414-SR/CI7/KU1414-SR	2
KU1414-SR/CML328/KU1414-SR	1
POP66SR/ACR91SUWAN1-SRC1/ACR91SUWAN1-SRC1-1/SYN-Y-STR-34-1-1-1-1-2-1-B*3	2
POP66SR/ACR91SUWAN1-SRC1/ACR91SUWAN1-SRC1-4/4001/KI21-4-1-1-1-1	2
POP66SR/ACR91SUWAN1-SRC1/ACR91SUWAN1-SRC1-6/(MP420/4001/MP420)	3
POP66SR/ACR91SUWAN1-SRC1/ACR91SUWAN1-SRC1-8/POP61-SR-11-2-3-3-1-B	2
POP66SR/ACR91SUWAN1-SRC1/ACR91SUWAN1-SRC1-9/(9450/CM116/9450)-3-3-1-2-1	1
SYN-Y-STR-34-1-1-1-1-2-1-B*3/(DE3/CI7)/SYN-Y-STR-34-1-1-1-1-2-1-B	1
SYN-Y-STR-34-1-1-1-1-2-1-B*5/NC354/SYN-Y-STR-34-1-1-1-1-2-1-B*5	1
SYN-Y-STR	1
4205/CI7/4205	1
(ACR97TZL-CCOMP1-Y-S3-13-1-B*2/CI7/ACR97TZL-CCOMP1-Y-S3-13-1-B*2)	1
ACR97TZL-CCOMP1-Y	2
SC55/KU1414-SR/KU1414-SR	1
TZE-COMP5-Y-C7	1
Z.Diplo BC4	1

The inbred lines included in the current study were selected based on a broad range of endosperm provitamin A carotenoid profile from previous preliminary tests. Many of the inbred lines included in the present study were also shown to have high level of genetic diversity using SSR and AFLP markers (Adeyemo et al., 2011a, 2011b).

2.2.2 Field trials

The inbred lines were evaluated at IITA's research station, Ibadan, Nigeria ($7^{o}29'11.99''N$, $3^{o}54'2.88''E$, altitude 190 m) for two seasons, in 2010 and 2011. The trial was arranged in a 13 x 10 alpha-lattice design with two replications. Planting space of 0.75 m between rows and 0.25 m between plants within a row was used. Different fields were used within IITA research station in each season. The fields were managed as per the recommended agronomic practices which included fertilization with NPK at planting, and top dressing with urea during flowering. Weeds were controlled using Primextra and Gramazone herbicides applied as pre-emergence, with supplemental hand weeding when necessary during the rest of the growing seasons.

Agronomic data that included major foliar diseases, date of male and female flowering, plant height, ear height, kernel colour, plant aspect and ear aspects were recorded. Ear and plant aspects were scored on 1 to 5 scales corresponding to the best to the worst. Kernel colour was scored based on HarvestPlus standard colour stripes for total carotenoid content, which was converted to a 1 to 9 scale reflecting pale yellow to deep orange colours as observed within the set of inbred lines evaluated (Appendix 2 – Figure 1).

2.2.3 Carotenoid analysis

Seed samples for carotenoids analyses were produced by controlled self pollination of all typical plants in each plot. Ears of each self pollinated line in each plot were harvested, well dried under ambient temperature with minimal exposure to direct sun light, and shelled separately. Samples of one hundred kernels were drawn and sent to the University of Wisconsin, USA for carotenoid analysis.

Carotenoids were extracted and quantified using HPLC at the University of Wisconsin, USA. The extraction protocol was the method of Howe and Tanumihardjo (2006) for analysis of dried maize kernels. Finely ground 0.5 g sample of each entry was transferred into a 50 ml glass centrifuge tube to which 6 ml of Ethanol (Decon Labs Inc. King of Prussia, Pennsylvania, United States) plus 0.1% butylated hydroxyl toluene (ICN Biomedicals Inc. Aurora, Ohio, United States) were added. After vortexing for 15 seconds tubes were placed in 85°C water bath for 5 min and 500 µl of 80% potassium hydroxide (w:v) (Mallinckrodt Chemicals, Phillipsburg, New Jersey, United States) was added. The samples were vortexed for 15 seconds, and put back in the water bath for 10 min with vortexing at 5 min interval approximately. They were then immediately placed on ice and 3 ml ice cold deionized water was added, vortexed for 15 seconds, followed by addition of 200µl of the internal standard β-Apo-8'-carotenal (Sigma-Aldrich St. Louis Missouri United States) and 4 ml hexane (Fisher Scientific, Fair Lawn, New Jersey, United States). After vortexing and centrifugation, the top hexane layer formed was transferred into a new test tube. The hexane extraction was repeated twice, adding 3 ml hexane each time. Samples were allowed to dry down completely under nitrogen gas using a Turbovap LV concentrator (Caliper Life Sciences) and reconstituted in 500 µl of 50:50 Methanol:Dichloroethane (both Fisher Scientific). Following vortexing and centrifuging, the extracts were transferred to HPLC vials placed in the autosampler tray and 50 µl aliquots of each extract were injected into an HPLC system (Waters Corporation, Milford Massachusetts, United States). The Waters HPLC system was operated with Empower 1 software and included a 717 Plus auto sampler with temperature control set at 5°C, a Waters 1525 binary HPLC pump, and a 2996 photodiode array detector for carotenoid quantification. Carotenoids were separated on a 3 µm C30 YMC Carotenoid Column (4.6 × 250 mm) eluted with a mobile phase of methanol/water (92:8 v/v) with 10 mM ammonium acetate as solvent A, and 100 % methyl tertiary butyl ether as solvent B. The gradient was applied for 30 minutes from 70% solvent A:30% solvent B, to 40% solvent A:60% solvent B. The flow rate was 1.0 mL/min. To maximize detection of carotenoids, the absorbance was measured at 450 nm. Alpha-carotene, β-carotene (*cis* and *trans* isomers), β-cryptoxanthin, lutein, and zeaxanthin were quantified based on calibrations using the respective external standards. Total carotenoid was calculated

as the sum of concentrations of α-carotene, lutein, β-carotene, β-cryptoxanthin, zeaxanthin). Provitamin A was calculated by adding the concentrations of β-carotene, and half concentrations of each of β-cryptoxanthin and α-carotene, since β-cryptoxanthin and α-carotene can provide only one molecule of retinol each as opposed to two molecules of retinol for β-carotene (US Institute of Medicine, 2001). Other derived carotenoid traits were also calculated as indicated in Harjes et al. (2008) and Yan et al. (2010): i.e., ratio of carotenoids in the β versus α branches of the carotenoid pathway, ratio of β-carotene to β-cryptoxanthin and ratio of β-carotene to all carotenoids (β-carotene + α-carotene + lutein + zeaxanthin + β-cryptoxanthin). The data for the ratio traits were transformed with natural logarithm before subjecting to statistical analysis since they followed non-normal distribution. All carotenoids concentrations were measured in µg/g dry weight (DW).

2.2.4 Genotyping with allele specific PCR-based markers

Maize leaf samples were collected from 3 to 4 randomly selected typical plants of each inbred line from one of the replications of the trial described above ~40 days after planting. DNA samples were isolated from freeze dried leaf samples of each genotype using either a CTAB (cetyl trimethyl ammonium bromide) based DNA extraction protocol or QIAGEN DNeasy® Plant Mini kit (Qiagen Inc., Hilden, Germany) following the company's protocol.

PCR based functional markers of three genes *lcy*E, and *crtRB1*, *PSY1* were tested across the 130 inbred lines. Primers, PCR conditions and thermal cycling profiles were as reported by Harjes et al. (2008) for *lcy*ε, Yan et al. (2010) for *crtRB1* and Fu et al. (2013) for *PSY1* (Appendix 1 – Table 1). The original two pairs of primers used by Harjes et al. (2008) to amplify the *lcyE*-3'TE indel marker were replaced by a single pair primer developed and tested by CIMMYT, as the former did not give consistent result in the present study. The replaced primer sequences were: forward ACCCGTACGTCGTTCATCTC and reverse ACCCTGCGTGGTCTCAAC (Babu et al., 2013). Primers were ordered from Integrated DNA Technology Inc (IDT, Belgium). All PCRs were run using BIOTAQ^{TM} DNA polymerase kit (Bioline Ltd, UK) with a mixture composed of 2 µl 10x NH_4 PCR buffer, 1 µl of each primer, 1 or 1.5 µl (depending on the marker) of 50 mM $MgCl_2$, 0.15 µl of BIOTAQ^{TM} polymerase, 1 µl of Dimethyl Sulfoxide (DMSO) to enhance specificity, and ultra

pure water making up to 25 µl total volume. PCR fragments were confirmed by sequencing three samples representing each allele of the 6 functional markers. PCR product sequences were aligned with sequences of the three genes, downloaded from GenBank of NCBI or MaizeGDB, using CLC genomics workbench (CLC Bio, Denmark) sequence analysis software. Fragments in the PCR products were resolved using 2% super fine resolution (SFR™) agarose gel. Names of polymorphic sites of each gene and the nature of polymorphisms are indicated in Table 2.2 according to their respective references.

Table 2.2 Nomenclature of functional markers and their allelic series as per their references

Gene	Polymorphic site/Marker(gene name-polymorphism)	Nature of Polymoprphism	Allelic series and notations*
PSY1 (Fu et al., 2013)[a]	*PSY* -SNP7	A-C substitution	**A**, C
	PSY1-IDI	378 bp indel	0, **378**
LCYE (Harjes et al., 2008)	*LCYE*-5'TE	285 indel	**1**, 2, 3, **4**
	LCYE-SNP (216)	G-C SNP	**G**, T
	LCYE-3'indel	8 bp indel	8, **0**
crtRB1 (Yan et al., 2010)	*crtRB1*-5'TE	397/206 bp indel	1, **2**, 3
	crtRB1-InDel4	12 bp indel	**12**, 0
	crtRB1-3'TE	325/1250 bp indel	**1**, 2, 3

*Allelic variants denoted in bold face underlined letters represent the best favourable alleles as described in the references. In the current study *lcyE 5'TE* yielded no amplification for 73% of the inbred lines invariably, thus scored as a '0' allele to mean 'no amplification').

2.2.5 Statistical analysis

The carotenoid data were analyzed using the PROC MIXED procedure of SAS® version 9.3 (SAS Institute, 2012) based on alpha lattice design in which lines were treated as fixed effects, while blocks, replications, years and year by line interaction were treated as random effects. Repeatability (r) of each carotenoid was estimated as described in Fu et al. (2013). Spear man rank correlation coefficient was calculated using SAS 9.3 (SAS Institute, 2012) to test the consistency of ranking of the inbred lines for carotenoid content across seasons (Menkir et al., 2008). In addition, correlations between pairs of different carotenoids, kernel colour and some agronomic traits were calculated using PROC CORR of SAS 9.3 (SAS Institute, 2012).

Associations between variation in carotenoid concentration and markers of each gene were calculated using the mixed linear model (MLM) (Yu et al., 2006) implemented in TASSEL version 3.0 (Bradbury et al., 2007). MLM incorporates population structure and kinship in the analysis to control spurious association results (Yu et al., 2006). Linkage disequilibrium between markers was also calculated using the same software. For across location analysis, best linear unbiased estimates (BLUEs) were calculated from across year data within Tassel using the generalized linear model (GLM) option by selecting only the phenotype data (Bradbury et al., 2007). Population structure (principal component analysis, PCA) and kinship of the 130 inbred lines were estimated within TASSEL 3.0 using 62, 0000 SNPs that covered the 10 maize chromosomes generated by genotyping by sequencing (GBS) method at the Institute for Genomic Diversity (IGD), Cornell, USA, according to Elshire et al. (2011). The SNPs were filtered out from the GBS pipeline output using threshold of 5% minimum minor allele frequency and 20% maximum missing data. In addition, 2328 SNPs were filtered using 0% missing and 5% minimum allele frequency and used for hierarchical clustering of SNP data for 26 inbred lines that harbored the favourable alleles of the two most significantly associated markers (*crtRB1-5'TE* and *-3'TE*) to assess their relationship and differentiation. The unweighted pair-group method with arithmetical averages (UPGMA) provided in PowerMarker version 3.25 (Liu and Muse, 2005) was employed to construct a dendrogram from Nei's 1972 frequency based distance matrix (Nei, 1972). Single and two-way ANOVA were conducted to determine genotype effects of the functional polymorphisms using PROC GLM and PROC MIXED of SAS software version 9.3 (SAS Institute, 2012).

2.3. Results

2.3.1 Carotenoid profiling

The analysis of variance (ANOVA) combined over the two years revealed highly significant variation among the maize inbred lines for all carotenoids, except for α-carotene (Table 2.3). The effects of year, and year by line, interactions were significant on all carotenoids, but not for β-cryptoxanthin. High repeatability estimates ranging from 62 to 89%, were recorded for all carotenoids excluding α-carotene, demonstrating the importance of the genetic component of the total variation observed for the traits. Replication did not have a significant effect on all carotenoids. Spearman's rank correlation coefficients across years were significant (p<0.001) for β-carotene (r = 0.83), β-cryptoxanthin (r = 0.75) zeaxanthin (r = 0.86) α-carotene (r = 0.30) and lutein (r = 0.49). Zeaxanthin was the dominant carotenoid with average mean value of 9.66 µg/g followed by β-carotene, 4.21 µg/g and lutein, 3.58 µg/g. The average mean values for provitamin A and total carotenoid were 5.87 and 20.78 µg/g, respectively. The α-carotene contents of most of the maize inbred lines were very low and not significantly different from zero (apart from some 16 inbred lines), which can be an explanation for the low repeatability value for this particular carotene. Estimated means averaged over the two years varied from 0.45 to 13.51 µg/g for lutein, from 0.04 to 25.90 µg/g for zeaxanthin, from 0.08 to 8.55 µg/g for β-cryptoxanthin, from 0 to 16.38 µg/g for β-carotene, from 0 to 17.25 µg/g for provitamin A, and from 4.43 to 42.71 µg/g for total carotenoids (Table 2.4, Figure 2.1a and 2.1b; Appendix 1 – Table 3).

Table 2.3 Combined ANOVA of carotenoid content of 130 inbred lines evaluated in 2010 and 2011 combined across

Sources of variation	DF	Mean squares of Carotenoids						
		lut	zeax	βcry	αcar	βcar	pva	tcar
Line	129	26.81***	133.99***	12.37***	0.43	40.71***	41.49***	226.72***
Year	1	519.32***	878.82**	7.95	51.98***	48.63**	84.07**	332*
Rep (Year)	2	2.81	22.07	0.69	0	0.57	1.29	23.58
Block(Rep*Year)	36	3.75**	13.43***	0.44	0.01*	0.88	1.36	31.32***
Line*Year	129	10.32***	14.21***	1.62***	0.37***	5.21***	6.63***	43.45***
Residual	222	2.15	4.08	0.48	0	0.76	1.01	11.28
r		0.62	0.89	0.87	0.15	0.87	0.84	0.81

Carotenoids are abbreviated as lut = Lutein, zeax = Zeaxanthin, βcry = β-cryptoxanthin,
αcar = alpha-carotene, βcar = β-carotene, pva = provitamin A, tcar = Total carotenoid, r = repeatability
*, **, *** = significant at P ≤ 0.05, 0.01, and 0.001, respectively.

Table 2.4 Means of selected inbred lines based on their provitamin A and genotypes of functional polymorphisms representing *lcyE and crtRB1* genes

	Carotenoids* (µg/gdry weight)							Genotype**					
								lcyE			*crtRB1*		
Entry No.	**lut**	**zeax**	**βcry**	**αcar**	**βcar**	**tpva**	**tcar**	**5'TE**	**SNP (216)**	**3'InDel**	**5'TE**	**InDel4**	**3'TE**
39	1.70ab	2.72c	0.44c	1.14a	16.38a	17.25a	22.88abc	'0'	**G**	8	**2**	0	**1**
45	1.03b	2.39c	0.34c	0.97a	14.79ab	15.45ab	19.53bc	'0'	**G**	8	***2***	0	**1**|3
124	0.99b	1.12c	0.16c	1.18a	13.64abc	14.33ab	17.08bc	3	T	**0**	2	**12**|0	**1**|3
106	1.80ab	0.26c	1.30c	1.21a	11.80abc	13.02abc	16.30bc	'0'	**G**	8	**2**	**12**	**1**
99	5.80ab	2.32c	0.39c	1.20a	11.30abc	12.11abc	21.10abc	3	T	**0**	**2**	0	**1**
23	5.76ab	4.83bc	0.75c	1.02a	10.77abc	11.63abc	23.04abc	'0'	**G**	8	**2**	**12**	**1**
101	12.09a	5.02bc	0.83c	1.39a	9.76abcd	10.91abc	29.19abc	'0'	**G**	8	**2**	**12**	**1**
107	1.90ab	1.05c	0.33c	0.80a	9.07abcd	9.65abc	13.17 c	'0'	**G**	8|**0**	**2**	**12**	**1**
11	2.03ab	10.80abc	1.99bc	0.83a	7.61bcd	9.06abc	23.48abc	'0'	**G**	8	1	0	3
50	7.28ab	18.74a	8.55a	0.57a	7.54bcd	12.10abc	42.71a	3	T	8|**0**	1	0	3
92	3.96ab	15.56ab	6.02ab	0.70a	6.72cd	10.10abc	32.98abc	'0'	G	8	1	**12**	3
98	3.22ab	9.43abc	2.59bc	0.77a	6.30cd	7.98bc	22.27abc	**1**	**G**	**0**	1	0	**1**|3
120	1.37ab	10.36abc	1.92bc	0.60a	6.27cd	7.52bc	20.46bc	'0'	**G**	8	1	0	3
96	8.89ab	19.82a	3.28bc	0.37a	2.93d	4.76c	35.41ab	'0'	**G**	**0**	1	0	3
SE	1.95	2.36	0.67	0.45	1.23	1.40	3.53						
Min.	0.45	0.04	0.08	0	0.03	0.06	4.43						
Max.	13.51	25.90	8.55	1.68	16.38	17.25	42.71						
Grand Mean	3.58	9.66	2.92	0.4	4.21	5.87	20.78						

*Means within column followed by same letters are not significantly different using least square means and Tukey's multiple comparison test (p<0.05). Letters for mean separation were generated using SAS macro by (Saxton, 1998). Abbreviations of carotenoids are described under Table 2.3.

Allelic series of genotypes composed of six polymorphisms: *lcyE* 5'TE, *lcyE* SNP (216), *lcyE* 3'indel, *crtRB1* 5'TE, *crtRB1* InDel4, and *crtRB1* 3'TE from left to right (discussed in later sections). Heterozygousity of the T allele of the dominant marker SNP(216) is not determined. Heterozygous alleles are separated by '|**'. Favourable alleles are bolded and underlined; description of the markers is presented in Table 2.2. The haplotype 3, T, 8**|**–, 1, 0, 3 represent the worst allelic combination. The haplotype 0, **G**, 8**|0**, **2**, **12**, **1** represent the best allelic combination.

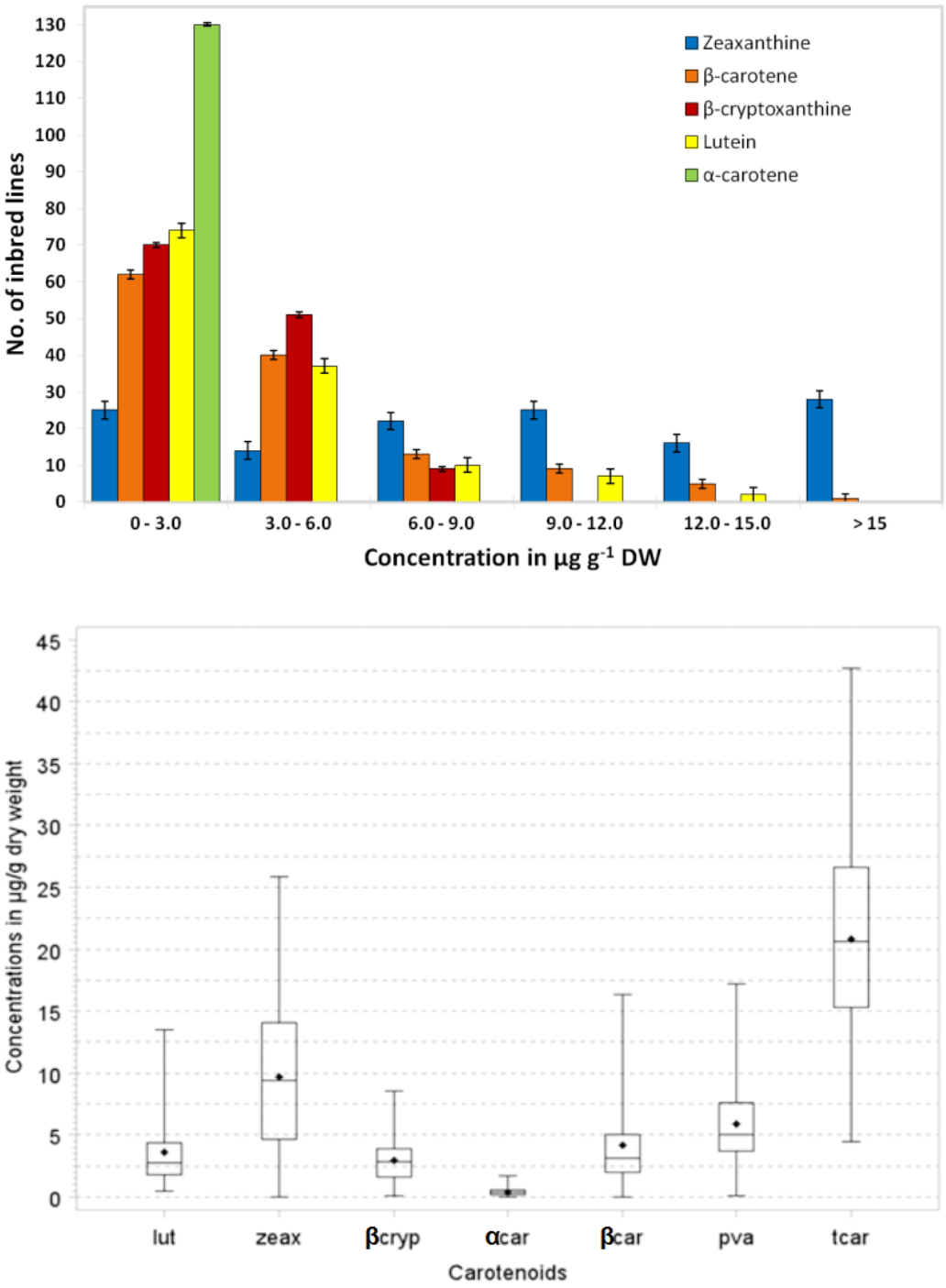


Figure 2.1 Distribution of mean concentration estimates of carotenoids in µg/g DW across 130 inbred lines.

(a) Histogram, error bars represent standard error of least square means of the respective carotenoid concentration. (b) Box plots; endpoints of upper and lower whiskers represent maximum and minimum concentrations, respectively; upper and lower edges of boxes represent third and first quartiles, respectively; line inside box represent median; symbol ◆, represent mean. Abbreviations for carotenoids are described in Table 2.3.

Pearson pairwise correlation analysis showed that β-carotene had significant negative associations with β-cryptoxanthin and zeaxanthin, which is in agreement with the positions of the carotenoids in the biosynthesis pathway. Kernel colour had significant (p<0.0001) positive correlation with Zeaxanthin (r = 0.64), β-cryptoxanthin (r = 0.49) and total carotenoid (r = 0.63) (Appendix 1 – Table 3; Appendix 2 – Figure 2, Appendix 2 Figure 3). Correlation of kernel colour with the rest of carotenoids was non-significant. The lack of significant correlation between kernel colour and β-carotene and provitamin A justifies the importance of use of markers to assist selection (Harjes et al., 2008).

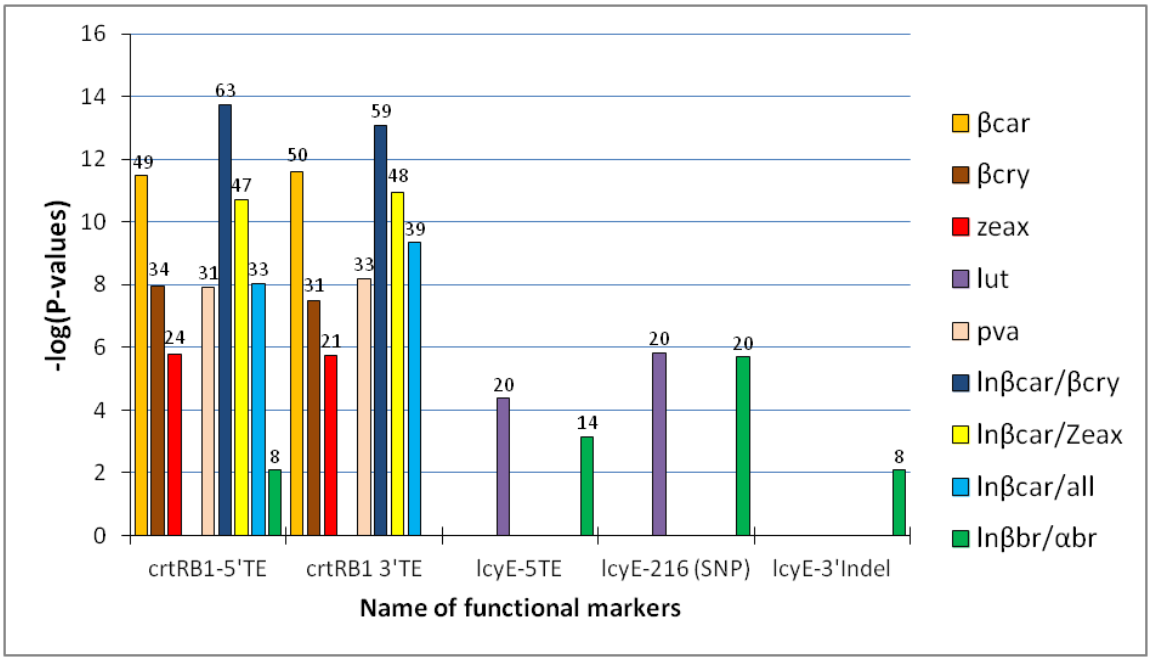


Figure 2.2 Significant marker-trait associations estimated using BLUEs for the two years carotenoid data.
Significance thresholds: –log (1.7E-04) = 4 at α = 0.01 and –log (8.3E-3) = 2 at α = 0.05. Carotenoids abbreviations described in Tables 2.3 and 2.7. Values indicated on top of each bar are proportion of variation accounted for by each marker (R^2).

2.3.2 Genotyping for *PSY*, *lcyE*, and *crtRB1*

The two markers of *PSY1* (Fu et al., 2013) were monomorphic for the favourable allelic variants across all 130 inbred lines, and were thus not considered for further analysis. In

contrast, all the PCR markers for *lcyE* and *crtRB1* were polymorphic across the inbred lines (Appendix 2 – Figure 4). Sequences of all sampled PCR fragments were aligned to their corresponding gene sequences (data not shown) confirming their identity. Alleles 2 and 4 of the 5'TE polymorphic site of *lcyE*, and allele 2 of the 3'TE and allele 3 of the 5'TE polymorphisms of *crtRB1* were not detected in this study (Table 2.5). The frequencies of the *lcyE* favourable alleles varied from 12% to 83%, while those of *crtRB1* varied from 18% to 19% (Table 2.5).

Table 2.5 Observed alleles of *PSY1*, *lcyE* and *crtRB1* functional markers in 130 yellow maize endosperm inbred lines and frequencies of the favourable allelic class for each locus

Marker	Expected allelic series	Allelic variants observed**	Favourable allele	Frequency of the favourable allele (%)
PSY1 SNP7	A, C	A	A	100
PSY1 InDel1	0, 378	0	0	100
lcyE 5'TE*	1, 2, 3, 4	'0', 1, 3	1	12
lcyE SNP (216)	G, T	G, T	G	83
lcyE 3'indel	8, 0	8, 0	0	38
crtRB1 5'TE	1, 2, 3	1, 2	2	18
crtRB1 indel4	12, 0	12, 0	12	19
crtRB1 3'TE	1, 2, 3	1, 3	1	18

*For *lcyE* 5'TE the vast majority of the inbred lines (> 70%) did not yield any amplification, and thus scored as '0' alleles to mean no amplification, not deletion. This was not observed by Harjes et al. (2008).

**There were individuals that were heterozygous for some of the marker classes. Description of markers and expected alleles are presented in Table 2.2.

The favourable alleles identified by the crtRB1-5'TE and crtRB1-3'TE markers were present in 26 inbred lines, co-occurring in 20 of the inbred lines. The two polymorphisms showed high linkage disequilibrium ($R^2 = 0.76$). However, both crtRB1-5'TE and crtRB1-3'TE markers were not in linkage disequilibrium with the crtRB1-indel4 marker (Table 2.6). Linkage disequilibrium values between markers of *lcyE* and *crtRB1* genes were low with the coefficients of pairwise allele frequencies (R^2 values) ranging from 0.004 to 0.188.

Table 2.6 Linkage disequilibrium between allele specific functional markers of *crtRB1* and *lcyE* genes in 130 tropical adapted diverse yellow maize inbred lines

R^2*	lcyE 5'TE	lcyE SNP (216)	lcyE 3'indel	crtRB1 5'TE	crtRB1 indel4	crtRB1 3'TE
lcyE 5'TE		7.29E-11	7.68E-13	1.09E-03	3.11E-01	8.50E-03
lcyE SNP (216)	0.3810		3.92E-07	1.43E-05	5.66E-01	1.82E-04
lcyE 3'indel	0.3990	0.2050		1.53E-01	1.13E-01	1.01E-01
crtRB1 5'TE	0.1010	0.1880	0.0201		1.55E-01	7.88E-19
crtRB1 indel4	0.0115	0.0041	0.0208	0.0166		2.48E-01
crtRB1 3'TE	0.0658	0.1350	0.0262	0.7570	0.0144	

*Upper triangle contains *p*-value and lower triangle contains R^2

Five different donor lines were determined to have contributed the favourable alleles of crtRB1-5'TE and 3'TE to the 26 inbred lines carrying either one or both of these favourable alleles. The vast majority of the inbred lines having favourable alleles of crtRB1-5'TE and/or -3'TE were derived from backcrosses containing a temperate inbred line DE3 as a donor parent. These inbred lines were among those lines that exhibited the highest levels of provitamin A. In addition, three tropical lines which were the recurrent parents of the best inbred lines also harboured favourable alleles for *lcyE*. Cluster analysis based on Nei's 1972 frequency based distance (Nei, 1972) using UPGMA separated the 26 best favourable-allele-carrying inbred lines into three major groups with one line separated from the major groups (Figure 2.3). Even though lines originating from the same backcross were grouped together, they showed sizable within group variability.

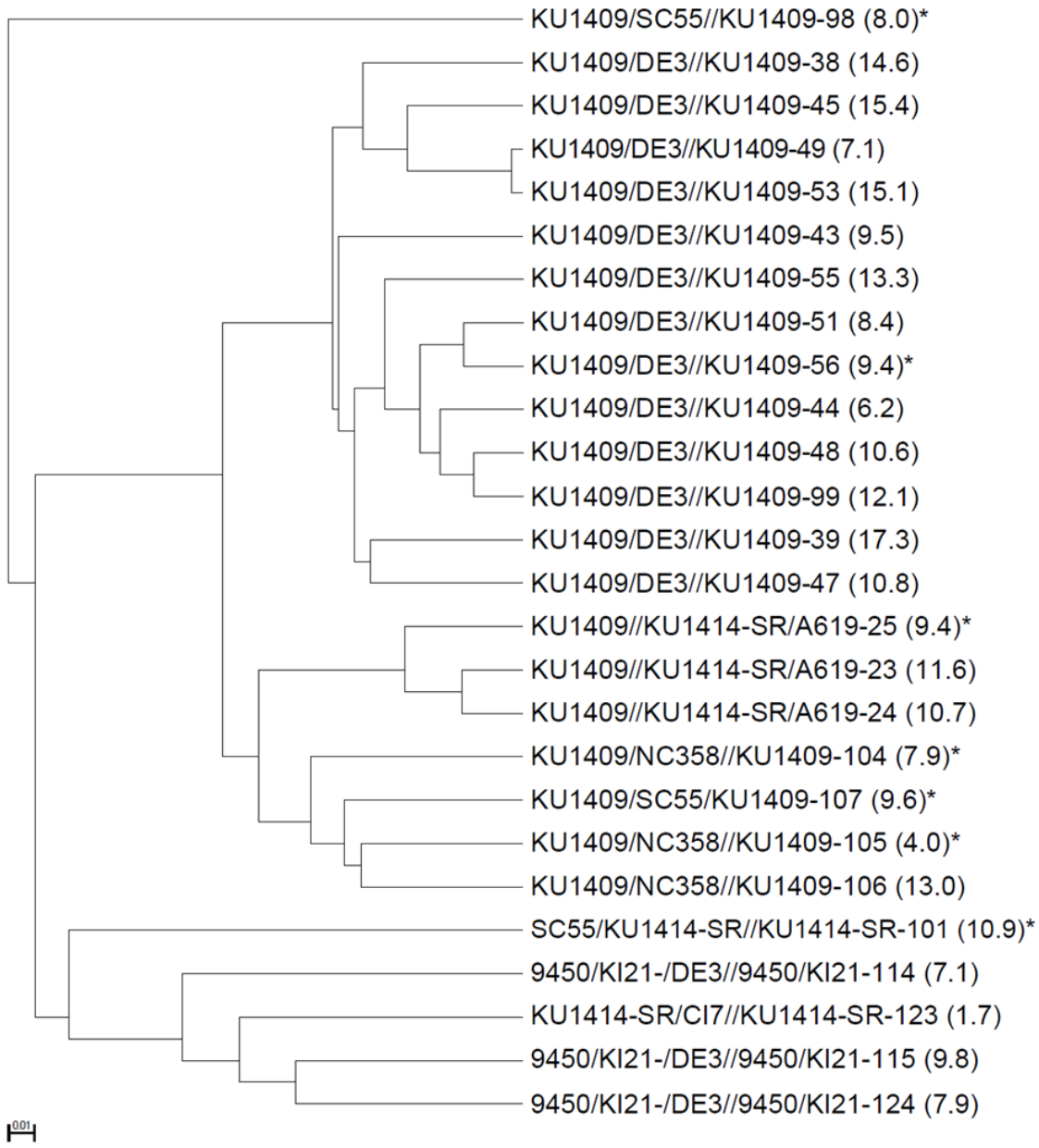


Figure 2.3 Dendrogram constructed using UPGMA based on Nei's distance for 26 inbred lines carrying the best favourable alleles of *crtRB1-5'TE* and *crtRB-3'TE* markers.

The pedigree refers to the sources of the inbred lines. The numbers after the pedigrees are inbred line entry numbers. Numbers in parenthesis are mean β-carotene concentration in µg/g DW. Entry 99 is the line used in Babu et al. (2013) for developing segregating populations. Twenty of the inbred lines contained favourable alleles of both markers except for those that are marked with the symbol *, which contain only one of the two favourable alleles.

2.3.3 Validation of the marker-trait associations

Association analyses were conducted for each year mean and for means averaged over the two years. Alpha-carotene was excluded from the analysis for its extremely low concentrations and insignificant variability across the inbred lines. The 3'TE and 5'TE polymorphic sites of the *crtRB1* candidate gene were significantly associated with all carotenoids and all the derived traits consistently over the two years ($\alpha = 0.01$), except lutein which was not affected by both markers, and the α to β branch carotenoids ratio which was not affected by *crtRB1*-5'TE in the second year. These two markers explained large proportions of the variations in the carotenoids and derived traits that ranged from 13 to 53% in the first year and 17 to 63% in the second year (Table 2.7). *CrtRB1*-indel4 was associated with provitamin A accounting for 9% of the variation ($\alpha = 0.5$). The functional markers of *lcyE*, though not consistent, were significantly associated only with lutein and ratio of β to α branch carotenoids explaining 15 to 21 % of the variations. However, *lcyE*-5'TE did not affect β to α branch ratio in the first year and *lcyE*-3'indel was not associated with any of the traits. None of the markers of each gene had a significant effect on total carotenoid content in both years; this is in line with the association analyses results of Harjes et al. (2008) and Yan et al. (2010). However, previous studies had detected significant reduction of total carotenoids for genotypes having favourable alleles of *lcyE* and *crtRB1* using segregating mapping populations. This may be due to differences in results obtained from association analyses using diverse versus narrow genetic base segregating populations (Babu et al., 2013; Yan et al., 2010).

2.3.4 Genotypes and combined effects of functional markers for the two genes on carotenoid content and composition

Fourteen unique genotypes were observed for *lcyE* and eight unique genotypes for *crtRB1* (Table 2.8). The two-way ANOVA combining the *lcyE* and *crtRB1* alleles revealed highly significant interaction effects for each carotenoid type and the derived traits. The combined effects of the alleles of the two genes were stronger than their separate effects. The two genes model explained 38 to 89% of the total variation in carotenoid concentration. Individual effects of the alleles were also highly significant

for almost all carotenoids. The combined *lcyE* markers explained the least variation in the β-branch carotenoids (β-cryptoxanthin, β-carotene, zeaxanthin), while the *crtRB1* markers explained the least variation in the α-branch carotenoid (lutein). The combined *crtRB1* markers had larger effects on individual and total provitamin A carotenoids in comparison to the effects of the *lcyE* markers.

Table 2.7 Marker-trait association for functional markers of *crtRB1* and *lcyE*

		Functional DNA (PCR) Markers**											
Season	**Carotenoids***	***crtRB1* 5'TE**		***crtRB1* Indel4**		***crtRB1* 3'TE**		***lcyE* 5TE**		***lcyE* 216(SNP)**		***lcyE* 3'Indel**	
		P	**R^2**	**P**	**R^2**	**P**	**R^2**	**P**	**R^2**	**P**	**R^2**	**P**	**R^2**
Year 1	**βcar**	6.6E-13	0.53			1.2E-12	0.51						
	βcry	2.8E-08	0.32			1.5E-07	0.28						
	zeax	3.7E-08	0.31			1.5E-07	0.28						
	lut							3.1E-03	0.11	1.6E-03	0.08		
	pva	1.2E-07	0.26	2.3E-03	0.09	6.5E-07	0.23						
	lnβcar/βcry	8.6E-12	0.47			7.7E-11	0.42						
	lnβcar/Zeax	4.1E-11	0.45			1.7E-10	0.42						
	lnβcar/all	5.4E-08	0.30			3.7E-08	0.30						
	lnββr/αbr	1.3E-04	0.15			4.7E-04	0.13			6.7E-04	0.10		
Year 2	**βcar**	4.4E-10	0.38			7.9E-11	0.42						
	βcry	4.6E-08	0.30			7.7E-08	0.29						
	zeax	1.2E-05	0.20			9.5E-06	0.20						
	lut							1.2E-04	0.18	2.7E-05	0.15	6.9E-04	0.12
	tpva	2.2E-05	0.17			4.8E-06	0.20						
	lnβcar/βcry	2.1E-14	0.63			1.3E-13	0.58						
	lnβcar/Zeax	8.1E-11	0.44			1.3E-11	0.47						
	lnβcar/all	5.4E-09	0.34			2.5E-10	0.41						
	lnββr/αbr							2.6E-04	0.16	9.1E-07	0.21	5.8E-05	0.17
Combined***	**bcar**	3.4E-12	0.49			2.5E-12	0.50						
	bcryp	1.1E-08	0.34			3.2E-08	0.31						
	zeax	1.6E-06	0.24			1.8E-06	0.23						
	lut							4.3E-05	0.20	1.5E-06	0.20		
	pva	1.2E-08	0.31			6.3E-09	0.33						
	lnbcar/bcryp	1.8E-14	0.63			8.4E-14	0.59						
	lnbcar/Zeax	1.9E-11	0.47			1.1E-11	0.48						
	lnbcar/all	9.3E-09	0.33			4.5E-10	0.39						
	lnbbr/abr	8.1E-03	0.08					7.2E-04	0.14	2.0E-06	0.20	8.3E-03	0.08

*Carotenoid abbreviations are described under Table 2.3. βbr/αbr = β-branch/α-branch carotenoids, βcar/βcry = β-carotene/β-cryptoxanthin, βcar/Zeax = β-carotene/zeaxanthin; βcar/all = β-carotene/total carotenoids

** Only significant values are indicated based on significance thresholds of 1.67E-04 and 8.3E-3 at 1% and 5%, respectively, after Bonferroni multiple test correction (Wright, 1992) (alpha/number of markers analyzed per trait, i. e., 0.01/6 = 1.7E-04 and 0.05/6=8.3E-3). Names of functional markers for each gene are described in Table 2.2.

***Combined association analysis over the two years using BLUEs, illustrated using bar graph in Figure 2.3

Table 2.8 Phenotypic variation explained (R^2) by individual and combined effects of *lcyE* and *crtRB1* alleles

Genotype combination*	No. of loci per genotype	No. of unique genotypes	R^2**									
			lut	**zeax**	**βcry**	**βcar**	**pva**	**tcar**	**βbr/αbr**	**βcar/βcry**	**βcar/zeax**	**βcar/all**
lcyE:5'TE, SNP(216), 3'TE; + crtRB1: 5'TE, Indel4, 3'TE	6	34	0.68	0.42	0.44	0.83	0.71	0.38	0.52	0.89	0.77	0.77
lcyE: 5'TE, SNP(216), 3'TE	3	14	0.42	0.22	NS	0.24	0.23	0.24	0.30	0.23	0.22	0.20
crtRB1: 5'TE, Indel4, 3'TE	3	10	0.24	0.35	0.28	0.63	0.48	NS	0.23	0.71	0.60	0.50

* Names and symbols of polymorphic sites and their allelic variants are described in Table 2.2. **R^2 from combined ANOVA computed for each of the 3 haplotype scenarios across 2 years. NS, non-significant at alpha 0.5, the rest are significant at alpha < 0.0001. Abbreviation of carotenoids' names described under Table 2.2 and 2.3.

Analysis of combinations of all of the six markers identified 34 unique genotypes (Table 2.9). The vast majority of these genotypes were represented by only one inbred line each. The most common genotype, '0', G, 8|-, 1, 0, -|3 (corresponding to *lcyE*-5'TE, -SNP (216), -3'InDel, *crtRB1-5'TE, -indel4,* 3' TE) was present in 49% of the inbred lines. The next most frequent genotype, namely '0', G,-|0, 1, 0, -|3, was present in 14% of the inbred lines. The average estimated effect of the most frequent genotypes on beta-carotene was 3.54 µg/g, (which was 2.45 µg/g less than the average effect of all the genotypes and 6.9 µg/g less than the average effect of those genotypes containing the favourable alleles of both *crtRB1* 5'TE and 3'TE) (Table 2.9).

Genotype, '0', G, 8|0, 2, 12, 1 contained the most optimal allelic combinations as it carried favourable alleles for 5 of the 6 loci and was present only in one inbred line (Entry 107) derived from a backcross KU1409/SC55/KU1409 (Table 2.1, Table 2.4). Its estimated effects were 9.05 µg/g for β-carotene, 0.33 µg/g for β-cryptoxanthin, 0.96 µg/g for zeaxanthin and 13.29 µg/g for total carotenoid. Although this genotype was predicted to be the best in terms of its allelic composition for the 6 markers, seven other genotypes were found to be superior to this genotype in their estimated levels of β-carotene. The presence of an unfavourable *lcyE* insertion in the homozygous state did not alter the effect of this genotype significantly, based on the observation of the effects on carotenoids observed in another genotype '0', G, 8, 2, 12, 1 (N = 4), which lacked the *lcyE*-3' insertion.

Genotype '0', G, 8, 2, 0, 1|- (N = 3) showed significantly better effect ($p < 0.01$) than the genotype with the predicted best allelic composition ('0', G, 8|0, 2, 12, 1), and had the strongest positive effect with an estimated average concentration of 15.03 µg/g for β-carotene and 15.08 µg/g for provitamin A. The major difference between the two genotypes is the lack of the favourable 12 bp insertion at *crtRB1*-indel4 in the former, which shows the negligible effect of this marker as was observed in the association analysis. Three inbred lines derived from KU1409/DE3/KU1409 contained the genotype'0', G, 8, 2, 0, 1|-.

The favourable alleles of 5'TE and 3'TE markers of crtRB1 (alleles 2 and 1) were present in almost all genotypic combinations that had large positive effects on β-carotene concentration ranging from 6.0 to 15.28 µg/g. The genotype 3, T, 8|0, 1, 0, 3 did not have any of the favourable alleles except the deletion allele representing *lcyE*-3'TE. Only one inbred line (Entry number 50) derived from a backcross KU1409/DE3/KU1409 carrying this genotype had estimated average effects of 7.56 µg/g for β-carotene, 8.55 µg/g for beta-cryptoxanthin, 12.09 µg/g for provitamin A and 43.3 µg/g for total carotenoid. The total carotenoid concentration of this genotype exceeded that of the average total carotenoid of those genotypes carrying the favourable alleles of *crtRB1 5'TE* and *3'TE* by 23.48 µg/g. This genotype also had 7.32 µg/g higher β-cryptoxanthin than the average of those carrying the above mentioned allelic classes. These results were corroborated by the low ratio values of beta-carotene to β-cryptoxanthin, β-carotene to zeaxanthin and β-carotene to all carotenoids.

Another exceptional genotype 3, G, 8, 1|2, 0, 3 containing the favourable allele of crtRB1 5'TE in the heterozygous state showed a very weak effect on beta-carotene (1.05 µg/g) and provitamin A (1.65 µg/g) content. The weakest effect was detected from genotype '0', G, 0, 1, 12, 3 (N = 3), which was devoid of the two best favourable alleles of crtRB1 5'TE and 3'TE (Table 2.4), which had relatively low level of total carotenoids (10.18 µg/g). Overall, the average effects of the genotypes harbouring both favourable alleles of crtRB1-5'TE and -3'TE (N = 23) resulted in 7.2 µg/g increase or 3.22 fold increase in β-carotene as compared to the effects of genotypes without any of the favourable alleles (N = 103). There was a 23% reduction in total carotenoid between the two sets of allelic classes (from 23.5 to 18 µg/g).

2.3.5 Agronomic performance of high provitamin A inbred lines

The top high levels of β-carotene and total provitamin A exhibiting inbred lines were all moderately tolerant to the major foliar diseases and had good plant aspect but poor ear aspect scores. Inbred line (KU1409/DE3/KU1409)S2-18-2-B-B-B-B was the best in terms of its ear aspect, plant aspect and β-carotene and provitamin A contents. It had

good levels of resistance to the major foliar disease (Appendix 1 – Table 5), and accumulated 11.30 µg/g β-carotene and 12.11 µg/g provitamin A. The inbred line carried the favourable alleles of the two best provitamin A associated functional polymorphisms (crtRB1 3'TE and 5'TE) making it an ideal donor parent of these favourable alleles for use in a marker assisted maize breeding program.

2.4 Discussion

The carotenoid profile and content of one hundred thirty tropically adapted inbred lines developed through conventional breeding at IITA were investigated across two seasons. Menkir colleagues (2008) previously reported 15 tropical adapted yellow maize inbred lines accumulating 5 to 7.8 µg/g of provitamin A. In the current study, 28 inbred lines with provitamin A content of 8.0 to 17.25 µg/g were identified which is a major improvement highlighting the success of the IITA breeding program in introgressing the best temperate favourable alleles of provitamin A into tropical adapted inbred lines. The maximum average estimated level of total carotenoids was 42.71 µg/g. Recently, high carotenoid inbred lines that accumulated as high as 100 µg/g were reported (Burt et al., 2011b). Such inbred lines can be crucial if the high influx of the substrates into the carotenoid pathway can be effectively tapped for elevating the provitamin A carotenoids fraction in maize endosperm (Vallabhaneni and Wurtzel, 2009).

Functional markers are powerful gene targeted molecular markers designed to identify nucleotide polymorphisms underlying variations in the phenotype of interest (Andersen and Lübberstedt, 2003; Varshney et al., 2005). Functional markers of three key carotenoid genes *PSY1* (Fu et al., 2013), *lcyE* (Harjes et al., 2008), and *crtRB1* (Yan et al., 2010) were investigated for their effects on a diverse population of inbred lines that displayed substantial variability in their endosperm carotenoid content and composition. The markers were developed using different association panels composed of diverse temperate, sub-tropical and/or tropical yellow maize inbred lines. The functional markers of the two genes *lcyE* and *crtRB1* have been reported as one of the most exciting discoveries in recent maize endosperm provitamin A improvement endeavors

(Pixley et al., 2013). Yet, independent marker validation studies showed inconsistencies in the efficiency of the markers (Babu et al., 2013; Burt et al., 2011b; Vignesh et al., 2012). Moreover, the panels and populations used to develop, assess and/or validate the functional markers were mainly temperate germplasm in which the favourable alleles of the most significant markers were present at very low frequencies. In the present study, the functional markers were further assessed using a panel of inbred lines that were found to harbor the best favourable alleles at relatively higher frequency under diverse and mixed genetic background of both the tropics and temperate germplasm.

The functional markers for *PSY1* were monomorphic for the favourable allelic variants across all the inbred lines possibly because of the conserved nature of *PSY1* in yellow maize lines (Palaisa et al., 2003). Fu and colleagues (2013) also found *PSY1* to be fixed within the tropical background panel used in their study for the favourable alleles. The variation in total carotenoid content of the 130 inbred lines is either due to the presence of some rare functional variation within the *PSY1* gene or other genes that are involved in the carotenoid biosynthesis (Vallabhaneni and Wurtzel, 2009). The markers for *lcyE* and *crtRB1* were polymorphic and the 3' and 5'TE polymorphisms of *crtRB1* were strongly associated with high levels of β-carotene in the current set of inbred lines. In line with previous results (Babu et al., 2013; Burt et al., 2011b) the effects of *lcyE* markers were shown to be weak and inconsistent.

Yan and colleagues (2010) detected the favourable allele of *crtRB1*-5'TE only in the temperate yellow maize germplasm with frequency of less than 3%. Here, a relatively high frequency (18%) of the allele was recorded across the 130 tropical adapted inbred lines. The 3'TE and 5'TE polymorphisms of *crtRB1* were in high linkage disequilibrium (R^2 = 0.76) which was different from what was reported by Yan et al. (2010). This seems likely as the two favourable alleles were most probably introgressed into the tropical adapted materials together in one step from donor temperate inbred lines that contained both alleles. Seventy five percent of the inbred lines harboring favourable alleles of *crtRB1* 5'TE and 3'TE were results of back crosses that used the temperate line DE3 as donor.

Nearly all of the inbred lines with high levels of β-carotene and total provitamin A harboured the favourable alleles of the most significant functional markers of *crtRB1*-3T' and *crtRB1*-5'TE. However, similar to previous studies (Burt et al., 2011b; Vignesh et al., 2012), slight divergences were observed, but was limited to a few inbred lines. For example, an inbred line with favourable allele of *crtRB1* 5'TE polymorphism yet very low in β-carotene and total provitamin A (< 2.0 µg/g DW) was observed. This inbred line had relatively low total carotenoid content (14.14 µg/g DW), suggesting high influx of substrates into the carotenoid biosynthesis pathway is perhaps one of the important factors to realize the desired action of the favourable alleles (Burt et al., 2011b; Fu et al., 2013; Yan et al., 2010). Therefore, introgression of these favourable alleles into adapted maize germplasm with high total carotenoid genetic background can be expected to lead to increased levels of β-carotene and total provitamin.

Higher levels of provitamin A and β-carotene concentrations were observed with combinations of the six markers of *lcyE* and *crtRB1* genes. The variations explained were also higher for the combined markers. Previous studies by Yan et al. (2010) and Babu et al. (2013) also demonstrated more significant effects for the functional polymorphisms of the two genes than when each was considered individually. The decision on use of either all markers of both genes or only the most significant markers perhaps depends on the availability of resources. Yet, the results reported here suggests it is more ideal to focus on the two most significant markers, *crtRB1*-5'TE and -3'TE. Babu and colleagues (2013) indicated the large and significant effect of the *crtRB1*-3'TE irrespective of the *lcy*E markers allelic states, and thus recommended the use of *crtRB1* locus alone for marker assisted selection in tropical breeding programs for faster genetic gain in β-carotene (Babu et al., 2013).

2.5 Conclusions

The study confirms the strong association the favourable alleles of the *crtRB1*-5'TE and -3'TE functional markers have with high levels of β-carotene. Since these two markers were found to be in strong linkage disequilibrium within the inbred lines studied, they can be predictive of one another, and thus only one of the polymorphic sites can be

targeted to reduce costs associated with genotyping. As the efficiency of PCR amplification of the markers can vary with genetic background the marker that gives reliable PCR amplification can be selected. The best inbred line combining desirable agronomic traits with high levels of provitamin A was the genotype KU1409/DE3/KU1409)S2-18-2-B-B-B-B and can be used as a donor parent of the favourable alleles of provitamin A. The most rapid strategy for high provitamin A maize breeding in IITA breeding program can be introgression of the favourable alleles from agnonomically good performing inbred lines into inbred lines that can accumulate high level of non-provitamin A carotenoids in their endosperm.

The first generations of high provitamin A accumulating hybrids have now been developed for Nigeria and Zambia, which will most likely be disseminated across other African countries with similar agro-ecologies (http://www.HarvestPlus.org; http://www.iita.org). The high provitamin A and diverse inbred lines that are harboring the best combinations of the *crtRB1*-5'TE and 3'TE markers reported in this study will speed up the development of the next generation high provitamin A hybrids for production in Sub-Saharan Africa thus contributing to the effort of alleviating hidden hunger due to vitamin A undernutrition.

Chapter 3: Genetic analysis of maize endosperm carotenoids in diverse yellow maize inbred lines using high density genome-wide SNP-trait association scan

3.0 Abstract

The genetic potential of maize for accumulation of provitamin A can be better harnessed through discovery and use of markers associated with loci across the genome which are underlying the phenotypic variation. A genome-wide association study (GWAS) was conducted using 130 diverse yellow maize inbred lines which were genotyped with high density SNPs. Extensive linkage disequilibrium was estimated ranging from 0.59 Mbp (megabase pair) on chromosome 2 to 1.6 Mbp on chromosome 3, with a genome-wide average of 0.85 Mbp. Numerous significant SNPs co-localizing with the known carotenoid biosynthesis genes *crtRB1*, *lcyE* and *ZEP1* were identified. The large effects of the two major genes *lcyE* and *crtRB1* were re-confirmed by the GWA study. In addition, significant novel associations were identified for several transcription factors that may be potentially involved in regulation of expression of carotenoid biosynthesis. An association of a transcription factor (called auxin response factor 20) with β-carotene and provitamin A was discovered when the GWAS was re-conducted by including the major effects of *lcyE* and *crtRB1* as covariates. Pathway level association analysis of about 40 isoprenoid biosynthesis genes demonstrated the significant contribution of some genes such as *hyd4*, *CCD7*, and *IPPI2* to the variation in carotenoid content variation across the panel. These results can be directly used for marker assisted provitamin A maize selection, or can serve as a starting point to design genomic selection strategies for provitamin A maize breeding. In addition, the research advances efforts towards identification of the genes (and allelic variants) involved in regulation of carotenoid biosynthesis.

3.1 Introduction

Genetic variation is essential for crop improvement. This variation is chiefly a reflection of functional DNA sequence variations across individuals in genomic regions that directly or indirectly affect the target phenotypes. The identification and tagging of such genomic regions with easily detectable diagnostic markers is a key initial step in marker assisted crop breeding (Soto-Cerda and Cloutier, 2012). Incorporation of molecular markers in breeding scheme has traditionally been accomplished using genetic linkage based QTL (quantitative trait loci) mapping approaches (Tanksley, 1993). However, QTL mapping can be performed more effectively using genome-wide association analysis (GWA) mapping approaches. GWA is an efficient genetic analysis technique that systematically tests for association between a large number of genome-wide markers and complex traits to identify (nucleotide sequence based) allelic variants associated with the variation in the concerned trait on a genome-wide scale (Zhu et al., 2008). GWA has the potential to localize QTLs with high resolution at the level of single gene to nucleotide sequences, within a relatively shorter period of time as compared to the classical linkage mapping approach (Yu and Buckler, 2006; Zhu et al., 2008). A variant of GWA mapping is the candidate gene association mapping approach, which capitalizes on prior knowledge of genes from biochemical, mutation or linkage analysis for a specific trait to detect functional nucleotide variants associated with the variation in the trait of interest (Zhu et al., 2008; Yan et al., 2010). This approach has the subjectivity bias drawback of missing unknown genes but is advantageous from perspectives of saving time and cost (Zhu et al., 2008).

Association mapping relies on the principle of linkage disequilibrium (LD), which is defined as the non-random co-occurrence of alleles at different loci (Tanksley, 1993; Flint-Garcia et al., 2003; Ersoz et al., 2007). The occurrence of LD between molecular markers such as single nucleotide polymorphism (SNPs) of known location and functional genes allows the identification of the positions and estimation of the effects of the genes underlying a phenotype. LD can occur between markers and/or QTLs and is influenced by phenomena such as recombination rate, strong selection, gene flow, mating system and/or genetic drift (Myles et al., 2009).

The major advantage of association mapping over the linkage mapping approach lies in its use of populations composed of diverse individuals or group of individuals derived from different natural populations, land races, varieties or diverse breeding germplasm. Such populations have potentially accumulated numerous independent recombination events creating many LD breakpoints across the target genome over generations or different breeding cycles, leading to formation of gene pools with a broader range of allelic variants. This is contrary to linkage mapping populations which are often constructed from bi-parental progenies that have experienced just a small number of recombination events allowing identification of two allelic variants (Yu and Buckler, 2006; Zhu et al., 2008; Soto-Cerda and Cloutier, 2012). The use of diverse population in association mapping, however, comes at a cost of the possible presence of non-random population structure and familial relatedness among the individuals arising from local adaptation, selection, and recent breeding history, which altogether can compromise the reliability of association mapping by inflating false positives, unless controlled by statistical methods (Yu and Buckler, 2006; Yu et al., 2006).

One of the most common statistical approaches that are used to account for the confounding effect of population structure and familial relatedness using genomic tools is the mixed linear model (MLM). MLM incorporates both population structure (Q) and relative kinship (K) inferred from marker data into the genome-wide association study (GWAS) to minimize spurious associations due to both type I and type II errors (Yu et al., 2006). The model treats population structure as fixed effects and familiar relationship or relative kinship among individuals as random effects (Yu et al., 2006; Lipka et al., 2012). Methods such as compressed MLM (CMLM), efficient mixed model association (EMMA) and population parameters previously determined (P3D), which build upon the mixed model but with improved statistical accuracies and computational speed have also been developed and integrated into association mapping softwares such as GAPIT and PLINK (Kang et al., 2010; Zhang et al., 2010; Lipka et al., 2012).

The availability of numerous markers distributed throughout the genome of a crop species is a prerequisite for GWAS for successful discovery of all the relevant recombination events in a diverse population. This may presents challenges as it can be

a daunting task to find sufficient number of markers with reasonable genome coverage in order to precisely locate target QTLs (Zhu et al., 2008). This challenge has now been circumvented by recent advances in next generation sequencing (NGS) technologies that have made whole-genome high-density genotyping across large number of individuals feasible. Genotyping using NGS has facilitated the discovery of countless single nucleotide polymorphisms (SNPs) linked to crucial regions of genomes underlying traits of interest (Davey et al., 2011). Genome-wide polymorphisms associated with traits of economic/agricultural value can be collectively used as genetic markers for genomics assisted breeding, which will have potential routine application in crop breeding as NGS prices continue to decline (Varshney et al., 2014, 2009; Xu et al., 2012; Bevan and Uauy, 2013; Bolger et al., 2014). For instance, the cost of sequencing 1Mbp DNA is reported to have dropped from $10K in 2001 to less than $0.1 in 2013 (www.genome.gov/sequencingcosts, accessed May 2014)

Genotyping by sequencing (GBS) is one of the NGS based genotyping approaches that are currently revolutionizing large-scale genome-wide marker development with relatively cheap and simplified methods (Elshire et al., 2011). The GBS approach employs restriction enzymes to reduce genome complexity by excluding high repeat regions while enriching low copy regions of the genome. GBS is particularly suitable for species with high level of genetic diversity, such as maize. The technique has been promoted as a quick, simple, highly specific, reproducible and capable of reaching important genomic regions that are inaccessible to other sequence capture methods (Elshire et al., 2011). The bioinformatics tools for processing of raw genotyping sequence data generated through GBS protocol are described in Glaubitz et al. (2014).

Genome-wide association analysis has been widely utilized to understand the genetics of human diseases (Zhu et al., 2008; Bush and Moore, 2012), and increased and better applications have been developed in crop plants for identification of QTLs associated with important agricultural traits (Brachi et al., 2011). Association studies have identified novel and confirmed known loci controlling agronomic traits such as plant height, yield and yield components, flowering time and plant architecture in a range of crops, including barley (Pasam et al., 2012), tomato (Shirasawa et al., 2013), wheat

(Edae et al., 2014; Wang et al., 2014), and maize (Thornsberry et al., 2001; Wang et al., 2012); as well as grain quality traits including oil content in maize (Li et al., 2013a), protein contents in wheat (Wang et al., 2014) and essential micronutrient such as α-tocopherol (vitamin E) and β-carotene in maize (Harjes et al., 2008; Yan et al., 2010; Li et al., 2012; Lipka et al., 2013b).

Maize is one of the most important staple crops in the developing world and has wide range of utilizations as animal feed and industry raw material (Yan et al., 2011). Dissecting the genetic architecture of maize endosperm carotenoids has been a major goal of maize breeders and geneticist, to exploit the nutritional significance of the diverse pigments that accumulate in yellow maize endosperm. Carotenoids such as β-carotene and β-cryptoxanthin have been major targets of crop biofortification efforts. These particular species of carotenoids can serve as precursors of vitamin A, an essential micronutrient which is lacking in the diets of millions of people in developing countries leading to global health problem. Genetic studies can facilitate the development of improved varieties with high levels of provitamin A, which can contribute to the fight against the problem of vitamin A deficiency in developing countries where maize is a staple crop (Wurtzel et al., 2012).

Recent studies that included gene targeted association analysis approach have led to the discovery of allele specific markers for two key genes that are involved in carotenoid biosynthesis and provitamin A accumulation, namely lycopene epsilon cyclase (*LCYE*) and a carotenoid hydroxylase *crtRB1* or *HYD3*) (Harjes et al., 2008; Yan et al., 2010). Diagnostic markers of specific alleles of these genes have been used by CIMMYT for speeding up the breeding of maize for increased provitamin A (Pixley et al., 2013), and their suitability to provitamin A maize breeding program at IITA has also been validated (Azmach et al., 2013/Chapter 2). Although these markers represent the major genes significantly affecting provitamin A, there may still be several other potential loci with small to medium effects on accumulation of carotenoids in maize endosperm. This can be expected under a diverse genetic background as there are many genes involved in carotenoid biosynthesis in maize endosperm.

The studies on *lcyE* and *crtRB1/HYD3* genes (Harjes et al., 2008; Yan et al., 2010) demonstrated the possibility of selective substrate channeling through the carotenoid pathway (Figure 1.2) in favour of more potent provitamin A carotenoids such as β-carotene. However, concentrations of a specific carotenoid may also vary with the level of total carotenoids synthesized. This is a function of substrate flux into the carotenoid pathway and downstream steps that will lead to degradation of carotenoids by cleavage enzymes (Rodríguez-Concepción, 2010; Vallabhaneni et al., 2010). Effort for provitamin A carotenoid improvement can benefit from the understanding of regulation of all rate limiting steps involved in the upstream substrate supplying pathways, through the carotenoid biosynthetic pathway, and the downstream carotenoids depleting steps. Expression profiling by Vallabhaneni and Wurtzel (2009) found that expression of several key carotenogenic genes correlated with the level of total carotenoids in maize endosperm, suggesting a possible functional role in the regulation of carotenoid accumulation in endosperm tissue. The genes identified in the expression profiling encode enzymes representing upstream pathways for isopentenyl diphosphate (IPP) and geranylgeranyl diphosphate (GGPP) synthesis and the downstream carotenoid biosynthetic pathway, involving conversion to abscisic acid (Figure 1.1). The authors demonstrated that the upstream gens *DXS3*, *DXR*, *HDR*, and *GGPPS1* were positively correlated with carotenoid accumulation at 25 days after pollination. In addition *PSY1* and *CRTISO1* were positively and negatively correlated with total carotenoids, respectively, at 20 days after pollination. Transcripts of *ZEP1 and ZEP2*, the genes contributing to metabolism of carotenoids toward abscisic acid, were also reported to be inversely associated with carotenoid content (Vallabhaneni and Wurtzel, 2009).

Carotenoid cleavage dioxygenase gene family (*CCD*) is known to catabolize multiple carotenoid substrate through cleavage of linear and cyclic carotenoids producing volatile non-carotenoid compounds (Vogel et al., 2008). Vallabhaneni et al. (2010) found significant negative correlation between copy number of the *CCD1* and endosperm carotenoid content. Therefore, identification of allele specific markers for multiple genes is required in order to accumulate favourable alleles of different loci in a single maize cultivar containing high levels of provitamin A.

In addition to the structural genes affecting carotenoid biosynthesis, there can be regulatory DNA sequences located upstream or downstream far from of the genes with possible significant impact on carotenoid biosynthesis (e.g. distal enhancer or repressor elements). There is evidence for the presence of such regulatory regions located as far as 150 to 250Kb away from the structural genes underlying agronomic traits, implying that marker discovery and development efforts should also consider non-coding DNA regions (Elshire et al., 2011; Lipka et al., 2013b). GWAS can help unravel key loci involved in carotenoid biosynthesis, knowledge which can be exploited for positively altering provitamin A carotenoid compositions. Aside from the association mapping studies that have led to the development of allele specific markers for provitamin A, few studies have to date reported GWAS for carotenoid composition and content in maize endosperm (Lipka et al., 2013a; Suwarno, 2012).

In this research, a GWA study was conducted on a collection of diverse and partially related yellow maize inbred lines with mixed genetic background of both tropical and temperate germplasm in order to identify multiple loci across the maize genome with significant effects on carotenoid biosynthesis. With this aim, GBS was used to generate high density SNPs covering the whole genome. This allowed analysis of diversity, kinship and linkage disequilibrium across the genome. The results of this study will add to that of the ongoing efforts in gene specific allele mining and genomic selection studies aimed at enhancing the genetic potential of maize for high provitamin A content.

3.2 Methods

3.2.1 Germplasm

The panel of 130 diverse yellow maize inbred lines which was used to validate the allele specific markers of *lcyE* and *crtRB1* was used in the GWAS (Chapter 2.2.1). This panel was composed of inbred lines developed from eight bi-parental crosses, four broad based populations, and 28 backcrosses of tropical inbred lines, involving five temperate lines as donors of high β-carotene alleles (Table 2.1). Since these inbred lines contained both tropical and temperate maize germplasm in their genetic backgrounds, they represent allelic diversity underlying the variation in carotenoid composition and content in the temperate and tropical gene pools.

3.2.2 Carotenoid analysis

The two years of carotenoid data used for allele specific markers study reported in Chapter 2 was also employed for the GWA study presented here. The inbred lines were evaluated at IITA's research station, Ibadan, Nigeria ($7^{o}29'11.99''N$, $3^{o}54'2.88''E$, altitude 190 m) for two seasons, in 2010 and 2011. The trial was arranged in a 13 x 10 alpha-lattice design with two replications. Each line was planted in 5 m row plot spaced 0.75 m between rows and 0.25 m within a row. Different fields were used within IITA research station in each season. The fields were managed as per the recommended agronomic practices which included fertilization with nitrogen, phosphorus and potassium, and weed control using Primextra and Gramazone herbicides with supplemental hand weeding when necessary during the growing seasons.

Seed samples for carotenoids analyses were produced by controlled self pollination of all typical plants in each plot. Ears of each self pollinated line in each plot were harvested, well dried under ambient temperature with minimal exposure to direct sunlight, and shelled separately. Samples of one hundred kernels were drawn and sent to the University of Wisconsin, USA for carotenoid analysis.

Carotenoids were extracted and quantified with HPLC at the University of Wisconsin, USA. The extraction protocol was the method of Howe and Tanumihardjo (2006) for

carotenoid analysis of dried maize kernels (Chapter 2 materials and methods). Alpha-carotene, β-carotene (*cis* and *trans* isomers), β-cryptoxanthin, lutein, and zeaxanthin were quantified .Total carotenoid was calculated as the sum of concentrations of α-carotene, lutein, β-carotene, β-cryptoxanthin, zeaxanthin). Provitamin A was calculated by adding up the concentrations of β-carotene, and half concentrations of each of β-cryptoxanthin and α-carotene, since β-cryptoxanthin and α-carotene can provide only one molecule of retinol each as opposed to two molecules of retinol for β-carotene (US Institute of Medicine, 2001). Other derived carotenoid traits were also calculated as indicated in Harjes et al. (2008) and Yan et al. (2010): i.e., ratio of carotenoids in β to α branch of the carotenoid pathway, ratio of β-carotene to β-cryptoxanthin and ratio of β-carotene to all carotenoids (β-carotene + α-carotene + lutein + zeaxanthin + β-cryptoxanthin). The data for the ratio traits were transformed using natural logarithm ($\log_e$) before subjecting to statistical analysis to correct for the non-normal distribution of the data. All carotenoids concentrations were measured in µg/g dry weight (DW). BLUEs (best linear unbiased estimates) calculated for each trait based on the two season carotenoid data were used in the GWAS. BLUEs were calculated using the GLM option of TASSEL software version 4 (Bradbury et al., 2007).

3.2.3 Genome-wide SNP marker generation using GBS

DNA samples were isolated from freeze dried leaf samples of each inbred line using Qiagen™ DNeasy™ plant mini kit following the protocol supplied with the product. Samples having at least 10 ng/µl DNA each were plated and sent to the Institute for Genomic Diversity (IGD), Cornel University, USA, for genotyping. GBS libraries were prepared and analyzed according to Elshire et al. (2011) and sequenced at the Institute. To briefly describe the procedure employed: Genomic DNA samples were digested with a selective restriction enzyme known as ApeKI, a partially methylation sensitive type II restriction endonuclease that recognizes a degenerate 5 bp sequence (GCWGC, where W can be A or T) and creates a 3 bp 5' overhanging. This restriction enzyme has relatively few recognition sites in the majority of maize retrotransposons thus excluding genomic regions populated with such repetitive sequences. The digested DNA fragments were then ligated in the same reaction to two types of adapters, common adapters and

barcoded adapters. DNA samples were pooled and PCR amplified using primers complementary to the adapters' sequences, where by fragments with combination of common-barcoded adapters were selectively amplified excluding adapter dimmers. The PCR products were finally purified and subjected to DNA sequencing using Illumina, Inc. (San Diego, CA) NGS platforms (Elshire et al., 2011).

SNPs were called from the sequenced GBS library using GBS production pipeline (Version: 3.0.134) an extension of the Java program TASSEL (Bradbury et al., 2007; Glaubitz et al., 2014), which used aligned short reads of GBS called tags on species level. The GBS pipeline options for calling SNPs included: 0.1 Minimum locus coverage, $1x10^6$ maximum number of SNPs per chromosome, duplicate SNPs above 0.05 mismatch rate were not merged, and 0.8 cutoff frequency between heterozygot vs. homozygote calls . Tags were aligned to the reference genome B73 refgen_v2 (Schnable et al., 2009). The detailed GBS pipeline options are described in Appendix 3 B, a report on the GBS pipeline by IGD for the association panel employed in this GWAS.

The GBS pipeline generated a data set containing a total of 619,596 unfiltered SNPs. This SNP dataset had a total of 51% missing data points which might be caused by biological presence-absence of sequences between the reference and each test genome, or errors introduced in the GBS procedures (Poland and Rife, 2012; Glaubitz et al., 2014; Personal communication with Dr Sharon E Mitchel, Cornell University, USA). The dataset was further filtered in TASSEL 4 on the bases of missing data proportion and minor allele frequency cutoff thresholds (Bradbury et al., 2007). The cutoff thresholds used to filter the dataset for the GWAS allowed only those SNPs showing a maximum of 20% missing data, and 1% minimum minor allele frequency (MAF). This resulted in a dataset of 109,937 SNPs. On the other hand, the diversity and genome-wide LD analysis were performed using datasets obtained by filtering with criteria of no missing data points and 1% and 10% minor allele frequencies which resulted in 3,532 and 1,658 genome-wide SNPs, respectively. SNP data summary and basic diversity parameters were calculated using TASSEL 4 (Bradbury et al., 2007) and PowerMarker 3.25 (Liu and Muse, 2005) softwares.

3.2.4 Linkage Disequilibrium

To determine the degree of resolution the association study could achieve (Yu and Buckler, 2006), both genome and chromosome wide linkage disequilibrium (LD) were estimated using squared allele frequency correlation coefficient (R^2) for all possible pairs of SNPs in a dataset. For genome-wide LD, SNP datasets of the 10 maize chromosomes were combined and filtered with cut-off threshold of no missing data and 10% minimum minor allele frequencies yielding 1,658 SNPs typed across all inbred lines. On the other hand, LD estimation within each chromosome was performed using the SNP data of each chromosome filtered at 10% maximum missing data per marker and 10% minimum minor allele frequencies. Missing data in all the SNP datasets used for chromosome wide LD analysis were not imputed. The software used to estimate LD was TASSEL 3 (Bradbury et al., 2007), which uses permutation tests to determine the *p*-values for each pairwise correlation. LD estimates having $P \geq 0.001$ were considered non-significant (Pasam et al., 2012). Genome-wide and chromosome wide rate of LD decays were estimated by plotting localized regression curves (LOESS) of the r^2 values versus the corresponding physical distances between the SNP pairs, followed by observation of the intersection point between the fitted LOESS curve and a critical r^2 values (Breseghello and Sorrells, 2006). Two background critical R^2 values for estimating LD decays within and across chromosomes were considered in the present study to offer comparison. The first baseline critical R^2 was determined by taking the parametric 95 percentile of distribution of R^2 values for unlinked SNPs, taking SNPs on different chromosomes and SNPs beyond 50Mbp apart on the same chromosome as unlinked (Breseghello and Sorrells, 2006; Pasam et al., 2012). The second baseline R^2 value was 0.2, an arbitrary value often used to describe LD decay (Zhu et al., 2008). Scatter plots and fitted smooth curves for estimating LD decay were plotted using a base scatter plot function of R version 3.0.3, 'scatter.smooth' (R Core Team, 2014). The function plots and adds a smooth curve to a scatter plot, computed according to LOESS (R Core Team, 2014). LD patterns of all SNPs significantly associated with carotenoids and local LD patterns in regions surrounding significant genes were visualized using LD plots generated with HaploView (Barrett et al., 2005).

3.2.5 Genome-wide SNP-trait association

Association between genome-wide SNPs and carotenoid content was performed using an R (R Core Team, 2014) package known as GAPIT, Genetic Association and Prediction Integrated Tools, written by Lipka et al. (2012). GAPIT package applies the state of the art statistical methods based on unified mixed linear model (MLM) to calculate genome-wide association between traits and large number of markers in a computationally efficient and statistically accurate strategy (Lipka et al., 2012; Zhang et al., 2010; Kang et al., 2010). Since the panel used in this study was composed of groups of inbred lines that were extracted from many back crosses and single crosses involving diverse parental germplasm, multiple level relative kinship was expected and thus the K+Q model of GWAS was applied to compute accurate association. The analysis was executed with the 'basic scenario' of GAPIT software which automatically calculated both K and Q using the SNP marker data. This scenario implements VanRaden's algorithm option (VanRaden, 2008) to calculate the K matrix, and uses principal component analysis (PCA) to define Q. Both PCA and K were calculated from randomly sampled SNPs representing the whole genome. It applies optimum compression levels using default kinship clustering and grouping values 'average' and 'mean', respectively. The model selection option was used to estimate the optimum number of principal components (PC) covariates using Bayesian Information Criterion (BIC). The variation explained for a trait by the model and a particular SNP in question were determined using the likelihood R^2 statistics calculated in the software (Lipka et al., 2013b).

SNP data used for GWAS was filtered in TASSEL 4 with maximum missing data of 20% and minimum MAF of 1%. Missing data was imputed automatically within GAPIT using the conservative option of 'major allele', which replaces missing data points with the major allele of the SNP. Different significance cut-off thresholds were used to assess the effect of the SNPs on carotenoids. The statistical significances of the SNPs were evaluated at 5 and 1% critical thresholds of the false discovery rate (FDR) adjusted *p*-values (Benjamini and Hochberg, 1995) and the family wise error correction term, Bonferroni adjustment, which corrects inflated *p*-values in a single step by dividing α or the critical *p*-value (0.05 or 0.01) with the total number of tests (Wright, 1992); i.e., α

divided by the number of SNPs tested. FDR values generated with GWAS result in GAPIT were used.

3.2.6 GWAS including allele specific markers of *lcyE* and *crtRB1* as covariates

Variations in carotenoid content and composition of the association panel caused by allelic variants in the two key genes, *lcyE* (Harjes et al., 2008) and *crtRB1* (Yan et al., 2010) were accounted for by including marker score data of the three allele-specific markers of each gene as covariates. To accomplish this, the six allele specific marker data were first transformed to principal components. Components explaining the largest proportion of the variation was then included as covariates in the unified mixed model for calculating the GWA using GAPIT.

3.2.7 Pathway level association analysis

GWAS may overlook loci with small to moderate effects owing to adjustment of significance levels to control the bias associated with simultaneous multiple hypothesis testing (Wang et al., 2007, 2010) and the focus given to only the most significantly associated SNPs. Thus, pathway-level association analysis was conducted to complement the GWAS as described in (Lipka et al., 2013b). Genomic information of 42 genes that are known to be involved in carotenoid biosynthesis were retrieved from the literature (Vallabhaneni and Wurtzel, 2009; Wurtzel et al., 2012), and the maize genome database, http://www.maizegdb.org (Lawrence et al., 2004) . SNPs and the corresponding *P*-values from the first GWAS result were compiled for regions flanking these genes within the range of +/- 1 Mbp, assuming a SNP distally linked to a significant gene as far as 1Mb can have statistically significant effect given the extensive average genome-wide LD estimates in the association panel. The *P*-values were then adjusted for multiple testing using FDR based on the method of Benjamini and Hochberg (1995). This was computed in SAS® software version 9.3 (SAS Institute, 2012) using the MULTTEST procedure.

3.3 Results

3.3.1 Carotenoid profile of maize diversity panel

The carotenoid composition and content of the panel used for the GWAS is discussed in Chapter 2. The panel displayed considerable diversity in carotenoid profile with mean carotenoid concentration averaged over two seasons varying from 0.45 to 13.51 µg/g for lutein, from 0.04 to 25.90 µg/g for cryptoxanthin, from 0.08 to 8.55 µg/g for β-cryptoxanthin, from 0.00 to 16.38 µg/g for β-carotene, from 0 to 17.25 µg/g for provitamin A, and from 4.43 to 42.71 µg/g for total carotenoids. Since the concentration of α-carotene was significantly low across the inbred lines and had poor repeatability, it was not included in the association study.

3.3.2 Genotyping and SNP diversity

The summary of the 110 k SNP data set used for the GWAS and its basic diversity parameters are presented in Table 3.1. The average missing data for this data set was 10%. SNP distribution across the genome was not uniform but attained vast coverage (Figure 3. 1). Minor allele frequencies displayed uniform distribution across the 10 chromosomes (average = 0.13 to 0.14, median = 0.06 to 0.8). The rare allele frequencies (< 0.05) represented the largest proportion of the minor allele frequencies (Figure 3.2).

The average inbreeding coefficient (f) estimates per locus ranged from below zero to 1 indicating the variable level of inbreeding across the genome, but the genome-wide mean f was 0.82. The heterozygosity (H) of the lines varied from 0.02 to 0.13 with an average of 0.05. More than half of the inbred lines showed less than 0.04 H. Both H and f had more or less uniform values across the chromosomes. The genome-wide polymorphic information content (PIC) of the SNPs ranged from 0.02 to 0.38, while the average was 0.18. The average PIC calculated across the genome in 10 Mbp intervals showed variable polymorphism across the genome (Appendix 2 – Figure 5).

Table 3.1 SNP data summary and basic diversity parameters for SNP dataset used in the GWAS

		Chr1	Chr2	Chr3	Chr4	Chr5	Chr6	Chr7	Chr8	Chr9	Chr10	GW
Total SNPs	-	17,743	13,023	11,905	11,606	12,941	8,557	9,424	9,373	8,008	7,357	109,937
Missing (%)	-	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
Avg. distance between successive SNPs in Mbp	Max	1.7	2.0	1.4	1.8	1.4	1.3	1.5	2.9	1.3	0.8	2.9
	Mean	0.017	0.018	0.019	0.021	0.017	0.020	0.019	0.019	0.020	0.020	0.019
MAF	Mean	0.13	0.14	0.13	0.13	0.13	0.13	0.12	0.13	0.14	0.13	0.13
	median	0.07	0.08	0.06	0.06	0.06	0.06	0.06	0.07	0.08	0.07	0.07
PIC	Min	0.02	0.02	0.02	0.06	0.02	0.02	0.02	0.02	0.02	0.02	0.02
	Max	0.38	0.38	0.38	0.20	0.38	0.38	0.38	0.38	0.38	0.38	0.38
	Mean	0.18	0.19	0.17	0.18	0.19	0.16	0.15	0.18	0.19	0.17	0.18
	Median	0.16	0.17	0.13	0.16	0.17	0.13	0.11	0.17	0.17	0.13	0.15
f	Min	-0.26	-0.28	-0.67	-0.13	-0.70	-0.78	-0.44	-0.92	-0.83	-0.44	-0.92
	Max	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Mean	0.78	0.75	0.68	0.81	0.74	0.73	0.72	0.67	0.71	0.71	0.82
	Median	0.89	0.88	0.85	0.89	0.87	0.85	0.87	0.86	0.88	0.87	0.87
H	Min	0.02	0.02	0.01	0.02	0.01	0.02	0.01	0.02	0.02	0.02	0.02
	Max	0.12	0.14	0.15	0.13	0.12	0.13	0.14	0.14	0.15	0.16	0.13
	Mean	0.04	0.05	0.05	0.05	0.04	0.05	0.14	0.05	0.05	0.05	0.05
	median	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.05	0.04	0.03

Chr = Chromosome, GW = Genome-wide, MAF = MAF, Min = Minimum, Max = Maximum, PIC = polymorphic information content, f = inbreeding coefficient, and H = heterozygousity,
H was estimated in TASSEL, per inbred line basis
PIC , and **f** were calculated in PowerMarker using 3532 SNPs without missing data point and 1% MAF.

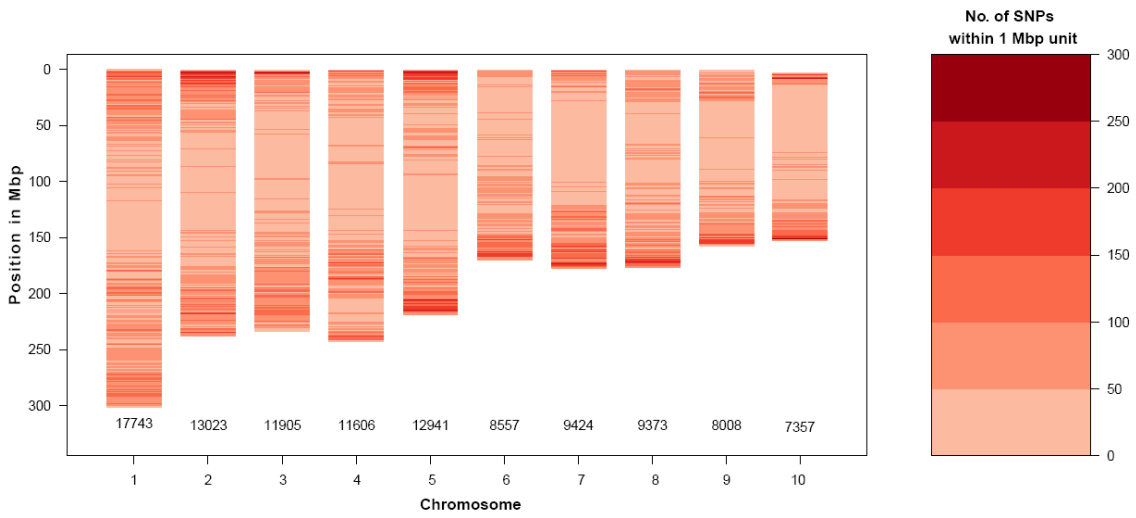


Figure 3.1 Schematic diagram illustrating the distribution of 110k SNPs
Obtained after filtering the raw GBS SNP data set with criteria of 20% minimum missing data and 1% minimum minor allele frequency (MAF) across the 10 chromosomes of maize (plotted with an R package Synbreed, Wimmer et al., 2012). Number of SNPs per chromosome is indicated above the x-axis.

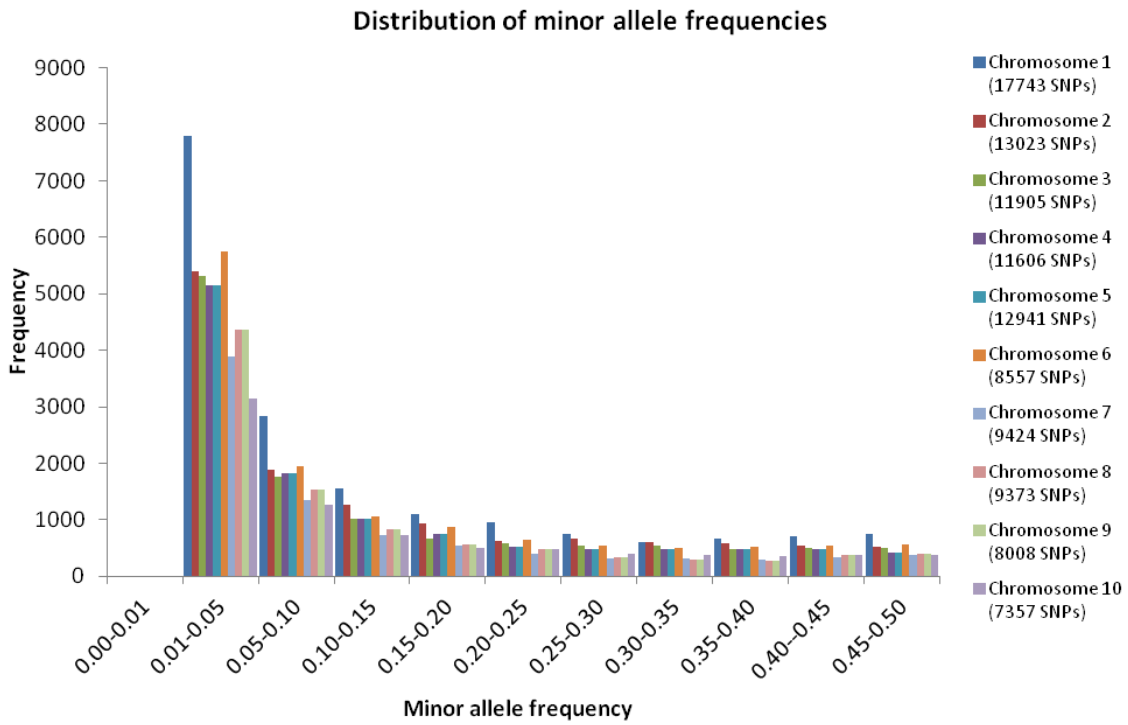


Figure 3.2 MAF distribution across the 10 maize chromosomes of 109,937 SNPs filtered from GBS SNPs.
Range values indicated on the X-axis represent ranges of minor allele frequencies (MAF) on each chromosome. Values on the Y-axis represent the frequencies of the MAF ranges. Alleles having frequencies below 50% are considered as minor alleles.

3.3.3 Population structure and kinship

The first three PCs accounted for 21% of the variation, thus the first 15 PCs had to be considered for the variance explained to reach 50% (Table 3.2). On the basis of BIC model selection the optimum number of principal components (PC) Q (population structure) to account for the variation in the various carotenoids was either no PC or two PCs (Table 3.3). This demonstrated the subtle contribution of genome-wide population structure to the variation in carotenoid profile of the panel based on PCA.

Table 3.2 Variances explained by the first 16 principal components

PC	Eigenvalues	Individual Proportion	Cumulative Proportion
1	6750.8	0.10854	0.10854
2	3892.3	0.062582	0.17112
3	2723.7	0.043793	0.21492
4	2082.6	0.033485	0.2484
5	1841.9	0.029615	0.27802
6	1671.6	0.026878	0.30489
7	1618.7	0.026026	0.33092
8	1595.6	0.025655	0.35658
9	1353.5	0.021762	0.37834
10	1308.7	0.021041	0.39938
11	1255.7	0.02019	0.41957
12	1220.3	0.01962	0.43919
13	1140.3	0.018334	0.45752
14	1133.6	0.018227	0.47575
15	1022.8	0.016445	0.49219
16	1010.4	0.016246	0.50844

Table 3.3 Optimum number of the first PCs selected based on BIC

Carotenoid	**No. of PCs selected based on BIC**
β-carotene	2
β -cryptoxanthin	0
Lutein	0
Zeaxanthin	0
Total carotenoid	0
Provitamin A	2
Lnβ-branch/α-branch	0
Lnβ-carotene/zeaxanthin	2
Lnβ-carotene/All	2
Lnβ-carotene/β-cryptoxanthin	2

As the association panel included inbred lines derived from several bi-parental and backcross progenies, relatedness was expected among individuals derived from crosses involving common parent(s) (Table 2.1). The kinship heat-map for the panel (Figure 3.3) illustrated that the majority of the inbred lines had kinship values below 0.5 suggesting the low level of overall relatedness in the panel. The counts of kinship values

started to rise from near zero, peaked when approaching kinship value of 0.5 and quickly declined before reaching 1 and stayed constant just above 0 till kinship value of 2, demonstrating, the variable level of relatedness across groups of inbred lines in the panel.

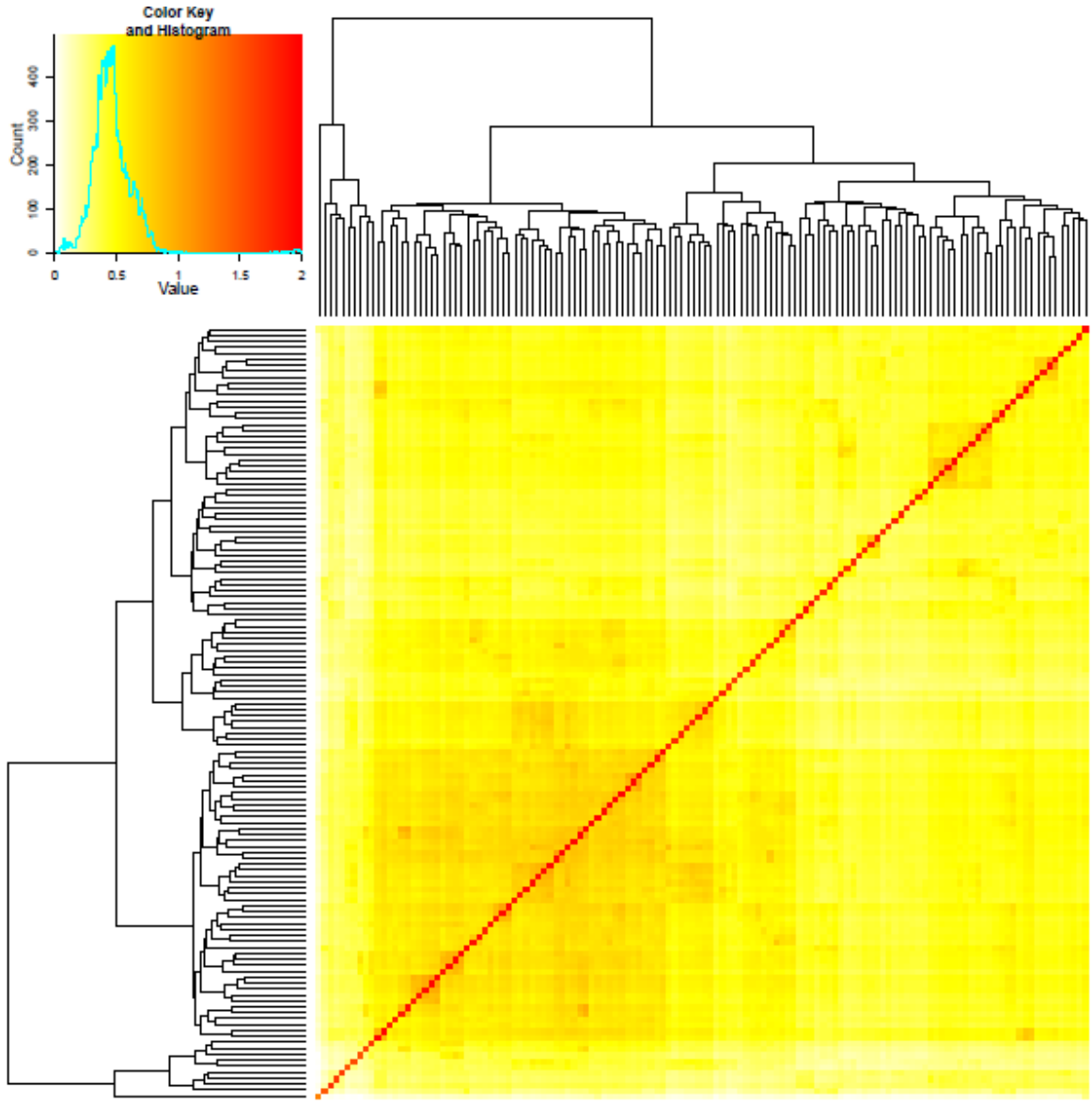


Figure 3.3 Kinship heat-map for 130 inbred lines

3.3.4 Linkage disequilibrium (LD)

Genome-wide LD analysis was performed using the 1658 SNPs filtered out from the large GBS dataset at 0% missing and 10% minimum MAF. Full matrix LD analysis of this dataset in TASSEL 3 involved 1.3 million pairwise correlations, 10% of which resulted in significant R^2 at $P<0.001$ (Table 3.4). The genome-wide extent of LD estimate was 0.83 Mbp at baseline $R^2 = 0.2$ and 0.65 Mbp at $R^2 = 0.25$ (Table 3.4, Figure 3.4). There was, however, heterogeneous distribution of LD as was evident from the pattern of LD heat-map created using the same SNP dataset (Figure 3.5) verifying the unpredictable nature of LD decay (Flint-Garcia et al., 2003), thus implying the variable association resolution attained across the genome.

LD for each chromosome persisted from 0.65 Mbp on chromosomes 2 to 1.6 Mbp on Chromosome 3 when considering 0.2 as a baseline critical R^2. On the other hand, at the calculated critical $R^2 = 0.25$, the LD decay across chromosomes was less variable ranging from 0.56 Mbp on chromosome 7, 0.59 on chromosome 2 and 8, 0.65 on chromosome 10, 0.71 on chromosomes 1, 3, 4, 5 and 6, to 0.77 on chromosome 9.

Table 3.4 Summary of extent of Chromosome and genome-wide LD estimates plus information on the SNP datasets used to calculate the LD

Chromosome	No. of SNPs	No. of pairwise Comparisons	R^2				LD decay*	
			% P<0.001	% >0.2	Avg	Max	R^2 = 0.2	R^2 = 0.25
1	3710	6878340	13.8	28.6	0.18	1	0.96	0.71
2	2790	2889260	12.4	30.8	0.19	1	0.65	0.59
3	2384	2839344	20.0	35.6	0.20	1	1.6	0.71
4	2294	2628924	13.2	29.8	0.19	1	0.71	0.71
5	2617	3423036	14.6	32.6	0.20	1	0.71	0.71
6	1644	1349724	12.1	31.0	0.19	1	0.83	0.71
7	1750	1529500	21.9	34.0	0.20	1	1.27	0.56
8	1851	1712175	16.0	32.5	0.20	1	0.71	0.59
9	1834	1679944	13.0	30.3	0.19	1	0.83	0.77
10	1562	1218360	13.0	32.5	0.20	1	0.71	0.65
Across genome	1658	1373654	10.25	15.40	0.25	1	0.83	0.65

*LD decay estimated at two baseline critical R^2 values for comparison purpose. Avg = average, Max = maximum, % *P*<0.001: percentage of significant LD (R^2) having their *P*-values < 0.001, %>0.2: percentage of R^2 values greater than 0.2 critical R^2

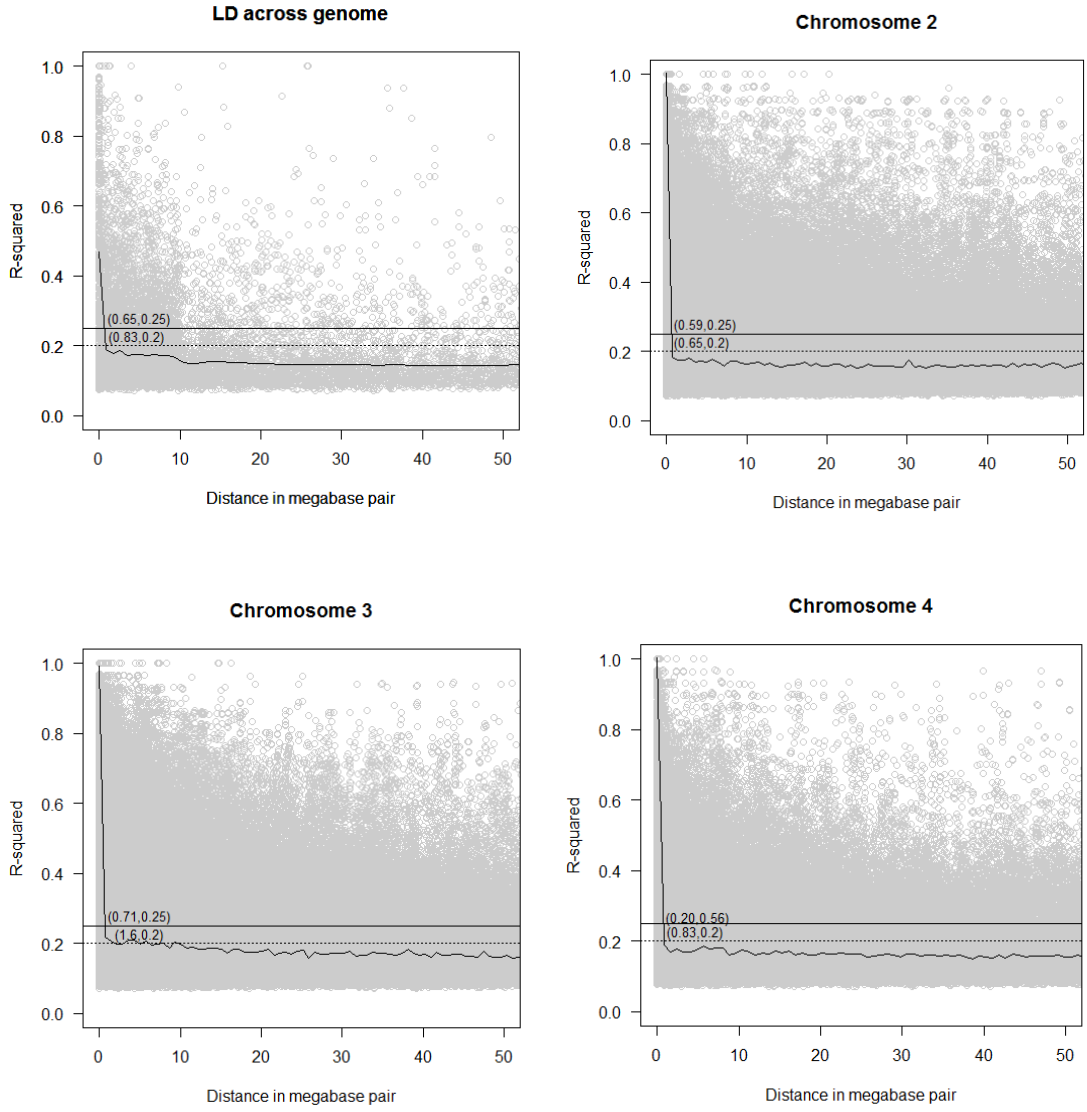


Figure 3.4 Genome-wide and chromosome wide LD decay plots

Estimated on the basis of pair-wise squared allele frequency correlation coefficients (R^2) among 1658 SNPs distributed across each of the 10 maize chromosomes. The dataset used for the LD analysis was filtered with criteria of no missing values and 10% minimum MAF. R^2 values with p<0.001 were considered significant. Horizontal axis shows distance in Mbp and vertical axis represents the R^2 values. Two estimates of LD are shown with the coordinate labels for the intersection points of the LOESS curves and lines drawn at R^2 = 0.2 (broken line) and R^2 = 0.25 (solid line).

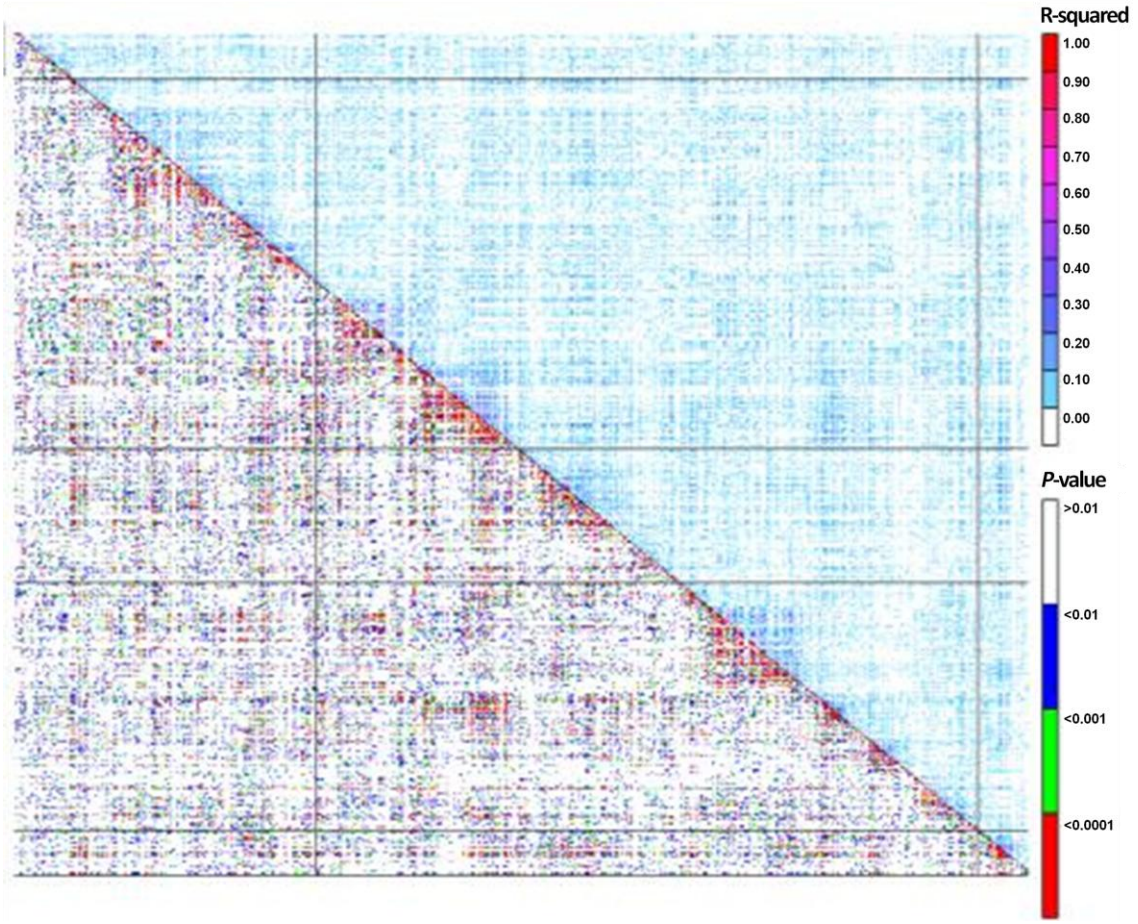


Figure 3.5 Genome-wide LD heat-map plotted using 1658 SNP dataset representing all the 10 chromosomes of maize in TASSEL 3 software.
The lower triangle represents *P*-values while the upper one represents R2. The colour legends indicate the level of significance and the corresponding strength of LD (R2).

3.3.5 Genome-wide Association Study (GWAS)

Out of 110 k SNPs tested using the unified MLM model that accounted for population structure and multiple level relatedness in the panel (Yu et al., 2006), 386 unique significant SNPs were detected at 5% FDR (Table 3.5). At this significance threshold, at least two significant SNPs were identified on each of the 10 chromosomes. The number of significant SNPs declined to 168 at 1% FDR correction rate, dropping all the SNPs on chromosomes 1, 5 and 7. Application of the conservative multiple comparison correction term, the Bonferroni test, at 5 and 1% levels further reduced the number of significant SNPs to 81 and 32, respectively. The vast majority of significant SNPs were found on chromosome 8 followed by chromosome 10, which were mainly associated with lutein and β-branch carotenoids, respectively. Except for significant SNPs on chromosome 6 and 9, the average minor allele frequencies of the significant SNPs at FDR 1 and 5 % were above 10%. Only 5 % of the significant SNPs at FDR 1% had their minor allele frequencies below 10%.

Table 3.5 Number of significant SNPs by chromosome with the corresponding average minor allele frequencies (MAF) at different cutoff thresholds

Chromosome	FDR		Bonferroni		Average MAF	
	5%	1%	5%	1%	FDR 5%	FDR 1%
1	7	0	0	0	0.15	-
2	14	4	3	0	0.21	0.20
3	7	3	0	0	0.16	0.18
4	4	3	1	1	0.13	0.14
5	5	0	0	0	0.14	-
6	7	1	0	0	0.13	0.03
7	3	0	0	0	0.10	-
8	226	120	53	27	0.20	0.19
9	4	1	0	0	0.09	0.08
10	109	36	24	13	0.21	0.21
All*	386	168	81	41	0.18	0.19

*SNPs associated with multiple trait counted only once

Lutein was the most significantly affected carotenoid in this GWA study (Table 3.6; Table 3.7). At the significance level of 1% FDR, a total of 129 SNPs distributed on chromosomes 2, 3, 4, 6, 8 and 9 were associated with lutein, with the largest fraction (>90%) coming from chromosome 8. The most significant SNPs associated with this carotenoid scored the lowest of all the *p*-value in the present GWAS result (SNPs S8_138938983 and S8_138938949, P = 9.81E-12). The model containing each of these SNP explained 53% of the variation in the carotenoid (Table 3.7). The majority of significant SNPs that survived the stringent significance threshold of 1% Bonferroni were also associated with lutein (a single SNP on chromosome 4 and 27 SNPs on chromosome 8). Many of these SNPs were also associated with the ratio of α to β branch carotenoids at FDR 5%. The second most significantly associated trait was the ratio of β-carotene to β-cryptoxanthin. Twenty six SNPs were associated with this derived trait at FDR 1%, the most significant SNP (S10_136007578) scoring *p*-value of 6.75E-10 and R^2 of 60%.

At 1% Bonferroni, 13 SNPs on chromosome 10 were associated with carotenoids of the β branch and some of the derived ratio traits (β-carotene, β-cryptoxanthin, β-carotene to β-cryptoxanthin and/or β-carotene to zeaxanthin). The ratio of α to β branch carotenoid was significantly affected by a single SNP on chromosome 4, and 10 SNPs on chromosome 8 at FDR 5%, the most significant SNPs in the group accounting for 33 and 36% of the variations in the derived trait, respectively. These SNPs were also significantly associated with lutein.

On the other hand, associations with zeaxanthin (12 SNPs) and provitamin A (3 SNPs) were detected only when relaxing the significance cutoff threshold to 5% FDR. The variances explained by the model involving the most significant SNPs were 33 % for zeaxanthin (SNP S10_136840488, P = 5.53E-07) and 51% for provitamin A (S10_134601800, P = 2.68E-02; Table 3.7). These SNPs were also associated with β-carotene and its derived ratio traits.

The GWAS result for each trait across chromosomes is illustrated with Manhattan plots (Figure 3.6), and with a chart that summarizes the association using the lowest *p*-values

attained at 5% FDR threshold (Figure 3.7). The allelic variants and effects for the selected most significant SNPs in the GWAS are indicated in Table 3.7.

Table 3.6 Carotenoids and number of associated SNPs by chromosome

Carotenoid	Significance Threshold	No. of SNPs (Chromosome)	Total
β-carotene	Bon 1%	5 (10)	5
	FDR 1%	17 (10)	17
β-cryptoxanthin	Bon 1%	2 (10)	2
	FDR 1%	12 (10)	12
Lutein	Bon 1%	27 (8), 1 (4)	28
	FDR 1%	1 (2), 3 (3), 3 (4), 1 (6) , 120 (8), 1 (9)	129
β-carotene to β-cryptoxanthin ratio	Bon 1%	12 (10)	12
	FDR 1%	26 (10)	26
α to β branch carotenoids ratio	FDR 1%	5 (8)	5
	FDR 5%	1 (4) 10 (8)	11
Zeaxanthin	FDR 5%	12 (10)	12
Provitamin A	FDR 5%	3 (10)	3
Total carotenoid	FDR 1%	3 (2)	3
β-carotene to zeaxanthin ratio	Bon 1%	5 (10)	5
	FDR 1%	12 (10)	12
β-carotene to all carotenoids ratio	FDR 5%	8 (10)	8

SNPs associated with multiple carotenoids were included in the counts here, thus sums of SNPs may not tally with those indicated in Table 3.5. FDR thresholds of 5% considered only for traits for which no significant SNPs could be obtained at the stringent thresholds. Bon = Bonferroni

Table 3.7 Summary table for the GWAS result

No.	SNP	Chr	Position	Alleles	Allelic Effect	*P*-value	MAF	R^2 of Model	R^2 of Model + SNP	FDR adjusted *P*-value	Carotenoid
1	S2_208672678	2	208,672,678	C/G	3.00	6.17E-06	0.03	0.16	0.31	6.40E-03	lut
2	S3_18632237	3	18,632,237	C/G	1.76	1.38E-06	0.23	0.16	0.33	2.10E-03	lut
3	S3_49624005	3	49,624,005	C/T	2.24	7.05E-06	0.13	0.16	0.31	6.85E-03	lut
4	S3_99107971	3	99,107,971	A/T	2.09	1.06E-05	0.18	0.16	0.30	9.02E-03	lut
5	S4_172435841	4	172,435,841	G/C	0.53	3.63E-09	0.15	0.16	0.42	3.99E-05	lut
6	S4_8900031	4	8,900,031	C/T	2.33	4.87E-06	0.17	0.16	0.31	5.41E-03	lut
7	S4_229316518	4	229,316,518	G/A	-2.75	9.16E-06	0.08	0.16	0.30	8.25E-03	lut
8	S6_165089413	6	165,089,413	T/G	3.27	9.27E-06	0.03	0.16	0.30	8.29E-03	lut
9	S8_16743428	8	16,743,428	C/T	2.18	8.65E-09	0.21	0.16	0.41	6.34E-05	lut
10	S8_111289041	8	111,289,041	C/T	2.67	5.82E-09	0.16	0.16	0.42	5.34E-05	lut
11	S8_118971709	8	118,971,709	C/A	-2.80	1.49E-08	0.12	0.16	0.40	1.03E-04	lut
12	S8_121485958	8	121,485,958	C/T	2.36	3.60E-08	0.14	0.16	0.39	1.65E-04	lut
13	S8_123786605	8	123,786,605	T/C	-2.79	2.86E-08	0.11	0.16	0.39	1.65E-04	lut
14	S8_124434722	8	124,434,722	G/T	-0.43	3.97E-10	0.25	0.16	0.46	7.28E-06	lut
15	S8_128541902	8	128,541,902	C/T	2.53	4.19E-08	0.13	0.16	0.38	1.71E-04	lut
16	S8_130212000	8	130,212,000	A/G	-0.47	1.67E-08	0.16	0.16	0.40	1.08E-04	lut
17	S8_131682022	8	131,682,022	G/T	1.93	8.38E-08	0.25	0.16	0.37	3.29E-04	lut
18	S8_138510292	8	138,510,292	C/G	-0.51	2.87E-10	0.15	0.16	0.47	7.28E-06	lut
19	S8_138938949	8	138,938,949	C/T	-0.60	9.81E-12	0.13	0.16	0.53	5.39E-07	lut
20	S8_141803960	8	141,803,960	G/A	-3.04	4.23E-09	0.10	0.16	0.42	4.23E-05	lut
21	S8_144458630	8	144,458,630	A/G	2.81	6.93E-09	0.11	0.16	0.41	5.44E-05	lut
22	S9_112005623	9	112,005,623	C/G	2.16	4.43E-06	0.08	0.16	0.31	5.23E-03	lut
23	S10_13600757	10	136,007,575	G/A	-2.39	1.19E-09	0.22	0.42	0.62	6.55E-05	βcar
24	S10_13987759	10	139,877,594	G/A	-3.20	5.13E-08	0.13	0.42	0.57	1.13E-03	βcar
25	S10_11697760	10	116,977,608	G/C	-3.11	2.74E-07	0.15	0.42	0.55	3.11E-03	βcar
26	S10_13683362	10	136,833,624	C/T	2.59	4.98E-07	0.15	0.42	0.55	4.98E-03	βcar
27	S10_10289734	10	10,289,734	T/G	-1.96	1.07E-06	0.20	0.42	0.54	9.00E-03	βcar
28	S10_12442759	10	124,427,599	C/T	2.76	1.31E-06	0.18	0.42	0.54	9.00E-03	βcar
29	S10_74479633	10	74,479,633	G/C	-1.89	1.53E-06	0.19	0.42	0.53	9.89E-03	βcar
30	S10_13465570	10	134,655,704	T/C	-2.98	6.98E-08	0.13	0.22	0.42	3.28E-03	βcryp
31	S10_13684048	10	136,840,488	T/C	1.42	9.59E-08	0.18	0.22	0.41	3.28E-03	βcryp
32	S10_59877496	10	59,877,496	G/C	1.51	3.19E-07	0.16	0.22	0.39	4.96E-03	βcryp
33	S10_13683362	10	136,833,624	C/T	2.59	3.61E-07	0.15	0.22	0.39	4.96E-03	βcryp
34	S10_13987759	10	139,877,594	G/A	-3.20	6.51E-07	0.13	0.22	0.38	7.43E-03	βcryp
35	S8_137468530	8	137,468,530	C/T	-0.46	2.23E-07	0.23	0.16	0.36	9.60E-03	lnβbr/α-br
36	S8_138938949	8	138,938,949	C/T	-0.60	2.71E-07	0.13	0.16	0.35	9.60E-03	lnβbr/α-br
37	S8_117876676	8	117,876,676	T/C	0.44	4.37E-07	0.36	0.16	0.34	9.60E-03	lnβbr/α-br
38	S10_13600757	10	136,007,578	G/T	2.39	6.75E-10	0.22	0.38	0.60	3.15E-05	lnβcar/βcry
39	S10_13987759	10	139,877,594	G/A	-3.20	9.60E-10	0.13	0.38	0.60	3.15E-05	lnβcar/βcry
40	S10_13683362	10	136,833,624	C/T	2.59	2.69E-08	0.15	0.38	0.56	3.67E-04	lnβcar/βcry
41	S10_11697760	10	116,977,608	G/C	-3.11	2.78E-08	0.15	0.38	0.55	3.67E-04	lnβcar/βcry
42	S10_14157461	10	141,574,617	T/C	-2.03	9.58E-08	0.13	0.38	0.54	8.10E-04	lnβcar/βcry
43	S10_4749679	10	4,749,679	G/C	-1.43	6.57E-07	0.38	0.38	0.52	3.44E-03	lnβcar/βcry
44	S10_13465098	10	134,650,981	A/T	3.03	1.65E-09	0.13	0.30	0.53	1.04E-04	lnβcar/zea
45	S10_71671890	10	71,671,890	A/T	2.03	2.52E-07	0.15	0.30	0.46	4.61E-03	lnβcar/zea
46	S10_13987759	10	139,877,594	G/A	-3.20	3.90E-07	0.13	0.30	0.46	5.36E-03	lnβcar/zea
47	S2_44473758	2	44,473,758	C/T	-5.61	1.78E-07	0.21	0.16	0.36	9.26E-03	tcar
48	S2_139644276	2	139,644,276	G/A	-1.10	2.53E-07	0.35	0.16	0.36	9.26E-03	tcar
49	S10_13460180	10	134,601,800	G/A	-2.09	6.39E-07	0.29	0.38	0.51	2.68E-02	pva
50	S10_13600757	10	136,007,578	G/T	2.39	7.32E-07	0.22	0.38	0.51	2.68E-02	pva
51	S10_13684048	10	136,840,488	T/C	1.42	5.53E-07	0.18	0.15	0.33	4.76E-02	zea
52	S10_13465570	10	134,655,704	T/C	-2.98	8.84E-07	0.13	0.15	0.32	4.76E-02	zea
53	S10_13563418	10	135,634,185	G/A	-1.52	2.92E-06	0.31	0.15	0.31	4.76E-02	zea
54	S10_13907594	10	139,075,941	A/C	2.03	4.00E-06	0.15	0.15	0.30	4.76E-02	zea

*Representative significant SNPs selected based on their positions and approximate LD decay. Significant SNPs were selected at FDR 1%, except for chromosome 8 SNPs that are associated with lutein which

were selected only at Bonferroni 1%. For zeaxanthin and total provitamin A the threshold was set at 5% FDR to detect significant SNPs. There are SNPs that appear 2 to 4 times as they were associated to multiple related traits. Bcar = β-carotene, βcryp = β-cryptoxanthin, lut = lutein, zea = zeaxanthin, pva = provitamin A, tcar = total carotenoid, β br/αbr = ratio of carotenoids on β to α branch, Chr = Chromosome, MAF = minor allele frequency, FDR = false discovery rate. Allelic effects of SNPs indicated refer to the second allelic variants.

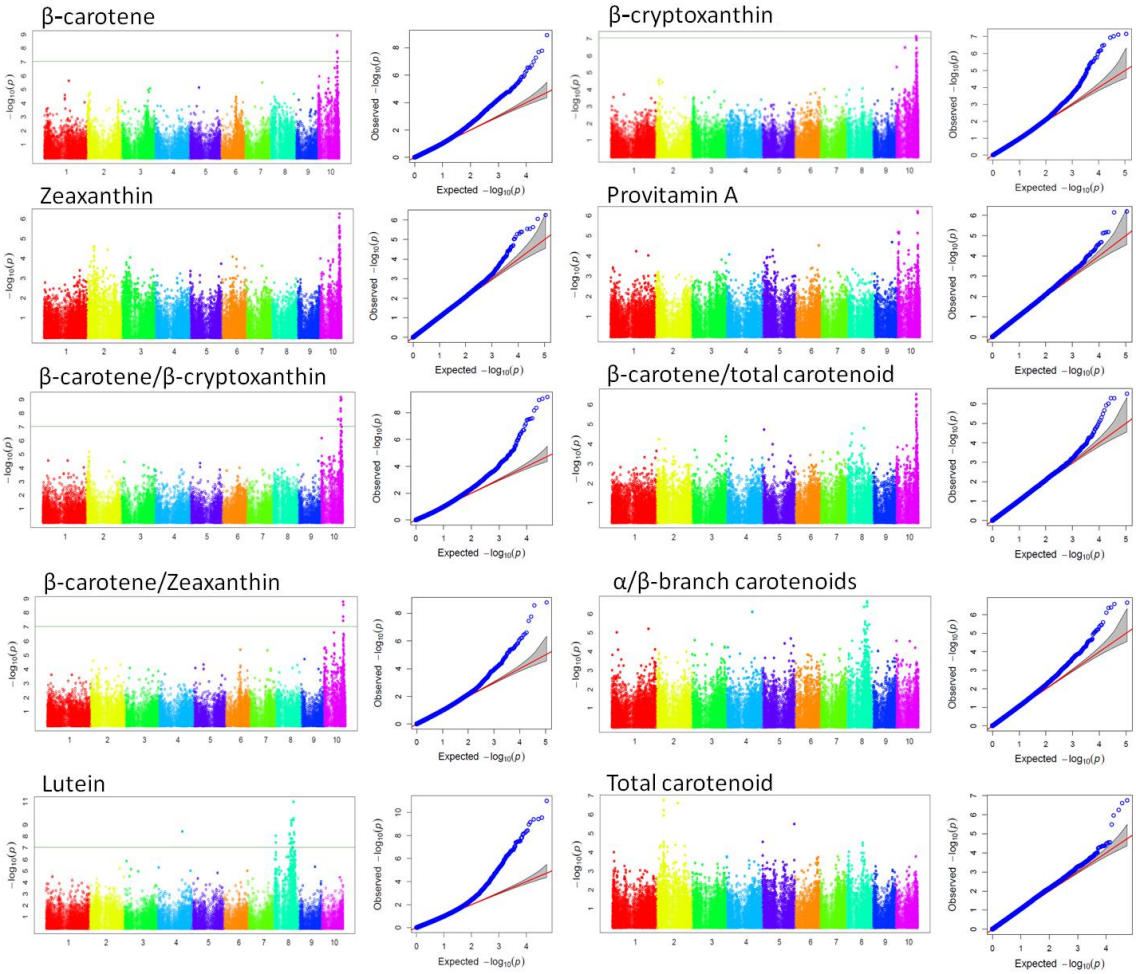


Figure 3.6 Manhattan and QQ-plots depicting the GWAS result.
For the Manhattan plots, horizontal axes represent chromosomes, vertical axes represent –log of the *p*-values to the base 10. Horizontal line at –log10(p) = 7.04 is 1% Bonferroni-adjusted cutoff threshold for highly significant associations. QQ-plots (quantile quantile-plots) show how well the MLM GWAS model fit expectation; they are the –log of *p*-values from the MLM GWAS in the y-axis plotted against their expected values under the null-hypothesis of no association between SNPs and the trait under consideration in the x-axis (Lipka et al., 2012).

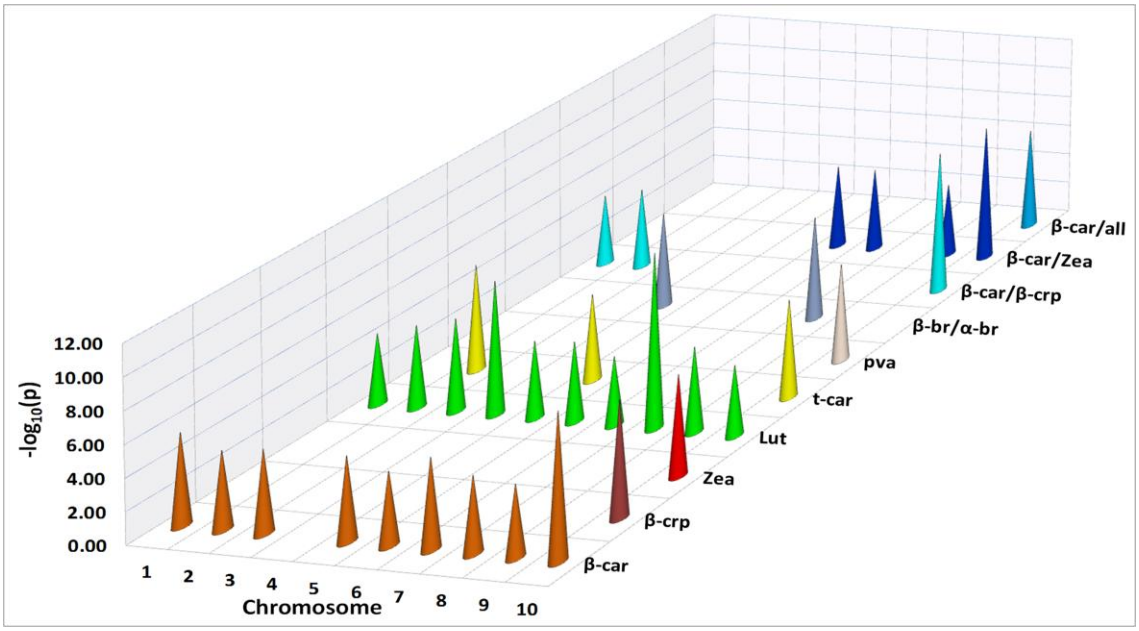


Figure 3.7 Significant association per chromosome per trait as represented by the most significant SNPs (the lowest *p*-value) at the cutoff threshold of FDR 5%.

The least significant SNP was on chromosome 7 (*p*-value = 4.77E-05 or $-\log_{10}(p) = 4.32$) and the most significant SNP was on chromosome 8 (*p*-value = 9.81E-12 or –log (*p*-value) = 11.01), which were both associated with lutein. Bcar = β-carotene, βcryp=β-cryptoxanthin, lut = lutein, zea = zeaxanthin, pva = provitamin A, tcar = total carotenoid βbr/αbr = ratio of carotenoids on β to α branch.

3.3.6 Genes in LD with significant SNPs

The genomic locations of prominent significant SNPs were investigated with respect to the genes they were linked to based on information retrieved from online databases for maize genome (http://www.maizegdb.org/ and http://ensembl.gramene.org/Zea_mays/). The list of all annotated genes, including those encoding uncharacterized proteins, within about 0.8 MB of the most significant SNPs are presented in Appendix 1 – Table 6. Here, only those genes that are the closest to the most significant SNPs are described.

The most significant SNP in the association signal for lutein around 16Mbp on chromosome 8 (SNP S8_16743428, *p*-value = 8.65E-09; Figure 3.6, Figure 3.8a, Table 3.7) was located within a putative gene GRMZM2G143211 (Table 3.9). This gene model contains a WD domain and displays homology to the yeast autophagy 18 (ATG18) gene in *Arabidopsis thaliana*. Two additional significant SNPs (S8_16444572 & S8_16444587) in this region were located in another candidate gene GRMZM2G380414, which encodes a protein called Ultraviolet-B-repressible which is likely involved in photosynthesis (Table 3.9). The complex association peak between 110 and 144MBp on the same chromosome for the same trait contained three highly significant SNPs, namely S8_111289041 (*p*-value = 5.82E-09); S8_124434722 (*P*-value = 3.97E-10), and S8_138938949/S8_138938983 (*p*-value = 9.81E-12) that were in strong LD, with R^2 value ranging from 0.31 to 0.67 (Figure 3.8). These SNPs were located within three different putative genes GRMZM2G333079, GRMZM2G330693 and GRMZM2G463133, respectively (Table 3.9) with the first and third genes having evidence for expression in maize endosperm (Sekhon et al., 2011) The proximal SNPs, S8_138938949 and S8_138938983 are 50K away from one of the major carotenoid biosynthesis genes, *lcyE*. Pairwise LD among these highly significant SNPs on chromosome 8 varied from R^2 = 0.23 to 0.67 (Figure 3.10).

On chromosome 10 the strong association peak surrounding the 138Mbp region for β-carotene (Figure 3.6, Figure 3.8b) contained two closely spaced and significant SNPs S10_136007575 and S10_136007578, p = 6.75E-10. These SNPs are the closest SNPs detected to the major candidate gene of carotenoid *crtRB1* (~ 40kb), but physically located within a putative RING zinc finger domain protein coding gene,

GRMZM2G397684 (Table 3.9, Figure 3.8b). The other significant SNPs in this region were S10_134650981 (*P*-value = 1.99E-08) and S10_139877594 (*P*-value = 5.12E-08), residing within candidate genes GRMZM2G018314 and GRMZM2G080516, respectively. The later encodes a transcription factor known as AP2-EREBP which is known to be expressed in maize endosperm (Sekhon et al., 2011). LD among the peak SNPs on chromosome 10 ranged from 0 between SNPs S10_134650981 and S10_139877594, to 1, between SNPs S10_136007575 and S10_136007578 (Figure 3.10).

The most significant SNP on chromosome 4 (SNP S4_172435841) that was associated with lutein (*p*-value = 3.63E-09) was located within a gene named JUMONJI-transcription factor-4 (jmj4), GRMZM2G027075 (Figure 3.8d). This gene is highly expressed in maize endosperm among other tissues (Sekhon et al., 2011). However, while this SNP was in LD with those significant SNPs on chromosome 8 (R^2 = 0.23 to 0.64, $P<0.001$, Table 3.8, Figure 3.10), it was in equilibrium with SNPs on chromosome 4 for which LD was estimated (data not shown), implying this significant SNP was mapped to the wrong chromosome.

A plausible small effect significant association was detected on the short arm of chromosome 2 coinciding with the gene involved in the conversion of carotenoids to abscisic acid, zeazanthine epoxidase 1 (*ZEP1,* GRMZM2G127139) (Vallabhaneni and Wurtzel, 2009). Two of the six SNPs that were significantly associated with total carotenoids at FDR 5% (S2_44448438, p = 1.11E-06) and S2_44448432 p = 1.11E-06) were physically located within this gene (Figure 3.8c). But, the most significant SNP, S2_44473758, p = 1.78E-07, was located about 33kb downstream of the gene in another putative gene GRMZM2G062559 that encodes uncharacterized protein. All of these SNPs were in high LD forming a haplotype block (Figure 3.8) when considered without the non significant SNPs in the region, and thus the significant effect most likely stemmed from *ZEP1* or it could be that they are linked with some regulatory regions. The other SNP on the same chromosome around position 139Mbp that showed strong association with total carotenoid (S2_139644276, *P*-value = 2.53E-07) was located in a putative gene GRMZM2G066213 (Figure 3.8c).

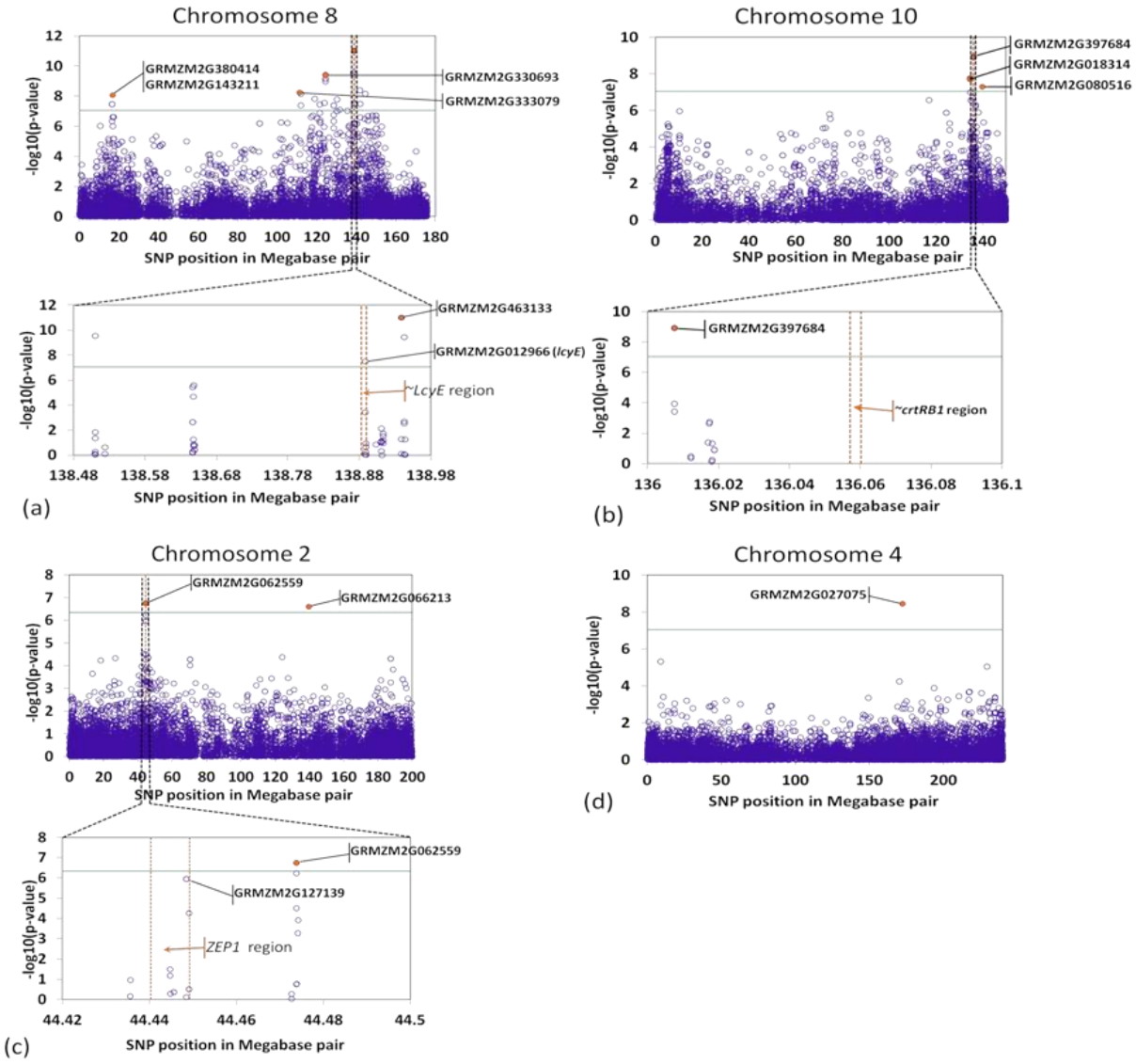


Figure 3.8 Scatter diagrams showing statistically significant association signals

(a) chromosome 8 for lutein, (b) chromosome 10 for β-carotene, (c) chromosome 2 for zeaxanthin and (d) chromosome 4 for lutein.The most significant or interesting SNPs are highlighted in orange colour and labeled with the IDs of the putative genes they are linked to. Light green horizontal lines represent 1% Bonferroni-adjusted significance threshold ($-\log_{10}(p) = 7.04$) except for chromosome 2, which refers to 5% Bonferroni significance threshold ($-\log_{10}(p) = 6.34$). Vertical orange lines show regions of the major carotenoid candidate genes *lcyE*, *crtRB1*, and *ZEP1*. Plots were made with Microsoft Excel chart.

Strong and extensive pairwise LDs were observed among the significant SNPs selected at 1% FDR (Figure 3.9, Table 3.8). Seventy percent of the pairwise comparison among the SNPs led to statistically significant LD ($P<0.001$) of which 21% was for inter-

chromosomal correlations. LD for within chromosome comparisons ranged from 0.37 to 1, both on chromosome 3, with genome-wide average of 0.42. For inter chromosomal comparisons, LD ranged from 0.18 in chromosome 10 to 0.5 in chromosome 3, with genome-wide average of 0.25. Significant SNPs on chromosomes 3 and 4 displayed strong inter-chromosomal LD with those on chromosome 8, but was negligible for SNPs on chromosome 10 (Figure 3.9a, Table 3.8). The confidence interval algorithm deployed in HaploView software generated eleven haplotype blocks based on LD of the significant SNPs on chromosome 8, five blocks for those on chromosome 10 and one block for those on chromosome 2. Haplotype blocks were identified for each of the three carotenoid genes *crtRB1*, *lcyE* and *ZEP1* when analyzing the significant SNPs in regions surrounding their corresponding genomic locations. Further analysis of LD for regions comprising these genes, with the inclusion of non-signifcant SNPs revealed heterogeneous LD (Figure 3.9e). This suggests that the LD among the significant SNPs residing in regions of these major genes could be functional, rather than only tight genetic linkage that occurred as a result of long range average LD decay in the association panel, as estimated in this study.

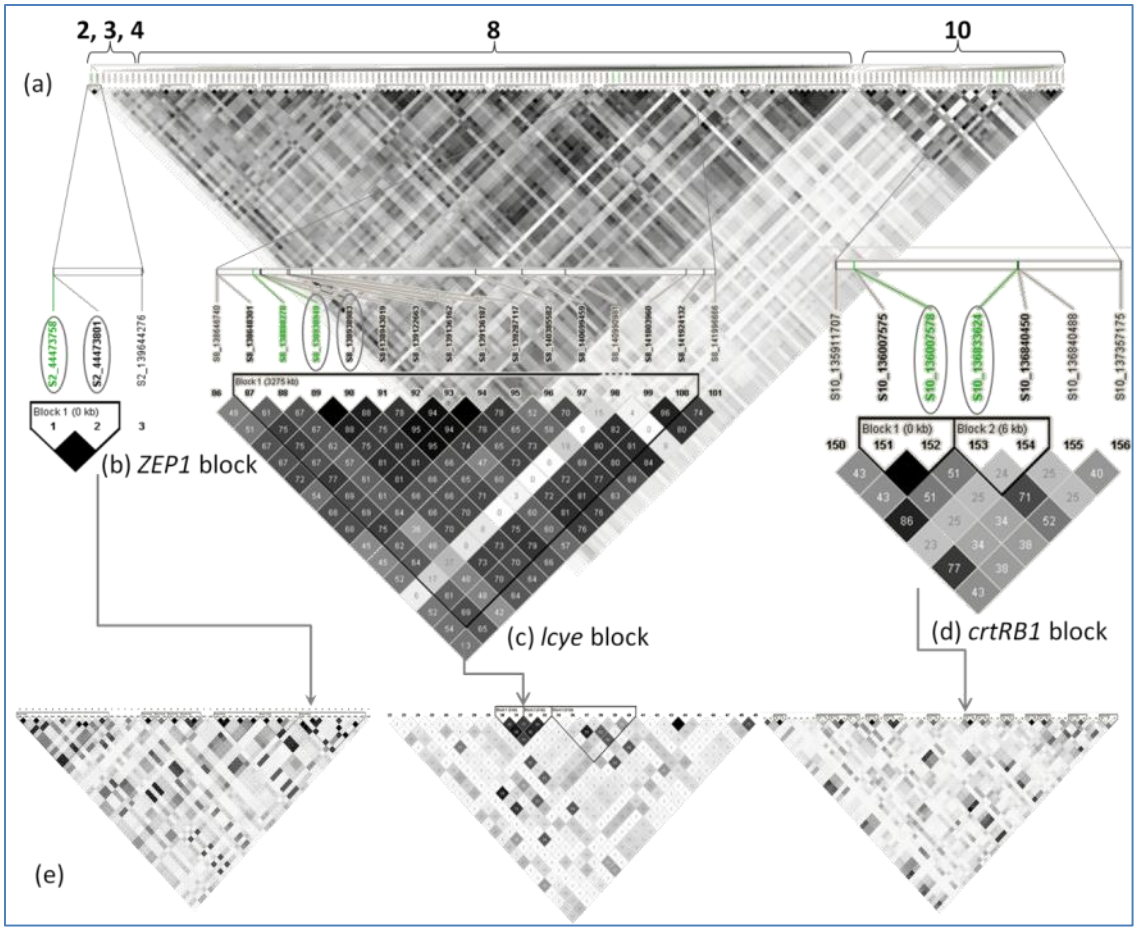


Figure 3.9 LD plots of significant SNPs, LD blocks surrounding *lcyE*, *crtRB1* and *ZEP1*

(a) LD plot of all significant SNPs selected at FDR 1%. Labels 2, 3, 4, 8 and 10 refer to the chromosomes of the SNPs that reached significance at this threshold. (b) an LD block on chromosome 2 surrounding the gene *ZEP1* (c) an LD block on chromosome 8 surrounding the gene *lcyE* (d) an LD block on chromosome 10 comprising the gene *crtRB1*. **(e)** LD plots that included non-significant SNPs in regions +/- of *crtRB1*, *lcyE* and *ZEP1* where significant associations were detected. Haplotype blocks were defined with the option of confidence interval (Gabriel et al., 2002). Green highlighted SNPs are the closest SNPs to the carotenoid genes indicated, with the most significant ones enclosed with oval shapes. The grayscale shading pattern of LD plot reflects the strength of linkage as it increases from the lightest to the darkest shaded cells paralleling the range of no LD (R^2 = 0%) to absolute LD (R^2 = 100%). Plots generated using HaploView software (Barrett et al., 2005).

Table 3.8 Summary of LD analysis among significant SNPs at 1% FDR

		Chr 2	chr 3	chr 4	chr 8	Chr 10	Total
No. of SNPs		3	3	2	119	36	163
% Significant LD (P < 0.001)		33	67	100	96	86	70
Haplotype blocks		1	-	-	11	5	17
Intra-chromosome LD	%P < 0.001						79
	minimum	1.00	0.34	0.59	0.09	0.13	0.09
	maximum	1.00	0.40	0.59	1.00	1.00	1.00
	average	1.00	0.37	0.59	0.42	0.45	0.42
Inter-chromosome LD	% *P*< 0.001						20.70
	min	0.10	0.13	0.15	0.10	0.10	0.10
	max	0.23	1.00	0.75	0.77	0.54	1.00
	average	0.15	0.40	0.37	0.28	0.18	0.25

Chr = chromosome, % *P*<0.001: percentage of significant LD (r^2) having their *P*-values < 0.001

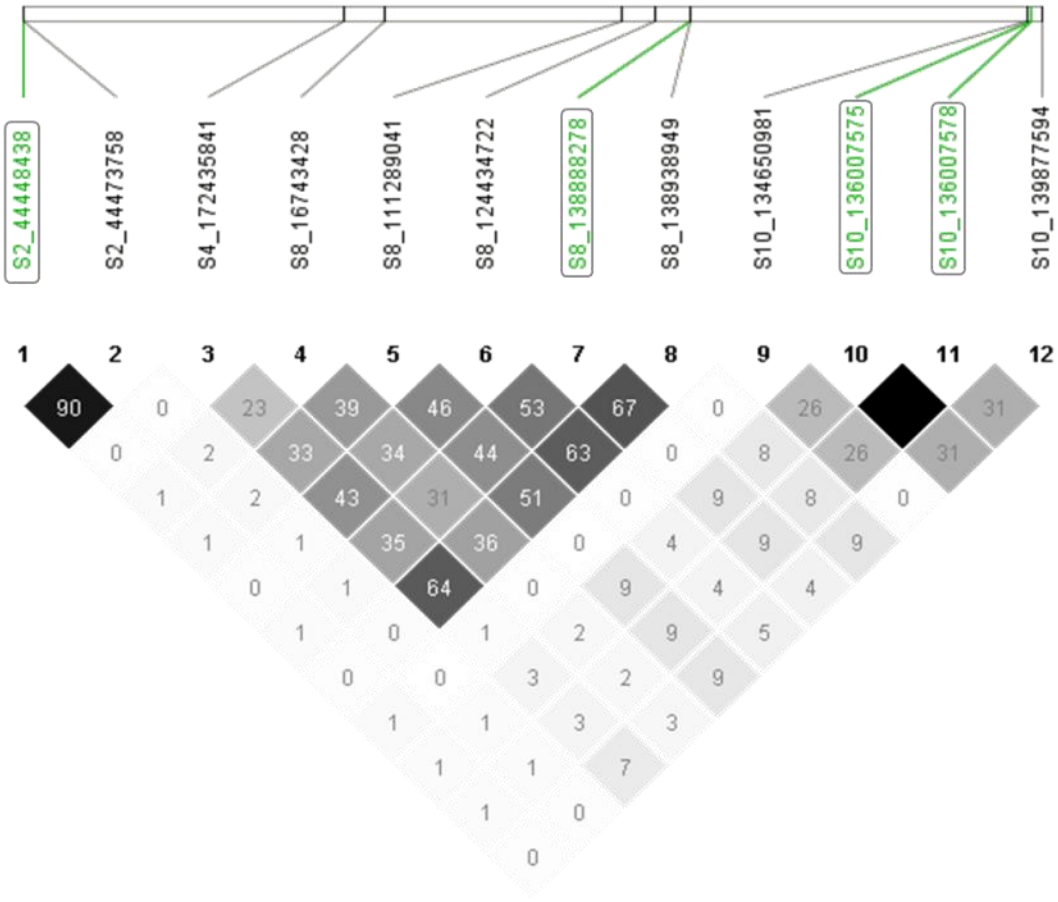


Figure 3.10 LD among selected peak SNPs.

Numbers in cells are R^2 values multiplied by 100, where the darkest gray scale shading denotes 100 and white denotes zero. The closest SNPs to the genes *ZEP1*, *lcyE*, and *crtRB1* (2 SNPs), appearing in the same order, are highlighted in green and enclosed with boxes.

Table 3.9 Selected candidate protein coding genes detected in the GWAS

SNP name	SNP position	lowest *P*-value	Trait affected	R^2 of Model without SNP	R^2 of Model with SNP	stable ID for the closest linked gene	Gene name	Gene description	Chr	strand	AGPv2 Gene start (bp)	AGPv2 Gene end (bp)
S10_136007575/ S10_136007579	136007575/ 136007578	6.7E-10	lnbcar/bcryp	0.38	0.60	GRMZM2G152135	crtRB1[a]	Beta-carotene hydroxylase 1	10	-1	136,057,100	136,060,219
S8_138888278[b]	138888278	3.2E-08	lut	0.16	0.39	GRMZM2G012966	lcyE	Lycopene epsilon cyclase1	8	1	138,882,594	138,889,812
S2_44448432/ S2_44448438	44448438/ 44448432	1.1E-06	tcar, zea[c]	0.16	0.33	GRMZM2G127139	ZEP1	Zeaxanthin epoxidase1	2	-1	44,440,299	44,449,237
S2_44473758/ S2_44473801	44473758	1.8E-07	tcar	0.16	0.36	GRMZM2G062559	-	Uncharacterized protein			44,471,623	44,474,212
S4_172435841	172435841	3.6E-09	lut	0.16	0.42	GRMZM2G027075	jmj4	Uncharacterized protein	4	1	172,430,260	172,439,306
S5_78384689[d]	78384689	7.1E-07	bcar, PVA	0.58	0.68	GRMZM2G102845	-	Auxin response factor 20	5	-1	78,381,834	78,389,884
S8_16743428	16743428	8.7E-09	lut	0.16	0.41	GRMZM2G143211	-	Uncharacterized protein	8	1	16,741,652	16,746,323
S8_16444572/ S8_16444587	16,444,572/ 16,444,587	3.5E-08	lut	0.16	0.39	GRMZM2G380414	-	Ultraviolet-B-repressible protein	8	-1	16,443,989	16,444,752
S8_111289041	111289041	4.0E-10	lut	0.16	0.42	GRMZM2G333079	-	Uncharacterized protein	8	-1	111,287,695	111,290,414
S8_124434722	124434722	5.8E-09	lut	0.16	0.46	GRMZM2G330693	-	Uncharacterized protein	8	1	124,434,479	124,435,152
S8_138938949/ S8_138938983 [b]	138,938,949 /138,938,98 3	9.8E-12	lut	0.16	0.53	GRMZM2G463133	-	Putative HLH DNA-binding domain superfamily protein	8	-1	138,938,542	138,943,955
S10_136007575/ S10_136007578	136007575/ 136007578	6.8E-10	lnbcar/bcryp	0.38	0.60	GRMZM2G397684	-	Putative RING zinc finger domain superfamily protein	10	1	136,006,849	136,007,871
S10_134650981	134650981	2.0E-08	bcar, lnbcar/zea	0.38	0.59	GRMZM2G018314	-	Uncharacterized protein	10	1	134,647,347	134,652,537
S10_139877594	139877594	5.1E-08	bcar, lnbcar/bcryp	0.38	0.60	GRMZM2G080516	ereb2	AP2-EREBP transcription factor	10	-1	139,875,910	139,877,865

[a] SNP linked to crtRB1 does not reside within the the gene

[b] S8_138888278 is not the most significant SNP linked with lcyE but was located within the gene, whereas SNPs S8_138938949/S8_138938983 were the most significant but were located in another gene about 50 Kb downstream of lcyE

[c] Association with zeaxanthin detected only when accounting for the allelic effects of lcyE and crtRB1

[d] S5_78384689 detected in GWAS that considered effects of lcyE and crtRB1 as covariates on the bases of their allele specific markers

Note: Genes and associated information retrieved from maizeGDB.org and gramene.org. Physical positions of SNPs and coordinates of genes given according to B73 RefGen_2. Bcar = β-carotene, βcryp = β-cryptoxanthin, lut = lutein, zea = zeaxanthin, pva = provitamin A, tcar = total carotenoid βbr/αbr = ratio of carotenoids on α to β branch

3.3.7 GWAS re-calculated with the allele specific markers of *crtRB1* and *lcyE* included as covariates

GWA was re-calculated by incorporating the allele specific markers of the two genes *lcyE* and *crtRB1* as additional fixed effect covariates in the MLM model. As expected, the number of SNPs significantly associated with any of the traits in this analysis was drastically reduced from 386 of the previous analysis to 38 SNPs, at a cut-off threshold of 5% FDR (Table 3.10). Specifically, numerous SNPs on chromosome 8 and 10 that were previously associated with lutein and β-carotene (plus its derived traits) were now turned statistically non-significant even at a lower significance threshold of 10% FDR (Table 3.10, Figure 3.11). Chromosome 10 was devoid of significant SNPs, and only two SNPs on chromosome 8 (SNP S8_138938949 & S8_138938983; *p*-value = 7.66E-08; R^2 = 0.53) were strongly associated with lutein. These SNPs were also the most significant SNPs in the initial GWAS result. The two SNPs were physically located within a putative gene GRMZM2G463133 encoding an HLH binding domain protein. However, since these SNPs were in high LD with SNPs in the *lcyE* region (Figure 3.9c), it is likely that the significant effect was originating from this known candidate carotenoid gene.

A notably different result in the re-run GWAS was the detection of new significant associations on chromosome 5 for β-carotene and provitamin A. SNPs S5_78384689 and S5_78427240 were associated with provitamin A at 5% FDR (*p*-value = 1.81E-07; R^2 = 68) and one of these SNPs S5_78384689, was associated with β-carotene at 10% FDR (*p*-value= 7.08E-07; R^2 0 = 81). The SNP S5_78384689 lied within a gene known as auxin-response factor 20 (GRMZM2G102845, 5:78381834 – 78389884, Table 3.9). In addition, seven SNPs on chromosome 2 were significantly associated with zeaxanthin content at FDR 1% (Table 3.10), which was also a different but expected association result as most of these SNPs were associated with total carotenoid in the first GWAS. Three of the zeaxanthin associated SNPs (44473748, S2_44473758, and S2_44473801) were located within the gene *ZEP1* and the other two were 23Kb upstream of the gene. In the first GWAS, only a few SNPs on chromosome 10 were associated with

zeaxanthin in regions surrounding *crtRB1* since this gene was strongly affecting the substrates of zeaxanthin.

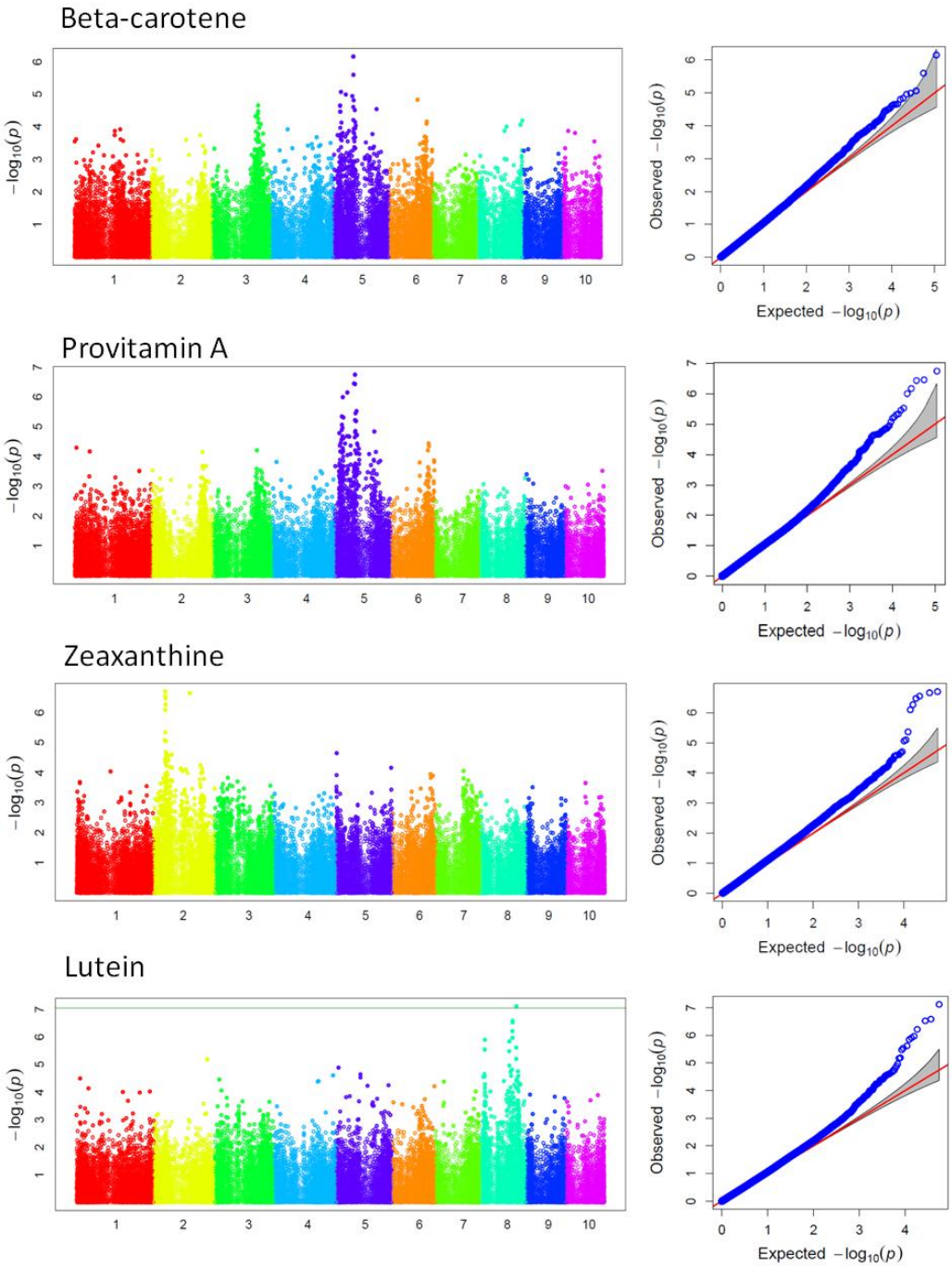


Figure 3.11 Manhattan and QQ-plots for GWAS conducted with allele specific markers of *crtRB1* and *lcyE* genes as covariates.

The association signals were significant after 5% FDR correction of the *p*-values.

Table 3.10 Summary of result for GWAS with allelele specific markers of *lcyE* and *crtRB1* included as covariates

No.	SNP	Chr	Position	Allele	Allelelic effect	*P*-value	MAF	R^2 of Model	R^2 of Model + SNP	FDR adjusted *P*-value	Carotenoid
1	S2_43342654	2	43342654	G/A	3.43	8.21E-07	0.33	0.33	0.48	0.011278	zea
2	S2_44448432	2	44448432	C/T	-3.99	2.89E-07	0.20	0.33	0.49	0.006272	zea
3	S2_44448438	2	44448438	T/G	3.99	2.89E-07	0.20	0.33	0.49	0.006272	zea
4	S2_44473748	2	44473748	T/G	3.86	3.42E-07	0.21	0.33	0.49	0.006272	zea
5	S2_44473758	2	44473758	C/T	-5.01	2.03E-07	0.21	0.33	0.49	0.006272	zea
6	S2_44473801	2	44473801	G/A	5.01	2.03E-07	0.21	0.33	0.49	0.006272	zea
7	S2_45967604	2	45967604	T/G	-0.44	5.55E-07	0.37	0.33	0.48	0.008711	zea
8	S2_139644276	2	139644276	G/A	-3.85	2.27E-07	0.35	0.33	0.49	0.006272	zea
9	S2_36077381	2	36077381	G/A	-0.97	8.27E-07	0.05	0.67	0.74	0.013309	lnβcar/zea
10	S2_43376157	2	43376157	C/T	0.95	3.40E-07	0.06	0.67	0.74	0.010586	lnβcar/zea
11	S2_44474088	2	44474088	G/A	-0.95	8.47E-07	0.05	0.67	0.74	0.013309	lnβcar/zea
12	S2_45967604	2	45967604	T/G	-0.44	2.20E-06	0.37	0.67	0.73	0.026892	lnβcar/zea
13	S2_47044902	2	47044902	A/C	0.82	1.85E-06	0.07	0.67	0.73	0.02546	lnβcar/zea
14	S2_47310378	2	47310378	C/G	0.93	3.85E-07	0.06	0.67	0.74	0.010586	lnβcar/zea
15	S2_47310382	2	47310382	G/C	-0.93	3.85E-07	0.06	0.67	0.74	0.010586	lnβcar/zea
16	S2_103681279	2	103681279	T/C	-0.76	7.17E-07	0.09	0.67	0.74	0.013309	lnβcar/zea
17	S2_109770228	2	109770228	A/C	0.90	3.75E-07	0.07	0.67	0.74	0.010586	lnβcar/zea
18	S7_108535010	7	108535010	A/G	0.62	3.55E-06	0.10	0.67	0.73	0.03903	lnβcar/zea
19	S2_208672678	2	208672678	C/G	2.67	6.67E-06	0.03	0.37	0.48	0.048918	lut
20	S8_16444572	8	16444572	T/C	-1.47	2.83E-06	0.29	0.37	0.49	0.025904	lut
21	S8_16444587	8	16444587	T/A	-1.47	2.83E-06	0.29	0.37	0.49	0.025904	lut
22	S8_16743428	8	16743428	C/T	1.69	1.25E-06	0.21	0.37	0.50	0.017159	lut
23	S8_111289041	8	111289041	C/T	2.00	3.35E-06	0.16	0.37	0.49	0.028289	lut
24	S8_111803908	8	111803908	A/G	1.90	1.45E-06	0.18	0.37	0.50	0.017674	lut
25	S8_124434722	8	124434722	G/T	1.84	6.13E-07	0.25	0.37	0.51	0.011226	lut
26	S8_124434723	8	124434723	C/G	1.79	1.08E-06	0.25	0.37	0.50	0.016903	lut
27	S8_124434725	8	124434725	G/A	-1.74	2.59E-07	0.28	0.37	0.52	0.008199	lut
28	S8_124434726	8	124434726	C/G	1.69	2.98E-07	0.29	0.37	0.52	0.008199	lut
29	S8_124434730	8	124434730	G/C	-1.84	6.13E-07	0.25	0.37	0.51	0.011226	lut
30	S8_138510292	8	138510292	C/G	2.28	2.38E-06	0.15	0.37	0.49	0.025904	lut
31	S8_138938949	8	138938949	C/T	2.82	7.66E-08	0.13	0.37	0.53	0.004208	lut
32	S8_138938983	8	138938983	C/T	2.82	7.66E-08	0.13	0.37	0.53	0.004208	lut
33	S8_138943019	8	138943019	C/T	2.49	6.36E-06	0.12	0.37	0.48	0.048918	lut
34	S2_44473758	2	44473758	C/T	-5.01	2.06E-06	0.21	0.20	0.35	0.075561	tcar
35	S2_44473801	2	44473801	G/A	5.01	2.06E-06	0.21	0.20	0.35	0.075561	tcar
36	S2_139644276	2	139644276	G/A	-3.85	9.62E-07	0.35	0.20	0.36	0.075561	tcar
37	S5_78384689	5	78384689	C/T	1.14	7.08E-07	0.27	0.76	0.81	0.077827	βcar
38	S5_30601081	5	30601081	C/G	1.45	1.03E-06	0.33	0.58	0.67	0.022747	tpva
39	S5_48678892	5	48678892	G/A	-1.56	6.97E-07	0.37	0.58	0.67	0.01915	tpva
40	S5_74462863	5	74462863	C/T	2.05	3.55E-07	0.20	0.58	0.67	0.013758	tpva
41	S5_78384689	5	78384689	C/T	1.14	1.81E-07	0.27	0.58	0.68	0.013758	tpva
42	S5_78427240	5	78427240	G/T	1.53	3.75E-07	0.27	0.58	0.67	0.013758	tpva

βcar = β-carotene, βcryp = β-cryptoxanthin, lut = lutein, zea = zeaxanthin, pva = provitamin A, tcar = total carotenoid, Chr = Chromosome, MAF = minor allele frequency, FDR = false discovery rate. Effects indicated are for the second allelic variants of SNPs.

3.3.8 Pathway-level association analysis

The initial analysis of MLM GWAS result was further augmented by a pathway-level analysis (Wang et al., 2007; Lipka et al., 2013b) which was performed on 42 genes and gene families that are known to be involved in carotenoid biosynthesis. The genes considered included those participating in upstream substrate-supplying steps of metabolic pathways, such as *GGPP1* and *HDR* through the carotenoid biosynthesis that included *Z-ISO*, *lcyb* and *HYD4*, to those steps involved in the catabolism of carotenoids, such as carotenoid carboxylase deoxygenase 1, *ccd1* and 9-cis-epoxycarotenoid dioxygenase 1, *9sisepd* (Vallabhaneni and Wurtzel, 2009; Wurtzel et al., 2012).

The total number of SNPs tested for each carotenoid was 6890, of which 237 were significant at 10% FDR in the pathway level analysis (Table 3.11). The largest percentage of significant SNPs in this analysis was for *crtRB1* (>60%) followed by *lcyE* (which was only 7%). For many of the genes only a single significant SNP was observed, although as many as 300 SNPs were tested within distances of 1Mbp in the flanking regions of the genes (Table 3.11). Positions of the significant SNPs relative to the genes tested ranged from those that resided within the genes (for *lcyE*, *ZEP1*, and *hyd4*) to SNPs that are as far as 1Mbp from the start or end site of the genes under consideration. The closest SNPs to the genes were not necessarily the most significant ones. The exact position of *PSY3* is not known but is estimated to be within a large stretch of genomic region between 143,407,361 and 147,088,644 on chromosome 7, as given in maizeGDB.org database relative to the positions of flanking genes. The position of the significant SNP relative to this gene therefore will depend on where exactly the gene is positioned.

Excluding *crtRB1*, *lcyE* and *ZEP1*, which were also highly significant in the initial GWAS, the pathway level analysis detected four additional carotenoid-associated genes *CCD7*, *hyd4* and *IPPI2* at 1% FDR (Table 3.11). Loosening the significance threshold to 5% revealed six more genes, *CCD1*, *hyd5*, *hyd6*, *PSY3*, ZEP2 and Z-*ISO*; and at 10% FDR eight more genes were spotted: *CRTISO2*, *cyp97a*, *cyp97c*, *DXS1*, *GGPPS4*, *hyd6*, *IPPI1* and *lcyb*. In total, about 20 genes were detected with 224 SNPs that reached the

cutoff threshold of 10% FDR out of the 6,890 total number of SNPs tested in the pathway level analysis. The distance from the most significant SNP (at 5% FDR) to the respective flanked gene ranged from 25 K for *ZEP1* to 838 K for *CCD7*. Although the most significant SNPs were not located within any of the 42 genes tested, there were significant SNPs that were located within *hyd4*, *ZEP1* and *lcyE*.

Table 3.11 Significant SNPs within 1Mbp flanking regions of candidate carotenoid genes, detected in the pathway-level association analysis at 10% FDR

Gene name	Gene description	Coordinate	Nr. of SNPs tested	Nr. of signif. SNPs	Position of the most signif. SNP	P value for the SNP	MAF	R^2 of Model	R^2 of Model + SNP	FDR Adjusted *P*-values	Pathway Adjusted FDR	Carotenoid(s) affected
CCD1	9,10-9,10 carotenoid cleavage dioxygenase 1	9:152086899-152092882	383	1	152,760,351	1.21E-04	0.15	0.16	0.27	5.28E-02	4.84E-02	lut
CCD7	Carotenoid cleavage dioxygenase	2:19458968-19461625	178	3	18,620,719	2.99E-05	0.21	0.38	0.48	4.98E-02	9.00E-03	βcar, βcar/βcryp
CRTISO2	carotene isomerase2	2:226366352-226371341	86	1	226,678,675	3.07E-04	0.08	0.42	0.48	1.22E-01	5.72E-02	βcar
crtRB1	B-carotene hydroxylase 1	10:136057214-136060201	228	153	136,007,578	1.19E-09	0.22	0.42	0.62	6.55E-05	1.00E-04	βcryp, βcar, zea, tcar, pva, βcar/βcryp, βcar/zea, βcar/all
cyp97a	cytochrome P450 13	5:215827224-215831730	307	1	216,160,900	4.85E-04	0.07	0.30	0.37	2.44E-01	8.79E-02	βcar/zea
cyp97c	cytochrome P450 14	1:86838343-86848726	68	1	85,896,637	1.91E-04	0.06	0.22	0.31	2.19E-01	5.26E-02	βcryp
DXS1	Uncharacterized	6:146378393-146382661	140	1	146,893,487	4.10E-04	0.20	0.42	0.48	1.35E-01	6.67E-02	βcar
GGPPS4	geranylgeranyl pyrophosphate synthase 4	5:193656291-193657722	178	1	193,547,691	5.96E-04	0.09	0.38	0.44	2.05E-01	9.13E-02	βcar/βcryp
hyd4	B-carotene hydroxylase 4	2:15865938-15868219	177	14	16,639,170	6.79E-06	0.11	0.38	0.49	2.11E-02	3.10E-03	βcar, βcryp, zea, tcar, βcar/βcryp, βcar/zea, βcar/all
hyd5	B-carotene hydroxylase 5	9:153692212-153694576	281	2	153,734,783	8.15E-05	0.13	0.16	0.27	3.81E-02	3.47E-02	lut
hyd6	B-carotene hydroxylase 6	1:5380196-5382574	233	1	4,547,997	1.32E-04	0.15	0.16	0.26	5.50E-02	4.97E-02	lut
IPPI1	Isopentenyl pyrophosphate isomerase 1	7:155559747-155562921	254	1	155,071,711	3.92E-04	0.07	0.38	0.45	1.73E-01	6.28E-02	βcar/βcryp
IPPI2	isopentenyl pyrophosphate isomerase 2	8:104659886-104663941	96	3	105,087,875	6.11E-07	0.17	0.16	0.34	1.17E-03	5.00E-04	lut
lcyb	Lycopene epsilon βcyclase	5:100700176-100702026	50	1	100,012,347	4.26E-04	0.07	0.42	0.48	1.38E-01	6.67E-02	βcar
lcye	Lycopene epsilon cyclase	8:138886838-138889812	116	18	138,938,949	9.81E-12	0.13	0.16	0.53	5.39E-07	1.00E-04	lut, βbr/α-br
PSY3	phytoene synthase 3	7:143407361-147088644	425	3	143,631,262	9.19E-05	0.42	0.30	0.39	1.06E-01	2.53E-02	βcar/zea, βcar/all
ZEP1	Zeaxanthin epoxidase 1	2:44440299-44449237	171	14	44,473,758	1.78E-07	0.21	0.16	0.36	9.26E-03	6.00E-04	zea, tcar, βcar/zea
ZEP2	Zeaxanthin epoxidase 2	10:120782243-120784775	163	7	120219787	5.04E-05	0.15	0.42	0.50	5.54E-02	1.74E-02	βcar, βcar/βcryp, βcar/zea
Z-ISO	Zeta-carotene isomerase	10:13607878-13611432	55	5	13,337,086	1.24E-04	0.12	0.42	0.49	8.00E-02	2.84E-02	βcar, βcar/βcryp, βcar/zea

βcar = β-carotene, βcryp = β-cryptoxanthin, lut = lutein, zea = zeaxanthin, pva = provitamin A, tcar = total carotenoid βbr/αbr = ratio of carotenoids on α to β branch.

3.4 Discussion

Genome-wide and gene-targeted association studies have been instrumental in the identification of functional nucleotide variants underlying important agronomic and nutritional traits that can be harnessed by crop breeders for developing improved cultivars through marker or genomics assisted selection (Hamblin et al., 2011; Babu et al., 2013). Association mapping using large population and markers sizes is advocated for successful and reliable prediction of LD and association between rare alleles and target phenotypes (Yu and Buckler, 2006; Zhu et al., 2008; Khatkar et al., 2008). However, in GWAS of plants, it has been been common practice to employ panels spanning from very small size diverse population of about, for example, only 90 individuals genotyped with a few hundred markers (Zhang et al., 2012), and a few hundred individuals genotyped with thousands of markers (Li et al., 2013a; Shirasawa et al., 2013) to even thousands of individuals genotyped with millions of markers (Lipka et al., 2013a).

Maize has been a target of candidate gene and genome-wide association studies for a number of agriculturally and nutritionally significant phenotypes (Yu and Buckler, 2006; Yan et al., 2010; Cook et al., 2012; Li et al., 2012, 2013a; Lipka et al., 2013b). In the present GWA study a panel of 130 diverse and partially related inbred lines that were genotyped with high density genome-wide GBS generated SNPs were used. The use of inbred lines combining the genome of both temperate and tropical germplasm allowed capturing of small to large effect carotenoid allelic variants that are present in the two gene pools under IITA's breeding program.

Despite its predisposition to large levels of missing data (Heslot et al., 2013), GBS still provides large number of SNPs with dense coverage and potentially less ascertainment bias - which is ideal for consistent GWAS (Elshire et al., 2011; Crossa et al., 2013). The SNP data set used in the GWAS had acceptable level of missing data of only 10% which was later predicted with conservative imputation criteria in GAPIT genetic analysis software so as to allow reliable genome-wide associations. Minimum minor allele frequency criteria of 1% was used to filter out potential spurious SNPs stemming from

sequencing error (Glaubitz et al., 2014) and at the same time to make sure that possible rare alleles were not lost (Brachi et al., 2011).

Frequency of minor alleles is a critical factor that can affect the accuracy of LD analysis and GWAS especially when using small samples (Yan et al., 2009; Tabangin et al., 2009). The filtered data set had large proportion of minor allele frequencies distributed uniformly across the genome, with minor allele frequencies between 1% and 5% accounting for the largest proportion. However, the minor allele frequency of the vast majority of the significant SNPs were above 10% which might be indicative of the positive detection power of the GWAS as the biasness associated with rare alleles when using small sized samples for association mapping was eliminated (Schnable et al., 2009). Besides, it implies that the alleles of the carotenoids might be segregating in the panel at frequencies higher than 10%, similar to the alleles of the SNPs they are associated with (Hamblin et al., 2011).

The two commonly used measures of LD are Lewontin's D and the squared pairwise correlation coefficient R^2 (Flint-Garcia et al., 2003; Chen et al., 2006). Although D' is a good measure of recombination history, it is severely affected with reduced sample size. R^2 summarizes both recombination and mutation history (Flint-Garcia et al., 2003). In the present study LD was estimated using R^2, since it helps detect LD with minimal error despite small sample size and low MAF (Yan et al., 2009; Khatkar et al., 2008). In addition R^2 is a more relevant measure of LD for conducting association analysis between genotype and traits (Flint-Garcia et al., 2003). The average genome-wide LD in the present study, which was estimated at about 830Kbp at a background critical R^2 = 0.2, was considered a little extensive. Previous LD studies in maize reported less than 1000 bp for land races, more than 2000 bp for diverse lines and about 100 kb for commercial elite inbred lines (Yu and Buckler, 2006). Although it can lack the power for high precision mapping, a mapping panel with persistent LD can be considered ideal, if low-resolution mapping is targeted (Flint-Garcia et al., 2003). The long range LD in this panel was expected since extensive LD is characteristics of advanced maize inbred lines that have experienced strong recent selection (Yu and Buckler, 2006). Besides, small size population are prone to genetic drift leading to loss of rare alleles and

increased LD (Flint-Garcia et al., 2003). There was, however, considerable localized variegation in LD structure across the genome suggesting that the mapping resolution also could vary accordingly. The relatively average slow LD decay in this study can imply that the significant SNPs detected might be either proxies of functional genes situated as far as thousands of kilo bases within the LD decay limits of their genomic locations; or they might also be located inside or very close to the genes relevant to the carotenoid biosynthesis, and thus could have a causative role or at least paves the way for the identification of those loci with causative roles.

Two MLM GWAS models plus a pathway level association analysis detected a number of small to large effect known carotenoid biosynthesis genes, as well as several putative genes encoding characterized or uncharacterized proteins. The first MLM GWAS considered population structure (Q) and relative kinship (K) only, while the second incorporated the allele specific markers of *lcyE* and *crtRB1* major carotenoid genes as additional fixed effect covariates.

The vast majority of highly significant hits in the first GWAS were on chromosome 8 associated with lutein, followed by chromosome 10 associated with β-carotene and the ratio of β-carotene to β-cryptoxanthin. These significant SNPs aggregated in the regions where the gene *lcyE* was located, on chromosome 8: 138,882,594 – 138,889,812; and *crtRB1* on chromosome 10:136,057,100 – 136,060,219 (Tables 3.7 and 3.11; Figure 3.8a and b). The large effects of these genes in the panel were expected, as the markers designed to detect the allelic variants of these genes (Harjes et al., 2008; Yan et al., 2010) were already confirmed to have significant impact in the same mapping panel (Azmach et al., 2013), indicating the successful introgression of the favourable alleles of these two genes into the tropical yellow maize genetic background.

Although the most significant SNP (p = 9.81E-12; Table 3.7) on chromosome 8 was located 49 Kb downstream off the *lcyE* gene, another significant SNP (SNP: S8_138888278, p = 3.19E-8) was detected within the gene about 1kb upstream of the 3'indel functional polymorphism that was reported by Harjes et al. (2008). Similarly, despite the fact that no SNP was found within the gene *crtRB1*, the closest SNPs

(S10_136007575 and S10_136007578) significantly associated with β-carotene to β-cryptoxanthin ratio (*p*-value = 6.75E-10) and other ratio involving β-carotene were located only 50 kb upstream of the gene.

The high level of LD among all the significant SNPs at FDR1% (average R^2 = 0.4, $p<0.001$) across the genome can be suggestive of the possible presence of spurious associations especially on chromosome 8 and 10, which was probably caused by inter or intra chromosomal long distance LD (Flint-Garcia et al., 2003).

The inclusion of the allele specific marker information for *lcyE* and *crtRB1* as additional fixed effect covariate allowed to control the large effects of the two genes clearing the complex associations detected on chromosome 8 and 10. Only 10% of the significant SNPs detected at 5% FDR in the first GWAS survived the correction for the allele specific markers. The two most significant SNPs on chromosome 8 still displayed strong association with lutein. The strong association of these two SNPs with lutein, despite accounting for the effect of allele specific markers of *lcyE*, could be due to a stronger LD of the SNPs with larger effect functional polymorphisms in the 3'TE untranslated region of *lcyE* not captured with the present genotyping and different from the ones described by Harjes et al. (2008) This can provide an explanation for the relatively low effect of the allele specific markers of *lcyE* in the marker validation study while the strongest association was detected in this region in the present GWA study.

A major outcome of controlling the effects of *lcyE* and *crtRB1* in the second GWAS was the detection of significant association for zeaxanthin on chromosome 2 at 1% FDR. The significant SNPs co-localized with a downstream carotenoid biosynthesis gene *ZEP1* (chromosome 2: 44,440,299 – 44,449,237). These SNPs were detected in the first MLM GWAS but in that case they were significantly associated with total carotenoid at 5% FDR. Zeaxanthin was then affected with significant SNPs only from the complex association signal on chromosome 10. This suggests that the rate of up or down regulation of conversion of β-carotene to zeaxanthin through β-cryptoxanthin was more pronounced than that of depletion of zeaxanthin to violaxanthin by *ZEP1* in the association panel, verifying the stronger impact of *crtRB1* on the level of zeaxanthin

than *ZEP 1* in the association panel. Which means that the reduced function of *crtRB1* lead to accumulation of β-carotene at the expense of zeaxanthin synthesis reflecting the larger effect of *crtRB1* on the concentration of zeaxanthin.

A recent association study also identified a SNP within the gene *ZEP1* on chromosome 2 significantly associated with zeaxanthin content (Suwarno, 2012). In addition a small effect QTL underlying kernel colour close to the gene *ZEP1* was reported and suggested as a target for allele mining by Chandler et al. (2013). This locus can therefore be considered as one of the loci potentially contributing to the variation in total carotenoid in the mapping panel used in this study and can be the next target gene for allele mining. Stacking of alleles of this and other genes in the biosynthesis pathway can increase accumulation of provitamin A and total carotenoid in maize endosperm.

The other interesting putative association that was uncovered in the second GWAS was the association between SNPs on chromosome 5 and provitamin A at 5% FDR and β-carotene at 10% FDR. These SNPs were co-localized with a gene encoding a protein called auxin-response factor 20 (arf20) (5:78381834 – 78389884). Auxin-response factors are known to have a role in conditioning carotenoid biosynthesis through coordinated regulation of transcription of genes involved in the pathway (Meier et al., 2011). Considering that the gene is also highly expressed in maize endosperm (Sekhon et al., 2011), it can be a potential target for further study towards unraveling the regulatory mechanism of carotenoid biosynthesis in maize endosperm.

The pathway level association analysis used to complement the GWAS showed the possible contribution of several isoprenoid biosynthesis genes to the variation of carotenoid composition and content observed in the mapping panel. Apart from the genes detected in the initial GWAS, *crtRB1*, *lcyE* and *ZEP*1, four additional isoprenoid biosynthesis genes, *CCD7*, *crF*, *hyd4* and *IPPI2*, may have contribution to the variation in carotenoid composition and content of the panel at 1% FDR after adjusting for pathway level analysis (Table 3.11).

A recent GWA study reported by (Lipka et al., 2013a), which employed thousands of maize inbred lines genotyped with high density markers, suggested a simple genetic

architecture for maize endosperm carotenoid biosynthesis. Our study, on the other hand, demonstrated that many small to medium effect loci are involved in the biosynthesis of carotenoids in maize endosperm.

3.5 Conclusions

The association study successfully detected SNPs co-localizing with known major and small effect carotenoid biosynthesis genes, demonstrating the accuracy of the data and its analyses. In addition, a number of putative associations were detected for novel candidate genes encoding transcription factors, which might have roles in regulation of the carotenoid biosynthesis in maize endosperm. This study paves the way for additional allele mining efforts and greater understanding of the genes involved in regulation of expression of carotenoid biosynthesis genes, which is necessary to further exploit the genetic potential of maize in accumulating provitamin A in the endosperm.

Chapter 4: Combining ability and heterosis of endosperm carotenoids and agronomic traits among yellow maize inbred lines developed for Sub-Saharan Africa

4.0 Abstract

Vitamin A deficiency is a serious public health problem in Sub-Sahara Africa which can potentially be reduced through maize provitamin A biofortification. Twenty four diverse yellow maize inbred lines recycled from maize populations having both temperate and tropical germplasm in their genetic background were evaluated for their combining ability in carotenoid content and agronomic traits. The inbred lines were inter-crossed in a factorial mating scheme, generating 80 hybrids that were evaluated across different environments in Nigeria. The study demonstrated that set, female and male parents and hybrids had significant effects on all carotenoids and agronomic traits measured. Environment and Genotype by Environment (GxE) interactions were also significant. Cross-over type of interaction between hybrids and environment was detected for agronomic traits but not for carotenoid content, indicating stability of carotenoid content across environments. Partitioning of the variance associated with the carotenoid and agronomic traits of the hybrids into its components revealed the predominance of general combining ability (GCA) effects. Significant positive and negative estimates of GCA effects for the inbred lines, and specific combining ability (SCA) effects for the crosses were detected for carotenoids and agronomic traits. Outstanding hybrids that had high levels of provitamin A and good agronomic performances were identified in the study. Hybrids that showed high parent heterosis for provitamin A content and good agronomic performance were also identified. The functional markers for the key provitamin A gene *crtRB1* showed correlation with hybrid provitamin A levels. Hence, classical breeding supported by marker assisted selection can potentially facilitate the development of high provitamin A maize hybrids.

4.1 Introduction

Global maize biofortification efforts have led to the development of maize inbred lines with high level of endosperm provitamin A (Babu et al., 2013; Pixley et al., 2013; Azmach et al., 2013). To effectively exploit the genetic potential of such inbred lines for

increasing the level of provitamin A in the endosperm of commercial maize cultivars, the mode of inheritance of carotenoid profiles and contents needs to be assessed. Breeders commonly achieve this by determining the general (GCA) and specific combining abilities (SCA) of inbred lines, both of which are genetic parameters useful for planning effective strategies to meet a particular breeding objectives. These parameters are inferred from field experiments that involve mating schemes which allow disaggregation of sources of variations into GCA and SCA effects (Singh et al., 2012). General combining ability (GCA) refers to the average performance of an inbred line as a male or female parent in its hybrid progenies measured as a deviation from the overall mean, while specific combining ability (SCA) refers to the deviation of a hybrid from its expected performance predicted on the basis of the GCA of its parents (Hallauer et al., 2010). GCA generally represents additive and additive by additive gene actions, while SCA emphasizes the non-additive (dominance or epistasis) components of gene actions.

Heritability estimate is another genetic parameter that is useful in planning selection and breeding strategies for improvement of populations and development of inbred lines and hybrids (Holland et al., 2003; Hallauer, 2007; Hallauer and Carena, 2009). Heritability estimates depend on the target trait, the population, the environment and testing situations, and can also change over time if significant changes occur in allele frequencies, requiring its estimation for every new factor (Acquaah, 2012; Hallauer and Carena, 2009).

Hybrid performance relies on the mode of gene action and heritability of the target trait which are governed by the extent and nature of the underpinning genetic variation. The diverse genetic variation (in maize endosperm carotenoid content and composition) of IITA's yellow maize germplasm (Menkir et al., 2008; Azmach et al., 2013) can serve as a tool for understanding the mode of inheritance of carotenoids in this germplasm. This is an essential step in planning and implementing efficient provitamin A maize breeding at IITA.

The carotenoid content in maize kernels has been a subject of inheritance studies since the early 20th century (Hauge and Trost, 1928, 1930; Hauge, 1930; Ford, 2000).

However, there have been few extensive studies to date that focus on combining ability of maize for the various endosperm carotenoids when compared to yield and yield components. Grogan and colleagues (1963), using F1 seeds of reciprocal crosses of 10 maize inbred lines, showed the importance of additive genetic effects and significant influences of both female and male parents on xanthophylls and carotenes. They also detected non additive gene action on xanthophylls and positive effects of increased yellow maize gene dosage on endosperm carotenoid content. Egesel et al. (2003a) found significant maternal effects for all carotenoids except for β-carotene. In another study, Egesel et al., (2003b) analyzed the combining abilities and breeding potential of 10 maize inbred lines in 45 diallele crosses for their carotenoid contents and found significant effects of GCA and SCA, with the GCA contributing the majority of the variations in the pigments. Using a complete diallel analysis of 8 inbred lines for β-carotene, Run et al. (2013) found significant GCA but non-significant SCA effects. Senete et al. (2011) detected significant GCA effect for all carotenoids they measured but found significant SCA effect only for lutein. Suwarno et al. (2014) also found significant GCA and weak or non-significant SCA effects for carotenoid content among hybrids generated using 21 factorially mated inbred lines. These studies demonstrated the importance of additive gene action in the inheritance of maize endosperm carotenoid content but were not consistent with respect to the effect of SCA. At the molecular level the significance of additive genetic effects for the major genes PSY1 lcyE and crtRB1 have been demonstrated (Yan et al., 2010; Fu et al., 2013). Considering the inconsistent effects of SCA and the relatively low and less variable concentrations of β-carotene and β-cryptoxanthin in parental lines used in some studies (e.g. Egesel et al., 2003b), additional studies that involve a broad range of carotenoid composition and content combined with diverse genetic backgrounds can help elucidate the mode of inheritance of these nutrients.

Heterosis is a phenomenon in which a hybrid outperforms its parents in terms of yield, size, vigor or other agronomic characters (Birchler et al., 2010; Hallauer et al., 2010). Heterosis can be inferred from high SCA estimates of crosses (Hallauer et al., 2010). The presence of heterosis for carotenoid content in maize endosperm has been demonstrated (Alfieri et al., 2014; Burt et al., 2011a). This is important because maize is

known for its strong manifestation of heterosis for grain yield and other important agronomic traits, and finding crosses that combine heterosis for both agronomic traits and carotenoids can have important economic and health benefits.

High provitamin A containing hybrids should also have high yield potential and desirable agronomic traits to be attractive to seed producer and farmers. It is therefore necessary to study the relationship between agronomic performance and carotenoid content in maize endosperm to determine the potential for development of maize hybrids that are both productive and nutritious.

Use of molecular markers associated with high provitamin A content of maize endosperm can facilitate the development of hybrids with high levels of provitamin A. Allele specific markers associated with endosperm provitamin A content of maize in inbred lines have been reported by Harjes et al. (2008) and Yan et al. (2010). The association of these markers with provitamin content of inbred lines has also been validated (Babu et al., 2013; Azmach et al., 2013). Yan et al. (2010) described a high provitamin A hybrid that also carried the most favourable alleles of *crtRB1*. Alfieri and colleagues (2014) reported the inconsistent correlation between markers originally reported by (Vallabhaneni et al., 2009). But, these particular markers had been shown to be not associated with provitamin A in larger size population (Yan et al., 2010). Thus, it is important to determine the effects of parental allelic compositions of the markers reported by Yan et al. (2010) and Harjes et al. (2008) in hybrids of different genetic backgrounds.

The objectives of this study were to (1) determine the mode of inheritance of endosperm carotenoids together with important agronomic traits, (2) assess the heterotic effects of inbred lines for carotenoid content, and (3) investigate the effects of the best diagnostic functional markers in hybrid combinations.

4.2 Materials and Methods

4.2.1 Plant materials and crossing scheme

Twenty four yellow maize inbred lines (Table 4.1) developed from tropical and temperate germplasm were crossed in a factorial mating scheme (Hallauer et al., 2010) at IITA's main research station, Ibadan (7°29′11.99″N, 3°54′2.88″E, altitude 190 m) during two dry seasons (from December 2010 to April 2011 and December 2011 to April 2012). Five sets of crossings were made by first grouping parental inbred lines into six different groups each containing four inbred lines. Except for two groups, inbred lines 13 to 16 which served only as female parents and 21 to 24 which served only as males in the entire crossing scheme, all the lines served as both male in one mating set and female parent in another set in a particular mating set (Table 4.1). The mating groups were crosses using the following arrangement: Set 1: G-I x G-II, Set 2: G-III x G-I, Set 3: G-II x G-IV, Set 4: G-V x G-III, and Set 5: G-VI x G-V. The criteria for grouping the inbred lines into these different sets of crossings were generally based on similarities in pedigree and carotenoid profile determined in previous carotenoid analyses. Thus, each set of crosses involved inbred lines with different genetic backgrounds to maximize combining ability and heterosis. This crossing scheme resulted in 80 single cross hybrids.

Table 4.1 Parental inbred lines used in a factorial mating scheme

Inbred line pedigree	Line/ tester No.	Group[1]	Parentage[2]	PVA [3]
SYN-Y-STR-34-1-1-1-1-2-1-B*3/(DE3/CI7)/SYN-Y-STR-34-1-1-1-1-2-1-B*11	1	I	F M	10.14
4205/CI7/4205-6-B*6	2	I	F M	8.39
KU1409/SC55/KU1409-4-B*4	3	I	F M	8.65
SC55/KU1414-SR/KU1414-SR-6-B*6	4	I	F M	8.34
(9450/KI 21-1-4-1-1-2-B/DE3/9450/KI 21-1-4-1-1-2-B) -29-B*3	5	II	F M	14.91
9450/KI21-3-2-2-1-3/KU1409/MO17LPA/KU1409-27-3-1-1-6-B*7	6	II	F M	15.72
9450/KI 21-1-5-3-2-2-B/DE3/9450/KI 21-1-5-3-2-2-B -32-B*4	7	II	F M	12.24
9450/KI 21-1-5-3-2-2-B/DE3/9450/KI 21-1-5-3-2-2-B -23-B*4	8	II	F M	11.31
(9450/KI 21-1-4-1-1-1-B/DE3/9450/KI 21-1-4-1-1-1-B) -33-B*3	9	III	F M	8.59
(9450 / KI 28)-1-2-1-1-B/DE3/(9450 / KI 28)-1-2-1-1-B -40-B*3	10	III	F M	12.72
(9450/KI 21-1-4-1-1-2-B/DE3/9450/KI 21-1-4-1-1-2-B) -30-B*3	11	III	F M	10.80
9450/KI 21-1-4-1-1-1-B/DE3/9450/KI 21-1-4-1-1-1-B -26-B*4	12	III	F M	10.26
(ACR97TZL-CCOMP1-Y-S3-13-1-B-B/CI7/ACR97TZL-CCOMP1-Y-S3-13-1-B-B)-13-B*3	13	IV	M	9.77
(KU1409/SC55/KU1409)-S2-19-1-B*4	14	IV	M	7.64
(POP66SR/ACR91SUWAN1-SRC1/ACR91SUWAN1-SRC1-6/(MP420/4001/MP420)-3-1-3-1-B)S2-5-B*6	15	IV	M	8.70
(9071 / 4058)-8-2-1-1-B*6	16	IV	M	7.87
(KU1414-SR/CI7/KU1414-SR) -49-B*3	17	V	F M	2.56
(KU1414-SR/CI7/KU1414-SR) -35-B*3	18	V	F M	3.97
(POP66SR/ACR91SUWAN1-SRC1/ACR91SUWAN1-SRC1-9 / (9450/CM 116/9450)-3-3-1-2-1)S2-5-B*4	19	V	F M	2.44
(POP66SR/ACR91SUWAN1-SRC1/ACR91SUWAN1-SRC1-4 / 4001/KI21-4-1-1-1-1)S2-3-B*4	20	V	F M	3.35
(9450/KI 21-1-4-1-1-1-B/DE3/9450/KI 21-1-4-1-1-1-B) -29-B*3	21	VI	F	1.99
(9450 / KI 28)-1-2-1-1-B/DE3/(9450 / KI 28)-1-2-1-1-B -27-B*3	22	VI	F	2.17
(9450 / KI 28)-1-2-1-1-B/DE3/(9450 / KI 28)-1-2-1-1-B -22-B*3	23	VI	F	2.91
9450/KI 21-1-5-3-2-1-B/DE3/9450/KI 21-1-5-3-2-1-B -27-B*4	24	VI	F	2.58
4001	Check			--
KU1414-SR	Check			--

[1] Group number referring to groups each containing four sets of lines

[2] Used either as F female (line) or M, male (tester) in different sets of crosses. Inbred lines 13 to 16 were used only as males, and inbred lines 21 to 24 were used only as females in the whole crossing scheme.

[3] PVA = provitamin A content data from preliminary carotenoid analyses.

4.2.2 Field trial

The 80 hybrids along with a commercial hybrid check were arranged in a 9 x 9 lattice design and planted at four testing sites in Nigeria in 2012 and at two testing sites in 2013. The testing sites were Ikenne $3^0$42′E, $6^0$54′N, altitude 30 m; Saminaka $8^0$39′E, $10^0$34′N, altitude 760 m; Bagauda $8^0$19′E, $12^0$01′N, altitude 520 m; and Zaria $7^0$45′E, $11^0$8′N, altitude 622 m. The parental inbred lines were also tested along with the hybrid trial separately in a 4 x 7 alpha lattice design at Zaria and Saminaka in both 2012 and 2013. Both hybrid and inbred line trials were planted in two replications and each entry was planted in a 5 m long single row plot with plant spacing of 75 cm between rows and 25 cm within a row. Two seeds per hill were planted but later thinned out to maintain one plant. The trials received all the recommended agronomic management practices at the respective testing sites.

4.2.3 Data recording

All the relevant data were recorded on a plot basis. Flowering dates were recorded by taking number of days at which 50% of the plants within a plot have set visible silks (as silking date) and started shedding sufficient pollens (as anthesis date). Ear and plant heights were measured in centimetres (cm) from the base of the plant to attachment line of the lower ear and the first tassel branch, respectively. Plant and ear aspects were scored on a 1 to 5 scale. Where, for ear aspect, 1 represented clean, well filled, uniform and larger ears, while 5 represented diseased, poorly filled, variable and smaller ears. For plant aspect 1 represented uniform, clean, vigorous and good overall phenotypic appeal, while 5 represented weak, diseased and poor overall phenotypic appeal. Shelled grain weight of each hybrid was measured in kilograms (kg) per plot and moisture content of the representative sample grain was measured for each shelled plot using a hand held moisture tester. The shelled grain weight per plot in kilograms was converted to grain yield in metric tons per hectare (t/ha) by first adjusting for 12.5 moisture content. Grain yield was not recorded for the inbred line trial, since several plants in each plot were self pollinated which affected seed setting. Five to six representative plants of each hybrid and all typical plants of each inbred lines planted at Zaria and Saminaka were self pollinated to produce seed samples for carotenoid analysis. Shelled

seeds from self pollinated ears were dried further at ambient temperature with minimal exposure to direct sunlight. Samples of about 100 seeds were drawn at random from each line or hybrid and sent to the University of Wisconsin for carotenoid analysis that used the method of Howe and Tanumihardjo (2006). Details of the procedures for carotenoid analysis were described in Chapter 2.2.3. Alpha-carotene, β-carotene (*cis* and *trans* isomers), β-cryptoxanthin, lutein, and zeaxanthin were quantified in µg/g dry weight (DW). Total carotenoid was calculated as the sum of concentrations of α-carotene, lutein, β-carotene, β-cryptoxanthin, zeaxanthin. Provitamin A was calculated by adding the concentrations of β-carotene, and half of each of β-cryptoxanthin and α-carotene concentrations, since β-cryptoxanthin and α-carotene concentrations can provide only one molecule of retinol each as opposed to β-carotene which can be converted to two molecules of retinols (US Institute of Medicine, 2001).

4.2.4 Molecular marker assessment

Genomic DNA samples were isolated from each inbred line. DNA samples were genotyped using the functional markers of the genes *lcyE* and *crtRB1*, which are described in detail in Chapter 2 of this PhD dissertation. Favourable or unfavourable alleles of the most significant functional markers (3'TE and 5'TE of *crtRB1*) detected in parental lines were analysed for their effect on concentration and composition of provitamin A carotenoids in hybrids. The genotypes of each hybrid were resolved for the functional markers of the two genes based on the marker data of their parental lines.

4.2.5 Statistical Analysis

Each location-year combination was considered an environment. PROC MIXED procedure of SAS® software (SAS Institute, 2012) was used to generate least square means for each of the carotenoids and agronomic traits on a single and across environments bases after correcting for block effects. For traits with significant genotype by environment interaction, Spearman's rank correlation coefficients were calculated between environments to determine whether the significant interactions of hybrids and parental inbred lines with the environment were of cross-over types. Combined analyses of variances (ANOVA) were conducted for each trait to obtain

mean squares for females, males, and females by males interactions, including their interactions with the environments. In the mixed model ANOVA, environments, replications within environment and blocks within replication were considered as random effects while mating sets, inbred lines (females and males) and hybrids were considered as fixed effects. General combining ability (GCA) effects were inferred from variances of the main effects of males and females, while specific combining ability (SCA) effects were inferred from that of female by male interactions (Hallauer et al., 2010).

Family means based heritability (H) estimates and the associated standard errors were calculated based on the method and code for SAS software provided by (Holland et al., 2003), assuming that the genotypes were random samples representing a reference population.

$$H = \frac{V(G)}{V(G) + \frac{V(GE)}{e} + \frac{V(E)}{er}}$$

Where, V(G) is variance due to genetic effect, V(GE) is variance due to genotype by environment interaction, V(E) is variance due to random errors, e is number of environments r is number of replications per environment

Simple correlation coefficients were also calculated among carotenoids and agronomic traits. Correlation analyses were performed using PROC CORR of SAS software (SAS Institute, 2012).

Mid and high parent heterosis and their significance were calculated as follow, according to the formula indicated in Hallauer et al. (2010) and Khorzoght et al. (2010).

$$mpH = \frac{F_1 - mp}{mp} x100, \qquad bpH = \frac{F_1 - bp}{bp} x100$$

Where mpH is mid-parent heterosis, bpH is best parent heterosis, F_1 is hybrid value, mp is an average value of the corresponding parents for the hybrid.

Significance of each hybrid value with respect to the corresponding mp and bp values was evaluated using *t* test (Khorzoght et al., 2010):

$$\text{Tmp} = \frac{\text{F}_1 - \text{mp}}{\sqrt{\frac{3}{2\text{r}}\text{EMS}}}, \qquad \text{Tbp} = \frac{\text{F}_1 - \text{bp}}{\sqrt{\frac{2}{\text{r}}\text{EMS}}}$$

Where Tmp is calculated t value for F_1 to mp comparison and Tbp is for that of F_1 to bp comparison, r is number of repeats. The probability associated with the t-test was calculated using Excel®.

The effects of the 3'TE and 5'TE markers of *crtRB1* (the most significantly associated diagnostic markers) on the provitamin A carotenoid content of the hybrids were assessed using mixed model analysis with PROC MIXED procedure of SAS software (SAS Institute, 2012).

4.3 Results

4.3.1 Endosperm carotenoid content in parents and hybrids

The mean carotenoid concentrations averaged over environments are presented in Table 4.2 for the inbred line trial and Appendix 1 – Table 7 for the hybrid trial. The inbred lines displayed significant variability in carotenoid content (Table 4.3). There were highly significant effects of the environment on concentrations of all carotenoid, but the effects were smaller compared to that of the variance among inbred lines. Inbred line by environment interaction was significant for α-carotene, β-cryptoxanthin, β-carotene and Provitamin A. Hybrids, environments, and hybrids by environment interaction had significant effects on concentrations of all carotenoids, except for zeaxanthin that showed a non-significant hybrid by environment interaction (Table 4.4). The effects associated with blocking were significant but had weak impact on carotenoids. Genetic variation among hybrids represented the largest proportion of the total variance for all the traits, followed by variation due to environments and hybrid by environment interaction. Spearman rank correlation analysis showed that the pairwise correlations between pairs of environments were significant ($p<0.0001$) for each carotenoid indicating that the environment by hybrid interaction was not of cross-over type.

High heritability estimates on the basis of family means were recorded for all carotenoids, ranging from 57 % for α-carotene to 95 % for lutein in the parent trial, and 57% for α-carotene to 92 % for β-cryptoxanthin in the hybrid trial. The heritability estimates for provitamin A in both inbred line and hybrid trials were 81% (Tables 4.3 and 4.4).

Table 4.2 Carotenoid concentrations of parental inbred lines averaged across four environments

Line/tester	Lutein	Zeaxanthin	α-carotene	β-crptoxanthin	β -carotene	Provitamin A	Total carotenoid	Marker[1]
1	2.43±1.12	10.78±1.29	0.37±0.07	1.75±0.44	3.92±0.68	4.96±0.73	19.1±1.74	3, 1
2	2.37±1.12	11.47±1.29	0.22±0.07	2.53±0.44	2.58±0.68	3.95±0.73	19.25±1.74	3, 1
3	2.86±1.12	14.00±1.29	0.25±0.07	2.71±0.44	2.48±0.68	3.96±0.73	22.26±1.74	3, 1
4	9.09±1.15	4.98±1.34	0.61±0.07	0.87±0.46	3.67±0.69	4.41±0.74	19.1±1.82	1, 2
5	2.4±1.12	2.89±1.29	0.41±0.07	0.60±0.44	6.15±0.68	6.65±0.73	12.13±1.74	1, 1
6	3.71±1.12	15.07±1.29	0.25±0.07	2.91±0.44	2.78±0.68	4.37±0.73	24.63±1.74	3, 1
7	5.10±1.12	3.21±1.29	0.43±0.07	1.14±0.44	4.63±0.68	5.43±0.73	14.74±1.74	1, 2
8	5.91±1.15	3.20±1.34	0.4±0.07	0.75±0.46	3.35±0.69	3.92±0.74	13.71±1.82	1, 2
9	6.04±1.12	9.14±1.29	0.32±0.07	3.59±0.44	1.8±0.68	3.73±0.73	20.93±1.74	3, 1
10	4.44±1.22	11.54±1.41	0.3±0.08	2.71±0.48	2.44±0.75	3.96±0.79	21.63±1.94	3, 1
11	9.14±1.15	13.07±1.34	0.44±0.07	2.46±0.46	2.40±0.69	3.86±0.74	27.52±1.82	3, 1
12	4.37±1.12	15.08±1.29	0.33±0.07	4.80±0.44	3.21±0.68	5.78±0.73	27.62±1.74	3, 1
13	7.10±1.12	16.31±1.29	0.38±0.07	2.47±0.44	3.39±0.68	4.79±0.73	29.5±1.74	3, 1
14	4.14±1.12	17.4±1.29	0.22±0.07	1.97±0.44	3.19±0.68	4.31±0.73	27.1±1.74	3, 1
15	4.50±1.12	12.2±1.29	0.4±0.07	3.03±0.44	2.67±0.68	4.39±0.73	22.88±1.74	3, 1
16	4.89±1.12	15.1±1.29	0.21±0.07	1.96±0.44	1.71±0.68	2.80±0.73	23.91±1.74	3, 1
17	2.79±1.15	9.65±1.34	0.11±0.07	1.05±0.46	1.10±0.69	1.73±0.74	14.74±1.82	3, 1
18	4.67±1.12	11.22±1.29	0.15±0.07	1.36±0.44	1.22±0.68	1.97±0.73	18.7±1.74	3, 1
19	5.71±1.12	8.18±1.29	0.2±0.07	1.05±0.44	1.24±0.68	1.88±0.73	16.34±1.74	3, 1
20	4.69±1.12	9.32±1.29	0.33±0.07	2.40±0.44	2.81±0.68	4.18±0.73	19.63±1.74	3, 1
21	13.06±1.12	7.07±1.29	0.26±0.07	0.94±0.44	1.19±0.68	1.78±0.73	22.49±1.74	3, 2
22	3.13±1.12	6.25±1.29	0.22±0.07	2.08±0.44	2.07±0.68	3.23±0.73	13.82±1.74	3, 1
23	19.19±1.12	6.97±1.29	0.43±0.07	1.04±0.44	2.18±0.68	2.9±0.73	29.74±1.74	3, 1
24	3.24±1.19	5.50±1.39	0.13±0.07	1.55±0.47	0.81±0.7	1.68±0.75	11.56±1.91	3, 1
Check1	5.77±1.12	12.77±1.29	0.235±0.07	2.485±0.44	1.71±0.68	3.075±0.73	22.915±1.74	3, 1
Check2	4.21±1.12	12.9±1.29	0.13±0.07	1.855±0.44	1.725±0.68	2.72±0.73	20.915±1.74	3, 1
Grand Mean	5.53±1.13	10.39±1.31	0.29±0.07	2.01±0.45	2.49±0.69	3.65±0.73	20.74±1.77	
SD	3.56	4.16	0.12	0.96	1.18	1.27	5.10	
Range	2.37-19.19	2.89-17.4	0.11-0.61	0.6-4.8	0.81-6.15	1.68-6.65	11.56-29.74	
LSD($p<0.05$)[2]	2.38	2.82	0.16	0.98	1.52	1.46	1.60	

[1] Haplotypes of the parents for the favorable/unfavorable alleles of the most significantly associated polymorphisms, 3'TE and 5'TE, of the gene *crtRB1* as determined in chapter 2. Allele 3 and 1 for 3'TE represent unfavorable and favorable alleles, respectively 2 and 1 represented favorable and unfavorable alleles for the 5'TE polymorphism.
[3]average LSD at $p<0.05$ presented

Table 4.3 Combined ANOVA for carotenoid content of 24 parents and 4 checks evaluated across four environments

Source	DF	Lutein	Zeaxanthin	β-cryptoxanthin	α-carotene	β-carotene	Provitamin A	Total carotenoid
Environment = Envt	3	116.69**	155.44**	0.28**	14.89**	39.55**	55.33**	210.74**
Replication=Rep(Envt)	4	7.35	7.96	0.02	0.5	0.4	0.56	32.72*
Block(Envt*Rep)	24	4.23	10.08*	0.01	0.96	0.35	0.97*	25.25**
Inbred line	27	87.36**	117.32**	0.09**	5.87**	9.35**	10.75**	165.47**
Envt*Inbred line	80	5.19	7.29	0.02**	0.9*	2.04**	1.9**	16.27
Error	77	3.59	5.72	0.01	0.6	0.22	0.54	12.08
Corrected Total SS	215	3922.05	5331.5	7.05	394.31	625.74	749.18	8957.56
R^2 %		0.93	0.92	0.88	0.88	0.97	0.94	0.90
CV%		33.89	22.70	38.51	36.10	18.92	20.12	16.61
Grand mean		5.59	10.53	2.02	0.29	2.49	3.65	20.92
Heritability (H)*		0.95	0.91	0.92	0.57	0.79	0.81	0.81
Standard error for H		0.02	0.02	0.08	0.01	0.04	0.04	0.04

*Heritability on a family mean basis for each trait was estimated using all random model for multi-environment lattice design, with a SAS software code suggested by Holland et al. (2003).

* significant at P < 0.05, ** significant at *p*<0.01

Table 4.4 Summary of combined ANOVA for carotenoid contents of **80 progeny hybrids** plus one commercial hybrid check evaluated across environments

Source	DF	Lutein	Zeaxanthin	α-carotene	β-cryptoxanthin	β-carotene	Provitamin A	Total carotenoid
Environment = Envt	3	367.84**	413**	1.04**	26.72**	132.64**	183.63**	1317.76**
Replication = Rep(Envt)	4	3	4.41	0.03	0.2	0.28	0.39	9.51
block(Envt*rep)	64	3.33*	11.9**	0.02**	0.31**	0.75**	1.21**	27.12**
Hybrid	80	22.72**	62.51**	0.05**	3.61**	5.98**	7.15**	78.25**
Envt*Hybrid	240	3.23**	6.35	0.02**	0.34**	1.41**	1.63**	15.54*
Set	4	56.11**	274.94**	0.31**	41.11**	100.07**	119.4**	366.52**
Envt*set	12	7.63**	6.36	0.06**	1.53**	18.39**	18.82**	39.19**
Female(set)	15	72.1**	144.23**	0.08**	5.01**	5.47**	7.95**	145.02**
Male(set)	15	37.91**	131.4**	0.07**	7.5**	5.4**	7.17**	146.04**
Female*Male(set)	45	5.31**	6.46	0.01	0.41**	0.64	0.98*	17.06
Envt*Female(set)	45	3.67*	7.05	0.04**	0.44**	0.91**	1.04*	15.79
Envt*Male(set)	45	4.27**	6.27	0.03**	0.51**	1.53**	1.96**	19.1
Envt*Female*Male(set)	135	3.14	7.29	0.02*	0.2	0.39	0.61	18.2
Error	294	2.23	5.37	0.01	0.16	0.37	0.54	11.95
Corrected total SS	633	4910.28	10573.79	18.29	600.57	1549.38	1985.17	19835.85
R^2 %		0.85	0.82	0.78	0.91	0.91	0.90	0.78
CV %		27.33	24.74	36.40	20.46	22.20	19.34	18.41
Grand mean		5.79	10.41	0.32	2.14	3.06	4.29	21.06
GCAf %		57	49	40	36	43	44	42
GCAm %		30	44	37	55	42	40	43
SCA %		13	7	23	9	15	16	15
Heritability (H±SE)[1]		0.87±0.02	0.91±0.02	0.92±0.01	0.57±0.08	0.79±0.08	0.81±0.04	0.81±0.04

[1]Heritability for each trait was estimated using all random model on family mean bases across environments. A SAS software code suggested by Holland et al. (2003) was used.

GCAm % = GCA male proportion, GCAf % = GCA female proportion

* Significant at P < 0.05, ** significant at *p*<0.01

As illustrated in Figure 4.1 the average proportions of the carotenoids in the inbred line and hybrid trials were uniform. Average concentrations of the carotenoids for the inbred line trial were 5.53 µg/g lutein, 10.29 µg/g zeaxanthin, 2.01 µg/g β-cryptoxanthin, and 2.49 µg/g β-carotene. For the hybrid trial the average concentrations were 5.79 µg/g for lutein 10.47 µg/g for zeaxanthin, 2.14 µg/g for β-cryptoxanthin and 3.04 µg/g for β-carotene. The highest average concentration was recorded for Zeaxanthin followed by lutein. Figure 4.2 shows the distribution of the carotenoids across the hybrids. Mean carotenoid concentrations averaged across environments in the inbred line trial varied from 1.68 to 6.65 µg/g for provitamin A and from 11.56 to 29.74 µg/g for total carotenoids. While in the hybrid trial they varied from 1.44 to 6.53 µg/g for provitamin A and 13.66 to 28.66 µg/g for total carotenoids. The checks in the parental inbred line trial (female and male parents of Oba super 2) accumulated mean concentration of 3.08 µg/g provitamin A and 22.92 µg/g total carotenoids for the female parent, and 2.72 µg/g provitamin A and 22.92 total carotenoid for the male parent.

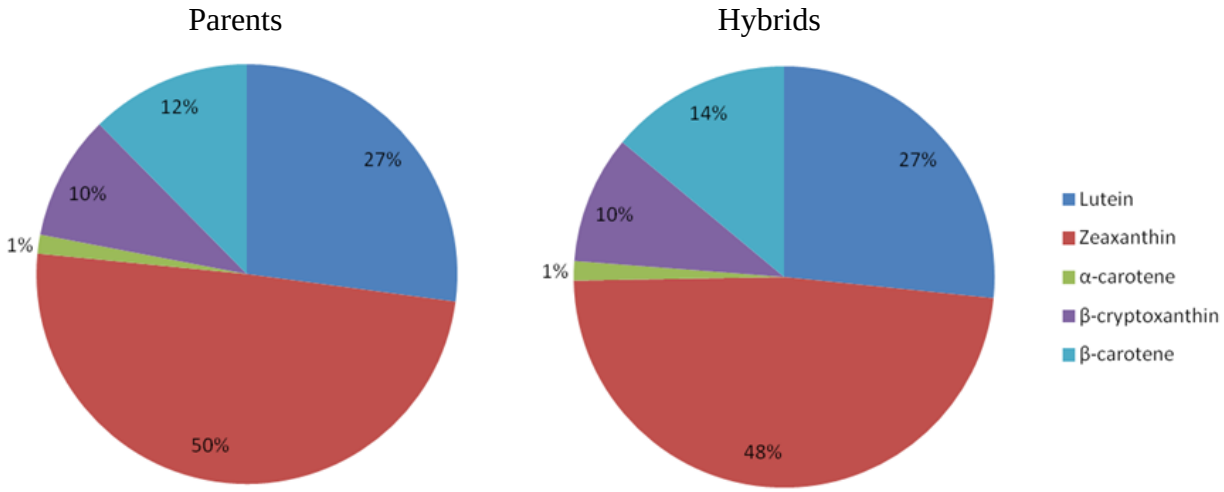


Figure 4.1 Proportions of the carotenoids measured for the inbred line and hybrid progeny trials averaged across environments

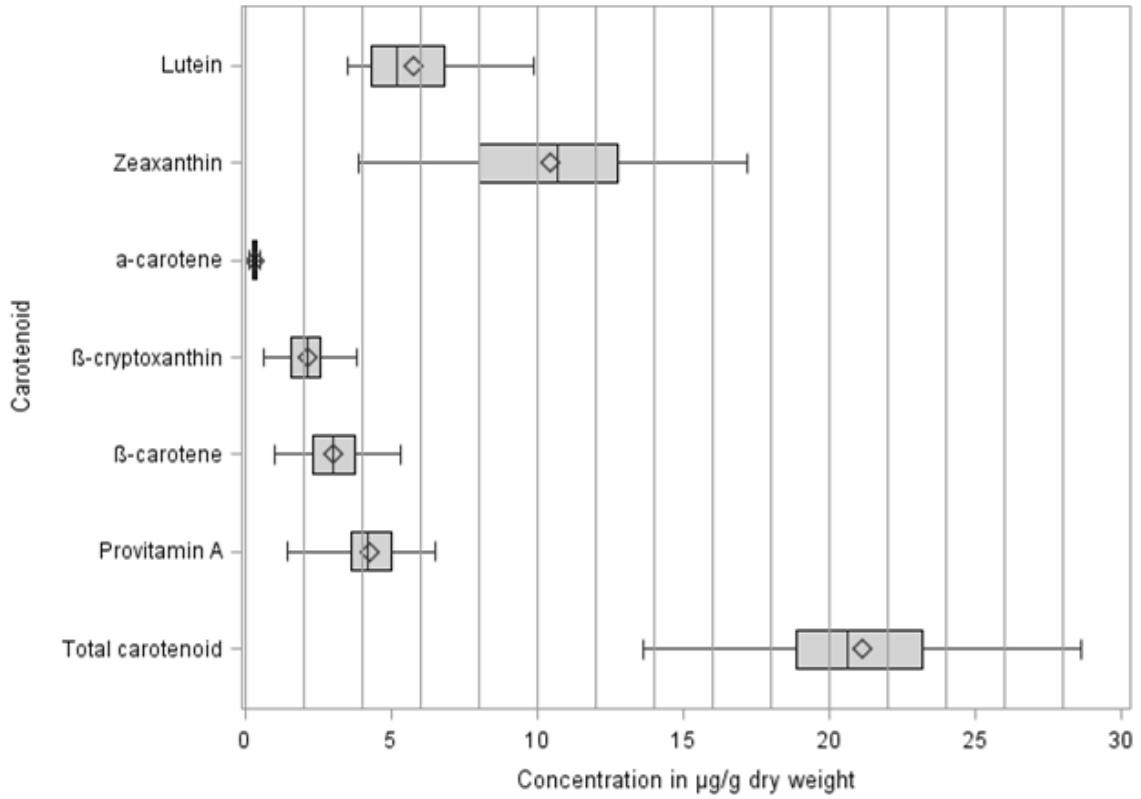


Figure 4.2 Distribution of carotenoid content among 81 hybrids (80 progenies plus a commercial check)

4.3.2 Combining ability for carotenoids

Mating sets had highly significant effects on all carotenoids (Table 4.4). As shown in Table 4.9, hybrids of set 2 (GII x GIV) showed the highest mean concentration for zeaxanthin (8.58 µg/g), β-cryptoxanthin (2.76 µg/g) and total carotenoids (23.54 µg/g) (Table 4.9). On the other hand, the mean concentrations of hybrids in set 1 (G-I x G-II) was the highest for β-carotene (4.16 µg/g) and provitamin A (5.17 µg/g). Hybrids in set 5 (G V x GVI) had the lowest provitamin A carotenoid concentrations of all the sets, except for lutein which was the highest (6.92 µg/g).

Further breaking down of the variability due to hybrids within the mating sets revealed that the females and males main GCA effects and their interaction with the environment were highly significant for most of the carotenoids. The female by male interaction (SCA) had significant effects on lutein, β-cryptoxanthin and provitamin A. However the contribution of SCA to the total sums of squares for hybrids (GCA female + GCA males

+ SCA) was very small ranging from 7 to 16%. The interaction of SCA with the environment was not significant for almost all the carotenoids.

4.3.3 GCA and SCA effects for carotenoids

The inbred lines showed significant ($P<$ 0.01 and 0.05) positive and negative GCA effects on concentrations of specific and multiple carotenoids (Table 4.7). The pattern of GCA effects generally followed the group assignments of the inbred lines in the mating sets. Inbred lines assigned to groups I (inbred lines 1 to 4) had significant positive male and female GCA effects for β-carotene and total provitamin A carotenoids, except for inbred line 4 which showed non-significant male GCA effect for the same carotenoids. These inbred lines generally showed negative male and female GCA effects for lutein, zeaxanthin and β-cryptoxanthin.

Inbred lines in group II (inbred lines 5 to 8) had positive female GCA effect for β-cryptoxanthin, and positive female and male GCA effects for β-carotene and total provitamin A carotenoids, except inbred line 5 which showed negative and significant female GCA effect for β-carotene and provitamin A. The inbred lines in this group also showed large positive female GCA effects for zeaxanthin and total carotenoids. However, inbred line 6 had positive GCA effects for all carotenoids excluding lutein.

Inbred lines of group III (inbred lines 9 to 12) had variable GCA effects with inbred lines 9 and 10 showing positive female GCA effects for zeaxanthin, β-cryptoxanthin and total carotenoids, 9 and 12 showing positive female GCA effects for β-carotene and provitamin A, and inbred lines 10, 11and 12 showing positive male GCA effects for zeaxanthin and β-cryptoxanthin. Inbred lines of group IV (inbred lines 13 to 16) showed positive male GCA effects for all carotenoids except lutein. However, inbred line 14 in this group had negative GCA effect for all carotenoids except for zeaxanthin and β-cryptoxanthin.

The Group V inbred lines showed negative GCA effects for most of the carotenoids. But all the inbred lines in this group had positive male GCA effect for lutein, positive female GCA effects for β-cryptoxanthin, and two inbred lines had positive female GCA effect

for zeaxanthin and total carotenoid. Similarly, group VI inbred lines (inbred lines 21 to 24), showed negative GCA effects for all carotenoids, except inbred lines 21 and 23 which showed large positive female GCA effect on lutein with inbred line 24 showing large positive GCA effect for zeaxanthin.

Overall, the results demonstrated that Group VI inbred lines as females and group V inbred lines as males were poor combiners for all carotenoids, except for lutein. Likewise, group V inbred lines as females and group III inbred lines as males were poor combiners for carotenoid content, but not for β-cryptoxanthin. Inbred lines of group I and II were generally good combiners both as female and male parents, especially for β-carotene and total provitamin A carotenoids.

A number of crosses showed significant (P<0.01 and 0.05) negative and positive SCA effects on carotenoids (Table 4.8). All the crosses formed in set 1 (GI x GII) had significant positive SCA effects on β-cryptoxanthin and many of them had large positive SCA effects on zeaxanthin. These crosses, on the other hand, displayed significant negative SCA effects on β-carotene and provitamin A. Apart from the negative SCA effects of set 2 crosses (G III x G I) for β-cryptoxanthin, the SCA effects for most of the crosses in sets 2 and 3 were not significant. Crosses in sets 4 and 5 (G V x G III and G VI x G V) displayed positive SCA effects for individual and total provitamin A carotenoids, except four crosses that showed negative SCA effects for β-cryptoxanthin.

The largest GCA effect on beta-carotene and provitamin A was recorded for inbred line 5 as a male. This inbred line accumulated the highest level of provitamin A (6.65 μg/g) but had one of the lowest total carotenoid content (12.13 μg/g). The largest GCA effect on zeaxanthin and total carotenoids were recorded for inbred line 6 as both female and male, and for inbred lines 7, 8, 9, 10 as female parents. The total carotenoid contents of these inbred lines ranged from 13.7 μg/g for inbred line 8, to 24.63 μg/g for inbred line 6.

4.3.4 Agronomic traits of parents and hybrids

Both the inbred line and hybrid trials exhibited significant differences in all agronomic traits measured (Tables 4.5 and 4.6). The effect of environment and its interaction with the inbred lines and hybrids were significant for all of the agronomic characters except for EH and PH among the inbred lines and EH among the hybrids. The means for agronomic traits of the hybrids are shown in Appendix 1 – Table 7. Hybrid means averaged across environments ranged from 56 to 63 days for days to anthesis (AD), from 0.66 to 5.87 t/ha for grain yield (GY), from 2.3 to 4.36 for ear aspect (EA), and from 2.45 to 4.34 for plant aspect (PA). Three hybrids, hybrid 24 (12 x 3), 32 (12 x 1) and 52 (20 x 12) had the highest grain yield (5.7 to 5.9 t/ha) which were not significantly different from about half of the hybrids evaluated, including the commercial check (Table 4.17, Appendix 1 – Table 7). The ear aspect and plant aspect of these top yielding hybrids were below 3.0, and their provitamin A contents were higher than 5 μg/g, which were not significantly different from the highest provitamin A concentration in the hybrid trial (6.5 μg/g).

The variances due to the female and male main effects were highly significant on all agronomic traits. Their interactions with the environment were also significant but not for EH and PH. The effects of female by male interaction (SCA effects), and its interaction with the environment were significant for most of the traits. The rank correlation coefficients were non-significant for all agronomic characters recorded, suggesting the presence of changes in ranking of the genotypes across environments. The proportion of female GCA effects relative to the total variation for the hybrids ranged from 26% for plant height to 59% for SD, and the proportion for male GCA effect ranged from 27% for AD to 43% for EA. The SCA component of variance ranged from 14% for SD to 35% for PA which were higher than the SCA component of variance observed for carotenoids, suggesting the greater importance of non-additive gene effects for the agronomic traits than for the endosperm carotenoid content.

Table 4.5 Mean squares for agronomic traits of the 24 parental inbred lines and 4 checks evaluated across environments

Source	AD	SD	ASI	EH	PH	HC	PA	EA
Environment = Envt	246.13**	268.97**	18.48**	2488.65**	10497.57**	20.37**	4.48**	9.63**
Replication(Envt)	14.28**	24.21**	2.54	308.86*	307.01	0.96**	0.25	0.1
Block(Envt*Rep)	5.33**	7.12*	2.05*	156.42*	398.78**	0.15	0.21	0.3
Inbred line	58.13**	86.01**	6.63**	371.41**	880.55**	0.53**	0.93**	0.63**
Envt*Inbred line	4.71**	7.3**	2.24**	72.21	174.43	0.33**	0.37**	0.28
Error	210.04	310.23	1.14	93.55	127.13	0.17	0.19	0.29
Corrected Total SS	3555.55	4769.83	562.92	39265.28	104487.73	133.84	101.25	80.99
R^2 %	0.94		0.84	0.80	0.90	0.89	0.84	0.78
CV %	2.43		42.07	16.20	8.42	22.61	13.66	15.33
Grand mean	64.96		2.54	59.72	133.88	1.85	3.19	3.49
Heritability (H)*	0.92±0.02	0.92±0.02	0.69±0.1	0.83±0.05	0.79±0.06	0.37±0.2	0.64±0.12	0.54±0.19

AD = no. of days to anthesis, SD = no. of days to silking ASI = anthesis silking interval (days to 50% silking minus days to 50% male flowering), EH = Ear height, PH = plant height, PA = plant aspect and EA = ear aspect

* significant at P < 0.05, ** significant at *p*<0.01

Table 4.6 Summary of ANOVA for agronomic traits of 80 hybrids plus one commercial hybrid check evaluated across five to six environments

Source	DF	SD	AD	PH	EH	GY	EA	PA
Environment = Envt	5	2045.94**	1958.78**	77377.7**	38520.3**	1157.84**	16.18**	9.08**
Replication = Rep (Envt)	6	13.09**	12.05**	926.49**	398.34**	4.39**	0.35	1.19**
Block (Envt*rep)	96	7.59**	4.84**	706.71**	354.61**	1.52**	0.4**	0.37**
Hybrid	80	30.9**	23.71**	1371.78**	889.44**	4.63**	1.2**	1.16**
Envt*Hybrid	400	3.08**	1.82**	229.64**	105.67	1.05**	0.35**	0.35**
Set	4	66.36**	68.64**	1917.41**	1939.31**	4.39**	2.83**	1.42**
Envt*set	20	5.69*	2.72	287.73	72.98	2.02**	0.87**	0.61**
Female (set)	15	108.67**	78.26**	2380.92**	2133.19**	7.07**	2.27**	2.1**
Male (set)	15	50.25**	39.13**	3765.76**	2426.06**	10.32**	1.02**	2.04**
Female*Male (set)	45	8.92**	6.18**	1009.79**	356.99**	2.91**	0.86**	0.75**
Envt*Female (set)	75	7.65**	3.4**	358.59	174.13	1.47**	0.62**	0.57**
Envt*Male (set)	75	4.5*	3*	343.62	181.58	1.93**	0.51**	0.63**
Envt*Female*Male (set)	225	2.74	1.96	265.75	113.88	0.99*	0.31	0.31**
Error	450	3.09	2.12	276.61	144.84	0.74	0.26	0.24
Corrected total SS	959	13663.71	11681.12	777988.00	406561.60	5767.09	479.28	440.68
Grand Mean		60.30	58.53	192.83	92.00	4.57	3.03	3.09
R^2 %		91.59	93.20	84.04	84.00	95.20	75.32	75.89
CV %		2.91	2.49	8.62	13.09	18.87	16.89	15.70
GCAf %		59	50	26	38	26	27	33
GCAm %		27	27	41	43	40	40	32
SCA %		14	23	33	19	34	33	35
Heritability (H±SE)		90±2	92±1	81±3	87±2	76±4	62±6	61±6

[1]Heritability for each trait was estimated using all random model for multi-environment lattice design, with a SAS software code suggested by Holland et al., 2003.

* significant at P < 0.05, ** significant at *p*<0.01

GCAm% = GCA male proportion, GCAf% = GCA female proportion, CV% = coefficient of variation in %, R^2 % = percentage of variation explained SE = Standard error, DF = degree of freedom, SD = no. of days to silking, AD = no. of days to anthesis, PH = plant height (in cm), EH = ear height (in cm), GY = grain yield (in t/ha), EA = ear aspect (1 to 5), PA = plant aspect (1 to 5).

Table 4.7 Estimates of general combining ability (GCA) effects for endosperm carotenoid contents in 24 parental inbred lines that were used to form 80 hybrids

		Lutein		Zeaxanthin		β-cryptoxanthin		β-carotene		Provitamin A		Total carotenoid	
Parent[1]	**Group[2]**	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**
1	I	-0.82**	-0.25	-2.07**	-0.75**	-0.55**	-0.06	1.1**	0.44**	1.1**	0.44**	-2.19**	-0.61
2	I	-1.63**	-0.95**	-2.56**	1.71**	-0.43**	-0.46**	0.72**	0.33**	0.72**	0.33**	-3.81**	0.74*
3	I	-0.78**	-0.97**	0.84**	-0.2	0.08	0.62**	0.95**	0.6**	0.95**	0.6**	0.98**	-0.05
4	I	1.67**	-0.81**	-3.6**	1**	-1.04**	0.03	1.67**	0.02	1.67**	0.02	-0.9**	0.17
5	II	0.44**	-1.16**	-0.63**	-2.2**	0.25**	-0.63**	-0.32**	1.99**	-0.32**	1.99**	-0.33	-1.82**
6	II	-0.61**	-0.65**	2.36**	3.77**	0.49**	0.31**	0.42**	0.59**	0.42**	0.59**	2.58**	3.94**
7	II	1.36**	0.09	2.04**	-4.14**	0.69**	-0.98**	0.54**	0.97**	0.54**	0.97**	4.42**	-3.77**
8	II	-0.63**	0.16	2.67**	-4.83**	1.05**	-0.63**	0.42**	0.9**	0.42**	0.9**	3.21**	-4.26**
9	III	-1.14**	-0.04	1.65**	-1.16**	0.17**	0.1	1.11**	-1.12**	1.11**	-1.12**	1.66**	-2.34**
10	III	-1.29**	-1.09**	4.41**	1.65**	0.65**	0.67**	-0.08	-0.5**	-0.08	-0.5**	3.57**	0.46
11	III	-0.85**	1.87**	-1.58**	1.91**	-0.49**	0.28**	0.13	-0.48**	0.13	-0.48**	-2.65**	3.62**
12	III	0.28	-1.19**	-2.72**	1.7**	-0.18**	0.9**	0.23**	-0.46**	0.23**	-0.46**	-2.33**	0.62
13	IV		-0.62**		2.24**		0.71**		0.45**		0.45**		2.61**
14	IV		-1.3**		0.48*		0.69**		-0.24*		-0.24*		-0.61
15	IV		-0.34*		3.22**		1.4**		0.23*		0.23		4.09**
16	IV		2.82**		0.52*		-0.33**		0.62**		0.62**		3.77**
17	V	-0.9**	0.67**	1.68**	-0.05	0.38**	-0.64**	-0.83**	-1.11**	-0.83**	-1.11**	0.06	-0.95*
18	V	0.44**	1.13**	2.98**	-0.38	0.49**	-0.61**	-0.7**	-1.21**	-0.7**	-1.21**	3.04**	-0.79*
19	V	-0.22	1.45**	-0.27	-2.64**	0.26**	-0.91**	-0.91**	-1.53**	-0.91**	-1.53**	-1.22**	-3.39**
20	V	0.22	1.2**	-0.29	-1.83**	0.82**	-0.46**	-0.12	-0.48**	-0.12	-0.48**	0.49	-1.45**
21	VI	3.56**		-3.36**		-1.2**		-1.49**		-1.49**		-2.21**	
22	VI	-1.28**		-0.93**		-0.3**		-0.7**		-0.7**		-3.04**	
23	VI	3.38**		-1.63**		-0.7**		-0.76**		-0.76**		0.65*	
24	VI	-1.21**		1.02**		-0.42**		-1.38**		-1.38**		-1.98**	
SE		0.15	0.16	0.20	0.19	0.20	0.19	0.05	0.06	0.07	0.10	0.08	0.11
GM		5.80	5.80	10.43	10.43	2.14	2.14	3.05	3.05	3.05	3.05	21.08	21.08

[1]Inbred lines13 to 16 and 21 to 24 served only as either males or females in the entire mating scheme. [2]Mating scheme indicated in materials and methods

* significant at P < 0.05, ** significant at *p*<0.01 SE = Standard error, GM = Grand mean

Table 4.8 Estimates of specific combining ability (SCA) effects for carotenoid contents in 80 progeny hybrids formed using 24 inbred lines mated in a factorial mating scheme

Hybrid	Set	Female	Male	Lut	Zea	βcrp	βcar	PVA	Tcar
1	1	1	8	-0.91	2.78**	0.48**	-1**	-1**	1.21
2	1	2	8	0.99	0.74	0.36*	-1.38**	-1.38**	0.52
3	1	3	8	1.46*	1.3	0.68**	-0.53*	-0.53*	2.86*
4	1	4	8	0.02	2.59**	0.42**	-1.54**	-1.54**	1.31
5	1	1	7	0	1.76	0.51**	-1.25**	-1.25**	0.8
6	1	2	7	0.02	2.12*	0.52**	-1.04**	-1.04**	1.55
7	1	3	7	0.28	1.45	0.34*	-1.18**	-1.18**	0.8
8	1	4	7	1.26*	2.07*	0.57**	-0.98**	-0.98**	2.76
9	1	1	6	1.14	0.77	0.43**	-1.01**	-1.01**	1.17
10	1	2	6	0.02	2.56**	0.52**	-1.21**	-1.21**	1.8
11	1	3	6	0.02	2.1*	0.42**	-1.74**	-1.74**	0.64
12	1	4	6	0.37	1.97*	0.57**	-0.5*	-0.5	2.3
13	1	1	5	1.33*	2.09*	0.52**	-1.2**	-1.2**	2.73
14	1	2	5	0.52	1.99*	0.54**	-0.82**	-0.82**	2.04
15	1	3	5	-0.2	2.55**	0.5**	-1**	-1**	1.61
16	1	4	5	-0.09	0.78	0.38*	-1.44**	-1.44**	-0.46
17	3	9	4	1.25*	-0.59	0.05	-0.43*	-0.43	0.16
18	3	10	4	1.49*	-0.9	-0.27	-0.58**	-0.58*	-0.11
19	3	11	4	1.34*	0.03	0.04	-0.32	-0.32	1.12
20	3	12	4	-1.1	-0.3	0.05	-0.06	-0.06	-1.41
21	3	9	3	0.58	-0.62	0.26	-0.12	-0.12	0.13
22	3	10	3	1.1	-0.21	0.17	-0.2	-0.2	0.76
23	3	11	3	0.24	-0.98	-0.36*	-0.83**	-0.83**	-1.77
24	3	12	3	1.06	0.04	-0.21	-0.24	-0.24	0.63
25	3	9	2	-0.13	-0.32	-0.35*	-0.35	-0.35	-1.11
26	3	10	2	0.22	0.62	0	-0.35	-0.35	0.53
27	3	11	2	0.72	-0.57	0.17	-0.25	-0.25	0.03
28	3	12	2	2.17**	-1.48	0.04	-0.45*	-0.45	0.3
29	3	9	1	1.28*	-0.22	-0.09	-0.5*	-0.5	0.58
30	3	10	1	0.17	-1.27	-0.04	-0.26	-0.26	-1.43
31	3	11	1	0.68	-0.24	0.01	0.01	0.01	0.37
32	3	12	1	0.86	-0.02	-0.02	-0.64**	-0.64*	0.23
33	2	5	16	-0.98	-1.48	-0.22	-0.42*	-0.42	-2.94*
34	2	6	16	-0.15	-2.67**	-0.64**	-0.16	-0.16	-3.5*
35	2	7	16	-0.12	-2.21*	-0.95**	-0.55**	-0.55*	-3.68*
36	2	8	16	0.69	-0.09	-0.67**	0.06	0.06	0.25
37	2	5	15	0.62	-0.99	-0.7**	-0.24	-0.24	-1.18
38	2	6	15	-0.96	-1.36	-0.5**	-0.37	-0.37	-3.08*
39	2	7	15	0.37	-1.3	-0.5**	0	0	-1.18
40	2	8	15	-0.59	-2.79**	-0.77**	-0.45*	-0.45	-4.43**
41	2	5	14	0.26	-1.07	-0.9**	0.04	0.04	-1.57
42	2	6	14	0.06	-1.34	-0.61**	-0.45*	-0.45	-2.05
43	2	7	14	-0.64	-1.73	-0.43**	-0.35	-0.35	-3*
44	2	8	14	-0.24	-2.3*	-0.54**	-0.3	-0.3	-3.25*
45	2	5	13	-0.47	-2.91**	-0.66**	-0.45*	-0.45	-4.19**
46	2	6	13	0.49	-1.07	-0.72**	-0.08	-0.08	-1.25
47	2	7	13	-0.17	-1.2	-0.6**	-0.16	-0.16	-2
48	2	8	13	-0.41	-1.27	-0.5**	-0.37	-0.37	-2.43
49	4	17	12	0.29	-1.03	-0.98**	0.5*	0.5	-0.97
50	4	18	12	-0.01	-1.38	-0.23	0.6**	0.6*	-0.99
51	4	19	12	0.33	-0.98	-0.37*	0.58**	0.58*	-0.25

Hybrid	Set	Female	Male	Lut	Zea	βcrp	βcar	PVA	Tcar
52	4	20	12	-0.15	-0.71	-0.36*	0.89**	0.89**	-0.16
53	4	17	11	-0.75	-2.51**	-0.65**	0.55**	0.55*	-2.92*
54	4	18	11	1.51*	-1.05	-0.58**	0.51*	0.51*	0.53
55	4	19	11	-0.67	-1.07	-0.46**	0.65**	0.65*	-1.54
56	4	20	11	0.37	0.52	-0.26	0.86**	0.86**	1.55
57	4	17	10	-0.17	0.25	0.24	0.65**	0.65*	0.63
58	4	18	10	-0.46	-0.36	-0.7**	0.7**	0.7**	-0.52
59	4	19	10	0.78	-1.2	-0.66**	0.75**	0.75**	-0.11
60	4	20	10	0.32	-2.81**	-0.83**	0.47*	0.47	-2.38
61	4	17	9	1.09	-0.82	-0.56**	0.87**	0.87**	0.88
62	4	18	9	-0.58	-1.32	-0.43**	0.76**	0.76**	-1.4
63	4	19	9	0.03	-0.85	-0.46**	0.59**	0.59*	-0.48
64	4	20	9	-0.08	-1.11	-0.5**	0.34	0.34	-1.38
65	5	21	20	-1.16	2.04*	0.7**	1.12**	1.12**	2.54
66	5	22	20	-1.42*	0.47	0.58**	0.62**	0.62*	-0.02
67	5	23	20	-1.5*	1.37	0.49**	1.17**	1.17**	1.37
68	5	24	20	-0.38**	1.02	0.86**	1.42**	1.42**	2.68
69	5	21	19	-1.03	1.28	0.6**	0.94**	0.94**	1.63
70	5	22	19	-0.12	2.16*	0.67**	1.33**	1.33**	3.84**
71	5	23	19	-1.56**	0.46	0.71**	1.02**	1.02**	0.39
72	5	24	19	-1.73**	1.01	0.64**	1.05**	1.05**	0.72
73	5	21	18	-0.91	0.74	0.64**	1.1**	1.1**	1.39
74	5	22	18	-1.29*	1.01	0.73**	1.19**	1.19**	1.45
75	5	23	18	-0.65	1.69	0.63**	1**	1**	2.42
76	5	24	18	-1.59**	1.46	0.61**	1.04**	1.04**	1.32
77	5	21	17	-1.34*	0.84	0.68**	1.17**	1.17**	1.02
78	5	22	17	-1.62**	1.26	0.64**	1.19**	1.19**	1.31
79	5	23	17	-0.74	1.39	0.79**	1.15**	1.15**	2.4
80	5	24	17	-0.75	1.42	0.51**	0.82**	0.82**	1.86
			SE	0.59	0.91	0.15	0.21	0.26	1.43

Table 4.9 Mean carotenoid concentrations (in μg/g) for the five mating sets

Set	Lutein	Zeaxanthin	α-carotene	β-cryptoxanthin	β-carotene	provitamin A	Total carotenoid
1. G-I x G-II	5.41_{B}	8.58_{C}	0.37_{A}	1.66_{C}	4.16_{A}	5.17_{A}	19.6_{B}
2. G-III x G-I	5.94_{AB}	12.04_{A}	0.37_{A}	2.76_{A}	3.4_{AB}	4.88_{A}	23.54_{A}
3. G-II X x-IV	5.06_{B}	10.87_{AB}	0.32_{AB}	2.18_{B}	3.32_{AB}	4.65_{A}	21.14_{B}
4. G-V x G-III	5.69_{B}	11.45_{A}	0.29_{AB}	2.63_{AB}	2.41_{B}	3.87_{AB}	21.67_{AB}
5. G-VI x G-V	6.92_{A}	9.2_{BC}	0.27_{B}	1.49_{C}	1.97_{B}	2.85_{B}	19.43_{B}
SE	0.81	0.89	0.04	0.23	0.57	0.64	1.53

Carotenoid means within a column followed by similar letter subscript were not significantly different at *P*< 0.01

4.3.5 GCA and SCA effects for agronomic traits

The inbred lines displayed significant (p<0.01 and 0.05) negative and positive GCA effects for all the agronomic traits (Table 4.10). Inbred lines with positive GCA effects were considered good combiners for GY, whereas inbred lines having negative GCA effects were considered good combiners for PA and EA. For AD and PH, both signs of the GCA effects can be desirable depending on early or late maturing, and taller or shorter cultivars are required. Inbred line 12 had positive GCA effects for GY and negative GCA effects for PA and EA as male and female, thus making it the best combiner for the three important agronomic traits. In addition, inbred lines 19 and 20 had significant positive GCA effects as females and males for GY. Inbred lines 2, 5, 7, 13 and inbred lines 6, 21, 22, 23, displayed positive GCA effects for GY as males and females, respectively. Inbred lines 4 and 7 had negative female and male GCA effects for PA. Inbred lines 6 and 21 had negative GCA effects as females, whereas inbred lines 15 and 16 had negative GCA effects as males for the same trait. Inbred lines 6, 11, 20 and 21 displayed negative GCA effects on EA when used as females, and inbred lines 2, 3, and 18 displayed negative GCA effects on EA when used as males. The majority of the inbred line showed negative female and male GCA effects for anthesis date (AD) and plant height (PH). But, inbred lines 2, 3, 5, 13, 17, and 18 had positive GCA effects for plant height as males. Inbred lines 4 and 6 showed positive GCA effects as males for AD and as female for PH. Inbred line 10, 22 and 23 had large positive GCA effects for both AD and PH.

Several hybrids showed significant SCA effects for important agronomic traits (Table 4.11). Hybrids generated in set 1 and 5 displayed most of the significant SCA effects for agronomic traits considered. Hybrids of set 1 had both positive and negative SCA effects on all the traits whereas hybrids in set 5 had only negative SCA effects for SD and AD dates. Hybrids 7 (3 x 7) and 12 (4 x 6) were attractive hybrids with large positive SCA effects for PH and GY, negative SCA effects for PA, EA, SD and AD.

The common parent of the three highest yielding hybrids 24, 32 and 55 was inbred line 12, which had positive GCA effect for provitamin A when crossed as a female parent (Table 4.7) and for grain yield when crossed as a female and male parent (Table, 4.10).

It also showed negative female and male GCA effects for PA and EA which is desirable. The other non-common parents of the high yielding hybrids were also good combiners for provitamin A except inbred line 20. Hybrid 38 (6 x 16) had the best EA and PA. The SCA estimates for all the carotenoids of this hybrid were negative (Table 4.8), but its two parents (inbred line 5 and 6) were good combiners for individual and total provitamin A carotenoids (Table 4.7).

Table 4.10 Estimates of general combining ability (GCA) effects for agronomic traits in 80 hybrids produced through factorial mating of 24 inbred lines

Parent[1]	Group	AD		PH		GY		PA		EA	
		Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
1	I	-0.55**	-0.81**	-4.93**	0.82	-0.39**	0.16	0.36**	0.13*	0.29**	-0.09
2	I	-0.96**	-0.6**	2.13	8.76**	-0.08	0.33**	0.24**	0.07	0.09	-0.09*
3	I	0.32*	-0.64**	-9.12**	10.53**	-0.72**	0.18	0.22**	0.01	0.27**	-0.2**
4	I	-0.35*	0.46**	6.51**	1.15	0.03	-0.46**	-0.32**	-0.11*	0.12*	0.04
5	II	-0.98**	-1.58**	-9.26**	3.68**	0.06	0.32**	0.02	0.03	0.25**	0.18**
6	II	-0.75**	1.38**	9.41**	-9.07**	0.57**	-1.82**	-0.19**	0.57**	-0.49**	0.58**
7	II	-1.7**	-1.72**	-8.74**	0.36	0	0.29**	-0.24**	-0.23**	0	-0.07
8	II	1.21**	0.37**	-7.66**	-0.39	-0.36**	0.06	0.18**	0.14*	0.36**	0.09
9	III	-0.46**	0.74**	1.64	-4.11**	-0.2*	-0.3**	0.08	-0.02	-0.06	0.03
10	III	1.57**	1.27**	14.11**	9.49**	-0.39**	0.06	0.24**	0.07	0.18**	0.08
11	III	-1.36**	-0.7**	-4.59**	-9.55**	0.03	0.2	0.06	-0.14*	-0.22**	-0.07
12	III	-1.33**	-0.38**	10.09**	-1.53	0.76**	0.34**	-0.28**	-0.24**	-0.25**	-0.23**
13	IV	-	-0.16	-	15.09**	-	0.31**	-	0.13*	-	0.02
14	IV	-	-1.01**	-	-14.64**	-	-0.02	-	0.07	-	0.04
15	IV	-	-1.33**	-	-9.78**	-	-0.17	-	-0.26**	-	0.03
16	IV	-	0.27*	-	-6.92**	-	0.16	-	-0.18**	-	0.04
17	V	0.71**	1.61**	1	6.87**	-0.47**	-0.25*	-0.09	-0.02	0.07	-0.06
18	V	1.08**	1.85**	-1.04	6.18**	0.04	-0.05	-0.25**	-0.27**	-0.01	-0.18**
19	V	-0.61**	0.64**	-6.3**	-8.62**	0.18*	0.35**	0.05	0.05	0.05	-0.09
20	V	-0.25	0.33*	0.64	1.68	0.53**	0.33**	-0.04	0.19**	-0.3**	-0.03
21	VI	-1.6**	-	-2.82*	-	0.6**	-	-0.23**	-	-0.32**	
22	VI	0.39**	-	2.86*	-	0.22*	-	0.01	-	-0.1	
23	VI	1.71**	-	9.48**	-	0.2*	-	0.17**	-	-0.06	
24	VI	3.92**	-	-3.41**	-	-0.63**	-	0.01	-	0.12*	
SE		0.13	0.12	1.19	1.17	0.83	0.85	0.08	0.10	0.05	0.05

[1]Inbred lines13 to 16 and 21 to 24 served only as male and female parents, respectively, in the entire mating scheme. * Significant at P < 0.05, ** significant at *p*<0.01, AD = no. of days to anthesis, PH = plant height (in cm), GY = grain yield (in t/ha), EA = ear aspect (1 to 5), PA = plant aspect (1 to 5)

Table 4.11 Estimates of specific combining ability (SCA) effects for agronomic traits in 80 hybrids generated through factorial mating of 24 inbred lines

Hybrid	Set	Female	Male	SD	AD	PH	GY	PA	EA
1	1	1	8	-0.31	-0.11	1.47	0.45	-0.41**	-0.47**
3	1	3	8	0.7	0.77	8.5	0.83**	-0.09	-0.21
4	1	4	8	1.6**	1.27**	-11.39*	-0.35	0.21	0.11
5	1	1	7	1.42**	1.43**	-3.56	-0.06	0.01	-0.01
6	1	2	7	0.69	0.55	3.38	0.31	-0.05	-0.35*
7	1	3	7	-1.1*	-1.12**	11.23*	1.18**	-0.4**	-0.63**
8	1	4	7	0.79	0.69	-5.64	-0.28	-0.06	0.21
11	1	3	6	2.78***	2.19**	-26.99**	-1.35**	0.46**	0.47**
12	1	4	6	-2.04***	-1.69**	31.34**	2.13**	-0.76**	-1.16**
15	1	3	5	-0.57	-0.3	12.67**	0.5	-0.47**	-0.41**
16	1	4	5	1.46**	1.28**	-8.89*	-0.34	0.11	0.06
21	3	9	3	0.56	0.56	-10.38*	-0.38	0.23	0.06
30	3	10	1	0.74	0.62	-6.04	-0.6*	-0.04	0.27
34	2	6	16	0.9	0.71	13.9**	-0.31	0.16	0.08
37	2	5	15	0.01	0.03	2.1	-0.23	0.43**	0.13
38	2	6	15	1.08*	0.99*	4.31	0.27	-0.19	-0.28
40	2	8	15	0.41	0.69	10.45*	-0.22	0.1	0.06
41	2	5	14	0.56	0.44	12.8**	-0.45	0.09	0.13
43	2	7	14	0.53	0.79	11.34*	0.09	-0.03	-0.24
45	2	5	13	1.46**	1.21**	-0.21	-0.05	-0.04	-0.09
51	4	19	12	0.01	0.01	4.46	-0.17	0.45**	0.12
53	4	17	11	-0.87	-0.49	6.63	0.31	0.12	-0.16
55	4	19	11	-0.72	-0.89*	-2.24	0	0.12	0.1
59	4	19	10	-0.96	-0.53	-0.48	0.55	-0.34*	-0.25
60	4	20	10	-0.31	-0.38	1.6	-0.5	0.31*	0.15
62	4	18	9	-1.52**	-1.24**	0.41	0.46	0.17	-0.16
63	4	19	9	0.57	0.47	3.95	-0.66*	0.11	0.21
65	5	21	20	-0.52	-0.98*	-2.02	-0.37	0.11	0.19
66	5	22	20	-1.31**	-1.19**	-7.5	-0.06	-0.09	0.09
67	5	23	20	-1.17*	-1.26**	-1.44	-0.29	-0.13	0.06
68	5	24	20	-1.38**	-0.99*	4.85	0.34	0.16	0.03
69	5	21	19	-1.37**	-1.18**	-2.35	0.17	-0.13	0.08
70	5	22	19	-0.97	-0.98*	-1.22	-0.29	-0.03	0.16
71	5	23	19	-0.65	-0.66	-2.86	-0.08	0.12	0.17
72	5	24	19	-1.39**	-1.6**	0.32	-0.18	0.08	-0.04
73	5	21	18	-0.97	-0.93*	2.46	0.06	-0.05	0.11
74	5	22	18	-1.41**	-1.21**	-1.23	-0.03	-0.02	-0.04
75	5	23	18	-1.31**	-1.32**	-3.44	-0.06	0.24	0.16
76	5	24	18	-0.68	-0.96*	-3.91	-0.36	-0.14	0.13
77	5	21	17	-1.52**	-1.33**	-4.2	-0.24	0.11	-0.02
78	5	22	17	-0.69	-1.04*	3.83	0	0.18	0.16
79	5	23	17	-1.24*	-1.18**	1.63	0.05	-0.19	-0.02
80	5	24	17	-0.94	-0.87*	-7.37	-0.19	-0.06	0.24
			SE	0.50	0.42	4.47	0.30	0.15	0.15

* Significant at P < 0.05, ** significant at *p*<0.01,

Note: only hybrids that had significant SCA for at least one trait are shown.

AD = no. of days to anthesis, PH = plant height (in cm), GY = grain yield (in t/ha), EA = ear aspect (1 to 5), PA = plant aspect (1 to 5)

4.3.6 Correlation and regression analyses of phenotypes in the parent and hybrid trials

The hybrid trial displayed significant positive and negative linear correlations for most of the measured carotenoids except for the non-significant correlations of β-carotene with lutein, zeaxanthin, and β-cryptoxanthin (Table 4.12). The correlations among carotenoids were stronger in the hybrid trial than in the inbred trial. In the inbred line trial, significant correlations among the measured carotenoids were detected only for α-carotene with lutein (r = 0.42, $p<0.05$), α-carotene with β-carotene (r = 0.49, $p<0.05$), and β-cryptoxanthin with zeaxanthin (r = 0.66, $p<0.01$). The strength and direction of correlation generally reflected the positions of the carotenoids in the carotenoid biosynthesis pathway and the enzymes involved in the synthesis of the corresponding carotenoid (Figure 1.2). The correlation between β-carotene and β-cryptoxanthin was not significant, whereas the correlation between α-carotene and β-carotene were strong and positive. Considering the calculated carotenoids (total carotenoid and provitamin A), in the inbred line trial provitamin A was strongly and positively correlated with β-carotene (r = 0.93, $p<0.01$) and α-carotene (r = 0.54, $p<0.01$), while total carotenoid was strongly and positively correlated with lutein (r = 0.55, $p<0.01$), zeaxanthin (r = 0.66, $p<0.01$) and β-cryptoxanthin (r = 0.46, $p<0.01$). In the hybrid trial, provitamin A and total carotenoid showed significant correlations with all the measured carotenoids, with the coefficient of correlation ranging from 0.24 ($p<0.05$) for total carotenoid with lutein correlation, to 0.94 ($P<0.01$) for provitamin A with β-carotene correlation.

In most cases the correlations between the carotenoids and agronomic traits were not significant (Table 4.13, Figure 4.3). Alpha-carotene showed significant negative association with all agronomic traits, except ASI and PA, whereas total carotenoid was positively correlated with SD, AD, PH, and GY. The scatter plots for the Pearson's coefficients of correlation in Figure 4.4 showed that β-carotene and beta-cryptoxanthin were not associated with GY, and its association with provitamin A appeared to be not definable with specific trend of linear correlation.

Table 4.12 Pearson correlation coefficients between mean values of carotenoid concentrations in the parent (upper matrix) and hybrid (lower matrix) trials

	Parents						
	Lut	**Zea**	**αcar**	**βcryp**	**βcar**	**PVA**	**Tcar**
Lut	1	-0.2	0.42*	-0.19	-0.07	-0.11	0.55**
Zea	-0.28*	1	-0.22	0.63**	-0.37	-0.14	0.66**
αcar	0.45**	-0.3*	1	0.02	0.49*β	0.54**	0.27
βcryp	-0.28*	0.66**	0.11	1	-0.19	0.18	0.46*
βcar	-0.02	0.05	0.66**	0.14	1	0.93**	-0.1
PVA	-0.13	0.27*	0.64**	0.45**	0.94**	1	0.08
Tcar	0.24*	0.84**	0.44**	0.61**	0.28*	0.45**	1
	Hybrids						

*significant at *P*<0.05, ** significant at *P*<0.01

Table 4.13 Pearson coefficient of correlation between carotenoids and agronomic traits averaged across environments and regression of hybrids carotenoid content on mid-parent values

	Lut	**Zea**	**αcar**	βcryp	βcar	**PVA**	**Tcar**
SD	0.06	0.17	-0.6**	-0.09	-0.12	0.17	0.41**
AD	0.06	0.17	-0.61**	-0.09	-0.12	0.16	0.41**
ASI	0.02	-0.04	0.05	0.06	0.09	0.01	-0.18
PH	0.01	0.11	-0.5**	-0.1	-0.12	0.28*	0.37**
EH	0.01	0.03	-0.39**	-0.08	-0.09	0.14	0.19
GY	-0.17	-0.17	-0.48**	-0.29**	-0.3**	0.23*	0.3**
EA	0.03	0.02	-0.29**	-0.01	0	0.12	0.03
PA	-0.09	-0.01	0.05	-0.08	0.1	0.06	-0.04
r	0.29**	0.39**	-0.04	0.39**	0.14	0.17	0.16
R^2%	8.23	15.59	0.18	15.49	2.07	2.74	2.52

*Significant at *p*<0.01, * Significant at *p*<0.05

Lut = Lutein, zea = Zeaxanthin, αcar = α-carotene, βcryp = β-cryptoxanthin, βcar = β-carotene, PVA = Provitamin A, tcar = Total carotenoid, SD = no. of days to silking, AD = no. of days to anthesis, PH = plant height (in cm), GY = grain yield (in t/ha), EA = ear aspect, PA = plant aspect

r = coefficient for regression of hybrids carotenoid content on mid-parent values, R^2 = percentage of variation in carotenoid content of the hybrids explained by that of the mid-parent values.

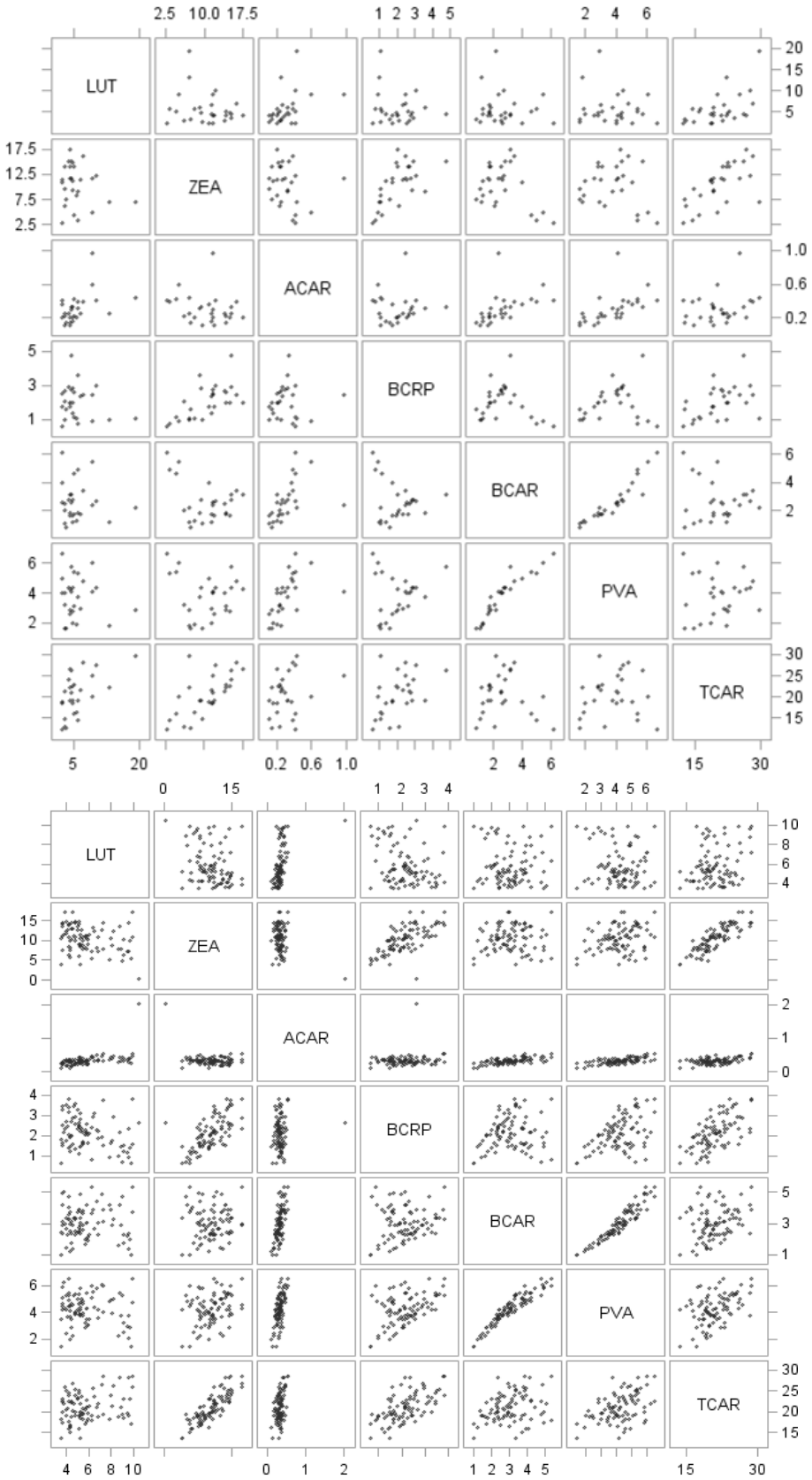


Figure 4.3 Scatter plots for simple correlation between mean values of carotenoids, in the parent trial (left panel) and hybrid trial (right panel).
LUT = Lutein, ZEA = Zeaxanthin, ACAR = α-carotene, BCRP = β-cryptoxanthin, BCAR = β-carotene, PVA = Provitamin, TCAR = Total carotenoid

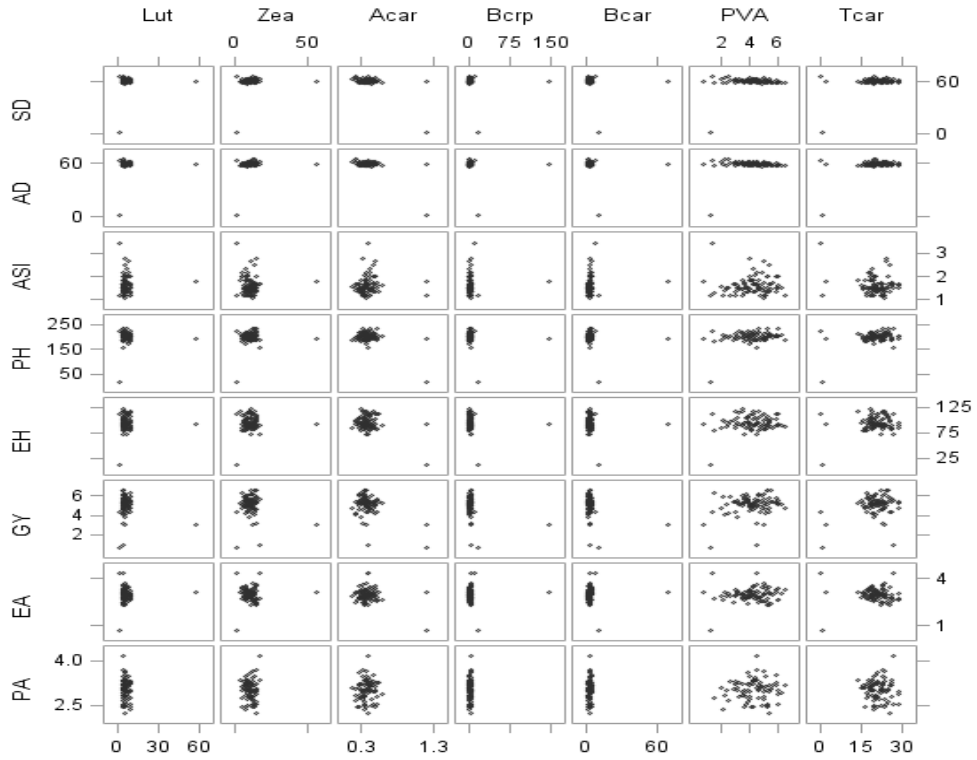


Figure 4.4 Scatter plots for correlations between carotenoids and agronomic traits.
Lut = Lutein, Zea = Zeaxanthin, Acar = α-carotene, Bcryp = β-cryptoxanthin, Bcar = β-carotene, PVA = Provitamin, Tcar = Total carotenoid, SD = no. of days to silking, AD = no. of days to anthesis, PH = plant height (in cm), EH = ear height (in cm), GY = grain yield (in t/ha), EA = ear aspect, PA = plant aspect.

Regression of hybrids carotenoid content on mid-parent values detected significant (p<0.01) but week associations for lutein (r = 0.29), zeaxanthin (r = 0.39), and β-cryptoxanthin (r = 0.39) (Table 4.13, Figure 4.5). The regression coefficients for β-carotene, provitamin A and total carotenoids were non-significant. This result suggested that the provitamin A content of hybrids may not be reliably predicted based only on the carotenoid content of their parents.

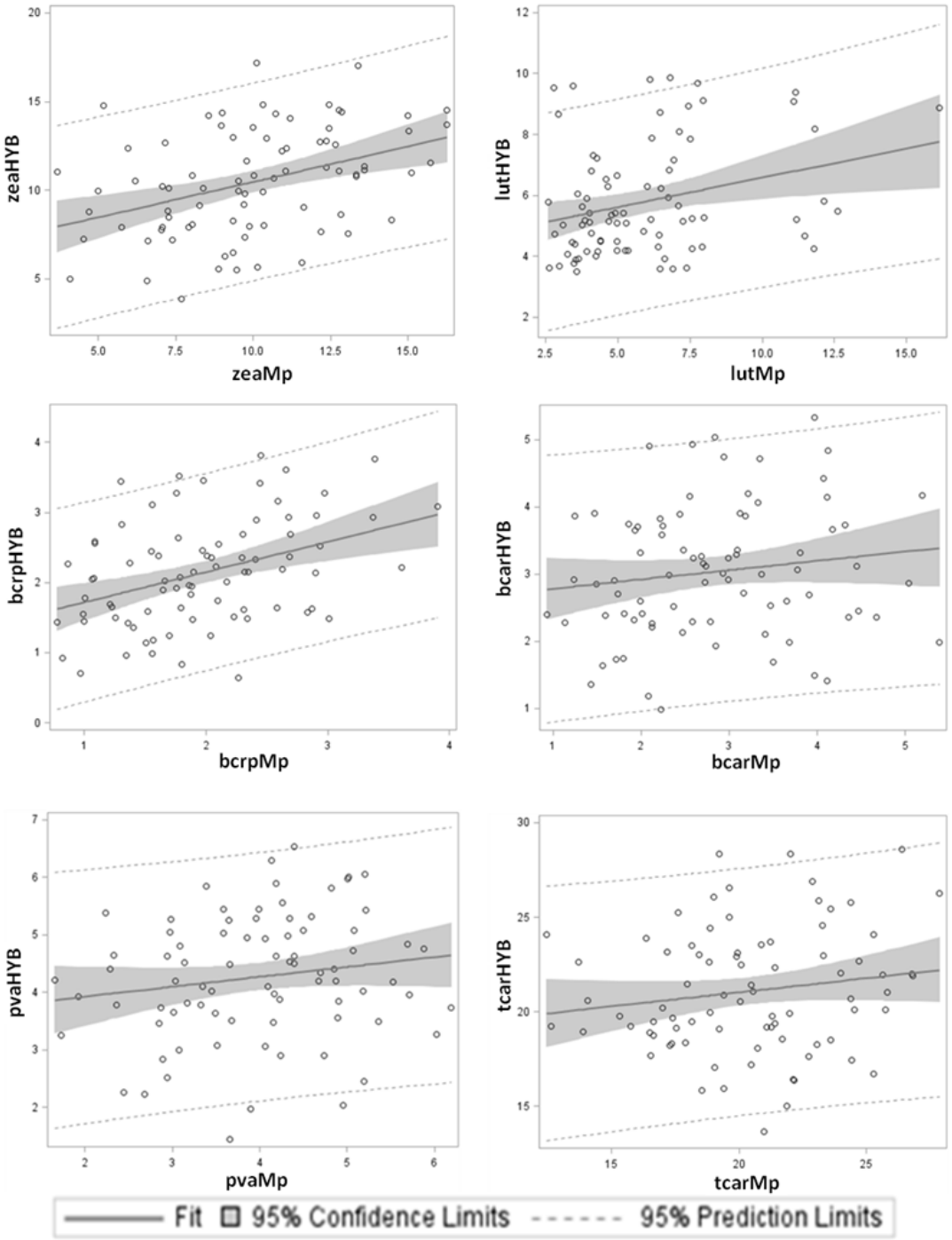


Figure 4.5 Regression of hybrids' carotenoid contents on mid-parent values.
lut = lutein, zea = zeaxanthin, bcrp = β-cryptoxanthin, bcar = β-carotene, pva = provitamin A, tcar = total carotenoid, suffix HYB = concentration of the carotenoid in hybrid, suffix Mp = mid-parent concentration value for the carotenoid.

4.3.7 Effects of the most significant functional markers *crtRB1* on carotenoid content of the hybrids

Haplotype assessment of the parental inbred lines for the best provitamin A markers designed to detect the allelic series of the 3'Indel and 5'TE polymorphic sites of the

gene *crtRB1* (Table 2.2) revealed four haplotype variants**:** '1, 1', '1, 2', '3, 1'and '3, 2'. Nineteen of the 24 parental lines carried haplotype '3, 1', three inbred lines carried haplotype '1, 2', and two inbred lines carried haplotypes '1, 1'or '3, 2'. The mating scheme generated hybrids that carried eight unique haplotypes (Table 4.14). Analysis of haplotypes effect on hybrid progenies showed that the markers had significant effects on β-carotene, provitamin A and total carotenoids. Haplotypes '1|3, 1', '3|1, 1' and '1, 2|1' rendered the highest mean carotenoid concentrations (4.22 to 5.23 µg/g). Haplotype pairs: '1|3, 1'and '3|1, 1', '1|3, 2|1' and '3|1, 1|2', which were basically of similar genotype apart from their differences in the parental source of the alleles, showed similar effect on mean concentrations, suggesting the absence of maternal or paternal effects.

Table 4.14 Effect of parental genotypes for the best provitamin A associated markers of *crtRB1* (3'TE and 5'TE) on the carotenoid content of the hybrid progenies

Haplotype[1]						
Female parent	**Male parent**	**Hybrid[1]**	**N[2].e.r**	**β-carotene[3]**	**provitamin A**	**Total carotenoid**
1, 1	3, 1	1|3,1	4	$4.22\pm0.96_{AB}$	$5.49\pm0.97_{A}$	$22.84\pm1.68_{AB}$
1, 2	1, 1	1, 2|1	1	$5.23\pm0.99_{A}$	$5.82\pm1.03_{A}$	$17.63\pm2.22_{ABC}$
1, 2	1, 2	1, 2	2	$4.49\pm0.97_{AB}$	$5.10\pm1.00_{A}$	$18.77\pm1.9_{ABC}$
1, 2	3, 1	1|3, 2|1	9	$3.40\pm0.95_{ABC}$	$4.48\pm0.96_{A}$	$19.45\pm1.55_{C}$
3, 1	1, 1	3|1, 1	3	$4.97\pm0.96_{A}$	$6.03\pm0.98_{A}$	$19.72\pm1.74_{ABC}$
3, 1	1, 2	3|1, 1|2	10	$3.78\pm0.95_{AB}$	$4.78\pm0.96_{A}$	$19.98\pm1.54_{BC}$
3, 1	3, 1	3, 1	47	$2.63\pm0.95_{BC}$	$4.04\pm0.95_{AB}$	$21.93\pm1.47_{A}$
3, 2	3, 1	3, 2|1	4	$1.57\pm0.96_{C}$	$2.22\pm0.97_{B}$	$18.84\pm1.68_{C}$
			R^2 %	76	66	38
			CV %	25	23	19

[1] Haplotypes refer to alleles for the 3'TE and 5'TE polymorphic loci in the same order, with the symbole '|' separating heterozygous alleles

[2]N.e.r = number of hybrids carrying the haplotype under consideration (replicated across e = 4 environments and r = 2 replications)

[3]Means followed by subscripts of same letter(s) are not significantly different

R^2 = percent explained variance by the model which included effects of environment and environment by markers interaction

4.3.8 Heterosis estimates

Forty two and 15 hybrids displayed significant positive mid-parent (mpH%) and high parent (hpH%) heterosis, respectively, for at least one of the carotenoids (Tables 4.15 and 4.16). At least one hybrid of each mating set displayed heterosis for at least one of the carotenoids. The largest number of hybrids with positive mpH%

heterosis was observed for β-carotene (22 hybrids) followed by β-cryptoxanthin (21 hybrids), zeaxanthin (17 hybrids) and lutein (14 hybrids). Significant mid-parent heterosis were detected for β-carotene, provitamin A and total carotenoids varying from 15% to 55%. Hybrid 30 (10 x 1) exhibited mpH for all carotenoids and had provitamin A content of 5.3 μg/g, but was poor in the agronomic traits measured. Mid-parent heterosis reached up to 77% for zeaxanthin and 114% for lutein. Hybrids 22 (10 x 3) and 23 (11 x 3) registered significant positive high parent heterosis for both β-cryptoxanthin and β-carotene, ranging from 30% to 56% with mean provitamin A concentrations of 5.27 and 6.0 μg/g, respectively. Hybrid 12 (4 x 6), which showed favourable SCA effects for the important agronomic traits considered in this study (GY, EA and PA), expressed high parent heterosis of 32% for provitamin A (6.06 μg/g) and 37% for total carotenoid (25.58 μg/g). This hybrid was comparable to the best performing hybrid in terms of its agronomic traits, having 4.9 t ha^{-1} for GY,and 2.6 for EA and PA. Hybrid 38 (6 x 15) also showed significant positive high parent heterosis for provitamin A (21%), high provitamin A concentration (5.32 μg/g) and good agronomic characters, 5.2 t/ha for GY, 2.3 for EA and 2.5 for PA.

Some hybrids displayed heterosis for multiple carotenoids. For example, hybrid 30 showed significant mid-parent heterosis for all carotenoids, while hybrids 15, 57 and 58 showed significant positive mid-parent heterosis for all β-branch carotenoids. On the other hand, most hybrids that displayed significant positive heterosis for carotenoids in the β-branch did not show significant positive heterosis for lutein (an α-branch carotenoid). Similarly, most of the hybrids that showed significant positive heterosis for at least one provitamin A carotenoid did not show significant heterosis for the non-provitamin A carotenoids, zeaxanthin and lutein.

Table 4.15 Percent mid-parent heterosis (mPH%) estimates for 42 hybrids out of the 80 hybrids generated from 24 inbred lines in a factorial mating design

				Heterosis (%)						PVA
Hybrid	**Set**	**Female**	**Male**	**Lut**	**Zea**	**βcryp**	**βcar**	**PVA**	**Tcar**	**(µg/g)**
3	**1**	**3**	**8**	51.56**	-6.81	31.85**	51.58***	47.56***	17.17	5.82
4	**1**	**4**	**8**	4.73	22.13	14.1	18.87*	16.03	10.96	4.83
7	**1**	**3**	**7**	36.05*	0.19	-17.77	7.62	2.47	4.73	4.81
8	**1**	**4**	**7**	25.8*	19.89	-29.41	13.83*	7.55	15.18	5.29
9	**1**	**1**	**6**	74.03**	-1.58	0.11	10.79	7.84	7.77	5.03
11	**1**	**3**	**6**	37	17.22*	2.83	10.65	8.83	14.7*	4.53
12	**1**	**4**	**6**	12	26.35*	3.88	49.94***	38.05***	21.55**	6.06
13	**1**	**1**	**5**	113.76***	21.86	26.55	-2.63	0.77	27.73*	5.85
15	**1**	**3**	**5**	43.08	36.99**	29.76*	16.92*	18.58**	28.21**	6.29
18	**3**	**10**	**4**	16.6	27.44*	-7.96	27.97**	18.28*	17.41*	4.95
19	**3**	**11**	**4**	8.2	20.11	-5.18	22.99*	15.21	11.06	4.76
20	**3**	**12**	**4**	29.51**	35.22**	-24.3**	20.59*	6.58	21.43**	5.43
21	**3**	**9**	**3**	46.4**	5.45	-1.39	26.01	14.46	9.72	4.41
22	**3**	**10**	**3**	7.43	16.23	29.87***	35.12**	32.9**	14.07	5.27
23	**3**	**11**	**3**	20.26	7.31	45.4***	58.18***	53.41***	14.92*	6.00
27	**3**	**11**	**2**	-10.33	-7.61	23.72**	20.86	20.94*	-6.37	4.72
30	**3**	**10**	**1**	48.48*	25.83*	18.03*	21.43*	20.81*	23.94**	5.39
31	**3**	**11**	**1**	8.87	11.69	39.22***	23.05*	26.24**	10.55	5.57
33	**2**	**5**	**16**	37.66	39.57**	84.45***	-4.75	6.73	28.28**	5.04
36	**2**	**8**	**16**	-22.49	-7.19	50.59**	29.84*	33.38**	-5.95	4.48
37	**2**	**5**	**15**	20.79	45.65**	74.53***	7.65	18.33**	29.58**	6.53
38	**2**	**6**	**15**	16.94	6.67	21.38**	23.32*	21.56*	7.1	5.32
40	**2**	**8**	**15**	19.74	-4.65	25.85*	21.59*	22.2*	4.86	5.07
41	**2**	**5**	**14**	10.45	32.85**	17.75	-10.23	-7.24	14.63	5.08
44	**2**	**8**	**14**	45.17**	-24.82*	10.28	-4.64	-2.11	-4.56	4.03
45	**2**	**5**	**13**	21.85	15.65	42.8**	-12.77*	-4.85	10.07	5.44
53	**4**	**17**	**11**	1.5	-0.54	21.71*	29.44	25.52*	1.3	3.51
54	**4**	**18**	**11**	38.89***	18.35*	23.95*	30.07	28.45*	22.72**	3.74
55	**4**	**19**	**11**	-7.83	2.37	26.1*	25.92	24.23*	0.04	3.56
56	**4**	**20**	**11**	18.24	10.61	21.76**	25.43*	23.89**	11.44	4.98
57	**4**	**17**	**10**	1.21	34.13**	83***	35.69*	48.13***	24.45*	4.22
58	**4**	**18**	**10**	3.49	29.63**	27.02**	38.74*	33.78**	19.41*	3.97
59	**4**	**19**	**10**	2.69	6.67	26.66*	29.19	27.62*	5.29	3.73
61	**4**	**17**	**9**	33.22*	8.99	-11.34	33.45	12.79	10.94	3.08
68	**5**	**24**	**20**	38.38*	45.97**	8.88	43.87**	29.27*	32.77**	3.78
70	**5**	**22**	**19**	33.95*	25.53	4.66	28.93	19.95	21.22	3.06
72	**5**	**24**	**19**	-4.78	44.83**	14.11	15.82	11.35	17.48	1.98
74	**5**	**22**	**18**	14.33	16.05	17.35	43.5*	34.69*	15.29	3.50
76	**5**	**24**	**18**	1.49	47.96**	19.05	47.15	34.26	27.17*	2.45
78	**5**	**22**	**17**	18.18	34.45*	16.85	52.72**	39.65**	28.73*	3.46
79	**5**	**23**	**17**	-17.16*	21.43	48.38*	39.13*	40.48*	3.62	3.25
80	**5**	**24**	**17**	49.37*	71.51***	23.8	42.12	31.33	53.55***	2.24

Note: only crosses with significant negative or positive heterosis for at least one of the carotenoids are shown Lut = Lutein, Zea = Zeaxanthin, βcryp = β-cryptoxanthin, βcar = β-carotene, PVA = Provitamin, Tcar = Total carotenoid

Table 4.16 High parent heterosis (hPH%) estimates, provitamin A content (PVA in µg/g), grain yield (GY in t/ha), ear aspect (EA) and plant aspect (PA) for 15 hybrids, out of the 80 hybrids generated from 24 inbred lines in a factorial mating design

				Heterosis						PVA	GY		
Hyb	**Set**	**F**	**M**	**Lut**	**Zea**	**βcryp**	**βcar**	**PVA**	**Tcar**	**(µg/g)**	**(t/ha)**	**EA**	**PA**
3	**1**	**3**	**8**	12.38	-42.77***	-15.77	31.89**	46.72***	-5.34	5.82	4.73	3.20	3.37
9	**1**	**1**	**6**	44.11*	-15.61	-19.83*	-5.26	1.4	-4.33	5.03	2.62	3.81	3.78
12	**1**	**4**	**6**	-21.14*	-15.95	-32.54***	31.82**	37.36***	7.91	6.06	4.89	2.59	2.58
13	**1**	**1**	**5**	112.22**	-22.74	-15.21	-20.29**	-12.01*	4.42	5.85	4.98	3.30	3.62
22	**3**	**10**	**3**	-11.71	6	29.84**	33.97*	32.9**	12.46	5.27	4.43	3.10	3.34
23	**3**	**11**	**3**	-21.08*	3.74	38.6***	55.63***	51.36***	3.93	6.00	4.65	2.77	2.96
27	**3**	**11**	**2**	-43.56***	-13.26	22.03*	16.68	19.46	-20.43**	4.72	4.90	2.83	3.30
30	**3**	**10**	**1**	14.97	21.67*	-2.89	-1.5	8.68	16.69	5.39	3.73	3.40	3.42
31	**3**	**11**	**1**	-31.07**	1.89	19.2*	-0.78	12.21	-6.37	5.57	4.74	2.88	3.18
38	**2**	**6**	**15**	6.63	-3.51	18.99*	20.78	21.3*	3.3	5.32	5.23	2.30	2.45
56	**4**	**20**	**11**	-10.55	-5.28	20.23*	16.36	19.15*	-4.54	4.98	4.79	2.89	3.06
57	**4**	**17**	**10**	-17.57	23.14*	26.81*	-1.66	6.35	4.63	4.22	3.93	3.43	3.29
58	**4**	**18**	**10**	0.89	27.86*	-4.62	4	0.07	11.33	3.97	4.54	3.16	3.07
79	**5**	**23**	**17**	-52.56***	4.57	47.66*	4.55	12.04	-22.51**	3.25	4.55	2.89	3.06
80	**5**	**24**	**17**	39.05	34.61*	3.6	23.49	29.36	37**	2.24	3.49	3.34	3.03

Note: only crosses with significant negative or positive heterosis for at least one of the carotenoids are shown

Lut = Lutein, Zea = Zeaxanthin, βcryp = β-cryptoxanthin, βcar = β-carotene, PVA = Provitamin A, tcar = Total carotenoid

Table 4.17 Summary of agronomic performance and provitamin A content for selected hybrids, out of the 80 hybrids evaluated across environments

				Agronomic traits				Caroteoids (µg/g)		hpH (%)
Hybrid	**Set**	**Female**	**Male**	**AD**	**GY**	**EA**	**PA**	**PVA**	**Tcar**	**for PVA**
37	2	5	15	56.26	4.21	3.44	3.29	6.53	22.68	-1.78
12	1	4	6	57.87	4.89	2.59	2.58	6.06	26.58	37.36**
14	1	2	5	56.64	5.29	3.08	3.09	5.89	17.47	-11.37
13	1	1	5	56.32	4.98	3.3	3.62	5.85	19.94	-12.01*
32	3	12	1	56.84	5.87	2.47	2.99	5.44	24.57	-5.78
45	2	5	13	58.61	4.88	3.23	3.21	5.44	22.91	-18.11**
38	2	6	15	57.45	5.23	2.3	2.45	5.32	25.44	21.3*
24	3	12	3	56.47	5.71	2.64	2.71	5.29	24.08	-8.46
52	4	20	12	57.37	5.68	2.34	2.73	5.25	22	-9.14
33	2	5	16	58.36	5.23	3.03	2.68	5.04	23.11	-24.14**
7	1	3	7	56.02	5.31	2.61	2.69	4.81	19.37	-11.34
27	3	11	2	57.2	4.9	2.83	3.3	4.72	21.9	19.46
47	2	7	13	57.24	4.93	3	3.23	4.64	18.31	-14.57*
46	2	6	13	57.48	5.35	2.42	3.09	4.63	22.65	-3.43
28	3	12	2	57.04	5.42	2.95	2.75	4.62	20.12	-19.96**
6	1	2	7	56.41	5.08	2.72	3.05	4.51	15	-16.8*
42	2	6	14	57.44	4.97	2.82	3.19	4.34	26.06	-0.68
51	4	19	12	57.55	4.91	2.98	3.35	3.88	20.11	-32.81**
25	3	9	2	58.15	4.92	2.82	3.15	3.84	18.59	-2.93
34	2	6	16	58.77	4.98	2.67	2.88	3.82	24.43	-12.56
67	5	23	20	59.32	4.79	3.01	3.33	3.79	21.02	-9.31
59	4	19	10	58.67	5.34	2.93	2.87	3.73	19.99	-5.94
66	5	22	20	58.07	5.05	2.99	3.21	3.64	16.71	-12.88
55	4	19	11	56.34	4.93	3.13	3.13	3.56	21.94	-7.63
64	4	20	9	58.62	5.22	2.75	2.99	3.48	17.64	-16.72
70	5	22	19	58.59	4.84	3	3.13	3.06	18.28	-5.14
65	5	21	20	56.29	5.11	2.87	3.16	2.99	19.78	-28.36**
71	5	23	19	60.22	5.03	3.06	3.44	2.52	18.53	-13.09
73	5	21	18	57.86	5.17	2.66	2.55	2.04	19.25	3.84
69	5	21	19	56.39	5.67	2.71	2.78	1.44	17.07	-23.31
11	1	3	6	62.43	0.66	4.36	4.34	4.53	26.89	3.79
81	Check	25	27	58.99	4.86	2.45	2.98	2.83	22.33	-48.01**
Grand Mean				58.54	4.56	3.03	3.09	4.27	21.10	
Min				56.02	0.66	2.30	2.45	1.44	13.66	
Max				63.34	5.87	4.36	4.34	6.53	28.60	
SE				1.63	1.25	0.22	0.20	0.72	2.06	
LSD (P<0.01)				1.64	1.26	0.66	0.67	1.34	4.12	

F = female M = male, GY = shelled grain weight (t/ha), EA = ear aspect, PA = plant aspect, βcar = β-carotene, PVA = provitamin A, Tcar = total carotenoid, hpH = high parent heterosis, Min = minimum, Max = maximum, SE = standard error, LSD = average least significant difference

4.4 Discussion

Combining ability analysis is an important research approach that is routinely applied in maize breeding to identify inbred lines with good breeding value for hybrid and synthetic formation. Combining ability helps to better understand the mode of gene action controlling the trait of interest and devise breeding strategies towards improving the trait. In this study, the combining ability of twenty four inbred lines for carotenoid and important agronomic traits was investigated. Both the parental inbred lines and their hybrids showed broad ranges of variation in carotenoids and agronomic traits. The provitamin A carotenoids profile observed in this study is comparable with the result reported for similar tropical diverse yellow maize germplasm by Menkir et al. (2008).

The β-carotene content in the hybrids of the present study varied from 0.10 to 5.33 µg/g averaged at 3.04 µg/g, and the provitamin A content of the hybrids varied from 1.44 to 6.53 µg/g with an average of 4.27 µg/g. Some previous combining ability studies reported slightly lower levels and ranges of beta-carotene and provitamin A than reported in the present study. Egesel et al. (2003b), Senete et al. (2011) and Run et al. (2013) studied hybrids with β-carotene concentration ranging from 0.5 to 4.4, and provitamin A concentration ranging from 1.3 to 6.3 µg/g. However, these studies used smaller numbers of hybrids compared to the present study: 80 hybrids in the present study vs. 52 by Run et al. (2013), 45 by Egesel et al. (2003b) and 25 by Senete et al. (2011). On the other hand, a recent combining ability study on maize carotenoid by Suwarno et al. (2014) used 21 inbred lines and 156 hybrids that registered high levels of provitamin A carotenoid ranging from 4.4 to 18 µg/g.

Zeaxanthin was the predominant carotenoid in both parent and progeny trials, constituting about 50% of their average total carotenoid content (Figure 4.1, Figure 4.2). Egesel et al. (2003b) also found zeaxanthin as the major carotenoid representing more than 50% of the total carotenoids. Suwarno et al. (2014), on the other hand, indicated that zeaxanthin represented only 33% of the total carotenoid in the hybrids they analyzed. These authors recorded hybrids that accumulated as high as 12 to 20 µg/g provitamin A, which might be the cause of the lower proportion of zeaxanthin in their hybrids due perhaps to substrate diversion in favour of the synthesis of upstream carotenoids. The contribution of β-carotene to the total

provitamin A was larger than that of β-cryptoxanthin, with β-carotene representing about 70% of the total provitamin A in both the inbred line and hybrid trials. This is in agreement with some previous results (Senete et al., 2011) but contrasts with other results that reported larger proportion of β-cryptoxanthin than β-carotene (Egesel et al., 2003b; Menkir et al., 2008, 2014; Suwarno et al., 2014)

The significant male and female GCA effects on carotenoids highlighted the importance of additive gene effects in controlling carotenoid content in maize endosperm. The significant SCA effect on lutein, β-cryptoxanthin and provitamin A, implied the presence of non-additive gene effects on the accumulation of these carotenoids in maize endosperm. These findings are in accordance with previous results that highlighted the preponderance of additive gene action in maize endosperm carotenoid accumulation (Grogan et al., 1963; Egesel et al., 2003b; Senete et al., 2011; Li et al., 2013b; Suwarno et al., 2014). The predominance of SCA effects on carotenoid content has been demonstrated in other crops (Singh et al., 2011 in cabbage; Peninah et al., 2014 in cassava; Dey et al., 2014 in cauliflower)

The line by tester analysis revealed a number of good combiner inbred lines for provitamin A carotenoids (inbred lines 1 to 8). The mean provitamin A concentrations for hybrids within sets generally followed the pattern of the GCA effects of the parents. Likewise, mean beta-carotene content in hybrids paralleled the provitamin A content of their parental inbred lines. High and medium provitamin A parents were good combiners for provitamin A, whereas low provitamin A parents were poor combiners and thus the crossing sets constituting those parents had low mean carotenoid concentrations. Thus, combining two high provitamin A parents with good GCA effects can generally result in hybrids with high provitamin A. Regression of the mid-parent provitamin A concentrations on those of the hybrids also suggested the provitamin A contents in hybrids may not necessarily be predicted on the bases of those of the parental inbred lines, possibly because of the involvement of non-additive gene actions in regulating accumulations of provitamin A carotenoids. It is therefore more ideal to consider both GCA and SCA when selecting inbred lines for provitamin A hybrid development (Dey et al., 2014).

The non-significant correlation between carotenoids and important agronomic traits such as GY, indicated that carotenoid content in maize endosperm could be

improved without affecting the agronomic performance of hybrids. Non-significant correlations between grain yield and provitamin A has also been reported by many researchers (Suwarno et al., 2014; Menkir et al., 2014; Egesel et al., 2003b). An exceptional result was that of Senete et al. (2011) which reported significant negative correlation (r = -0.4) between grain yield and provitamin A.

Assessment of effects of the functional markers for provitamin A in maize hybrids demonstrated that the favourable alleles of the best analytic markers for provitamin A were correlated with higher levels of provitamin A. This demonstrated the efficacy and usefulness of these markers not only in screening high provitamin A inbred lines but also for generation of the desired end products, which are high provitamin A commercial varieties.

A study by (Burt et al., 2011a) concluded that heterosis for maize carotenoid content was a rare phenomenon. However, a recent study by Alfieri et al. (2014) found several hybrids manifesting both mid and high parent heterosis. In the present study 42 hybrids expressed mid-parent heterosis, and 15 parents expressed high parent heterosis for at least one carotenoid. Mid-parent heterosis was detected on 20 hybrids for provitamin A and 17 hybrids for total carotenoids. High parent heterosis was observed only on six hybrids for provitamin A and a single hybrid for total carotenoids. Hybrids 24, 32, and 52 displayed the best in agronomic performances and provitamin A contents. However, none of these hybrids expressed significant mid or high parent heterosis. On the other hand, hybrids 12 and 38 showed significant high parent heterosis for provitamin A and performed well in terms of their both provitamin A and agronomic traits (GY, EA and PA). The heterosis detected for provitamin A in some hybrids can help establish heterotic patterns among the inbred lines and help the design of future hybrid breeding programs. The result suggests the possibility of exploiting the heterotic potential for provitamin content of yellow maize hybrids that also show desirable agronomic characters.

4.5 Conclusion

This study corroborated the predominance of additive genetic effect on the provitamin A and other carotenoids in maize endosperm, and also highlighted the significant contribution of non-additive genetic effects to variations in the trait

leading to heterosis. Good combining inbred lines that have high level of provitamin A can be used to generate hybrids with high levels of provitamin A. The availability of parental inbred lines that result in hybrids with both good agronomic performance and high level of provitamin A can be used to advance IITA's provitamin A maize development efforts. The result of assessment of provitamin A markers in hybrids can serve as a basis for planning future research that incorporates genomics assisted selection for development of maize hybrids with high provitamin A.

Chapter 5: General conclusions and future perspectives

5.1 General Conclusion

Maize is an ideal staple crop for provitamin A biofortification due to three major facts. Firstly, the crop can naturally synthesize and accumulate carotenoids in its endosperm, and diverse germplasm is readily available for this trait to serve as input for breeding; Secondly, it is an important scientific model crop that has enormous genetic, genomic and biochemical knowledge and databases that can be readily utilized to harness the crop f-or biofortification; Thirdly, the crop is a major staple crop that serves as a source of calories for hundreds of millions of people in Sub-Saharan Africa and other developing regions. Hence, diets from biofortified maize can provide a cheap and sustainable source of vitamin A to the poor of these regions.

The principal carotenoids that are naturally accumulated in yellow maize endosperm are the non provitamin A carotenoids zeaxanthin and lutein. However, there is a great potential for enhancing the fraction of provitamin A carotenoids, β-carotene and β-cryptoxanthin in maize endosperm to levels that are biologically sufficient to reduce levels of vitamin A deficiency and its deleterious effects on vulnerable communities. The exploitation of the genetic potential of maize has led to the identification of germplasm with high levels of provitamin A (Pixley et al., 2013; Suwarno et al., 2014). This resource has been exploited for the successful release of the first wave of commercial maize high vitamin A cultivars, even though these lines have so far attained only 50% of the initial target concentration set by health and nutrition experts (HarvestPlus, 2014).

Identifying and/or developing provitamin A rich germplasm, and breeding high vitamin A genetics into high yielding, biotic and abiotic stress tolerant/resistant and commercially acceptable varieties is a continuous processes that needs to be performed in an environment (in the context of time and space), and demand-specific manner. Such breeding efforts are typically time, money and effort consuming. Hence, it is necessary to plan and implement breeding approaches that help facilitate the entire cycle of maize germplasm enhancement and variety development, in addition to incorporating provitamin A content as one of the target phenotypes for selection. Marker assisted selection is a technique that has already found successful

applications in simplifying and speeding up maize provitamin A breeding (Yan et al., 2010; Pixley et al., 2013; HarvestPlus, 2014).

LcyE and *crtRB1* are the two major genes that have been targeted for marker-assisted provitamin A breeding. The gene *LcyE* is involved in the regulation of levels of lycopene, the carotenoid at the branching point of the pathway, towards the α-branch of the pathway. The *crtRB1* gene acts within the β-branch of the pathway converting β-carotene to a carotenoid that has half as much provitamin A activity, β-cryptoxanthin. These two genes have been shown to have allelic variants with reduced functions that can help channel substrates in favour of increased β-carotene biosynthesis (Harjes et al., 2008; Yan et al., 2010; Babu et al., 2013; Fu et al., 2013). These two genes were proposed to explain the largest proportion of the variation in provitamin A carotenoids of maize. However, some studies that tested these two functional markers under different genetic backgrounds reported inconsistent results (Burt et al., 2011b; Babu et al., 2013; Vignesh et al., 2012). Further investigation of these genes and their analytic functional markers was therefore necessary in order to understand the consistency and the "provitamin A potential" of these markers in populations with different and diverse genetic backgrounds. The inconsistency of these two functional markers could be suggestive of possible effects of other carotenogenic genes. In addition, the involvement of numerous genes in carotenoid biosynthesis indicates that additional genes (apart from *lcyE* and *crtRB1*) can clearly play significant roles in provitamin A accumulation in maize endosperm. Hence, it is necessary to use "objective" genome-scanning methods such as genome-wide association studies to identify all genomic loci with detectable effects on the provitamin A content of maize endosperm.

The development and identification of high provitamin A tropical adapted breeding lines that can accumulate more than 15 μg/g provitamin A (Pixley et al., 2013; Babu et al., 2013; Azmach et al., 2013; Suwarno et al., 2014) is an important achievement in the maize provitamin A biofortification endeavour. However, it is, necessary to advance these high provitamin A biofortified inbred lines to commercial varieties by first testing their breeding value for provitamin A through genetic studies. This is necessary to determine the mode of gene action governing provitamin A accumulation within a set of inbred lines, and for identifying inbred line

combinations with desirable GCA and SCA effects for provitamin A content along with important agronomic characters.

In the above context, the research in this PhD thesis was designed to attain the following major research objectives:

1. To investigate the applicability of markers previously designed to detect the desirable and undesirable alleles of the two genes, *lcyE* and *crtRB1* on a set of diverse inbred lines developed at IITA.
2. To detect loci associated with endosperm carotenoid content variations at genome-wide scale.
3. To investigate the combining ability of inbred lines for their carotenoid content and agronomic traits and estimate their heterotic effects.

HarvestPlus has designed and been following a three stage strategy to curb the problem of vitamin A and other micronutrient deficiencies in developing countries through crop biofortification (HarvestPlus, 2012). The first stage is where target micronutrient deficient populations are identified, and germplasm and associated genes relevant to the target micronutrient biosynthesis are discovered. In the second stage, the discoveries of the first stage are utilized to develop crop varieties with improved levels of provitamin A, which are then tested in the field and nutrition studies are also implemented. The third stage is the delivery stage where biofortified crops are released, promoted and disseminated in the target micronutrient deficiency affected countries and their health and livelihood impacts are assessed. The objectives of this PhD project fit within the first and second stages of HarvestPlus strategy for maize biofortification.

The germplasm studied to achieve the research objectives of this PhD thesis included a panel of 130 advanced yellow maize inbred lines developed from progenies of various types of crosses that involved tropical and temperate yellow maize germplasm. The temperate inbred lines served as donors of high provitamin A alleles. Towards meeting the first objective of the thesis, the panel was first analysed for its endosperm carotenoid contents and found to display a broad range of diversity for carotenoids. Several inbred lines that had provitamin A concentration higher than 10 µg/g were detected, the highest reaching 17.25 µg/g, which was a 120% increase

from the highest provitamin A concentration that had been reported by Menkir et al. (2008). The inbred lines showed a good level of polymorphism for the six markers of the two major genes, lcyE and crtRB1 (three markers for each).

The divergent and mixed genetic background of the yellow maize inbred lines used in this study was considered to provide a suitable model for testing the provitamin A potential of the functional markers proposed for provitamin A maize breeding. A number of inbred lines of the panel harboured favourable alleles of both *lcyE* and *crtRB1*. Markers of *crtRB1* explained a greater variation of carotenoid content more consistently than *lcyE* markers, yet the joint effects of the two genes were larger than their individual effects for each carotenoid. The best provitamin A associated markers were those detecting the 3'TE and 5'TE polymorphic sites of *crtRB1*. The relative proportion of beta-carotene to beta-cryptoxanthin was the most affected by these functional polymorphisms, which is in accord with the known role of this gene. Considering the larger effect of the *crtRB1* markers on provitamin A carotenoids as compared to those of *lcyE*, use of *crtRB1* markers alone is considered adequate for increasing the genetic gain in provitamin A maize breeding (Babu et al., 2013).

Two functional markers for the *PSY1* gene involved in the first committed step of the carotenoid biosynthesis pathway of maize endosperm (Fu et al., 2013) were also investigated. Unsurprisingly, these markers were monomorphic for the favourable alleles of the markers developed by Fu et al. (2013) across all the inbred lines, as *PSY1* gene is generally conserved in yellow maize (Palaisa et al., 2003). Screening the inbred lines on the bases of their ear aspect, plant aspect and provitamin A content as well as haplotype content for the favourable allele of the best analytic markers identified one outstanding inbred line, i.e. (KU1409/DE3/KU1409) S2-18-2-B*4).

Having confirmed the effect of lcyE and crtRB1 and identified their best diagnostic markers in Chapter 2, in Chapter 3 the research focus was on identifying additional loci on a genome-wide scale with significant effects on maize endosperm carotenoid composition and content. Initially, a standard MLM GWAS that took the population structure and cryptic familial relatedness into account was applied. This not only identified numerous significant SNPs that co-localized with the already known major effect genes (*lcyE* and *crtRB1*) and a minor effect gene (*ZEP1*), but also detected a

number of other significant SNPs associated with genes of previously unknown carotenoid functions, many of which were transcription factors. The largest fraction of the significant SNPs concentrated on chromosome 8 and 10 in regions where the major effect genes *lcyE* and *crtRB1* were located.

Re-conducting of the GWA by incorporating the allele specific marker data for *lcyE* and *crtRB1* genes as covariates revealed new gene associations for β-carotene/provitamin A and zeaxanthin. The loci associated with zeaxanthin were also detected in the initial GWAS, but they were then associated with total carotenoids only. The new locus associated with β-carotene and provitamin A was a transcription factor called auxin-response factor 20 which is implicated in regulation of carotenoid biosynthesis at transcription level in plants. Hence, it will be important to consider this locus for future investigation as carotenoid biosynthesis regulation is an important intervention point for exploiting the genetic potential of maize for elevated provitamin A biosynthesis.

As the focus in GWAS is directed primarily at those SNPs at the peaks of association signals and also due to genome-wide multiple testing, small effect genes with significant contribution to the observed variation likely escape detection. The pathway level association analysis employed in this study complemented the GWAS by FDR adjusting the *p*-values of only those SNPs in the flanking regions of genes involved in the isoprenoid biosynthesis pathway. Out of 42 genes tested with about 6.8 K SNPs, significant association of 20 genes were detected at 10% FDR. Genes detected by this method included *IPPI2* and *CCD1* with lutein, *CCD7* with beta-carotene and ratio of beta-carotene to beta-cryptoxanthin, *hyd4* with total carotenoids and beta-branch carotenoids and their ratio traits.

Once the high provitamin A germplasm is identified and bred into adapted inbred line, this has to be converted to the germplasm form that can reach the next-user (farmer), namely as commercial hybrids and populations. To achieve the third objective of this PhD thesis, 24 advanced yellow maize inbred lines (with similar genetic background to those used for the first and second objectives) were deployed in a factorial mating to generate hybrids for line by tester analysis and heterosis effect studies. This study corroborated previous combining abilities studies showing the preponderance of GCA effects in the inheritance of provitamin A and non -----

provitamin A carotenoids. Since there was no significant correlation between important agronomic traits and provitamin A, it was confirmed that maize can be improved for the two traits simultaneously which is important for developing varieties that meet the objective of biofortification and at the same time are acceptable by farmers and seed producers. The identification of agronomically good performing and high provitamin A accumulating hybrids in this study supports a conclusion that the two traits can be improved independently. These hybrids displayed significant heterosis for provitamin A in the desired direction and had good combining parents.

5.2 Future directions

In the second objective an inbred line (KU1409/DE3/KU1409)S2-26-B*5) that accumulated the highest level of total carotenoids but did not carry the favourable alleles for the best diagnostic provitamin A markers (3'TE and 5'TE markers of *crtRB1*) was observed. This inbred line can be now targeted for backcrossing with KU1409/DE3/KU1409) S2-18-2-B*4, the inbred line that showed good agronomic performance and harboured the favourable alleles of provitamin A markers (*crtRB1*).

As a follow up to the above point, introduction of high total carotenoid accumulating germplasm can be considered which can be used to introgress the favourable alleles of *crtRB1*.

Many genes distributed across the genome contributed to the variation in endosperm carotenoid content of the association panel studied. Considering the large number of these loci and their wide genome coverage, inbred line selection using genome selection can be considered for future provitamin A maize breeding.

Heterotic patterns can be established on the basis of the genetic background of the good combiner parental inbred lines that resulted in positive heterosis for provitamin A. The heterotic groups for provitamin A content established by taking the agronomic traits into account will facilitate development of hybrid and synthetic varieties that have high levels of provitamin A.

References

Acquaah, G. (2012). Principles of Plant Genetics and Breeding 2nd ed. (Blackwell Publishing: Malden).

Adeyemo, O., Menkir, A., Gedil, M., and Omidiji, O. (2011a). Carotenoid and molecular marker-based diversity assessment in tropical yellow endosperm maize inbred lines. J. Food Agric. Environ. **9**: 383–392.

Adeyemo, O., Menkir, A., Gedil, M., and Omidiji, O. (2011b). Genetic diversity assessment and relationship among tropicalyellow endosperm maize inbred lines using SSR markers. Maydica **56**: 43–50.

Alfieri, M., Hidalgo, A., Berardo, N., and Redaelli, R. (2014). Carotenoid composition and heterotic effect in selected Italian maize germplasm. J. Cereal Sci. **59**: 181–188.

Andersen, J.R. and Lübberstedt, T. (2003). Functional markers in plants. Trends Plant Sci. **8**: 554–560.

Azmach, G., Gedil, M., Menkir, a, and Spillane, C. (2013). Marker-trait association analysis of functional gene markers for provitamin A levels across diverse tropical yellow maize inbred lines. BMC Plant Biol. **13**: 227.

Azmach, G., Worku, M., Wolde, L., Abera, W., Tadesse, B., Keno, T., Chibsa, T., Spillane, C., and Menkir, A. (2011). Development of improved yellow maize germplasm in Ethiopia. In Meeting the Challenges of Global Climate Change and Food Security Through Innovative Maize Research. Proceedings of the National Maize Workshop of Ethiopia, B.M. Worku, M.; Twumasi Afriyie, S.; Wolde, L.; Tadesse, B.; Demisie, G.; Bogale, G.; Wegary, D.; Prasanna, ed (CIMMYT: Addis Ababa), pp. 58–65.

Babu, R., Rojas, N.P., Gao, S., Yan, J., and Pixley, K. (2013). Validation of the effects of molecular marker polymorphisms in LcyE and CrtRB1 on provitamin A concentrations for 26 tropical maize populations. Theor. Appl. Genet. **126**: 389–399.

Bagge, M., Xia, X., and Lubberstedt, T. (2007). Functional markers in wheat. Curr. Opin. Plant Biol. **10**: 211–216.

Ballottari, M. and Alcocer, M. (2014). Regulation of photosystem I light harvesting by zeaxanthin. Proc. Natl. Acad. Sci. U. S. A. **111**: E2431–E24318.

Barrett, J.C., Fry, B., Maller, J., and Daly, M.J. (2005). Haploview: Analysis and visualization of LD and haplotype maps. Bioinformatics **21**: 263–265.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Statisitical Soc. Ser. B **57**: 289–300.

Bevan, M.W. and Uauy, C. (2013). Genomics reveals new landscapes for crop improvement. Genome Biol. **14**: 206.

Biesalski, H. (2013). First international conference on hidden hunger, Hohenheim, Stuttgart, Germany March 6 – 9, 2013. Food Secur. **5**: 457–473.

Birchler, J.A., Yao, H., Chudalayandi, S., Vaiman, D., and Veitia, R.A. (2010). Heterosis. Plant Cell **22**: 2105–2112.

Black, R.E., Allen, L.H., Bhutta, Z. a, Caulfield, L.E., de Onis, M., Ezzati, M., Mathers, C., and Rivera, J. (2008). Maternal and child undernutrition: Global and regional exposures and health consequences. Lancet **371**: 243–260.

Blaner, W.S. (2013). The fat-soluble vitamins 100 years later: where are we now? J. Lipid Res. **54**: 1716–1718.

Bloem, M.W.M., Semba, R.D.R., and Kraemer, K. (2010). Castel Gandolfo workshop: an introduction to the impact of climate change, the economic crisis, and the increase in the food prices on malnutrition. J. Nutr. **140**: 132S–135S.

Blomhoff, R. and Blomhoff, H.K. (2006). Overview of retinoid metabolism and function. J. Neurobiol. **66**: 606–630.

Bolger, M.E., Weisshaar, B., Scholz, U., Stein, N., Usadel, B., and Mayer, K.F.X. (2014). Plant genome sequencing - applications for crop improvement. Curr. Opin. Biotechnol. **26**: 31–37.

Bouis, H.E., Hotz, C., McClafferty, B., Meenakshi, J. V., and Pfeiffer, W.H. (2011). Biofortification: A new tool to reduce micronutrient malnutrition. Food Nutr. Bull. **32**: S31–S40.

Bouis, H.E. and Welch, R.M. (2010). Biofortification – A sustainable agricultural strategy for reducing micronutrient malnutrition in the Global South. Crop Sci. **50**: S20–S32.

Brachi, B., Morris, G.P., and Borevitz, J.O. (2011). Genome-wide association studies in plants: the missing heritability is in the field. Genome Biol. **12**: 232.

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**: 2633–2635.

Breseghello, F. and Sorrells, M.E. (2006). Association mapping of kernel size and milling quality in wheat (Triticum aestivum L.) cultivars. Genetics **172**: 1165–1177.

Burt, A.J., Grainger, C.M., Shelp, B.J., and Lee, E.A. (2011a). Heterosis for carotenoid concentration and profile in maize hybrids. Genome **1004**: 993–1004.

Burt, A.J., Grainger, C.M., Smid, M.P., Shelp, B.J., and Lee, E.A. (2011b). Allele mining of exotic maize germplasm to enhance macular carotenoids. Crop Sci. **51**: 991–1004.

Bush, W.S. and Moore, J.H. (2012). Genome-wide association studies. PLoS Comput. Biol. **8**: e1002822.

Cazzonelli, C.I. (2011). Carotenoids in nature: insights from plants and beyond. Funct. Plant Biol. **38**: 833–847.

Cazzonelli, C.I. and Pogson, B.J. (2010). Source to sink: Regulation of carotenoid biosynthesis in plants. Trends Plant Sci. **15**: 266–274.

Chandler, K., Lipka, A.E., Owens, B.F., Li, H., Buckler, E.S., Rocheford, T., and Gore, M. a. (2013). Genetic analysis of visually scored orange kernel color in maize. Crop Sci. **53**: 189–200.

Chen, Y., Lin, C.-H., and Sabatti, C. (2006). Volume measures for linkage disequilibrium. BMC Genet. **7**: 54.

Cook, J.P., McMullen, M.D., Holland, J.B., Tian, F., Bradbury, P., Ross-Ibarra, J., Buckler, E.S., and Flint-Garcia, S. a. (2012). Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels. Plant Physiol. **158**: 824–834.

Copenhagen Consensus (2012). Third copenhagen consensus outcome document. Retreived from URL, http://www.copenhagenconsensus.com/copenhagen-consensus-iii/outcome . Last accessed in August 2014.

Crossa, J. et al. (2013). Genomic prediction in maize breeding populations with genotyping-by-sequencing. G3 (Bethesda). **3**: 1903–1926.

Cunningham, F.X. (2002). Regulation of carotenoid synthesis and accumulation in plants. Symp. A Q. J. Mod. Foreign Lit. **74**: 1409–1417.

Cunningham, F.X. and Gantt, E. (1998). Genes and enzymes of carotenoid biosynthesis in plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. **49**: 557–583.

Cuttriss, A.J., Cazzonelli, C.I., Wurtzel, E.T., and Pogson, B.J. (2011). Carotenoids. Adv. Bot. Res. **58**: 1–36.

D'Ambrosio, D.N., Clugston, R.D., and Blaner, W.S. (2011). Vitamin A metabolism: An update. Nutrients **3**: 63–103.

Davey, J.W., Hohenlohe, P. a, Etter, P.D., Boone, J.Q., Catchen, J.M., and Blaxter, M.L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat. Rev. Genet. **12**: 499–510.

Dey, S.S., Singh, N., Bhatia, R., Parkash, C., and Chandel, C. (2014). Genetic combining ability and heterosis for important vitamins and antioxidant pigments in cauliflower (Brassica oleracea var. botrytis L.). Euphytica **195**: 169–181.

Edae, E. a, Byrne, P.F., Haley, S.D., Lopes, M.S., and Reynolds, M.P. (2014). Genome-wide association mapping of yield and yield components of spring wheat under contrasting moisture regimes. Theor. Appl. Genet. **127**: 791–807.

Egesel, C.O., Wong, J., Lambert, R., and Rocheford, T.R. (2003a). Gene dosage effects on carotenoid concentration in maize grain. Plant Cell **48**: 183–190.

Egesel, C.O., Wong, J.C., Lambert, R.J., and Rocheford, T.R. (2003b). Combining ability of maize inbreds for carotenoids and tocopherols. Crop Sci. **43**: 818–823.

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**: 1–10.

Eroglu, A. and Harrison, E.H. (2013). Carotenoid metabolism in mammals, including man: formation, occurrence, and function of apocarotenoids. J. Lipid Res. **54**: 1719–30.

Ersoz, E.S., Yu, J., and Buckler, E.S. (2007). Applications of linkage disequilibrium and association mapping in crop plants. Genomics **1**: 97–119.

FAO/WHO (2004). Vitamin and Mineral Requirements in Human Nutrition 2nd ed. (World Health Organization: Geneva).

FAOSTAT (2012). FAOSTAT crop production database: Maize producation in tonnes and area coverage in hecatares, World and Africa data. Queried from URL, http://faostat3.fao.org/browse/Q/*/E. Last accessed in August 2014.

FAOSTAT (2011). FAOSTAT Food balances database: crops primary equivalent in kilocalorie per capita for maize in Africa. Queried from URL, http://faostat3.fao.org/browse/FB/*/E. Last accessed in August 2014.

Farre, G., Sanahuja, G., Naqvi, S., Bai, C., and Capell, T. (2010). Travel advice on the road to carotenoids in plants. Plant Sci. **179**: 28–48.

Flint-Garcia, S. a, Thornsberry, J.M., and Buckler, E.S. (2003). Structure of linkage disequilibrium in plants. Annu. Rev. Plant Biol. **54**: 357–374.

Ford, R.H. (2000). Inheritance of Kernel Color in Corn: Explanations & Investigations. Am. Biol. Teach. **62**: 181–188.

Fraser, P.D. and Bramley, P.M. (2004). The biosynthesis and nutritional uses of carotenoids. Prog Lipid Res **43**: 228–265.

Fu, Z., Chai, Y., Zhou, Y., Yang, X., Warburton, M.L., Xu, S., Cai, Y., Zhang, D., Li, J., and Yan, J. (2013). Natural variation in the sequence of PSY1 and frequency of favorable polymorphisms among tropical and temperate maize germplasm. Theor. Appl. Genet. **126**: 923–935.

Gabriel, S.B. et al. (2002). The structure of haplotype blocks in the human genome. Science **296**: 2225–2229.

Gannon, B., Kaliwile, C., Arscott, S., Schmaelzle, S., Masi, C., Chileshe, J., and Tanumihardjo, S. (2014). Biofortified orange maize is as efficacious as a vitamin A supplement in Zambian children even on the background of high liver reserves of vitamin A : a community-based , randomoised placebo-controlled trial. Am. J. Clin. Nutr. **28**: 1541–1550.

Giuliano, G. (2014). Plant carotenoids: genomics meets multi-gene engineering. Curr. Opin. Plant Biol. **19**: 111–117.

Glaubitz, J.C., Casstevens, T.M., Lu, F., Harriman, J., Elshire, R.J., Sun, Q., and Buckler, E.S. (2014). TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. PLoS One **9**: e90346.

Grogan, C.O., Blessin, C.W., Dimler, R.J., and Campbell, C.M. (1963). Parental influence on xanthophylls and carotenes in corn. Crop Sci. **3**: 213–214.

De Groote, H. and Kimenju, S.C. (2008). Comparing consumer preferences for color and nutritional quality in maize: Application of a semi-double-bound logistic model on urban consumers in Kenya. Food Policy **33**: 362–370.

De Groote, H., Kimenju, S.C., and Morawetz, U.B. (2011). Estimating consumer willingness to pay for food quality with experimental auctions: the case of yellow versus fortified maize meal in Kenya. Agric. Econ. **42**: 1–16.

Grune, T., Lietz, G., Palou, A., Ross, a C., Stahl, W., Tang, G., Thurnham, D., Yin, S., and Biesalski, H.K. (2010). Beta-carotene is an important vitamin A source for humans. J. Nutr. **140**: 2268S–2285S.

Gupta, P.K. and Rustgi, S. (2004). Molecular markers from the transcribed/expressed region of the genome in higher plants. Funct. Integr. Genomics **4**: 139–162.

Hallauer, A.R. (2007). History, contribution, and future of quantitative genetics in plant breeding: Lessons from maize. Crop Sci. **47**: S4–S19.

Hallauer, A.R. and Carena, M.J. (2009). Maize Breeding. In Cereals, M.J. Carena, ed, Handbook of Plant Breeding. (Springer US: Fargo), pp. 3–98.

Hallauer, A.R., Filho, J.B.M., and Carena, M.J. (2010). Quantitative Genetics in Maize Breeding 3rd ed. J. Prohens, F. Nuez, and M.J. Carena, eds (Springer Science+Business Media: New York).

Hamblin, M.T., Buckler, E.S., and Jannink, J.L. (2011). Population genetics of genomics-based crop improvement methods. Trends Genet. **27**: 98–106.

Harjes, C.E., Rocheford, T.R., Bai, L., Brutnell, T.P., Vallabhaneni, R., Williams, M., Wurtzel, E.T., Kandianis, C.B., Sowinski, S.G., Stapleton, A.E., Yan, J., and Buckler, E.S. (2008). Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. Science **319**: 330–333.

Harrison, E.H. (2012). Mechanisms involved in the intestinal absorption of dietary vitamin A and provitamin A carotenoids. Biochim. Biophys. Acta - Mol. Cell Biol. Lipids **1821**: 70–77.

HarvestPlus (2014). Biofortification Progress Briefs. Retreived from URL, http://www.harvestplus.org/sites/default/files/Biofortification_Progress_Briefs_August2014_WEB_0.pdf. Last accessed in August 2014.

HarvestPlus (2012). Bridging the Delta 2012 Annual Report. Retreived from URL, http://www.harvestplus.org/content/harvestplus-2012-annual-report-released. Last accessed in August 2014.

Haskell, M.J. (2012). The challenge to reach nutritional adequacy for vitamin A: β-carotene bioavailability and conversion—evidence in humans. Am. J. Clin. Nutr. **96**: 1193S–1203S.

Hauge, S.M. (1930). An inheritance study of the distribution of vitamin a in maize: II. Vitamin a in hybrid red maize. J. Biol. Chem. **86**: 161–165.

Hauge, S.M. and Trost, J.F. (1928). An inheritance study of the distribution of vitamin A in maize. J. Biol. Chem. **80**: 107–114.

Hauge, S.M. and Trost, J.F. (1930). An inheritance study of the distribution of vitamin a in maize: III. vitamin a content in relation to yellow endosperm. J. Biol. Chem. **86**: 167–172.

Heslot, N., Rutkoski, J., Poland, J., Jannink, J.L., and Sorrells, M.E. (2013). Impact of marker ascertainment bias on genomic selection accuracy and estimates of genetic diversity. PLoS One **8**: e74612.

Hirschberg, J. (2001). Carotenoid biosynthesis in flowering plants. Curr. Opin. Plant Biol. **4**: 210–218.

Hirschberg, J., Cohen, M., Harker, M., Lotan, T., Mann, V., and Pecker, I. (1997). Molecular genetics of the carotenoid biosynthesis pathway in plants and algae. Pure Appl. Chem. **69**: 2151–2158.

Holland, J.B., Nyquist, W.E., and Cervantes-Martinez, C.T. (2003). Estimating and interpreting heritability for plant breeding.pdf. Plant Breed. Rev. **22**: 9–112.

Howe, J.A. and Tanumihardjo, S.A. (2006). Carotenoid-biofortified maize maintains adequate vitamin a status in Mongolian gerbils. J. Nutr. **136**: 2562–2567.

Jlali, M., Graulet, B., Chauveau-Duriot, B., Godet, E., Praud, C., Nunes, C.S., Le Bihan-Duval, E., Berri, C., and Duclos, M.J. (2014). Nutrigenetics of carotenoid metabolism in the chicken: a polymorphism at the β,β-carotene 15,15'-mono-oxygenase 1 (BCMO1) locus affects the response to dietary β-carotene. Br. J. Nutr. **111**: 2079–2088.

Kang, H.M., Sul, J.H., Service, S.K., Zaitlen, N. a, Kong, S.-Y., Freimer, N.B., Sabatti, C., and Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. Nat. Genet. **42**: 348–354.

Kean, E.G., Hamaker, B.R., and Ferruzzi, M.G. (2008). Carotenoid bioaccessibility from whole grain and degermed maize meal products. J. Agric. Food Chem. **56**: 9918–9926.

Khatkar, M.S., Nicholas, F.W., Collins, A.R., Zenger, K.R., Cavanagh, J. a L., Barris, W., Schnabel, R.D., Taylor, J.F., and Raadsma, H.W. (2008). Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high-density SNP panel. BMC Genomics **9**: 187.

Khorzoght, E.G., Soltanloo, H., Ramezanpour, S.S., and Arabi, M.K. (2010). Combining ability analysis and estimation of heterosis for resistance to head blight caused by Fusarium graminearum in spring wheat. Aust. J. Crop Sci. **4**: 626–632.

Krinsky, N.I. and Johnson, E.J. (2005). Carotenoid actions and their relation to health and disease. Mol. Aspects Med. **26**: 459–516.

Lawrence, C.J., Dong, Q., Polacco, M.L., Seigfried, T.E., and Brendel, V. (2004). MaizeGDB, the community database for maize genetics and genomics. Nucleic Acids Res. **32**: D393–D397.

Li, H. et al. (2013a). Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. Nat. Genet. **45**: 43–50.

Li, Q., Yang, X., Xu, S., Cai, Y., Zhang, D., Han, Y., Li, L., Zhang, Z., Gao, S., Li, J., and Yan, J. (2012). Genome-wide association studies identified three independent polymorphisms associated with α-tocopherol content in maize kernels. PLoS One **7**: e36807.

Li, R., Xiao, L.H., Wang, J., Lu, Y.L., Rong, T.Z., Pan, G.T., Wu, Y.Q., Tang, Q., Lan, H., and Cao, M.J. (2013b). Combining ability and parent-offspring correlation of maize (Zea may L.) grain β-Carotene content with a complete diallel. J. Integr. Agric. **12**: 19–26.

Li, S., Nugroho, A., Rocheford, T., and White, W.S. (2010). Vitamin A equivalence of the β-carotene in β-carotene- biofortified maize porridge consumed by women. Am. J. Clin. Nutr. **92**: 1105–1112.

Li, S., Tayie, F.A.K., Young, M.F., Rocheford, T., and White, W.S. (2007). Retention of provitamin A carotenoids in high beta-carotene maize (Zea mays) during traditional African household processing. J. Agric. Food Chem. **55**: 10744–10750.

Lipka, A.E. et al. (2013a). High-resolution mapping of tocochromanol and carotenoid grain traits via NAM-GWAS reveals distinct genetic architectures in maize more QTL found for tocochromanols results suggest simpler genetic architecture for carotenoids more carotenoid QTL coincided. Maize Genet. Conf. Abstr. **55**: 282.

Lipka, A.E., Gore, M. a, Magallanes-Lundback, M., Mesberg, A., Lin, H., Tiede, T., Chen, C., Buell, C.R., Buckler, E.S., Rocheford, T., and DellaPenna, D. (2013b). Genome-wide association study and pathway-level analysis of tocochromanol levels in maize grain. G3 Genes| Genomes| Genet. **3**: 1287–1299.

Lipka, A.E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P.J., Gore, M. a., Buckler, E.S., and Zhang, Z. (2012). GAPIT: Genome association and prediction integrated tool. Bioinformatics **28**: 2397–2399.

Liu, K. and Muse, S. V (2005). PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics **21**: 2128–2129.

Lu, Y., Shah, T., Hao, Z., Taba, S., Zhang, S., Gao, S., Liu, J., Cao, M., Wang, J., Prakash, a. B., Rong, T., and Xu, Y. (2011). Comparative SNP and haplotype analysis reveals a higher genetic

diversity and rapider LD decay in tropical than temperate germplasm in maize. PLoS One **6**: e24861.

M'mboyi, F., Mugo, S., Mwimali, M., and Ambani, L. (2010). Maize Production and Improvement in Sub-Saharan Africa 1st ed. (African Biotechnology Stakeholders Forum (ABSF): Nairobi).

Mamoon, A., Subauste, A., Subauste, M.C., and Subauste, J. (2014). Retinoic acid regulates several genes in bile acid and lipid metabolism via upregulation of small heterodimer partner in hepatocytes. Gene **550**: 165–170.

Mangelsdorf, P.C. and Fraps, G.S. (1931). A direct quantitative relationship between vitamin A in corn and the number of genes for yellow pigmentation. Science (80-.). **73**: 241–242.

Mayer, J.E., Pfeiffer, W.H., and Beyer, P. (2008). Biofortified crops to alleviate micronutrient malnutrition. Curr. Opin. Plant Biol. **11**: 166–170.

Maziya-Dixon, B., Kling, J.G., Menkir, a., and Dixon, a. (2000). Genetic variation in total carotene, iron, and zinc contents of maize and cassava genotypes. Food Nutr. Bull. **21**: 419–422.

Meenakshi, J. V., Banerji, A., Manyong, V., Tomlins, K., Hamukwala, P., Zulu, R., and Mungoma, C. (2010a). Consumer acceptance of provitamin A orange maize in rural Zambia (International Food Policy Research Institute (IFPRI): Washington DC).

Meenakshi, J. V., Johnson, N.L., Manyong, V.M., Degroote, H., Gonzalez, C., Garci, J., and Meng, E. (2010b). How cost-effective is biofortification in combating micronutrient malnutrition? An ex ante assessment. World Dev. **38**: 64–75.

Meier, S., Tzfadia, O., Vallabhaneni, R., Gehring, C., and Wurtzel, E.T. (2011). A transcriptional analysis of carotenoid, chlorophyll and plastidial isoprenoid biosynthesis genes during development and osmotic stress responses in Arabidopsis thaliana. BMC Syst. Biol. **5**: 77.

Menkir, A., Gedil, M., Tanumihardjo, S., Adepoju, A., and Bossey, B. (2014). Carotenoid accumulation and agronomic performance of maize hybrids involving parental combinations from different marker-based groups. Food Chem. **148**: 131–137.

Menkir, A., Liu, W., White, W.S., Maziya-Dixon, B., and Rocheford, T. (2008). Carotenoid diversity in tropical-adapted yellow maize inbred lines. Food Chem. **109**: 521–529.

Menkir, A. and Mazia-Dixon, B. (2004). Influence of genotype and environment on β-carotene content of tropical yellow-endosperm maize genotypes. Maydica **49**: 313–318.

Messias, R.D.S., Galli, V., e Silva, S.D.D.A., and Rombaldi, C.V. (2014). Carotenoid biosynthetic and catabolic pathways: Gene expression and carotenoid content in grains of maize landraces. Nutrients **6**: 546–563.

Meyers, K.J. et al. (2014). Genetic evidence for role of carotenoids in age-related macular degeneration in the Carotenoids in Age-Related Eye Disease Study (CAREDS). Invest. Ophthalmol. Vis. Sci. **55**: 587–599.

Miller, B. and Welch, R. (2013). Food system strategies for preventing micronutrient malnutrition. Food Policy **42**: 115–128.

Mishra, P. and Singh, N.K. (2010). Spectrophotometric and TLC based characterization of kernel carotenoids in short duration maize. Maydica **55**: 95–100.

Moise, A.R., Al-Babili, S., and Wurtzel, E.T. (2014). Mechanistic aspects of carotenoid biosynthesis. Chem. Rev. **114**: 164–193.

Moose, S.P. and Mumm, R.H. (2008). Molecular plant breeding as the foundation for 21st century crop improvement. Plant Physiol. **147**: 969–977.

Moran, N.A. and Jarvik, T. (2010). Lateral transfer of genes from fungi underlies carotenoid production in aphids. Science **328**: 624–627.

Msangi, S., Sulser, T.B., Bouis, A., Hawes, D., and Batka, M. (2010). Integrated economic modeling of global and regional micronutrient security (International Food Policy Research Institute (IFPRI): NW, Washington).

Muzhingi, T., Langyintuo, A.S., Malaba, L.C., and Banziger, M. (2008a). Consumer acceptability of yellow maize products in Zimbabwe. Food Policy **33**: 352–361.

Muzhingi, T., Yeum, K.J., Russell, R.M., Johnson, E.J., Qin, J., and Tang, G. (2008b). Determination of carotenoids in Yellow Maize, the effects of saponification and food preparations. Int. J. Vitam. Nutr. Res. **78**: 112–120.

Myles, S., Peiffer, J., Brown, P.J., Ersoz, E.S., Zhang, Z., Costich, D.E., and Buckler, E.S. (2009). Association mapping: critical considerations shift from genotyping to experimental design. Plant Cell **21**: 2194–2202.

Nei, M. (1972). Genetic distance between populations. Am. Nat. **106**: 283–292.

Nestel, P., Bouis, H.E., Meenakshi, J. V, Pfeiffer, W., and Food, S. (2006). Biofortification of staple food crops. J. Nutr. **136**: 1064–1067.

Newell-McGloughlin, M. (2008). Nutritionally improved agricultural crops. Plant Physiol. **147**: 939–953.

Nguyen, P.H., Nguyen, H., Gonzalez-Casanova, I., Copeland, E., Strizich, G., Lowe, A., Pham, H., Truong, T. V., Nguyen, S., Martorell, R., and Ramakrishnan, U. (2014). Micronutrient intakes among women of reproductive age in Vietnam. PLoS One **9**: e89504.

Nuss, E.T. and Tanumihardjo, S. a. (2010). Maize: A paramount staple crop in the context of global nutrition. Compr. Rev. Food Sci. Food Saf. **9**: 417–436.

Olson, J. a (1989). Provitamin A function of carotenoids: the conversion of beta-carotene into vitamin A. J. Nutr. **119**: 105–8.

Otten, J.J., Hellwig, J.P., and Meyers, L.D. eds (2006). Dietary reference intakes: The essential guide to nutrient requirements (National Academies Press: Washington DC).

Palaisa, K. a, Morgante, M., Williams, M., and Rafalski, A. (2003). Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci. Plant Cell **15**: 1795–1806.

Palmer, A., Chileshe, J., Schulze, K., Siamusantu, W., Klemm, R., Molobeka, N., Barffour, M., Kalungwana, N., Arguello, M., and West, K. (2014). Impact of provitamin A biofortified maize on vitamin A status in Zambian children. FASEB J. **28**.

Pasam, R.K., Sharma, R., Malosetti, M., van Eeuwijk, F. a, Haseneyer, G., Kilian, B., and Graner, A. (2012). Genome-wide association studies for agronomical traits in a world wide spring barley collection. BMC Plant Biol. **12**: 16.

Peninah, N., Richard, E., and Joseph, K. (2014). Combining ability for beta-carotene and important quantitative traits in a cassava. J. Plant Breed. Crop Sci. **6**: 24–30.

Perumalsamy, S., Bharani, M., Sudha, M., Nagarajan, P., Arul, L., Saraswathi, R., Balasubramanian, P., and Ramalingam, J. (2010). Functional marker-assisted selection for bacterial leaf blight resistance genes in rice (Oryza sativa L.). Plant Breed. **129**: 400–406.

Pfeiffer, W.H. and McClafferty, B. (2007). HarvestPlus: Breeding crops for better nutrition. In Crop Science, pp. S88–S105.

Pillay, K., Derera, J., Siwela, M., Veldman, F.J., and Nutrition, H. (2011). Consumer acceptance of yellow, provitamin a-biofortified maize in KwaZulu-Natal. South African J. Clin. Nutr. **24**: 186–191.

Pixley, K., Rojas, N.N., Babu, R., Mutale, R., Surles, R., and Simpungwe, E. (2013). Biofortification of maize with provitamin a carotenoids. In Carotenoids and Human Health, S.A. Tanumihardjo, ed, Nutrition and Health. (Humana Press), pp. 271–292.

Poland, J. a and Rife, T.W. (2012). Genotyping-by-sequencing for plant breeding and genetics. Plant Genome J. **5**: 92–102.

Poletti, S., Gruissem, W., and Sautter, C. (2004). The nutritional fortification of cereals. Curr. Opin. Biotechnol. **15**: 162–165.

Qaim, M., Stein, A.J., and Meenakshi, J. V. (2007). Economics of biofortification. Agric. Econ. **37**: 119–133.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. URL, http://www.R-project.org. Last accessed August 2014.

Ranum, P., Peña-Rosas, J.P., Garcia-Casal, M.N., and Pe, J.P. (2014). Global maize production, utilization, and consumption. Ann. N. Y. Acad. Sci. **1312**: 105–112.

Reboul, E. (2013). Absorption of vitamin A and carotenoids by the enterocyte: Focus on transport proteins. Nutrients **5**: 3563–3581.

Rodríguez-Concepción, M. (2010). Supply of precursors for carotenoid biosynthesis in plants. Arch. Biochem. Biophys. **504**: 118–122.

Ruel, M.T. (2003). Operationalizing dietary diversity: A review of measurement issues and research priorities. J. Nutr. **133**: 3911S–3926S.

Saari, J.C. (2012). Vitamin A metabolism in rod and cone visual cycles. Annu. Rev. Nutr. **32**: 125–145.

SAS Institute (2012). SAS system for windows. Release 9.3. SAS Institute Inc. Cary, North Carolina, USA

Saxton, A. (1998). A macro for converting mean separation output to letter groupings in Proc Mixed. In Proceedings of the 23rd SAS Users Group International (SAS Institute, Cary, NC), pp. 1243–1246.

Schnable, P.S. et al. (2009). The B73 maize genome: complexity, diversity, and dynamics. Science (80-.). **326**: 1112–1115.

Sekhon, R.S., Lin, H., Childs, K.L., Hansey, C.N., Robin Buell, C., De Leon, N., and Kaeppler, S.M. (2011). Genome-wide atlas of transcription during maize development. Plant J. **66**: 553–563.

Dela Seña, C., Riedl, K.M., Narayanasamy, S., Curley, R.W., Schwartz, S.J., and Harrison, E.H. (2014). The human enzyme that converts dietary provitamin A carotenoids to vitamin A is a dioxygenase. J. Biol. Chem. **289**: 13661–13666.

Senete, C.T., Guimarães, P.E.D.O., Paes, M.C.D., and Souza, J.C. De (2011). Diallel analysis of maize inbred lines for carotenoids and grain yield. Euphytica **182**: 395–404.

Sherwin, J.C., Reacher, M.H., Dean, W.H., and Ngondi, J. (2012). Epidemiology of vitamin A deficiency and xerophthalmia in at-risk populations. Trans. R. Soc. Trop. Med. Hyg. **106**: 205–214.

Shetty, P. (2011). Addressing Micronutrient Malnutrition to Achieve Nutrition Security. In Combating Micronutrient Deficiencies: Food-Based Approaches, L. Thompson, B.;Amoroso, ed (CABI: Wallingford), pp. 28–39.

Shinopoulos, K.E., Yu, J., Nixon, P.J., and Brudvig, G.W. (2014). Using site-directed mutagenesis to probe the role of the D2 carotenoid in the secondary electron-transfer pathway of photosystem II. Photosynth. Res. **120**: 141–152.

Shirasawa, K., Fukuoka, H., Matsunaga, H., Kobayashi, Y., Kobayashi, I., Hirakawa, H., Isobe, S., and Tabata, S. (2013). Genome-wide association studies using single nucleotide polymorphism markers developed by re-sequencing of the genomes of cultivated tomato. DNA Res. **20**: 593–603.

Shumskaya, M. and Wurtzel, E.T. (2013). The carotenoid biosynthetic pathway: Thinking in all dimensions. Plant Sci. **208**: 58–63.

Singh, B.K., Sharma, S.R., and Singh, B. (2011). Combining ability for antioxidants and economic traits in cabbage. **68**: 490–497.

Singh, M., Gupta, S., and Parsad, R. (2012). Genetic crosses experiments. In Design and Analysis of Experiments: Special Designs and Applications, K. Hinkelmann, ed (John Wiley & Sons, Inc.).

Smale, M., Byerlee, D., and Jayne, T. (2013). Maize Revolutions in Sub-Saharan Africa. In An African Green Revolution, K. Otsuka and D.F. Larson, eds, Policy Research Working Paper 5659. (Springer Netherlands), pp. 165–195.

Soto-Cerda, B.J. and Cloutier, S. (2012). Association Mapping in Plant Genomes. In Genetic Diversity in Plantslants, M. Caliskan, ed (InTech), pp. 29–54.

Stamatakis, K., Tsimilli-michael, M., and Papageorgiou, G.C. (2014). On the question of the light-harvesting role of β-carotene in photosystem II and photosystem I core complexes. Plant Physiol. Biochem. **81**: 121–127.

Stevens, R. and Winter-Nelson, A. (2008). Consumer acceptance of provitamin A-biofortified maize in Maputo, Mozambique. Food Policy **33**: 341–351.

Suwarno, W.B. (2012). Combining ability, association mapping, and genomic predictions for provitamin A carotenoid concentrations in tropical maize (Zea mays L.). PhD dessertation.

Suwarno, W.B., Pixley, K. V., Palacios-Rojas, N., Kaeppler, S.M., and Babu, R. (2014). Formation of heterotic groups and understanding genetic effects in a provitamin a biofortified maize breeding program. Crop Sci. **54**: 14–24.

Tabangin, M.E., Woo, J.G., and Martin, L.J. (2009). The effect of minor allele frequency on the likelihood of obtaining false positives. BMC Proc. **3**: S41.

Tang, G. (2010). Bioconversion of dietary provitamin A carotenoids to vitamin A in humans. Am. J. Clin. Nutr. **91**: 1468S–1473S.

Tanksley, S.D. (1993). Mapping polygenes. Annu. Rev. Genet. **27**: 205–233.

The World Bank (2006). Repositioning Mutrition as Central to Development: A Strategy for Large-Scale Action (The World Bank: Washington DC).

Thornsberry, J.M., Goodman, M.M., Doebley, J., Kresovich, S., Nielsen, D., and Buckler, E.S. (2001). Dwarf8 polymorphisms associate with variation in flowering time. Nat. Genet. **28**: 286–289.

Tracewell, C.A., Cua, A., Stewart, D.H., Bocian, D.F., and Brudvig, G.W. (2001). Characterization of carotenoid and chlorophyll photooxidation in photosystem II. Biochemistry **40**: 193–203.

Tulchinsky, T. (2010). Micronutrient deficiency conditions: global health issues. Public Heal. Rev **32**: 243–255.

Underwood, B.A. (2000). Overcoming micronutrient deficiencies in developing countries: Is there a role for agriculture? Food Nutr. Bull. **21**: 356–360.

Unnevehr, L., Pray, C., and Paarlberg, R. (2007). Addressing micronutrient deficiencies: alternative interventions and technologies. AgBioForum **10**: 124–134.

UNSCN (2010). Progress in Nutrition, 6th Report on the World Nutrition Situation (United Nations Standing Committee on Nutrition: Geneva).

US Institute of Medicine (2001). Dietary Reference Intakes for Vitamin A, Vitamin K, Arsenic, Boron, Chromium, Copper, Iodine, Iron, Manganese, Molybdenum, Nickel, Silicon, Vanadium, and Zinc (The National Academies Press: Washington, DC).

Vallabhaneni, R., Bradbury, L.M.T., and Wurtzel, E.T. (2010). The carotenoid dioxygenase gene family in maize, sorghum, and rice. Arch. Biochem. Biophys. **504**: 104–111.

Vallabhaneni, R., Gallagher, C.E., Licciardello, N., Cuttriss, A.J., Quinlan, R.F., and Wurtzel, E.T. (2009). Metabolite sorting of a germplasm collection reveals the hydroxylase3 locus as a new target for maize provitamin A biofortification. Plant Physiol. **151**: 1635–1645.

Vallabhaneni, R. and Wurtzel, E.T. (2009). Timing and biosynthetic potential for carotenoid accumulation in genetically diverse germplasm of maize. Plant Physiol. **150**: 562–572.

VanRaden, P.M. (2008). Efficient methods to compute genomic predictions. J. Dairy Sci. **91**: 4414–4423.

Varshney, R.K., Graner, A., and Sorrells, M.E. (2005). Genomics-assisted breeding for crop improvement. Trends Plant Sci. **10**: 621–630.

Varshney, R.K., Mahendar, T., Aggarwal, R.K., and Börner, A. (2007). Genic molecular markers in plants: development and application. In Genomics Assisted Crop Improvement, R. Varshney and R. Tuberosa, eds (Springer Netherlands), pp. 13–29.

Varshney, R.K., Nayak, S.N., May, G.D., and Jackson, S. a. (2009). Next-generation sequencing technologies and their implications for crop genetics and breeding. Trends Biotechnol.

Varshney, R.K., Terauchi, R., and McCouch, S.R. (2014). Harvesting the promising fruits of genomics: Applying genome sequencing technologies to crop breeding. PLoS Biol. **12**: e1001883.

Vignesh, M., Hossain, F., Nepolean, T., Saha, S., Agrawal, P.K., Guleria, S.K., Prasanna, B.M., and Gupta, H.S. (2012). Genetic variability for kernel β-carotene and utilization of crtRB1 3'TE gene for biofortification in maize (Zea mays L.). Indian J. Genet. Plant Breed. **72**: 189–194.

Vogel, J.T., Tan, B.C., McCarty, D.R., and Klee, H.J. (2008). The carotenoid cleavage dioxygenase 1 enzyme has broad substrate specificity, cleaving multiple carotenoids at two different bond positions. J. Biol. Chem. **283**: 11364–11373.

Wang, G., Leonard, J.M., von Zitzewitz, J., James Peterson, C., Ross, A.S., and Riera-Lizarazu, O. (2014). Marker-trait association analysis of kernel hardness and related agronomic traits in a core collection of wheat lines. Mol. Breed. **34**: 177–184.

Wang, K., Li, M., and Bucan, M. (2007). Pathway-based approaches for analysis of genomewide association studies. Am. J. Hum. Genet. **81**: 1278–1283.

Wang, K., Li, M., and Hakonarson, H. (2010). Analysing biological pathways in genome-wide association studies. Nat. Publ. Gr. **11**: 843–854.

Wang, M., Yan, J., Zhao, J., Song, W., Zhang, X., Xiao, Y., and Zheng, Y. (2012). Genome-wide association study (GWAS) of resistance to head smut in maize. Plant Sci. **196**: 125–131.

Wang, Q. and Dooner, H.K. (2006). Remarkable variation in maize genome structure inferred from haplotype diversity at the bz locus. Proc. Natl. Acad. Sci. U. S. A. **103**: 17644–17649.

Welch, R.M. and Graham, R.D. (1999). A new paradigm for world agriculture: meeting human needs: Productive, sustainable, nutritious. F. Crop. Res. **60**: 1–10.

West JR, Keith, Darnton-Hill, I. (2008). Vitamin A deficiency. In Nutrition and Health in Developing Countries, R.D. Semba and M.W. Bloem, eds, Nutrition and Health Series. (Humana Press: Totowa, NJ), pp. 377–433.

West, K.K.P. and Mehra, S. (2010). Vitamin A intake and status in populations facing economic stress. J. Nutr. **140**: 201S–207S.

WHO (2003). Diet, nutrition and the prevention of chronic diseases: Report of a joint WHO/FAO expert consultation. World Health Organ. Tech. Rep. Ser. **916**.

WHO (2009). Global prevalence of vitamin A deficiency in populations at risk 1995-2005: WHO global database on vitamin A deficiency (World Health Organization: Geneva).

Wimmer, V., Albrecht, T., Auinger, H.-J., and Schön, C.-C. (2012). Synbreed: A framework for the analysis of genomic prediction data using R. Bioinformatics **28**: 2086–2087.

Wright, S. (1992). Adjusted p-values for simultaneous inference. Biometrics **48**: 1005–1013.

Wurtzel, E.T. (2007). Genomics, genetics and biochemical of maize carotenoid biosynthesis. Recent Adv. Phytochem. **38**: 85–110.

Wurtzel, E.T.E., Cuttriss, A., and Vallabhaneni, R. (2012). Maize provitamin a carotenoids, current resources, and future metabolic engineering challenges. Front. Plant Sci. **3**.

Xu, Y., Lu, Y., Xie, C., Gao, S., Wan, J., and Prasanna, B. (2012). Whole-genome strategies for marker-assisted plant breeding. Mol. Breed. **29**: 833–854.

Xu, Y., Skinner, D.J., Wu, H., Palacios-Rojas, N., Araus, J.L., Yan, J., Gao, S., Warburton, M.L., and Crouch, J.H. (2009). Advances in maize genomics and their value for enhancing genetic gains from breeding. Int. J. Plant Genomics **2009**.

Yan, J. et al. (2010). Rare genetic variation at Zea mays crtRB1 increases β-carotene in maize grain. Nat. Genet. **42**: 322–327.

Yan, J., Shah, T., Warburton, M.L., Buckler, E.S., Mcmullen, M.D., and Crouch, J. (2009). Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. PLoS One **4**: e8451.

Yan, J., Warburton, M., and Crouch, J. (2011). Association mapping for enhancing maize (Zea mays L.) genetic improvement. Crop Sci. **51**: 433.

Young, P.R., Lashbrooke, J.G., Alexandersson, E., Jacobson, D., Moser, C., Velasco, R., and Vivier, M. a (2012). The genes and enzymes of the carotenoid metabolic pathway in Vitis vinifera L. BMC Genomics **13**: 243.

Yu, J. and Buckler, E.S. (2006). Genetic association mapping and genome organization of maize. Curr. Opin. Biotechnol. **17**: 155–160.

Yu, J., Pressoir, G., Briggs, W.H., Vroh Bi, I., Yamasaki, M., Doebley, J.F., McMullen, M.D., Gaut, B.S., Nielsen, D.M., Holland, J.B., Kresovich, S., and Buckler, E.S. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. **38**: 203–208.

Zhang, Q., Wu, C., Ren, F., Li, Y., and Zhang, C. (2012). Association analysis of important agronomical traits of maize inbred lines with SSRs. Genetics **6**: 1131–1138.

Zhang, Z., Ersoz, E., Lai, C.-Q., Todhunter, R.J., Tiwari, H.K., Gore, M. a, Bradbury, P.J., Yu, J., Arnett, D.K., Ordovas, J.M., and Buckler, E.S. (2010). Mixed linear model approach adapted for genome-wide association studies. Nat. Genet. **42**: 355–360.

Zhu, C., Gore, M., Buckler, E.S., and Yu, J. (2008). Status and prospects of association mapping in plants. Plant Genome J. **1**: 5–20.

.

Appendix 1 – Supplemental Tables

Appendix 1 – Table 1 Primer sequences for functional markers of *lcy*E, *crt*RB1 and *PSY1* tested across 130 diverse inbred lines developed at IITA

	Polymorphic site/Marker name	**Primer sequences**
LCYE (Harjes et al., 2008)	LCYE 5'TE	CGCTAGCAAGCCCATTATTTTTA CGGTATGGTTTTTGGTATACGG GAGAGGGAGACGACGAGACAC AAGCATCCGACCAAAATAACAG
	LCYE SNP (216)	GCGGCAGTGGGCGTGGAT TGAAGTACGGCTGCAGGACAACG
	LCYE 3'indel*	ACCCGTACGTCGTTCATCTC ACCCTGCGTGGTCTCAAC
crtRB1 (Yan et al., 2010)	crtRB1 5'TE	TTAGAGCCTCGACCCTCTGTG AATCCCTTTCCATGTTACGC
	crtRB1 InDel4	ACCGTCACGTGCTTCGTGCC CTTCCGCGCCTCCTTCTC
	crtRB1 3'TE	ACACCACATGGACAAGTTCG ACACTCTGGCCCATGAACAC ACAGCAATACAGGGGACCAG
PSY1 (Fu et al., 2013)	SNP7	CACGACGTTGTAAAACGACCTGCTA CTGCTAGCCTGTGAGAGCTCAT TGCGAACCAACTCACGAAC
	IDI	AGACATCACACACACACGACAC GTAACTCACCAGGCTCACTTGT

* The four primers reported by Harjes et al., (2008) did not give reliable results under our condition thus a single primer developed and tested by CIMMYT, indicated in Babu et al. (2012), was used.

Appendix 1 – Table 2 Thermocycler programs used for PCR amplification of the allele specific markers of *PSY1* (Fu et al 2013) *lcyE* (Harjes et al 2008), and *crtRB1* (Yan et al 2010)

(a) *lcyE 5'TE* PCR thermocycling profile

Step	Temperature (^{0}C)	Time (min)	Note
1	94	3	
2	94	1	
3	64	1	-1℃/cycle
4	72	1.5	From step 2:5X
5	94	1	
6	58	1	
7	72	1.5	From step 5:28X
8	72	15	
9	4	hold	

(b) *lcyE* SNP (216) *and* crtRB1 InDel4 PCR thermocycling profile

Step	Tempreture (^{0}C)	Time (min)	Note
1	94	3	
2	94	1	
3	68	0.75	-1℃/cycle
4	72	1	From step 2:10X
5	94	1	
6	58	0.75	
7	72	1	From step 5:28X
8	72	15	
9	4	hold	

(c) *lcyE* Indel PCR thermocycling profile

Step	Temperature (^{0}C)	Time (min)	Note
1	94	5	
2	94	1	
3	64	0.75	- 0.5℃/cycle
4	72	0.5	From step 2:19X
5	94	1	
6	54	0.66	
7	72	0.75	From step 5:19X
8	72	15	
9	4	hold	

(d)*crtRB15'TE* and *3'TE*PCR thermocycling profile

Step	Tempreture(^{0}C)	Time(min)	Note
1	94	5	
2	94	1	
3	64	1	- 0.5℃/cycle
4	72	1	From step 2:19X
5	94	1	
6	54	1	
7	72	1	From step 5:19X
8	72	15	
9	4	hold	

(e) *PSY1* InDel 1

Step	Tempreture(^{0}C)	Time(min)	Note
1	94	3	
2	94	1	
3	55	1	From step 2: 35X
4	72	1.5	
5	75	10	
6	4	hold	

(f) *PSY1* SNP7

Step	Tempreture(^{0}C)	Time(min)	Note
1	94	3	
2	94	1	
3	72	1	-1℃/cycle
4	72	1.5	From step 2:17X
5	94	1	
6	55	1	
7	72	1.5	From step 5:18X
8	72	10	
9	4	hold	

Appendix 1 – Table 3 Least square means (LSmeans) of carotenoid concentrations for 130 inbred lines

Entry	Lut	Zea	βcar	βcry	αcar	PVA	Tcar	Kernel colour
1	7.09	9.53	1.83	3.3	0.59	3.77	22.35	5.50
2	3.43	3.9	4.98	2.61	0.32	6.48	15.55	8.00
3	4.46	12.84	3.06	3.6	0.32	5.01	24.34	8.00
4	3.33	14.64	3.39	2.8	0.05	4.83	24.19	8.50
5	4.16	11.75	2.57	3.13	0.27	4.28	21.84	6.50
6	2.57	14.29	2.81	3.92	0.31	4.91	23.87	8.00
7	1.01	22.3	3.62	4.31	0.39	6	31.73	9.00
8	3.42	17.59	2.95	5.07	0.35	5.61	29.13	9.00
9	1.99	12.03	3.12	3.26	0.01	4.76	20.4	6.50
10	3.74	10.89	3.49	3.92	0.38	5.6	22.26	7.00
11	2.03	10.8	7.61	1.99	0.83	9.06	23.48	8.50
12	6.44	14.69	3.13	3.46	0.38	5.06	28.12	8.00
13	11.91	3.09	0.58	0.34	0.18	0.86	16.22	3.00
14	3.07	1.49	-0.04	0.08	0	-0.06	4.43	1.00
15	3.09	1.95	0.29	1.21	0.03	0.87	6.36	3.00
16	2.97	7.98	1.73	2.23	0.35	2.95	14.97	3.00
17	2.6	1.97	1.12	1.18	0.09	1.71	6.82	2.00
18	2.99	9.04	2.69	3.53	0.33	4.62	18.54	8.00
19	2.87	6.58	3.16	3.53	0.32	5.04	16.19	8.00
20	2.26	6.93	3.27	3.11	0.16	4.86	15.54	5.00
21	2.12	7.47	2.04	4.3	0.2	4.29	16.06	8.00
22	1.27	6	3.91	4.04	0.35	6.06	15.34	7.00
23	5.76	4.83	10.77	0.75	1.02	11.63	23.04	5.50
24	4.96	5.56	9.8	0.66	1.23	10.74	22.09	7.50
25	5.31	5.34	8.65	0.72	0.97	9.44	20.8	8.00
26	2.43	7.68	2.31	5.04	0.19	4.95	17.85	6.50
27	2.48	11.02	2.61	5.68	0.22	5.56	21.9	5.50
28	4.65	12.71	4.07	3.76	0.13	6.03	25.37	6.50
29	4.12	4.6	0.55	0.58	0	0.91	10.11	5.00
30	2.05	10.79	0.84	0.96	0.1	1.41	14.88	6.50
31	1.94	22.67	3.18	4.39	0.04	5.4	32.24	8.50
32	2.88	10.55	3.06	1.86	1.11	4.51	19.37	8.50
33	4.37	22.15	2.66	2.81	0.31	4.31	32.7	8.50
34	1.46	5.45	2.53	3.55	0.02	4.34	13.09	5.50
35	3.33	13.65	2.87	7.75	0.09	6.78	27.66	6.00
36	2.14	9.23	1.68	5.04	0.1	4.29	18.35	7.50
37	6.65	18.36	4.41	2.87	0.47	6.14	33.12	7.00
38	4.36	1.47	14.49	0.2	0.09	14.62	20.61	5.50
39	1.7	2.72	16.38	0.44	1.14	17.25	22.88	2.00
40	1.22	17.45	4.78	5.41	0.49	7.73	29.23	7.50
41	2.26	11.38	5.04	4.46	0.55	7.57	23.67	8.50
42	2.73	7.59	4.7	7.03	0.44	8.38	22.26	7.50
43	4.34	0.04	8.96	0.18	0.81	9.47	14.39	5.50
44	11.67	3.95	4.72	2.63	0.26	6.18	23.27	5.50
45	1.03	2.39	14.79	0.34	0.97	15.45	19.53	4.50
46	2.51	20.42	5.92	7.03	0.54	9.66	36.27	8.00
47	11.17	3	9.93	0.67	1.05	10.76	25.79	5.00
48	7.21	1.87	9.96	0.17	1.12	10.63	20.49	6.50
49	1.39	0.91	6.03	1.71	0.34	7.1	10.5	3.50
50	7.28	18.74	7.54	8.55	0.57	12.1	42.71	7.50
51	13.51	6.04	7.12	1.47	0.95	8.36	29.2	6.50

Entry	Lut	Zea	βcar	βcry	αcar	PVA	Tcar	Kernel colour
52	6.16	16.95	3.45	4.28	0.4	5.75	30.97	7.00
53	0.76	0.99	14.19	0.25	1.68	15.1	17.65	2.50
54	3.88	12.28	5.2	6.11	0.66	8.64	28.47	7.50
55	2.66	3.62	12.3	0.5	1.29	13.28	20.7	7.50
56	9.11	13.17	6.1	5.64	0.94	9.42	35.02	8.00
57	2.45	13.74	1.79	2.68	0.13	3.23	20.99	7.50
58	3.48	13.93	2.35	3.42	0.04	4.07	23.09	6.50
59	1.61	6.48	1.85	2.23	0.39	3.18	12.72	6.00
60	4.54	19.37	3	4.78	0.07	5.39	31.5	9.00
61	4.1	19.16	2.49	3.13	0.04	4.06	28.84	8.00
62	2.39	4.3	1.14	2.18	0.1	2.28	10.13	5.00
63	5.19	6.35	1.17	1.81	0.03	2.09	14.5	6.00
64	2.99	8.82	2.49	4.59	0.17	4.87	19.04	8.50
65	4.83	6.69	3.2	4.91	0.18	5.72	19.71	8.00
66	1.11	4.02	2.62	3.93	0.2	4.71	11.93	5.50
67	2.09	7.09	3.74	2.25	0.37	5.04	15.64	6.00
68	2.59	9.58	2.94	2.69	0.34	4.51	18.51	7.50
69	1.91	10.64	3.71	2.91	0.38	5.34	19.48	7.00
70	1.88	9.41	3.81	2.76	0.34	5.43	18.5	6.50
71	4	17.74	2.29	2.31	0.23	3.6	26.66	8.00
72	1.4	6.89	1.34	1.82	0.14	2.33	11.58	5.00
73	2.88	10.37	2.62	3.06	0.38	4.37	19.49	5.00
74	1.13	6.92	1.65	2.02	0.11	2.71	11.83	5.50
75	1.35	2.17	1.67	2.3	0.14	2.88	7.66	6.50
76	0.64	0.68	1.52	1.81	0.12	2.45	4.62	3.00
77	1.48	2.42	2.28	2.09	0.23	3.46	8.49	3.00
78	0.5	1.05	1.62	1.62	0.17	2.58	5.25	3.00
79	0.93	0.71	1.5	1.86	0	2.43	5.01	2.00
80	1.12	1.41	2.85	6.34	0.28	6.12	11.81	5.50
81	1.61	2.06	3.17	6.05	0.3	6.4	13.45	8.00
82	1.86	11.55	1.39	1.15	0	1.98	16.01	7.00
83	1.69	10.74	1.21	1.6	0.12	2.09	15.38	5.00
84	1.79	8.95	7.1	2.7	0.87	8.87	21.27	8.00
85	2.8	15.24	4.03	3.2	0.01	5.59	25.12	8.50
86	2.55	22.04	4.8	5.9	0.53	8.01	35.81	8.50
87	3.77	6.42	1.71	3.54	0.23	3.6	15.78	5.00
88	5.61	18.63	3.84	3.01	0.03	5.33	30.97	7.50
89	6.82	16.08	4.77	3.85	0.57	7.04	32.53	6.50
90	2.38	15.81	4.27	3.52	0.39	6.18	26.21	8.00
91	3.09	17.92	3.64	3.48	0.38	5.53	28.39	8.50
92	3.96	15.56	6.72	6.02	0.7	10.1	32.98	9.00
93	3.7	25.9	4.27	4.18	0.44	6.57	38.41	8.50
94	2.95	11.04	2.74	4.26	0.41	5.05	21.25	6.00
95	1.47	8.18	1.84	2.14	0.15	2.93	13.58	7.00
96	8.89	19.82	2.93	3.28	0.37	4.76	35.41	8.50
97	1.61	18.69	3.49	3.87	0.44	5.63	27.94	9.00
98	3.22	9.43	6.3	2.59	0.77	7.98	22.27	6.00
99	5.8	2.32	11.3	0.39	1.2	12.11	21.1	6.00
100	5.9	15.93	5.02	5.89	0.59	8.23	33.2	7.50
101	12.09	5.02	9.76	0.83	1.39	10.91	29.19	9.00
102	1.65	8.73	1.99	3.5	0.2	3.85	16.14	8.00
103	1.61	9.14	1.77	4.04	0.27	3.93	16.9	7.00
104	0.61	13.17	6.49	2.83	0	7.88	22.92	2.00
105	0.51	0.8	2.84	1.9	0.4	3.96	6.32	3.50
106	1.8	0.26	11.8	1.3	1.21	13.02	16.3	3.00

Entry	Lut	Zea	βcar	βcry	αcar	PVA	Tcar	Kernel colour
107	1.9	1.05	9.07	0.33	0.8	9.65	13.17	6.50
108	11.17	10.48	2.43	2.53	0.58	3.97	27.13	7.50
109	9.6	16.22	2.92	2.99	0.41	4.62	32.1	6.50
110	2.62	14.06	1.97	3.19	0.24	3.71	22.1	8.00
111	2.69	16.7	2.04	2.36	0.28	3.35	24.06	7.50
112	2.15	6.99	2.03	3.41	0.21	3.85	14.78	5.50
113	5.83	12.54	1.47	1.65	0.11	2.3	21.42	4.00
114	3.65	1.27	6.63	0.66	0.33	7.13	12.53	5.50
115	0.45	0.59	9.21	0.39	0.89	9.8	11.27	6.00
116	5.07	11.03	0.58	1.24	0.04	1.19	17.87	4.50
117	1.9	14.06	3.97	7.58	0.33	7.9	27.76	8.00
118	1.7	17.86	4.68	3.43	0.43	6.63	28.16	8.50
119	1.32	7.92	2.37	1.7	0.19	3.32	13.49	5.50
120	1.37	10.36	6.27	1.92	0.6	7.52	20.46	9.00
121	3.4	18.25	5.23	2.78	0.61	6.89	30.03	8.50
122	3.14	14.4	1.43	2.07	0.12	2.46	21	5.50
123	1.76	9.74	1.07	1.11	0.08	1.67	13.74	4.50
124	0.99	1.12	13.64	0.16	1.18	14.33	17.08	2.00
125	6.1	16.67	3.52	4.86	0.46	6.24	32	8.00
126	11.16	6.53	2.36	1.4	0.33	3.21	21.64	4.00
127	1.36	5.98	1.59	2.42	0.17	2.98	11.9	4.00
128	1.93	11.42	2.46	3.29	0.09	4.12	19.04	5.00
129	6.84	4.69	0.63	0.51	0.26	0.99	12.8	4.00
130	2.94	9.8	2.02	4.9	0.36	4.65	19.98	6.00
Mean	**3.58**	**9.66**	**4.21**	**2.92**	**0.40**	**5.87**	**20.78**	6.36
Median	**2.77**	**9.42**	**3.09**	**2.81**	**0.33**	**5.06**	**20.66**	6.50
SE	**1.95**	**2.36**	**1.23**	**0.67**	**0.45**	**1.40**	**3.53**	

Zea = cryptoxanthin, βcryp = β-cryptoxanthin, βcar = β-carotene, tpva = total provitamin A, tcar = total carotenoid. See Appendix Figure 2.

Appendix 1 – Table 4 Pearson's pairwise correlation for carotenoids and kernel colour of 130 inbred lines

	lut	zeax	βcryp	βcar	tpva	tcar	kcol
lut	1	0.09	-0.08	0.08	0.07	0.44	0.11
p		0.2854	0.382	0.3624	0.416	<.0001	0.1949
zeax	0.09	1	0.55	-0.26	-0.12	0.81	0.64
p	0.2854		<.0001	0.003	0.1797	<.0001	<.0001
bcryp	-0.08	0.55	1	-0.28	-0.02	0.50	0.49
p	0.382	<.0001		0.0014	0.8315	<.0001	<.0001
bcar	0.08	-0.26	-0.28	1	0.97	0.22	-0.04
p	0.3624	0.003	0.0014		<.0001	0.011	0.692
tpva	0.07	-0.12	-0.02	0.97	1	0.37	0.10
p	0.416	0.1797	0.8315	<.0001		<.0001	0.2406
tcar	0.44	0.81	0.50	0.22	0.37	1	0.63
p	<.0001	<.0001	<.0001	0.011	<.0001		<.0001
kcol	0.11	0.64	0.49	-0.04	0.10	0.63	1
	0.1949	<.0001	<.0001	0.692	0.2406	<.0001	

Zea = cryptoxanthin, βcryp = β-cryptoxanthin, βcar = β-carotene, tpva = total provitamin A, tcar = total carotenoid. See Appendix Figure 2. p = probability

Appendix 1 – Table 5 Agronomic performance of the 130 inbred lines

Entry	ASI	EH	PH	PA	EA	Rust	Blight	Curv
1	0.75	60.50	113.33	2.50	2.75	1.50	2.25	1.00
2	1.00	57.00	127.00	2.50	3.00	1.50	1.75	1.00
3	1.25	55.50	112.75	2.50	3.50	1.50	1.75	1.25
4	1.50	59.17	120.00	2.50	3.50	1.75	1.50	1.00
5	-0.50	75.83	152.00	1.75	2.25	1.50	1.50	1.00
6	0.25	68.17	141.50	2.00	2.75	2.00	1.50	1.00
7	1.50	67.00	137.67	2.50	2.75	2.00	1.50	1.00
8	-3.25	49.17	117.33	2.75	3.75	2.25	1.75	1.25
9	0.33	62.50	111.83	2.75	3.25	2.50	2.25	1.00
10	-1.25	62.50	107.75	3.00	3.75	1.75	2.50	1.00
11	0.50	89.00	159.83	2.00	3.25	1.50	1.75	1.25
12	0.00	82.67	150.00	2.50	2.75	2.00	1.50	1.25
13	4.25	66.83	144.17	2.75	3.50	1.25	2.50	1.00
14	0.00	73.33	158.33	1.50	2.75	1.00	1.50	1.00
15	-0.25	68.17	137.00	2.25	2.75	1.50	2.00	1.25
16	4.00	64.17	148.67	2.25	2.50	1.50	1.50	1.25
17	1.75	45.83	109.33	2.75	3.50	2.50	1.75	1.25
18	0.25	52.17	114.17	3.25	3.75	3.25	1.75	1.25
19	1.25	72.83	147.67	1.75	2.50	2.50	1.50	1.00
20	0.25	67.50	148.50	2.00	3.25	2.25	1.75	1.25
21	-0.25	62.17	138.50	2.50	3.50	2.00	1.50	1.00
22	0.25	67.33	140.83	2.25	2.75	2.00	2.00	1.50
23	-1.50	62.33	135.00	3.25	3.75	1.50	3.50	1.25
24	0.00	63.17	138.00	2.50	3.50	2.50	2.00	1.00
25	-0.75	51.00	130.00	2.25	3.50	1.50	2.00	1.00
26	1.00	91.33	176.33	2.75	3.50	1.75	2.00	1.50
27	1.00	65.33	140.67	2.75	3.00	1.50	1.75	2.00
28	1.00	81.17	154.67	2.50	3.25	1.75	2.25	1.00
29	0.75	57.83	121.83	2.50	3.50	1.50	1.75	1.25
30	-1.50	66.33	142.17	2.25	3.00	2.50	1.50	1.00
31	-0.25	78.67	152.17	1.75	2.25	2.00	1.50	1.25
32	1.75	78.67	155.83	2.25	3.00	2.25	1.50	1.00
33	0.75	60.33	136.33	2.50	3.25	1.75	2.00	1.25
34	-0.25	75.00	145.33	2.25	3.25	2.50	1.75	1.25
35	1.75	55.17	128.67	3.00	4.00	1.50	1.75	1.50
36	0.75	62.67	135.83	2.50	3.00	1.50	2.25	1.75
37	0.25	58.83	134.83	2.25	3.50	1.75	1.75	1.25
38	0.00	57.00	133.33	2.50	3.50	1.75	1.50	1.00
39	-0.50	54.67	128.83	2.50	3.25	2.25	1.50	1.00
40	-1.00	61.50	128.67	3.00	3.75	1.75	3.00	1.00
41	0.00	73.50	155.17	2.50	2.75	1.75	1.75	1.50
42	1.50	63.67	132.67	2.25	3.25	1.75	1.50	1.00
43	0.25	74.17	173.83	2.00	3.50	2.00	1.75	1.25
44	-0.25	62.17	148.33	2.25	3.75	2.75	1.75	1.00
45	1.50	57.00	141.83	2.25	3.50	2.00	1.50	1.00
46	1.25	53.67	115.00	2.50	3.50	1.50	2.00	1.00
47	0.50	69.83	141.33	2.75	4.00	2.25	2.00	1.25
48	0.25	70.33	146.50	3.00	4.00	2.50	2.00	1.50
49	-0.25	66.67	150.83	2.50	3.00	2.00	2.00	1.00
50	-0.25	54.33	137.83	2.75	4.00	2.00	3.50	1.25
51	1.25	46.50	130.50	2.50	3.25	2.00	2.00	1.25
52	1.25	58.17	119.83	2.75	3.75	1.75	3.25	1.00
53	0.00	66.00	149.67	1.75	3.25	1.75	2.25	1.25
54	0.25	69.83	157.00	2.25	3.50	2.00	2.00	1.00
55	1.25	54.17	140.33	2.50	3.50	2.00	1.75	1.50
56	1.25	43.50	120.83	2.75	3.75	2.00	1.50	1.25
57	0.25	82.00	175.83	1.75	3.25	2.00	1.75	1.25

Entry	ASI	EH	PH	PA	EA	Rust	Blight	Curv
58	0.00	65.83	156.00	1.50	3.00	2.00	1.50	1.00
59	0.00	73.67	163.67	2.25	2.25	1.75	2.25	1.00
60	0.25	50.83	124.50	2.25	3.50	1.75	2.00	1.00
61	-0.50	61.00	150.00	2.00	3.25	2.00	2.25	1.00
62	0.25	64.00	157.83	2.25	3.25	2.50	2.00	1.00
63	1.25	70.50	144.00	2.00	3.00	2.00	1.50	1.25
64	-1.00	76.00	163.83	2.00	2.25	1.50	1.50	1.50
65	1.50	68.50	156.50	2.00	2.75	1.75	1.50	1.25
66	-1.00	73.17	133.50	2.50	3.00	2.00	2.00	2.00
67	0.25	67.00	139.17	2.75	3.00	1.75	1.75	1.50
68	-0.50	68.17	139.50	2.25	3.75	2.25	1.75	1.25
69	0.25	73.17	150.17	3.00	3.50	2.00	1.75	1.25
70	1.00	68.00	147.00	3.00	3.25	2.25	1.75	1.50
71	-1.25	49.83	114.83	2.50	3.50	2.25	1.50	1.25
72	0.25	74.33	144.33	2.25	2.75	2.75	1.50	1.50
73	0.50	87.33	152.33	2.50	3.50	1.50	1.75	1.25
74	1.00	59.83	139.83	2.25	3.25	1.50	2.00	1.50
75	2.75	59.17	130.33	2.75	3.50	2.50	1.50	1.00
76	1.75	60.83	124.50	2.75	3.75	2.75	1.50	1.00
77	-0.50	88.33	175.17	2.25	3.00	1.50	1.50	1.50
78	1.00	59.00	148.67	2.25	3.25	2.50	2.00	1.25
79	0.50	62.83	144.33	3.00	3.75	2.25	1.75	1.25
80	2.00	51.83	125.83	2.50	3.75	2.00	1.50	1.50
81	0.50	73.00	136.50	2.50	3.25	1.50	1.75	1.50
82	0.25	72.00	135.83	2.50	3.25	2.50	1.75	1.50
83	1.50	60.17	130.00	2.50	3.00	1.75	1.50	1.25
84	0.75	85.00	163.00	2.50	3.00	2.00	1.50	1.25
85	2.00	43.17	118.50	3.00	3.75	1.50	1.50	2.50
86	0.75	68.00	131.33	2.25	3.00	1.50	1.75	1.00
87	2.50	90.83	181.17	1.75	2.50	1.50	1.75	1.50
88	0.00	71.50	148.50	2.25	2.75	1.50	2.50	1.00
89	1.50	77.33	153.83	2.50	3.50	1.50	2.25	1.00
90	1.50	62.17	137.67	2.50	2.75	2.75	1.75	1.25
91	0.75	80.33	152.00	1.75	3.00	1.50	1.50	1.00
92	4.25	86.67	169.50	2.25	2.50	2.50	1.50	1.00
93	0.25	50.83	106.08	3.00	3.75	2.50	1.50	1.00
94	0.25	38.33	106.50	3.00	3.75	2.75	1.75	1.25
95	1.00	67.92	146.92	2.75	3.50	2.50	2.25	1.50
96	1.50	73.67	149.33	2.50	2.75	2.00	2.00	1.00
97	2.50	54.83	140.67	2.50	3.25	1.75	1.75	1.00
98	1.75	68.00	141.50	2.50	3.25	1.50	1.50	1.50
99	1.00	65.83	155.67	2.25	2.75	2.50	1.75	1.00
100	1.00	63.83	153.33	2.25	3.25	2.00	1.75	1.50
101	2.75	66.33	153.67	3.00	3.75	3.25	1.75	1.25
102	2.25	44.17	114.83	2.75	3.25	1.50	1.75	1.00
103	2.00	79.00	144.33	1.75	3.25	1.50	1.50	1.25
104	0.25	76.00	163.83	1.75	3.25	1.75	1.75	1.00
105	0.00	82.67	155.83	2.00	3.25	1.50	2.00	1.00
106	-0.50	90.00	172.33	2.25	3.25	2.00	2.25	1.50
107	-2.50	71.67	143.33	2.00	3.25	1.50	2.00	1.50
108	4.25	58.67	129.33	2.00	3.25	1.50	1.75	1.00
109	4.00	71.50	131.33	2.50	3.25	2.50	1.75	1.00
110	-0.25	57.67	131.67	2.50	3.25	1.50	1.50	1.50
111	2.00	76.00	159.67	1.75	2.50	2.00	1.50	1.00
112	1.75	57.67	139.50	3.00	3.25	1.75	3.00	1.25
113	0.00	48.50	114.67	2.75	3.50	1.75	2.00	1.00
114	2.75	79.00	136.83	2.50	3.50	2.25	1.50	1.50
115	2.00	41.33	116.17	2.75	3.25	1.75	2.25	1.00
116	2.00	65.33	151.33	2.25	3.25	2.50	1.50	1.00

Entry	ASI	EH	PH	PA	EA	Rust	Blight	Curv
117	0.50	48.00	122.50	2.75	3.50	2.25	1.75	1.00
118	-0.25	66.50	157.67	2.25	3.00	2.50	1.75	1.00
119	-0.25	74.33	161.50	2.75	3.00	3.00	1.75	1.25
120	2.00	50.50	135.33	2.50	3.50	2.00	1.50	1.25
121	1.25	48.83	121.50	4.25	4.00	1.50	4.00	1.25
122	2.00	69.67	143.17	2.25	3.00	1.50	1.50	1.00
123	3.75	73.33	123.50	2.25	3.00	1.50	1.50	1.25
124	2.00	52.33	142.67	2.75	4.00	1.50	2.50	1.00
125	2.25	39.00	112.50	3.25	4.00	2.00	2.25	1.50
126	3.00	64.17	142.83	2.75	3.50	3.00	2.00	1.25
127	2.25	58.08	119.25	3.25	3.25	2.25	1.75	1.00
128	4.25	63.00	138.33	2.50	3.00	2.25	1.50	1.25
129	2.50	48.00	113.00	2.75	3.00	1.50	1.75	1.25
130	2.50	49.00	126.50	2.00	3.75	2.00	1.50	1.00
Grand Mean	0.87	64.87	139.97	2.45	3.26	1.96	1.85	1.21

ASI = Anthesis silking interval, EH = Ear height, PH = Plant height, PA = plant aspect, EA = Ear aspect, Curv = curvilaria

Appendix 1 – Table 6 Genes within LD range of the most significant SNPs detected by the GWAS (with/without allele specific markers as covariates)

Gene stable ID	Gene name	Gene description	Chr	Strand	Coordinate start	Coordinate end	Most significant SNP	Distance from gene (Mbp)
GRMZM2G021710			2	-1	35,273,097	35,277,507		799,874
GRMZM2G023204			2	-1	35,324,876	35,326,785		750,596
GRMZM2G389789		Uncharacterized protein	2	1	35,355,453	35,360,661		716,720
AC211676.4_FG001			2	-1	35,482,452	35,482,829		594,552
GRMZM2G448126			2	1	35,482,483	35,484,426		592,955
GRMZM2G099817		Uncharacterized protein	2	1	35,568,928	35,570,039		507,342
GRMZM2G703008			2	-1	35,620,406	35,620,868		456,513
GRMZM2G152000		Zinc finger, C3HC4 type family protein	2	1	35,674,894	35,676,093		401,288
GRMZM2G429662		Zinc finger, C3HC4 type family protein	2	-1	35,746,325	35,748,257		329,124
AC214613.3_FG003			2	-1	35,790,943	35,792,868		284,513
GRMZM2G128800		Zinc finger, C3HC4 type family protein	2	-1	35,841,329	35,842,533		234,848
GRMZM2G054046			2	1	35,876,928	35,877,532		199,849
GRMZM2G703013			2	1	35,877,646	35,878,504		198,877
AC193754.3_FG008			2	-1	35,880,818	35,882,857		194,524
GRMZM2G053929		Uncharacterized protein	2	-1	35,884,125	35,885,404		191,977
GRMZM2G099183		Uncharacterized protein	2	1	35,949,719	35,954,263		123,118
AC193754.3_FG005			2	-1	36,034,684	36,035,349		42,032
AC186613.4_FG003			2	-1	36,039,757	36,040,371		37,010
GRMZM2G021619			2	-1	36,079,677	36,082,153	S2_36077381	2,296
GRMZM2G030598		Kelch motif family protein	2	-1	36,133,981	36,154,824		56,600
GRMZM2G102183	*LIP*	Malate synthase, glyoxysomal	2	1	36,317,292	36,319,768		239,911
GRMZM2G102238		Acid phosphatase/vanadium-dependent haloperoxidase related	2	-1	36,320,023	36,330,061		242,642
GRMZM2G112072		Putative uncharacterized protein	2	1	36,358,387	36,360,987		281,006
GRMZM2G130109		Uncharacterized protein	2	1	36,445,644	36,449,142		368,263
GRMZM2G130224			2	-1	36,450,557	36,453,156		373,176
GRMZM2G130230		Glucose-6-phosphate 1-dehydrogenase	2	-1	36,456,199	36,464,613		378,818
GRMZM2G142712		Uncharacterized protein	2	-1	36,563,910	36,568,978		486,529
GRMZM2G093832			2	1	36,573,079	36,579,034		495,698
GRMZM2G093781			2	-1	36,582,037	36,585,511		504,656
GRMZM2G093720			2	1	36,586,610	36,591,326		509,229
GRMZM2G093557		26S proteasome non-ATPase regulatory subunit 6	2	-1	36,591,474	36,594,714		514,093
GRMZM2G100872		Uncharacterized protein	2	1	36,597,705	36,602,732		520,324
GRMZM2G100815			2	-1	36,605,205	36,613,857		527,824

Gene stable ID	Gene name	Gene description	Chr	Strand	Coordinate start	Coordinate end	Most significant SNP	Distance from gene (Mbp)
GRMZM2G372930			2	-1	36,617,989	36,618,665		540,608
GRMZM2G063262			2	1	36,763,535	36,768,059		686,154
GRMZM2G063163		Uncharacterized protein	2	-1	36,767,527	36,778,587		690,146
GRMZM2G045117		BCL-2 binding anthanogene-1	2	-1	36,824,114	36,825,732		746,733
GRMZM2G045070		Topoisomerase-like protein	2	1	36,833,209	36,838,151		755,828
GRMZM2G044989		Uncharacterized protein	2	-1	36,838,775	36,842,537		761,394
GRMZM2G094792		Uncharacterized protein	2	-1	42,618,631	42,623,189		752,968
GRMZM2G393762			2	-1	42,633,247	42,637,017		739,140
GRMZM2G427790			2	-1	42,693,055	42,695,872		680,285
GRMZM2G065073		Gamma-glutamyltranspeptidase 1	2	1	42,784,980	42,792,494		583,663
GRMZM2G522241			2	1	42,793,644	42,794,026		582,131
GRMZM2G065157		Putative uncharacterized protein	2	1	42,794,940	42,795,548		580,609
GRMZM2G065171		Uncharacterized protein	2	1	42,800,989	42,805,578		570,579
GRMZM2G476822			2	-1	42,919,265	42,931,636		444,521
GRMZM2G044322		Uncharacterized protein	2	-1	42,957,793	42,959,684		416,473
GRMZM2G148147		Putative uncharacterized protein	2	1	43,000,189	43,001,179		374,978
GRMZM2G119375			2	-1	43,091,374	43,093,513		282,644
GRMZM2G018275		Uncharacterized protein	2	-1	43,117,505	43,119,018		257,139
GRMZM2G071996		Putative uncharacterized protein	2	1	43,182,532	43,185,690		190,467
GRMZM2G071768		Serologically defined breast cancer antigen NY-BR-84	2	1	43,188,346	43,197,822		178,335
GRMZM2G071714	*LIP1*	Lipoyl synthase, mitochondrial	2	1	43,200,299	43,203,078		173,079
GRMZM2G071704			2	1	43,204,316	43,205,109		171,048
GRMZM2G092107		Putative uncharacterized protein	2	1	43,339,685	43,346,040		30,117
GRMZM2G442523		Uncharacterized protein	2	1	43,370,373	43,375,906		251
GRMZM2G442546		Uncharacterized protein	2	-1	43,375,177	43,378,906	S2_43376157	980
GRMZM2G031028		Plastid-lipid-associated protein 2	2	1	43,428,306	43,433,991		52,149
GRMZM5G843141		Uncharacterized protein	2	-1	43,434,582	43,441,085		58,425
GRMZM5G831135			2	-1	43,480,267	43,481,433		104,110
GRMZM2G360374			2	1	43,480,267	43,481,433		104,110
GRMZM2G061495			2	-1	43,482,588	43,483,408		106,431
GRMZM2G010338		Uncharacterized protein	2	1	43,523,290	43,532,367		147,133
GRMZM2G049568		Ubiquitin carboxyl-terminal hydrolase	2	1	43,549,463	43,557,488		173,306
GRMZM2G049549		Uncharacterized protein	2	-1	43,557,836	43,561,417		181,679
GRMZM5G837621			2	1	43,579,798	43,580,388		203,641
GRMZM2G171296		Putative uncharacterized protein	2	-1	43,633,707	43,634,200		257,550
GRMZM2G009125		Cytokinin-O-glucosyltransferase 2	2	-1	43,708,019	43,710,470		331,862
AC194705.3_FG004			2	1	43,736,695	43,738,206		360,538
GRMZM2G325139			2	-1	43,760,551	43,761,190		712,611
GRMZM5G870067		Uncharacterized protein	2	-1	43,834,734	43,836,570		637,231
GRMZM2G417770			2	1	43,943,602	43,945,578		528,223
GRMZM2G024119		Uncharacterized protein	2	-1	44,013,584	44,035,237		438,564
GRMZM2G703021		Uncharacterized protein	2	1	44,038,259	44,040,528		433,273
GRMZM2G145758		Histone H3	2	-1	44,167,696	44,170,031		303,770
GRMZM2G470882		Uncharacterized protein	2	-1	44,187,637	44,189,354		284,447
GRMZM2G171254		Uncharacterized protein	2	-1	44,190,672	44,202,611		271,190
GRMZM2G471039			2	1	44,193,630	44,195,023		278,778
GRMZM2G171277		Uncharacterized protein	2	1	44,200,748	44,202,483		271,318
GRMZM2G160966			2	-1	44,260,729	44,262,323		211,478
GRMZM2G160994		Putative uncharacterized protein	2	-1	44,262,632	44,265,116		208,685
GRMZM2G161012			2	-1	44,274,679	44,276,416		197,385
GRMZM2G167262		Signal recognition particle 19 kDa protein	2	-1	44,298,898	44,304,717		169,084
GRMZM2G046900		Putative uncharacterized protein	2	-1	44,330,496	44,336,261		137,540
GRMZM2G347488		Putative uncharacterized protein	2	-1	44,337,299	44,338,263		135,538
GRMZM2G347489			2	1	44,340,046	44,345,073		128,728
GRMZM2G015844		Uncharacterized protein	2	1	44,378,673	44,382,847		90,954
GRMZM2G015610		Uncharacterized protein	2	1	44,386,603	44,391,746		82,055
GRMZM2G127184			2	1	44,434,561	44,436,541		37,260
GRMZM2G127173			2	-1	44,436,649	44,439,381		34,420
GRMZM2G127139	*ZEP1*	Zeaxanthin epoxidase	2	-1	44,440,299	44,449,237	S2_44473801	24,564
GRMZM2G062559		Uncharacterized protein	2	-1	44,471,623	44,474,212	S2_44473758	411
GRMZM5G889338		OVERLAPPING	2	1	44,471,647	44,472,033		2,154
GRMZM2G345238			2	-1	44,547,174	44,552,558		73,373
GRMZM2G046402		Uncharacterized protein	2	-1	44,553,002	44,562,139		79,201

Gene stable ID	Gene name	Gene description	Chr	Strand	Coordinate start	Coordinate end	Most significant SNP	Distance from gene (Mbp)
GRMZM2G474783			2	-1	44,697,029	44,698,753		223,228
GRMZM2G053985			2	1	44,719,043	44,719,955		245,242
GRMZM2G355906		Uncharacterized protein	2	-1	44,720,554	44,751,914		246,753
GRMZM2G126062		Aldose reductase	2	-1	44,795,877	44,798,393		322,076
GRMZM2G051355		Uncharacterized protein	2	1	44,854,880	44,861,495		381,079
GRMZM2G103050			2	1	44,928,620	44,937,597		454,819
GRMZM2G408038			2	1	44,941,098	44,944,904		467,297
GRMZM2G101069		Glutamate decarboxylase	2	1	44,991,322	44,994,248		517,521
GRMZM2G104843		Lipoxygenase	2	-1	45,192,092	45,196,460		718,291
GRMZM2G105542		Uncharacterized protein	2	-1	45,219,705	45,222,039		745,904
GRMZM2G105608			2	-1	45,223,315	45,224,127		749,514
GRMZM2G105617			2	1	45,230,433	45,231,167		756,632
GRMZM2G090675		Uncharacterized protein	2	-1	102,897,151	102,905,974		784128
GRMZM2G019742			2	-1	103,027,499	103,028,664		653780
GRMZM2G019819			2	1	103,045,720	103,049,593		635559
GRMZM2G109873			2	1	103,171,373	103,172,289		509906
GRMZM2G109879			2	1	103,172,610	103,179,072		508669
GRMZM2G365957			2	-1	103,336,613	103,342,858		344666
GRMZM2G303995			2	-1	103,451,158	103,452,508		230121
GRMZM2G303993			2	-1	103,499,700	103,500,770		181579
AC196719.3_FG002			2	1	103,500,310	103,500,681		180969
GRMZM2G003796			2	1	103,532,762	103,534,648		148517
GRMZM2G334321			2	1	103,668,965	103,670,087		12314
GRMZM2G334336			2	-1	103,680,545	103,682,621	S2_103681279	734
GRMZM2G334338			2	-1	103,685,681	103,686,659		5,380
GRMZM2G397965			2	-1	103,825,178	103,826,833		145,554
GRMZM2G089010		Uncharacterized protein	2	1	103,960,946	103,965,539		284,260
AC190677.3_FG003			2	-1	104,040,064	104,040,735		359,456
GRMZM2G172686			2	-1	104,167,615	104,170,477		489,198
GRMZM2G172695			2	1	104,173,790	104,176,088		494,809
GRMZM2G132968		60S ribosomal protein L3	2	1	138,864,214	138,867,222		777,054
GRMZM2G132966			2	1	138,892,765	138,893,204		751,072
AC204953.3_FG002			2	-1	138,906,162	138,906,854		737,422
GRMZM5G854138		Uncharacterized protein	2	1	139,066,950	139,068,044		576,232
GRMZM2G486618			2	-1	139,175,046	139,175,672		468,604
GRMZM2G047292		Stachyose synthase	2	-1	139,185,152	139,187,472		456,804
GRMZM2G348151		Uncharacterized protein	2	-1	139,433,415	139,435,902		208,374
GRMZM2G099502			2	1	139,516,746	139,518,995		125,281
GRMZM2G066213			2	-1	139,643,865	139,645,573	S2_139644276	411
GRMZM2G366802			2	1	139,653,617	139,654,676		9,341
GRMZM2G522398			2	1	139,658,407	139,662,410		14,131
GRMZM2G039648			2	1	139,743,422	139,745,283		99,146
GRMZM5G898141			2	1	139,759,830	139,764,442		115,554
GRMZM2G144885			2	-1	139,902,080	139,902,696		257,804
AC205544.3_FG004			2	-1	140,056,166	140,057,095		411,890
GRMZM2G006765		Uncharacterized protein	2	1	140,057,140	140,059,599		412,864
GRMZM2G010282		Uncharacterized protein	2	-1	140,089,814	140,091,612		445,538
AC226536.2_FG001			2	1	140,306,755	140,307,437		662,479
GRMZM2G125352		Uncharacterized protein	2	-1	140,309,065	140,313,923		664,789
GRMZM5G821024		Uncharacterized protein	2	-1	140,322,684	140,326,990		678,408
GRMZM2G014400			2	1	207,955,688	207,967,618		705,060
GRMZM2G112954			2	1	207,974,613	207,977,249		695,429
GRMZM2G034927			2	-1	208,075,008	208,076,470		596,208
GRMZM2G034764		C2 domain containing protein	2	1	208,079,590	208,081,944		590,734
GRMZM2G113848			2	-1	208,114,241	208,115,156		557,522
GRMZM2G403800		Uncharacterized protein	2	-1	208,119,426	208,120,531		552,147
GRMZM2G105207		NADH-ubiquinone oxidoreductase 18 kDa subunit	2	1	208,123,671	208,129,714		542,964
GRMZM2G105302		Putative uncharacterized protein	2	-1	208,128,674	208,130,991		541,687
GRMZM2G351417			2	1	208,195,862	208,203,005		469,673
GRMZM5G881349			2	-1	208,203,708	208,204,361		468,317
GRMZM2G003937		Uncharacterized protein	2	1	208,205,739	208,206,806		465,872
GRMZM2G129575			2	1	208,223,142	208,227,514		445,164
GRMZM2G454474		Triacylglycerol lipase	2	-1	208,329,261	208,330,419		342,259
GRMZM2G575305		Putative uncharacterized protein	2	1	208,345,176	208,349,036		323,642
GRMZM2G021151			2	1	208,361,028	208,361,925		310,753

Gene stable ID	Gene name	Gene description	Chr	Strand	Coordinate start	Coordinate end	Most significant SNP	Distance from gene (Mbp)
GRMZM2G019060			2	1	208,376,585	208,378,677		294,001
GRMZM2G028905		Uncharacterized protein	2	-1	208,433,511	208,437,003		235,675
GRMZM2G090274		Eukaryotic peptide chain release factor subunit 1-1	2	-1	208,530,771	208,535,295		137,383
GRMZM2G090493		Uncharacterized protein	2	1	208,538,326	208,540,184		132,494
GRMZM2G307604			2	1	208,569,032	208,571,129		101,549
GRMZM2G007060		Uncharacterized protein	2	1	208,574,549	208,579,842		92,836
GRMZM2G082612		Uncharacterized protein	2	1	208,609,666	208,612,318		60,360
GRMZM2G383841			2	-1	208,616,426	208,617,726		54,952
GRMZM5G865483			2	1	208,653,141	208,653,828		18,850
AC211702.2_FG002			2	1	208,654,679	208,656,244		16,434
GRMZM5G823629		Uncharacterized protein	2	1	208,669,032	208,674,022	S2_208672678	3,646
GRMZM2G455889			2	-1	208,675,050	208,677,330		2372
GRMZM5G824534		Putative uncharacterized protein	2	1	208,845,445	208,851,248		172767
GRMZM2G177110		Seed specific protein Bn15D17A	2	1	208,854,344	208,855,480		181666
GRMZM2G125501			2	1	208,881,462	208,883,455		208784
GRMZM2G108040			2	-1	208,942,939	208,943,796		270261
GRMZM2G121868			2	-1	208,974,900	208,977,568		302222
GRMZM2G007324		Uncharacterized protein	2	-1	208,999,148	209,001,435		326470
GRMZM2G077845		GIR1	2	1	209,010,511	209,011,342		337833
GRMZM2G376432			2	-1	209,010,512	209,014,024		337834
GRMZM2G175718		Uncharacterized protein	2	-1	209,018,645	209,025,333		345967
GRMZM2G378106		Indole-3-acetic acid amido synthetase	2	-1	209,110,460	209,113,041		437782
GRMZM2G079067			2	1	209,119,338	209,119,949		446660
GRMZM2G079082			2	-1	209,124,511	209,128,692		451833
GRMZM2G378121		Uncharacterized protein	2	1	209,131,619	209,132,616		458941
GRMZM2G040230		Metal ion binding protein	2	-1	209,195,758	209,223,851		523080
GRMZM2G414002			2	-1	209,248,084	209,263,370		575406
GRMZM2G113512		Uncharacterized protein	2	-1	209,278,354	209,283,777		605676
GRMZM2G113696	*TIF5A*	Eukaryotic translation initiation factor 5A	2	-1	209,289,645	209,292,956		616967
GRMZM2G115817		Circadian clock coupling factor ZGT	2	1	209,300,195	209,304,000		627517
GRMZM2G146292		Uncharacterized protein	2	-1	209,337,124	209,345,160		664446
GRMZM2G146416		Uncharacterized protein	2	-1	209,349,729	209,354,569		677051
GRMZM2G703354			2	-1	209,470,409	209,472,160		797731
GRMZM2G084984		Uncharacterized protein	4	1	171,638,282	171,652,162		797559
GRMZM2G102163		Protein kinase superfamily proteinsssdf protein	4	1	171,665,330	171,672,226		770511
GRMZM2G102200		Uncharacterized protein	4	-1	171,673,499	171,674,430		762342
GRMZM2G402319		Uncharacterized protein	4	1	171,674,702	171,676,783		761139
GRMZM2G102216		Putative glutathione S-transferase family proteinssd protein	4	-1	171,676,838	171,678,581		759003
GRMZM2G102230		60S ribosomal protein L23s proteindf	4	-1	171,688,353	171,692,167		747488
GRMZM2G102815		(Csu525(RpL17)), mRNA	4	1	171,691,584	171,694,852		744257
GRMZM2G099097		Uncharacterized protein	4	1	171,730,311	171,735,318		705530
GRMZM2G099297		Uncharacterized protein	4	1	171,770,786	171,772,540		665055
GRMZM2G099376		Uncharacterized protein	4	-1	171,778,900	171,780,935		656941
GRMZM2G099420		Uncharacterized protein	4	-1	171,783,871	171,785,676		651970
GRMZM2G370815		Uncharacterized protein	4	-1	171,789,366	171,791,813		646475
GRMZM2G702697		Uncharacterized protein	4	1	171,801,352	171,802,121		634489
GRMZM2G702699		Uncharacterized protein	4	1	171,804,365	171,814,541		631476
AC189280.3_FG001		Uncharacterized protein	4	1	171,842,284	171,845,974		593557
GRMZM2G466532		Uncharacterized protein	4	-1	171,908,549	171,910,339		527292
GRMZM2G466545		Uncharacterized protein	4	1	171,917,079	171,922,792		518762
GRMZM2G164854		Uncharacterized protein	4	-1	171,921,876	171,925,757		513965
GRMZM2G034005		Protein binding proteinss proteins	4	-1	172,010,354	172,019,014		425487
GRMZM2G062084		Uncharacterized protein	4	1	172,055,498	172,060,332		380343
GRMZM2G075336		ATP-dependent transporter YFL028C; Putative ABC transporter family protein	4	1	172,080,242	172,084,381		355599
GRMZM2G075124		Putative casein kinase family proteinss sprotein	4	1	172,085,426	172,089,671		350415
GRMZM2G023899		Uncharacterized protein	4	1	172,109,058	172,110,448		326783
GRMZM2G140901		Putative NAC domain transcription factor superfamily	4	1	172,192,989	172,195,242		242852

Gene stable ID	Gene name	Gene description	Chr	Strand	Coordinate start	Coordinate end	Most significant SNP	Distance from gene (Mbp)
		protein						
GRMZM2G107109		Uncharacterized protein	4	1	172,257,209	172,265,943		178632
GRMZM2G140524		Uncharacterized protein	4	1	172,309,385	172,318,450		126456
GRMZM2G171163		Uncharacterized protein	4	1	172,357,234	172,360,388		78607
GRMZM2G027075	*jmj4*	Uncharacterized protein	4	1	172,430,260	172,439,306	S4_172435841	5,581
GRMZM2G027302		Uncharacterized protein	4	-1	172,444,330	172,446,310		8489
GRMZM2G364068		Uncharacterized protein	4	-1	172,522,239	172,530,660		86398
GRMZM2G130043		Starch synthase V	4	-1	172,606,253	172,677,186		170412
GRMZM2G130002		Putative RING zinc finger and VWF domain family proteins protein	4	1	172,685,051	172,689,834		249210
GRMZM2G129979		Putative G10 domain family protein	4	-1	172,692,344	172,695,319		256503
GRMZM2G178398		Uncharacterized protein	4	1	172,748,163	172,754,261		312322
GRMZM2G379005		Putative GATA transcription factor family protein	4	-1	172,760,157	172,765,319		324316
AC186156.3_FG005		Uncharacterized protein	4	1	172,794,270	172,794,881		358429
AC186156.3_FG004		Uncharacterized protein	4	1	172,821,941	172822552		386100
GRMZM2G162052		Uncharacterized protein	4	1	172,831,531	172,833,992		395690
GRMZM2G162007		Putative alcohol dehydrogenase superfamily proteins protein	4	-1	172,834,862	172,837,214		399021
GRMZM2G108364		PAP-specific phosphatase	4	1	172,992,607	172,995,763		556766
GRMZM2G410487		Uncharacterized protein	4	-1	172,996,195	172,997,403		560354
GRMZM2G108284		Uncharacterized protein	4	1	172,998,043	173,001,754		562202
GRMZM2G108712		Proliferating cell nuclear antigen	4	-1	173,033,396	173,035,374		597555
GRMZM2G054007		Uncharacterized protein	4	1	173,177,172	173,177,669		741331
GRMZM2G054012		40S ribosomal protein S30	4	-1	173,177,749	173,179,720		741908
GRMZM2G054065		Uncharacterized protein	4	1	173,181,274	173,182,106		745433
GRMZM2G054076		Chaperone protein dnaJ	4	-1	173,181,520	173,194,277		745679
GRMZM2G054210		Uncharacterized protein	4	-1	173,209,988	173,224,508		774147
GRMZM2G126772		Uncharacterized protein	5	1	77,657,450	77,659,533		725,156
GRMZM2G428393			5	-1	77,663,317	77,666,150		718,539
GRMZM2G405947			5	1	77,669,514	77,675,552		709,137
GRMZM2G033523			5	1	77,700,111	77,701,507		683,182
GRMZM2G010731			5	1	77,727,436	77,729,552		655,137
GRMZM2G010779		Vacuolar cation/proton exchanger 2	5	1	77,729,765	77,749,253		635,436
GRMZM2G010920		Uncharacterized protein	5	-1	77,748,895	77,751,969		632,720
GRMZM2G010944		Uncharacterized protein	5	-1	77,753,934	77,757,260		627,429
GRMZM2G011456		Uncharacterized protein	5	-1	77,757,345	77,777,262		607,427
GRMZM2G097207		Glucan endo-1,3-beta-glucosidase 6	5	1	77,916,981	77,922,073		462,616
GRMZM2G097275		Uncharacterized protein	5	1	77,940,666	77,945,714		438,975
GRMZM2G023798		Uncharacterized protein	5	-1	77,990,308	77,998,801		385,888
GRMZM2G435796			5	-1	78,022,476	78,023,096		361,593
GRMZM5G832780			5	-1	78,129,898	78,131,652		253,037
GRMZM2G425559			5	1	78,129,898	78,131,784		252,905
GRMZM2G006937		Uncharacterized protein	5	1	78,255,163	78,310,499		74,190
GRMZM2G102912		AIG2-like protein	5	1	78,380,304	78,381,593		3,096
GRMZM2G102845	*arf2o*	Auxin response factor 20	5	-1	78,381,834	78,389,884	S5_78384689	2,855
GRMZM5G865367			5	1	78,519,893	78,520,683		135,204
GRMZM2G322493		Uncharacterized protein	5	1	78,758,856	78,765,635		374,167
GRMZM2G167741		Uncharacterized protein	5	1	78,772,058	78,780,292		387,369
GRMZM2G410357		Uncharacterized protein	5	1	78,804,756	78,815,306		420,067
GRMZM2G410393		Putative uncharacterized protein	5	1	78,819,803	78,826,612		435,114
AC197118.3_FG005		Antigenic determinant of rec-A protein	5	-1	78,820,119	78,821,393		435,430
GRMZM2G133048			5	1	78,904,639	78,918,850		519,950
GRMZM2G018686		Putative HLH DNA-binding domain superfamily protein	5	-1	79,055,654	79,056,960		670,965
AC212103.3_FG002			5	1	79,162,381	79,163,631		777,692
AC193606.2_FG001			5	-1	79,175,814	79,176,557		791,125
GRMZM2G332749			5	1	79,183,135	79,188,028		798,446
GRMZM2G319798			7	-1	107,780,060	107,782,693		752,317
GRMZM2G074672		Protein CCC1	7	1	107,855,871	107,858,296		676,714
GRMZM2G019183		Uncharacterized protein	7	1	107,926,112	107,930,789		604,221
GRMZM2G043240		Putative uncharacterized protein	7	-1	108,007,980	108,011,772		523,238

Gene stable ID	Gene name	Gene description	Chr	Strand	Coordinate start	Coordinate end	Most significant SNP	Distance from gene (Mbp)
AC234163.1_FG002		Uncharacterized protein	7	1	108,253,198	108,254,882		280,128
GRMZM2G700188		Uncharacterized protein	7	-1	108,281,609	108,388,756		146,254
GRMZM2G153527		Receptor kinase	7	-1	108,480,222	108,482,455		52,555
GRMZM2G045638		Uncharacterized protein	7	1	108,534,749	108,535,532	S7_108535010	261
GRMZM2G169356			7	-1	108,698,427	108,699,794		163417
GRMZM2G126566		Typical P-type R2R3 Myb protein	7	1	108,787,033	108,790,136		252023
GRMZM2G126507		Uncharacterized protein	7	-1	108,843,243	108,846,250		308233
GRMZM2G446171			7	-1	109,047,698	109,054,636		512688
GRMZM2G146225		Uncharacterized protein	7	-1	109,056,610	109,058,446		521600
GRMZM2G066876			7	-1	109,149,609	109,158,470		614599
AC217843.3_FG001		Uncharacterized protein	7	-1	109,217,683	109,218,246		682673
GRMZM2G092550			7	1	109,309,083	109,325,401		774073
GRMZM2G092525		Uncharacterized protein	7	-1	109,326,537	109,329,414		791527
GRMZM2G700405		Uncharacterized protein	8	1	15,945,384	15,946,458		796970
GRMZM2G700407		Uncharacterized protein	8	1	15,950,199	15,951,272		792156
GRMZM2G048804		Uncharacterized protein	8	1	15,958,422	15,977,849		765579
GRMZM2G048763		Uncharacterized protein	8	-1	15,981,227	15,982,875		760553
GRMZM2G083328		Uncharacterized protein	8	-1	15,984,817	15,987,132		756296
GRMZM2G083394		Uncharacterized protein	8	-1	15,987,735	15,992,845		750583
GRMZM2G588728		Uncharacterized protein	8	-1	16,040,312	16,040,810		702618
AC211687.3_FG009		Uncharacterized protein	8	1	16,040,396	16,040,806		702622
AC212565.3_FG002		Uncharacterized protein	8	-1	16,076,551	16,077,390		666038
GRMZM5G869161		Uncharacterized protein	8	1	16,164,364	16,164,876		578552
AC212565.3_FG001	*H4C7*	Histone H4	8	-1	16,166,029	16,166,694		576734
GRMZM2G469298		Uncharacterized protein	8	1	16,219,186	16,246,258		497170
GRMZM2G341010		Uncharacterized protein	8	1	16,321,295	16,335,574		407854
GRMZM2G033017		Uncharacterized protein	8	-1	16,352,457	16,353,551		389877
GRMZM2G081848		Uncharacterized protein	8	-1	16,422,190	16,431,362		312066
GRMZM2G380414		Ultraviolet-B-repressible protein	8	-1	16,443,989	16,444,752		298676
GRMZM2G587368		Uncharacterized protein	8	-1	16,545,140	16,549,180		194248
GRMZM2G165354		Uncharacterized protein	8	1	16,632,417	16,640,500		102928
GRMZM2G061187		Uncharacterized protein	8	-1	16,690,779	16,693,942		49486
GRMZM2G141216		Putative RING zinc finger domain superfamily protein	8	1	16,700,565	16,704,075		39353
GRMZM2G143211		Uncharacterized protein	8	1	16,741,652	16,746,323	S8_16743428	1776
AC211474.3_FG006		Uncharacterized protein	8	1	16,747,527	16,754,983		4099
GRMZM2G143258		Uncharacterized protein	8	-1	16,755,280	16,757,481		11852
GRMZM2G143274		Putative MYB DNA-binding domain superfamily protein	8	-1	16,770,013	16,771,212		26585
GRMZM2G143278		Uncharacterized protein	8	-1	16,777,746	16,778,788		34318
GRMZM2G127308		Tryptophan aminotransferase	8	-1	16,850,861	16,855,009		107433
GRMZM5G853988		Uncharacterized protein	8	1	16,905,840	16,906,421		162412
GRMZM2G025997		Putative RING zinc finger domain superfamily protein	8	1	16,911,248	16,915,154		167820
GRMZM2G026015		Photosystem I reaction center subunit XI	8	1	16,916,336	16,918,485		172908
GRMZM2G025215		Putative DUF1421 domain family protein	8	-1	16,994,084	16,999,444		250656
GRMZM2G463493		Putative leucine-rich repeat receptor protein kinase family protein	8	1	17,058,192	17,062,569		314764
GRMZM2G162928		Uncharacterized protein	8	-1	17,066,051	17,071,072		322623
GRMZM2G095595		Uncharacterized protein	8	1	17,175,395	17,180,489		431967
GRMZM2G095655		Uncharacterized protein	8	1	17,188,772	17,190,196		445344
GRMZM2G473147		Uncharacterized protein	8	-1	17,242,891	17,244,208		499463
GRMZM2G067675		Putative leucine-rich repeat receptor-like protein kinase family protein	8	-1	17,301,645	17,306,811		558217
GRMZM2G119230		Uncharacterized protein	8	1	17,370,948	17,371,391		627520
GRMZM5G893381		Putative DEAD-box ATP-dependent RNA helicase family protein	8	-1	17,374,877	17,377,796		631449
GRMZM2G062069		Uncharacterized protein	8	-1	17,380,160	17,383,660		636732
GRMZM2G061969		Phospholipase D	8	-1	17,388,792	17,393,555		645364
GRMZM2G364129		Uncharacterized protein	8	-1	17,395,313	17,398,611		651885
GRMZM2G061932		Lipase/lipooxygenase, PLAT/LH2	8	1	17,420,296	17,421,579		676868
GRMZM2G061735		Uncharacterized protein	8	-1	17,422,102	17,426,460		678674

Gene stable ID	Gene name	Gene description	Chr	Strand	Coordinate start	Coordinate end	Most significant SNP	Distance from gene (Mbp)
GRMZM2G021069		Minichromosome maintenance protein	8	-1	110516644	110522956		766085
GRMZM2G387227		Cyclin superfamily protein, putative	8	1	110586176	110588594		700447
GRMZM2G303168		Uncharacterized protein	8	-1	110596224	110597303		691738
GRMZM2G130569		Uncharacterized protein	8	1	110612301	110612764		676277
GRMZM2G039017		Uncharacterized protein	8	1	110619879	110621612		667429
GRMZM2G150134		Uncharacterized protein	8	1	110686795	110688334		600707
GRMZM2G327741		Uncharacterized protein	8	1	111,014,304	111,015,047		788,861
GRMZM2G016655		Uncharacterized protein	8	-1	111,043,262	111,050,129		753,779
GRMZM2G072238		Putative HLH DNA-binding domain superfamily protein	8		111,192,460	111,199,551		604,357
GRMZM2G333079		Uncharacterized protein	8	-1	111,287,695	111,290,414		513,494
GRMZM2G416308		Putative prolin-rich extensin-like receptor protein kinase family protein	8	1	111,624,795	111,630,247		173,661
GRMZM2G025175		Uncharacterized protein	8	-1	111,704,094	111,705,090		98,818
GRMZM2G046537		Rhomboid family protein	8	-1	111,722,040	111,730,019		73,889
GRMZM2G157332		Coiled-coil domain-containing protein 25	8	-1	111,799,986	111,807,525	S8_111803908	3,922
GRMZM2G157564		CER5	8	-1	111,823,880	111,830,792		19,972
GRMZM2G053882		Uncharacterized protein	8	1	111,864,621	111,872,240		60,713
GRMZM2G362163		Uncharacterized protein	8	-1	112,001,191	112,002,329		197,283
GRMZM2G153569		Elongation factor 1-delta 1	8	-1	112,147,719	112,150,640		343,811
GRMZM2G153552		Uncharacterized protein	8	1	112,176,065	112,177,057		372,157
GRMZM2G050286		AGP20	8	-1	112,310,806	112,311,880		506,898
GRMZM2G700603		Uncharacterized protein	8	1	112,378,878	112,381,753		574,970
GRMZM2G308999		Uncharacterized protein	8	1	112,463,421	112,466,262		659,513
GRMZM5G854731		Cyclin-dependent kinase inhibitor 2	8	1	112,496,249	112,497,984		692,341
GRMZM2G179679		Uncharacterized protein	8	1	112,530,662	112,532,375		726,754
GRMZM2G180668		DNA-binding WRKY	8	1	123,810,520	123,811,466		623259
GRMZM2G104283		Uncharacterized protein	8	1	123,844,264	123,848,418		586307
GRMZM2G350023		Uncharacterized protein	8	1	123,899,080	123,900,573		534152
GRMZM2G136369		Putative homeodomain-like transcription factor superfamily protein	8	1	123,906,214	123,912,675		522050
GRMZM5G891056		Uncharacterized protein	8	1	123,944,098	123,967,891		466834
AC233864.1_FG002		Uncharacterized protein	8	1	124,013,979	124,014,615		420110
GRMZM5G892365		Uncharacterized protein	8	1	124,033,198	124,034,311		400414
AC233864.1_FG009		Uncharacterized protein	8	-1	124,042,906	124,043,298		391427
AC233864.1_FG014		Uncharacterized protein	8	-1	124,086,876	124,091,045		343680
GRMZM2G115346		Uncharacterized protein	8	-1	124,144,399	124,145,855		288870
GRMZM2G415891		Uncharacterized protein	8	1	124,164,647	124,165,273		269452
GRMZM2G115364		Uncharacterized protein	8	-1	124,165,959	124,168,715		266010
GRMZM2G169236		Uncharacterized protein	8	-1	124,188,697	124,190,213		244512
GRMZM2G162884		Uncharacterized protein	8	1	124,208,874	124,213,413		221312
GRMZM2G084477		MPPN domain containing protein	8	-1	124,223,433	124,225,424		209301
GRMZM2G127844		Uncharacterized protein	8	-1	124,282,351	124,287,196		147529
GRMZM2G135381		Putative GATA transcription factor family protein	8	1	124,357,134	124,359,274		75451
GRMZM2G038082		Uncharacterized protein	8	-1	124,433,434	124,434,146		579
GRMZM2G330693		Uncharacterized protein	8	1	124,434,479	124,435,152	S8_124434725	246
GRMZM5G802899		Uncharacterized protein	8	1	124,437,928	124,439,232		3203
GRMZM2G173700		Uncharacterized protein	8	-1	124,457,224	124,464,067		22499
GRMZM2G173710		Uncharacterized protein	8	-1	124,485,354	124,490,167		50629
GRMZM2G071223		Uncharacterized protein	8	1	124,583,741	124,586,190		149016
GRMZM2G071339		Uncharacterized protein	8	1	124,587,539	124,589,539		152814
GRMZM2G163546		Uncharacterized protein	8	1	124,622,026	124,623,523		187301
GRMZM2G058690		Uncharacterized protein	8	-1	124,645,482	124,648,145		210757
GRMZM2G058745		Uncharacterized protein	8	1	124,652,438	124,657,923		217713
GRMZM2G109842	*PRO2*	Profilin-2	8	-1	124,707,856	124,708,953		273131
GRMZM2G106819		Uncharacterized protein	8	1	124,796,660	124,797,582		361935
GRMZM2G106917		Uncharacterized protein	8	1	124,799,105	124,810,762		364380
GRMZM2G093418		Harpin-induced protein	8	1	124,839,636	124,840,524		404911
GRMZM2G093404		CCCH transcription factor	8	1	124,841,672	124,843,159		406947
GRMZM2G095807		Uncharacterized protein	8	-1	124,847,602	124,849,378		412877

Gene stable ID	Gene name	Gene description	Chr	Strand	Coordinate start	Coordinate end	Most significant SNP	Distance from gene (Mbp)
GRMZM2G082707		50S ribosomal protein L20	8	-1	124,930,000	124,932,025		495275
GRMZM2G319747		Proteasome maturation factor UMP1 family protein	8	-1	124,936,785	124,942,287		502060
GRMZM2G015433		Putative WRKY DNA-binding domain superfamily protein	8	-1	124,950,950	124,952,109		516225
GRMZM2G353822		Uncharacterized protein	8	-1	125,007,017	125,008,185		572292
GRMZM5G855375		Uncharacterized protein	8	1	125,007,024	125,008,185		572299
GRMZM2G052200		Uncharacterized protein	8	-1	125,008,668	125,012,511		573943
GRMZM2G408989		ER lumen protein retaining receptor	8	-1	125,111,092	125,114,558		676367
AC205471.4_FG003		Uncharacterized protein	8	1	125,180,331	125,182,094		745606
AC205471.4_FG008		Uncharacterized protein	8	-1	125,196,525	125,197,922		761800
AC205471.4_FG007		Uncharacterized protein	8	-1	125,225,855	125,226,494		791130
GRMZM2G018375	*THI1-1*	Thiamine thiazole synthase 1, chloroplastic	8	-1	138,142,847	138,144,617		794,332
AC199315.4_FG001		Protein transport protein Sec61 beta subunit	8	1	138,225,306	138,225,551		713,398
AC199315.4_FG002		Uncharacterized protein	8	1	138,255,217	138,255,384		683,565
GRMZM2G164341		Putative HLH DNA-binding domain superfamily protein	8	-1	138,262,239	138,264,389		674,560
GRMZM2G142984		Uncharacterized protein	8	-1	138,315,538	138,326,465		612,484
GRMZM2G132577		Uncharacterized protein	8	-1	138,390,391	138,391,121		547,828
GRMZM2G368556		Uncharacterized protein	8	-1	138,431,180	138,431,962		506,987
GRMZM2G700683		Uncharacterized protein	8	1	138,510,107	138,515,042		423,907
GRMZM2G106479		Uncharacterized protein	8	-1	138,519,520	138,524,207		414,742
GRMZM2G136765		Uncharacterized protein	8	1	138,556,689	138,562,724		376,225
GRMZM2G409343		Uncharacterized protein	8	1	138,645,560	138,647,354		291,595
GRMZM5G852338		Uncharacterized protein	8	1	138,647,887	138,650,204		288,745
GRMZM2G170628		Uncharacterized protein	8	-1	138,657,602	138,660,155		278,794
GRMZM2G170632		RNA-binding protein	8	-1	138,667,667	138,671,356		267,593
GRMZM2G116083		Uncharacterized protein	8	1	138,789,602	138,795,274		143,675
GRMZM2G013448		Uncharacterized protein	8	1	138,861,203	138,862,890		76,059
GRMZM5G851965		Uncharacterized protein	8	1	138,868,954	138,870,495		68,454
GRMZM2G012966	*lcyE*	Lycopene epsilon cyclase1	8	1	138,882,594	138,889,812		49,137
GRMZM2G165428		Putative leucine-rich repeat protein kinase family protein	8	1	138,909,363	138,913,061		25,888
GRMZM2G463133		Putative HLH DNA-binding domain superfamily protein	8	-1	138,938,542	138,943,955	S8_138938949	407
AC196426.3_FG007		Uncharacterized protein	8	-1	139,014,642	139,020,725		75693
GRMZM2G096655		Uncharacterized protein	8	1	139,104,748	139,106,331		165799
GRMZM2G395771		Uncharacterized protein	8	1	139,138,199	139,140,971		199250
GRMZM2G096764		Uncharacterized protein	8	-1	139,143,564	139,151,103		204615
GRMZM2G050553		Uncharacterized protein	8	1	139,286,853	139,300,161		347904
GRMZM2G027333		Uncharacterized protein	8	1	139,342,590	139,346,470		403641
GRMZM2G142705		Uncharacterized protein	8	-1	139,450,959	139,452,582		512010
GRMZM2G114895		LIN1 protein	8	1	139,487,349	139,489,298		548400
GRMZM2G420119		Uncharacterized protein	8	-1	139,592,349	139,595,685		653400
GRMZM2G046037		Uncharacterized protein	8	1	139,630,135	139,631,894		691186
GRMZM2G054900		Putative calmodulin-binding family protein	8	1	139,711,000	139,713,137		772051
GRMZM2G386430		Uncharacterized protein	10	1	133877805	133886124		764857
GRMZM2G404375		Uncharacterized protein	10	-1	133982300	133985927		665054
GRMZM2G404367		Uncharacterized protein	10	-1	133989048	133990408		660573
GRMZM2G102860		Uncharacterized protein	10	1	133992324	133993013		657968
GRMZM2G126665		Uncharacterized protein	10	1	134032517	134034788		616193
GRMZM2G427087		Putative homeodomain-like transcription factor superfamily protein	10	1	134039237	134040599		610382
GRMZM2G126742		Uncharacterized protein	10	-1	134043229	134046251		604730
GRMZM2G427097		Glutamate dehydrogenase	10	1	134053442	134057895		593086
GRMZM2G427106		Uncharacterized protein	10	-1	134057764	134059399		591582
GRMZM2G421212		Uncharacterized protein	10	-1	134094157	134098692		552289
GRMZM2G134523		Uncharacterized protein	10	-1	134251243	134254824		396157
GRMZM2G134517		Uncharacterized protein	10	-1	134255159	134255926		395055
GRMZM2G016819		Uncharacterized protein	10	1	134399292	134400918		250063
GRMZM2G016939		Protein kinase G11A protein	10	1	134410371	134414070		236911
GRMZM2G017789		Ubiquitin carboxyl-terminal hydrolase isozyme L3 protein	10	1	134414639	134418345		232636

Gene stable ID	Gene name	Gene description	Chr	Strand	Coordinate start	Coordinate end	Most significant SNP	Distance from gene (Mbp)
GRMZM2G027173		ELMO domain-containing protein 2 protein	10	-1	134431102	134434923		216058
GRMZM2G026983		Uncharacterized protein	10	-1	134436742	134439351		211630
GRMZM2G092599		Uncharacterized protein	10	-1	134439511	134441032		209949
GRMZM5G805382		Uncharacterized protein	10	-1	134457411	134458785		192196
GRMZM2G063972		Uncharacterized protein	10	1	134495450	134504022		146959
GRMZM2G018631		Putative RING zinc finger domain superfamily protein	10	-1	134566707	134567642		83339
GRMZM2G322586		Uncharacterized protein	10	1	134569581	134576295		74686
GRMZM2G322582		Putative WAK-related receptor-like protein kinase family protein	10	1	134583763	134585981		65000
GRMZM2G018485		Uncharacterized protein	10	-1	134600335	134602107		48874
GRMZM2G018464		Putative G10 domain family protein	10	1	134606660	134612036		38945
GRMZM2G018314		Uncharacterized protein	10	1	134647347	134652537	S10_134650981	1556
GRMZM2G322506		Uncharacterized protein	10	1	134653914	134659406		2933
GRMZM5G887529		Uncharacterized protein	10	-1	134707390	134707643		56409
GRMZM2G018027		Uncharacterized protein	10	-1	134717099	134718846		66118
GRMZM2G072121		Uncharacterized protein	10	1	134791598	134804631		140617
GRMZM2G149178	*H4C7*	Histone H4	10	-1	134805431	134806076		154450
GRMZM2G481249		Uncharacterized protein	10	1	134827823	134828576		176842
GRMZM2G181030		Putative MYB DNA-binding domain superfamily protein protein	10	-1	134830171	134832693		179190
GRMZM2G321239		CDPK protein; Putative calcium-dependent protein kinase family protein	10	1	134917888	134922506		266907
GRMZM2G097813		Uncharacterized protein	10	1	134969148	134971067		318167
GRMZM2G097848		Uncharacterized protein	10	-1	134974316	134977168		323335
GRMZM2G031280		Putative RING zinc finger domain superfamily protein	10	-1	134991961	134992953		340980
GRMZM2G031326		Nucleolar complex protein 4	10	-1	134999391	135006443		348410
GRMZM2G031453		Uncharacterized protein	10	1	135008283	135011071		357302
GRMZM2G322634		Uncharacterized protein	10	-1	135044840	135048155		393859
AC233979.1_FG006		Uncharacterized protein	10	-1	135056126	135058358		405145
GRMZM5G814481		Uncharacterized protein	10	-1	135125968	135126659		474987
AC233979.1_FG007		Uncharacterized protein	10	-1	135130731	135131021		479750
AC233979.1_FG008		Uncharacterized protein	10	-1	135135344	135135658		484363
AC233979.1_FG009		Uncharacterized protein	10	1	135164262	135169386		513281
AC233979.1_FG010		Uncharacterized protein	10	1	135171168	135172765		520187
AC233979.1_FG011		Putative RING zinc finger domain superfamily protein	10	1	135174868	135175986		523887
GRMZM2G413796		Uncharacterized protein	10	1	135273676	135318895		622695
GRMZM2G113818		Uncharacterized protein	10	1	135325713	135327992		674732
GRMZM2G113800		Uncharacterized protein	10	1	135330047	135331862		679066
GRMZM2G413774		Uncharacterized protein	10	1	135364950	135393105		713969
GRMZM2G086474		Putative HLH DNA-binding domain superfamily protein	10	1	135,677,365	135,680,963		326612
GRMZM2G086403		PLATZ transcription factors proteidn	10	-1	135,685,281	135,687,959		319616
GRMZM2G173429		Uncharacterized protein	10	1	135,800,706	135,802,992		204583
GRMZM2G162640		Uncharacterized protein	10	1	135,910,599	135,912,870		94705
GRMZM2G397684		Putative RING zinc finger domain superfamily protein	10	1	136,006,849	136,007,871	S10_136007575/ S10_136007578	293
GRMZM2G098676		Putative DUF604-domain containing/glycosyltransferase-related family protein	10	1	136,016,690	136,019,623		9112
GRMZM2G152135	*crtRB1*	Beta-carotene hydroxylase 1	10	-1	136,057,100	136,060,219		49522
GRMZM2G454399		Uncharacterized protein	10	-1	136,064,677	136,068,220		57099
GRMZM5G894619		Uncharacterized protein	10	-1	136,083,487	136,086,305		75909
GRMZM2G040673		Plant-specific domain TIGR01568 family protein	10	-1	136,103,462	136,105,185		95884
GRMZM2G040359		Uncharacterized protein	10	-1	136,105,398	136,107,763		97820
GRMZM2G016477		Putative leucine-rich repeat receptor-like protein kinase family protein	10	-1	136,114,579	136,118,747		107001
GRMZM2G105855		Uncharacterized protein	10	-1	136,179,354	136,184,210		171776
GRMZM2G305146		Uncharacterized protein	10	1	136,245,381	136,247,528		237803
GRMZM2G134703		Putative cytochrome P450	10	-1	136,271,138	136,273,515		263560

Gene stable ID	Gene name	Gene description	Chr	Strand	Coordinate start	Coordinate end	Most significant SNP	Distance from gene (Mbp)
		superfamily proteins proteind						
GRMZM2G414915		Putative subtilase family proteins protein	10	1	136,292,132	136,296,887		284554
GRMZM2G702514		Uncharacterized protein	10	-1	136,325,395	136,339,339		317817
AC199382.3_FG005		Uncharacterized protein	10	1	136,339,578	136,340,304		332000
GRMZM2G018018		Uncharacterized protein	10	-1	136,435,256	136,440,939		427678
GRMZM2G424112		Uncharacterized protein	10	-1	136,478,132	136,478,844		470554
GRMZM2G006212		Putative MATE efflux family protein	10	-1	136,615,140	136,617,256		607562
GRMZM2G077082		Uncharacterized protein	10	1	136,697,887	136,699,626		690309
GRMZM2G077069		Uncharacterized protein	10	-1	136,700,998	136,704,479		693420
GRMZM2G077036		Uncharacterized protein	10	1	136,709,655	136,713,380		702077
AC199370.4_FG006		Uncharacterized protein	10	-1	136,742,880	136,743,326		735302
GRMZM5G896748		Uncharacterized protein	10	1	136,742,977	136,743,532		735399
GRMZM2G148467		Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein	10	-1	139078940	139081895		795699
GRMZM2G124965		Uncharacterized protein	10	1	139115338	139120886		756708
GRMZM2G125023	*TIP2-3*	Aquaporin TIP2-3	10	1	139122197	139125216		752378
GRMZM2G392040		Uncharacterized protein	10	1	139191682	139193279		684315
GRMZM2G311059		Putative MYB DNA-binding domain superfamily protein	10	-1	139838952	139839963		37631
GRMZM2G007899		Uncharacterized protein	10	-1	139840305	139841290		36304
GRMZM2G303768		Putative leucine-rich repeat receptor-like protein kinase family protein	10	1	139852918	139855428		22166
GRMZM2G004709		EREBP-4 like protein protein	10	-1	139858081	139861295		16299
GRMZM2G080516		AP2-EREBP transcription factor	10	-1	139875910	139877865	S10_139877594	271
GRMZM2G129006		Uncharacterized protein	10	-1	139934291	139935167		56697
GRMZM2G555422		Uncharacterized protein	10	-1	139959580	139962869		81986
GRMZM2G128057		Uncharacterized protein	10	1	139976216	139979369		98622
GRMZM2G128074		Uncharacterized protein	10	-1	139980303	139984144		102709
GRMZM2G128078		Uncharacterized protein	10	-1	139984959	139987580		107365
GRMZM2G128092		Uncharacterized protein	10	-1	139995082	140003086		117488
GRMZM2G081557		Putative MYB DNA-binding domain superfamily protein protein	10	-1	140048664	140050182		171070
GRMZM2G119383		Uncharacterized protein	10	-1	140142031	140143150		264437
GRMZM2G119370		Uncharacterized protein	10	1	140145782	140146639		268188
GRMZM2G150107		Uncharacterized protein	10	1	140170361	140170979		292767
GRMZM2G397848		Uncharacterized protein	10	-1	140184269	140185171		306675
GRMZM2G097618		Uncharacterized protein	10	-1	140185478	140185945		307884
GRMZM2G097628		Uncharacterized protein	10	-1	140186581	140187542		308987
GRMZM2G097636		Putative MYB DNA-binding domain superfamily protein	10	-1	140190031	140191484		312437
GRMZM2G097638		Putative MYB DNA-binding domain superfamily protein	10	-1	140193073	140194565		315479
GRMZM2G104989		Putative RNA recognition motif containing family protein	10	-1	140223325	140225120		345731
AC198515.3_FG011		Uncharacterized protein	10	1	140268585	140271261		390991
GRMZM2G017752		Uncharacterized protein	10	1	140289208	140290291		411614
GRMZM2G017647		Uncharacterized protein	10	-1	140305834	140311238		428240
GRMZM2G159389		Uncharacterized protein	10	-1	140344785	140347607		467191
GRMZM2G458283		Putative translation elongation factor Tu family protein	10	-1	140351382	140355980		473788
GRMZM5G872750		Uncharacterized protein	10	-1	140357577	140358332		479983
GRMZM2G364060		SC3 protein protein	10	-1	140486168	140496625		608574
AC192330.1_FG004		Uncharacterized protein	10	1	140597198	140599033		719604
AC208110.2_FG001		Putative WRKY DNA-binding domain superfamily protein	10	1	140656185	140659057		778591

Note: Genes and associated information retrieved from maizeGDB.org and gramene.org. Physical positions of SNPs and coordinates of genes given according to B73 RefGen_2.

Appendix 1 – Table 7 Carotenoid concentration (µg/g) of 80 hybrids generated from 24 inbred lines in a factorial mating design and tested across environments at Saminaka, Zaria, Bagauda and Ikene in Nigeria

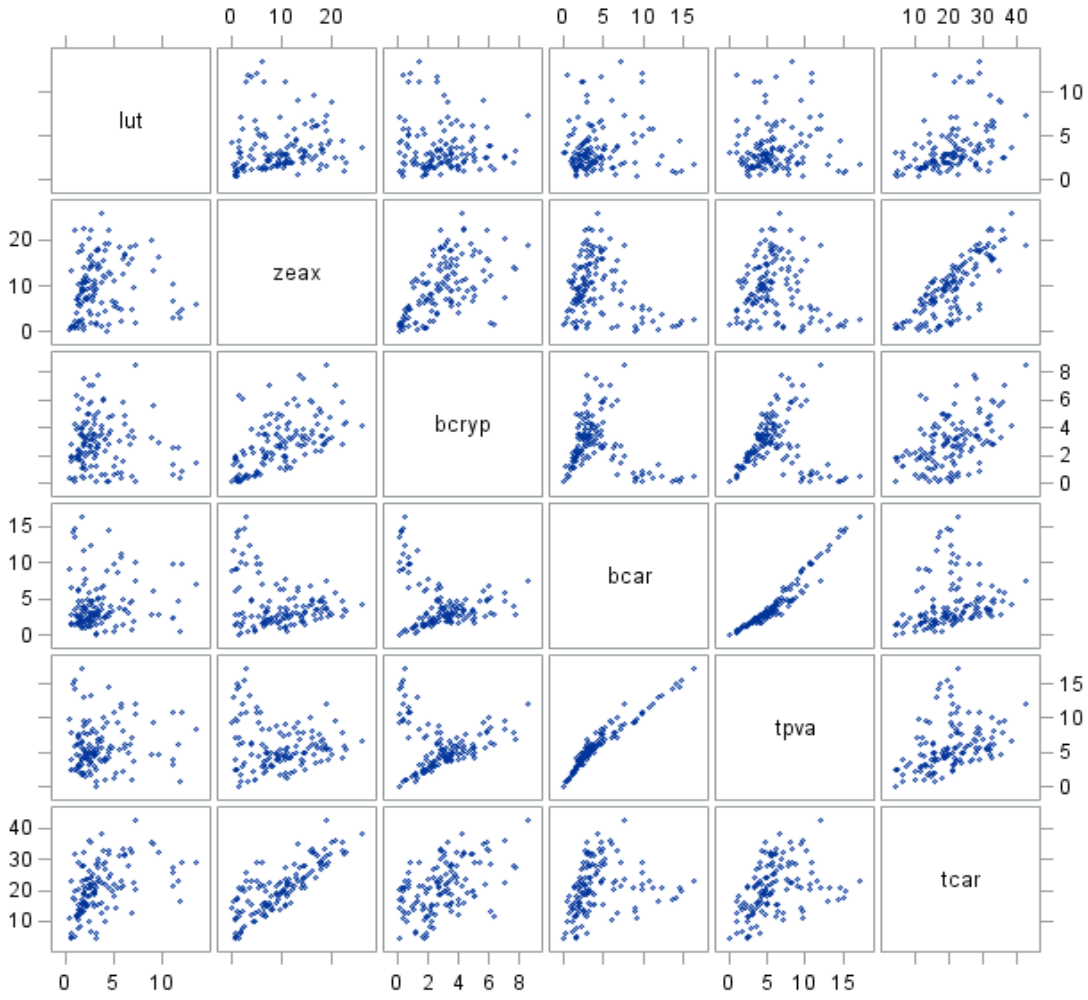
	Set	Female	Male	Lut	Zea	αcar	βcrp	βcar	pva	tcar
1	1	1	8	4.18	6.30	0.34	1.43	4.06	4.93	15.85
2	1	2	8	5.27	3.88	0.31	1.48	3.32	4.21	13.66
3	1	3	8	6.65	8.01	0.50	2.28	4.42	5.82	21.07
4	1	4	8	7.86	4.99	0.38	0.93	4.17	4.83	18.20
5	1	1	7	5.08	5.94	0.37	1.14	3.90	4.64	15.96
6	1	2	7	4.31	5.65	0.35	1.24	3.71	4.51	15.00
7	1	3	7	5.41	8.62	0.38	1.59	3.82	4.81	19.37
8	1	4	7	8.92	4.91	0.44	0.71	4.72	5.29	19.49
9	1	1	6	5.35	12.72	0.32	2.33	3.71	5.03	23.56
10	1	2	6	3.60	14.31	0.27	2.52	3.11	4.49	22.99
11	1	3	6	4.50	17.04	0.34	2.89	2.91	4.53	26.89
12	1	4	6	7.17	12.67	0.46	1.96	4.84	6.06	26.58
13	1	1	5	5.17	8.33	0.41	1.48	4.90	5.85	19.94
14	1	2	5	3.57	7.57	0.30	1.62	4.93	5.89	17.47
15	1	3	5	3.76	11.57	0.33	2.15	5.04	6.29	22.04
16	1	4	5	6.28	5.51	0.44	0.84	5.33	5.97	18.08
17	3	9	4	8.08	8.78	0.40	1.78	2.87	3.95	21.46
18	3	10	4	7.89	10.53	0.44	1.65	3.91	4.95	23.91
19	3	11	4	9.86	10.84	0.45	1.58	3.73	4.76	25.89
20	3	12	4	8.71	13.56	0.42	2.15	4.15	5.43	28.37
21	3	9	3	6.51	12.20	0.33	3.11	2.69	4.41	23.69
22	3	10	3	3.92	14.84	0.36	3.52	3.32	5.27	25.03
23	3	11	3	7.21	14.53	0.53	3.76	3.86	6.00	28.60
24	3	12	3	4.14	13.71	0.30	3.81	3.24	5.29	24.08
25	3	9	2	5.24	9.12	0.26	2.23	2.60	3.84	18.59
26	3	10	2	3.91	11.68	0.31	2.69	2.70	4.19	20.55
27	3	11	2	5.16	11.34	0.37	3.08	3.01	4.72	21.90
28	3	12	2	3.59	11.12	0.22	3.28	2.88	4.62	20.12
29	3	9	1	5.10	9.17	0.36	2.45	2.71	4.11	19.07
30	3	10	1	5.10	14.04	0.42	2.63	3.86	5.39	25.24
31	3	11	1	6.30	13.32	0.42	2.93	3.89	5.57	25.77
32	3	12	1	4.14	14.19	0.33	3.42	3.58	5.44	24.57
33	2	5	16	5.02	12.55	0.28	2.36	3.74	5.04	23.11
34	2	6	16	5.14	14.84	0.29	2.54	2.40	3.82	24.43
35	2	7	16	5.41	9.82	0.30	1.69	2.92	3.92	19.66
36	2	8	16	4.19	8.49	0.32	2.05	3.28	4.48	17.69
37	2	5	15	4.17	10.99	0.36	3.16	4.74	6.53	22.68
38	2	6	15	4.80	14.54	0.35	3.60	3.36	5.32	25.44
39	2	7	15	4.29	7.65	0.32	1.92	2.99	4.11	16.44
40	2	8	15	6.24	7.34	0.45	2.38	3.66	5.07	19.18
41	2	5	14	3.61	13.48	0.26	1.51	4.19	5.08	22.48
42	2	6	14	3.88	17.17	0.30	2.36	3.00	4.34	26.06
43	2	7	14	4.76	9.97	0.23	1.35	3.24	4.03	19.16
44	2	8	14	7.30	7.74	0.33	1.50	3.12	4.03	19.47
45	2	5	13	5.79	11.10	0.36	2.19	4.16	5.44	22.91
46	2	6	13	4.39	12.93	0.28	2.68	3.15	4.63	22.65
47	2	7	13	5.42	7.95	0.33	1.64	3.65	4.64	18.31
48	2	8	13	6.79	7.20	0.37	1.89	3.06	4.19	18.90
49	4	17	12	4.05	12.76	0.19	2.46	2.32	3.65	20.90
50	4	18	12	5.01	13.66	0.23	3.28	2.45	4.19	23.50
51	4	19	12	4.71	10.80	0.25	2.93	2.29	3.88	20.11
52	4	20	12	4.66	11.08	0.30	3.46	3.36	5.25	22.00
53	4	17	11	6.05	11.30	0.36	2.13	2.26	3.51	21.41
54	4	18	11	9.59	14.38	0.41	2.37	2.35	3.74	28.36
55	4	19	11	6.84	10.88	0.31	2.21	2.29	3.56	21.94
56	4	20	11	8.18	12.38	0.47	2.96	3.26	4.98	26.27
57	4	17	10	3.66	14.21	0.22	3.44	2.40	4.22	22.63
58	4	18	10	4.71	14.75	0.27	2.59	2.53	3.97	24.08
59	4	19	10	5.21	10.52	0.29	2.38	2.38	3.73	19.99
60	4	20	10	5.22	8.84	0.29	2.82	2.86	4.41	19.21
61	4	17	9	5.88	10.24	0.25	2.06	1.93	3.08	19.79
62	4	18	9	5.63	11.05	0.27	2.27	1.98	3.27	20.61
63	4	19	9	5.65	8.05	0.32	2.08	1.69	2.90	17.21
64	4	20	9	5.79	7.92	0.23	2.56	2.10	3.48	17.64
65	5	21	20	9.37	7.14	0.35	1.18	2.21	2.99	19.78
66	5	22	20	4.25	8.25	0.28	1.95	2.51	3.64	16.71
67	5	23	20	8.87	7.90	0.33	1.44	2.90	3.79	21.02
68	5	24	20	5.48	10.81	0.21	2.15	2.60	3.78	20.71

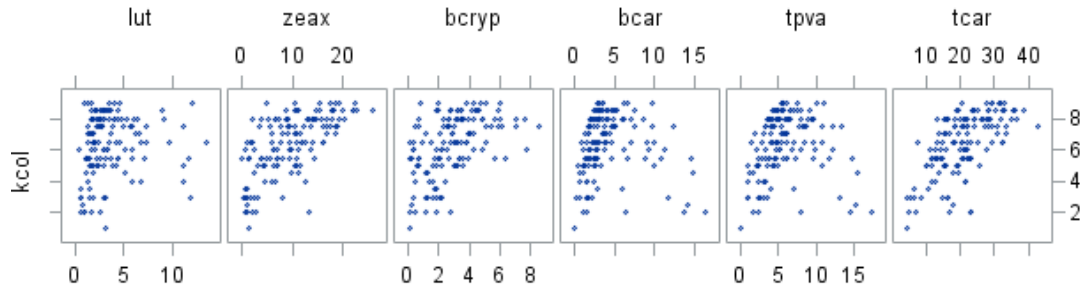
	Set	Female	Male	Lut	Zea	αcar	βcrp	βcar	pva	tcar
69	5	21	19	9.80	5.53	0.24	0.64	0.99	1.44	17.07
70	5	22	19	5.92	9.06	0.22	1.64	2.13	3.06	18.28
71	5	23	19	9.09	6.45	0.32	1.24	1.74	2.52	18.53
72	5	24	19	4.26	9.91	0.12	1.49	1.19	1.98	16.39
73	5	21	18	9.54	7.25	0.30	0.97	1.40	2.04	19.25
74	5	22	18	4.46	10.14	0.29	2.02	2.36	3.50	18.75
75	5	23	18	9.68	9.94	0.41	1.43	1.98	2.90	23.16
76	5	24	18	4.01	12.37	0.18	1.74	1.49	2.45	19.24
77	5	21	17	8.65	7.90	0.29	0.98	1.64	2.26	18.96
78	5	22	17	3.50	10.69	0.25	1.83	2.41	3.46	18.38
79	5	23	17	9.11	10.09	0.38	1.55	2.28	3.25	23.04
80	5	24	17	4.50	12.99	0.17	1.61	1.35	2.24	20.19
81.00	Check	25	27	4.47	14.40	0.18	2.01	1.73	2.83	22.33
Grand Mean				5.79	10.47	0.32	2.14	3.04	4.27	21.10
Average SE				1.02	1.23	0.07	0.30	0.64	0.72	2.06
Min				3.50	3.88	0.12	0.64	0.99	1.44	13.66
Max				9.86	17.17	0.53	3.81	5.33	6.53	28.60
Range				3.5-9.86	3.88-17.17	0.12-0.53	0.64-3.81	0.99-5.33	1.44-6.53	13.66-28.6
LSD average (p<0.01)				1.88	2.60	0.15	1.25	0.60	1.34	4.12

Appendix 2 – Supplemental Figures

Appendix 2 – Figure 1 The diverse colours of maize ears observed among the inbred liens studied

Appendix 2 – Figure 2 Scatter plot matrices for pairwise correlation between carotenoid traits and kernel colour





Zea = cryptoxanthin, bcryp = β-cryptoxanthin, bcar = β-carotene, tpva = total provitamin A, tcar = total carotenoid

Appendix 2 – Figure 3 Sample ears and their corresponding carotenoid profiles and genotype for the six allele specific markers

e) Entry 39

Genotype	lut	zea	βcry	βcar	tpva	tcar	
0, G, 8, 2, 0, 1	1.70	2.72	0.44	16.38	17.25	22.88	µg g^{-1}

f) Entry 45

Genotype	lut	zea	βcry	βcar	tpva	tcar	
0, G, 8, 2, 0, 1|3	1.03	2.39	0.34	14.79	15.45	19.53	µg g^{-1}

g) Entry 53

Genotype	lut	zea	βcry	βcar	tpva	tcar	
0, G, 8, 2, 0, 1	0.76	0.99	0.25	14.19	15.10	17.65	µg g^{-1}

h) Entry 64

Genotype	lut	zea	βcry	βcar	tpva	tcar	
0, G, 8, 1, 0, 3	2.99	8.82	4.59	2.49	4.87	19.04	µg g^{-1}

i) Entry 74

Genotype	lut	zea	βcry	βcar	tpva	tcar	
0, G, 8, 1, 0, 3	1.13	6.92	2.02	1.65	2.71	11.83	µg g^{-1}

j) iEntry 90

Genotype	lut	zea	βcry	βcar	tpva	tcar	
0, G, 0, 1, 0, 3	2.38	15.81	3.52	4.27	6.18	26.21	µg g^{-1}

k) Entry 92

Genotype	lut	zea	βcry	βcar	tpva	tcar	
0, G, 8, 1,12, 3	3.96	15.56	6.02	6.72	10.10	32.98	µg g^{-1}

l) Entry **106**

Genotype	lut	zea	βcry	βcar	tpva	tcar	
0, G, 8, 2,12, 1	1.80	0.26	1.30	11.80	13.02	16.30	µg g^{-1}

m) Entry 124

Genotype	lut	zea	βcry	βcar	tpva	tcar	
3, T|0, 2, 0|I, 1|3	0.99	1.12	0.16	13.64	14.33	17.08	µg g

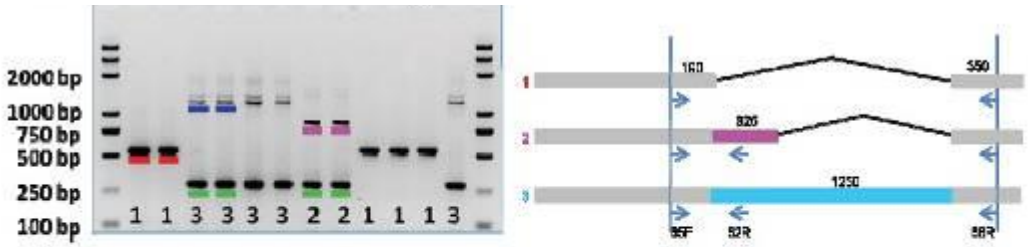
n) Entry 128

Genotype	lut	zea	βcry	βcar	tpva	tcar	
0, G, 0, 1,12, 3	1.93	11.42	3.29	2.46	4.12	19.04	µg g^{-1}

Appendix 2 – Figure 4 Illustrations of expected and observed fragments based on the 130 inbred lines assayed with the allele specific markers

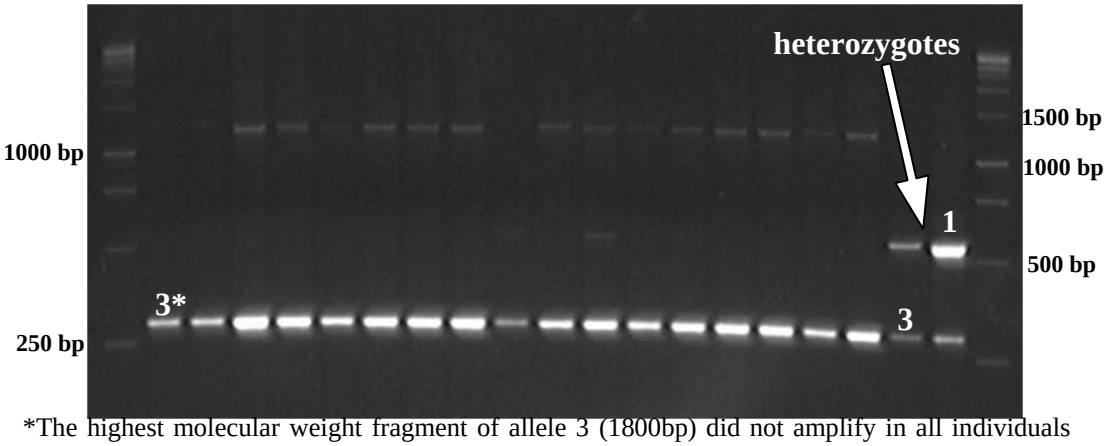
A) *crt*RB1 3'TE

Expected band patterns for each allele of *crt*RB13'TE polymorphic site marker assay (Yan et al 2010):



Allele hierarchy according to β-carotene concentrations: 1>3>2
1: 543 bp (without insertion)
2: 296+875 bp (with 325bp insertion)
3: 296+1221+1800 bp (with1250bp insertion)
Note: the largest fragment (1800 bp) by primers 65F and 66R was usually weak or not amplified.

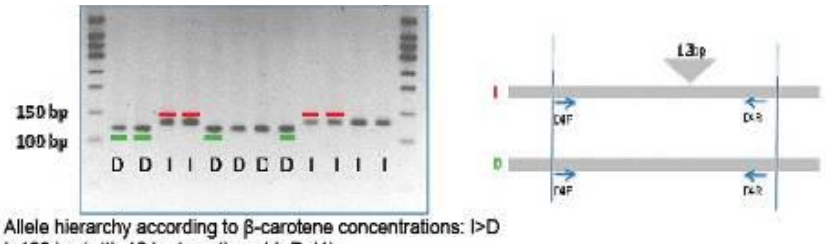
Sample agarose gel image of 3'TE polymorphic site of *crt*RB1:



*The highest molecular weight fragment of allele 3 (1800bp) did not amplify in all individuals assayed (also noted on Yan et al 2010) and in some individuals the fragment of 1221bp did not

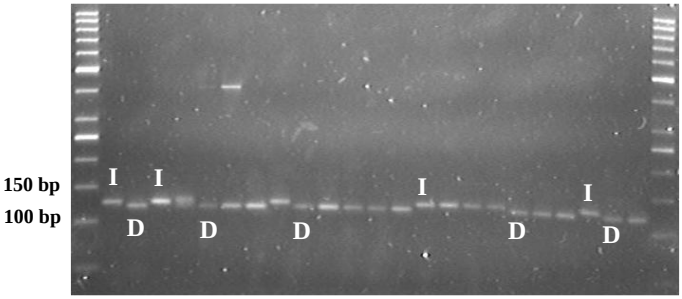
B) *crt*RB1 InDel4

Expected band pattern for *crt*RB1 InDel4 polymorphic site marker assay (Yan et al 2010):



Allele hierarchy according to β-carotene concentrations: I>D
I: 129 bp (with 12 bp insertion at InDel4)
D: 117 bp (with 12 bp deletion at InDel4)
Note: highly rare and non-functional 6bp insertion at InDel3 may interfere with banding pattern

Sample agarose gel image of *crt*RB InDel4 polymorphic site genotyping:

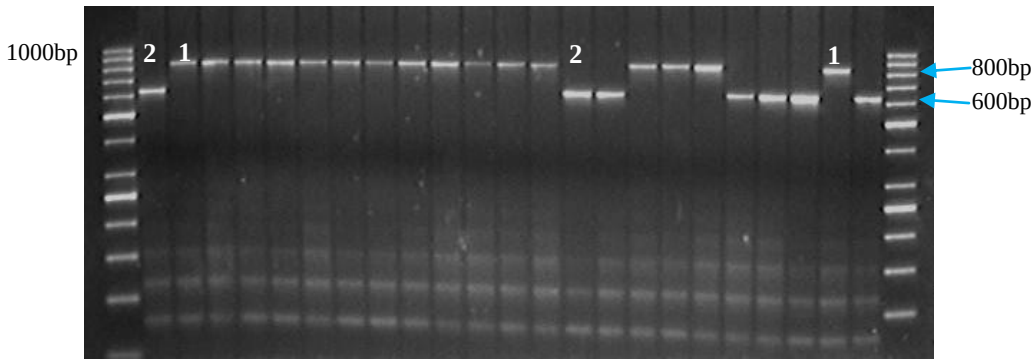


C) *crt*RB1 5'TE

Expected fragment sizes and band pattern for *crt*RB1 5'TE polymorphic site marker assay (Yan et al 2010)

Allele hierarchy according to β-carotene concentrations: 2>3>1
1: 800 bp (with 397bp insertion)
2: 600 bp (with 206bp insertion)
3: 400 bp (without insertion)

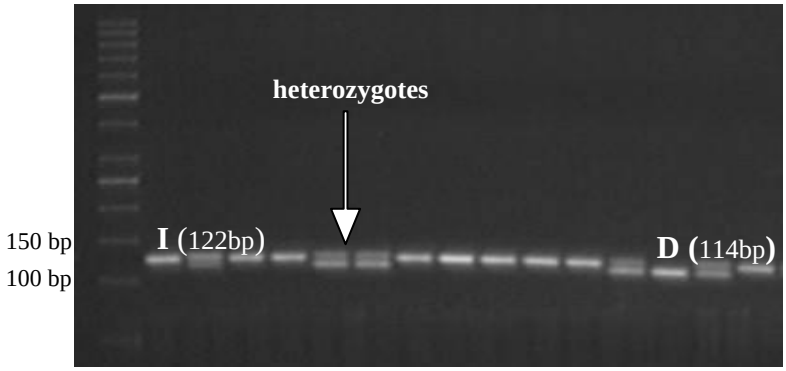
Sample agarose gel images of *crt*RB1 5'TE polymorphic site genotyping



D) *lcy*E 3'indel

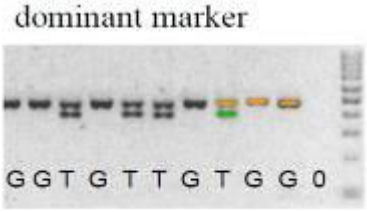
Expected fragment sizes for *lcy*E 3'indel:
D: 114 bp (8 bp deletion), I: 122 (8 bp insertion)
Allele hierarchy according to beta carotene concentration: D>I (Harjes et al, 2008)

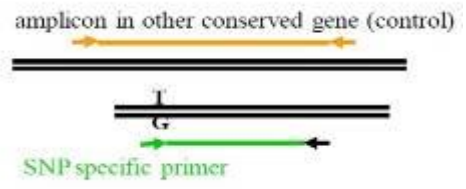
Sample agarose gel image for *lcy*E 3'Indel genotyping:



E) *lcy*E snp (216)

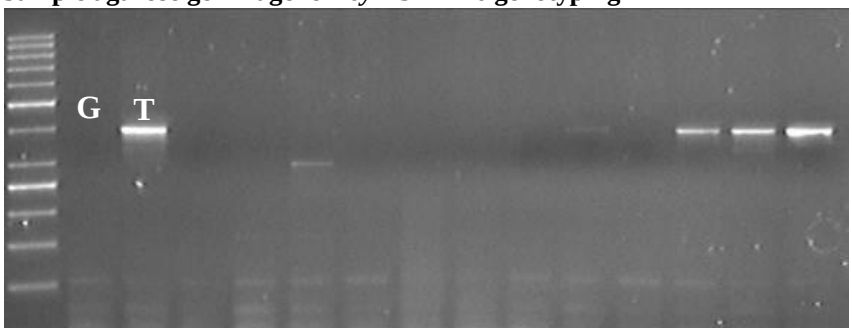
Expected fragment sizes/pattern for *lcy*E SNP216 polymorphism (Harjes et al 2008):





T: 395 bp
G: inbred line having the G SNP, no band if lines have this nucleotide substitution.
G class increases beta carotene

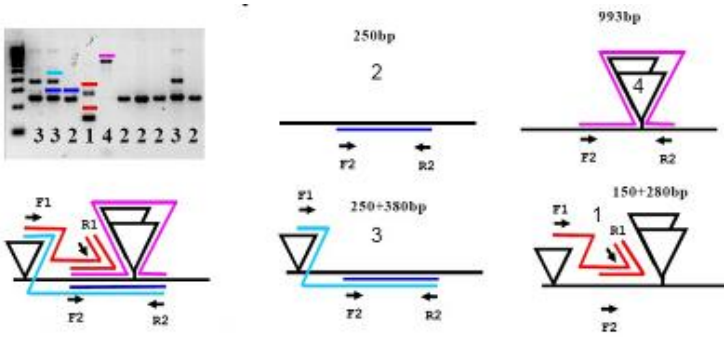
Sample agarose gel image for *lcy*E SNP216 genotyping



*Control PCR to amplify a different gene product as indicated in Harjes et al 2008 was not included

F) *lcyE* 5'TE

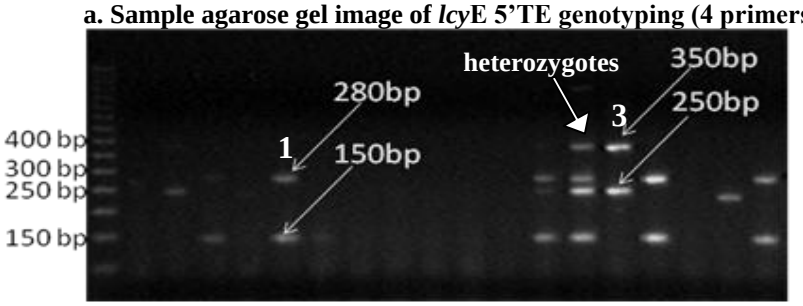
Expected fragment sizes of each allele of *lcy*E 5'TE polymorphic site (Harjes et al 2008)



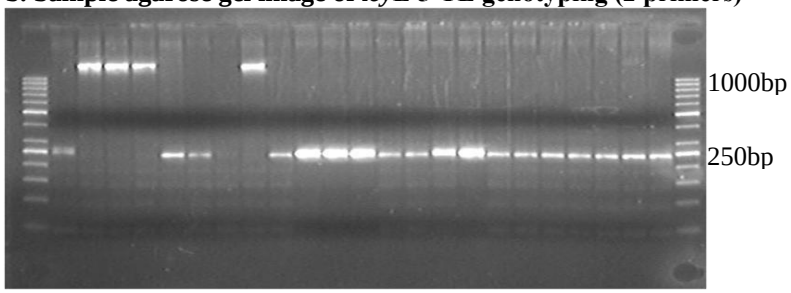
1:150+280bp,
2: 250bp,
3: 250+380bp
4. 993bp

Alleles 1 and 4, which represented the same allele, increase beta carotene

a. Sample agarose gel image of *lcy*E 5'TE genotyping (4 primers)

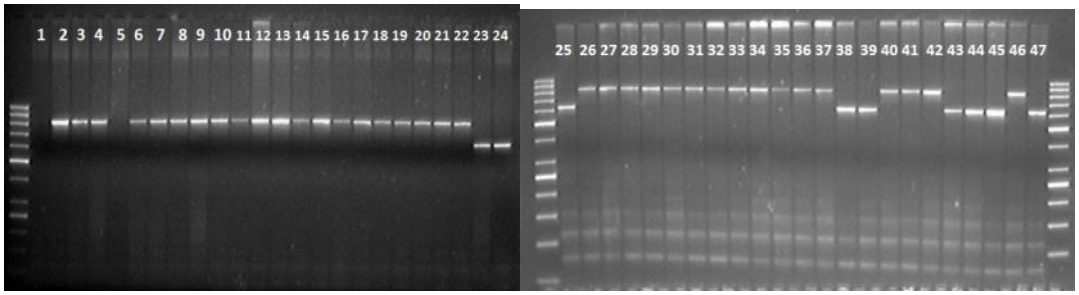


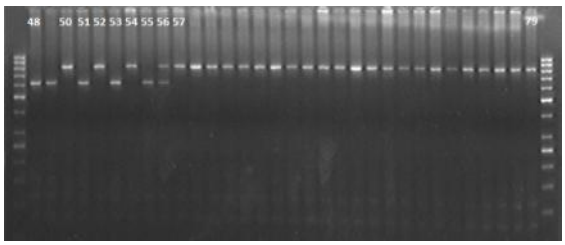
b. Sample agarose gel image of *lcy*E 5'TE genotyping (2 primers)

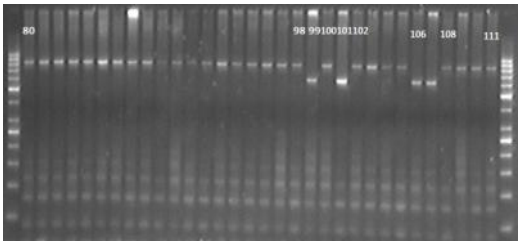


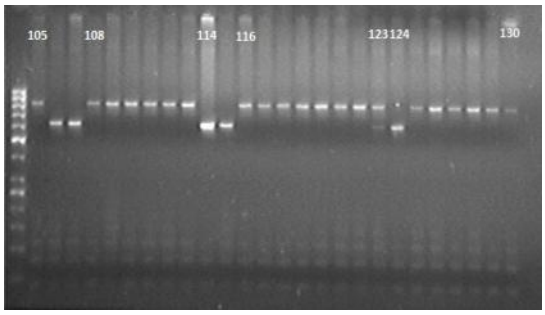
More sample gel images from the allele specific markers assay:

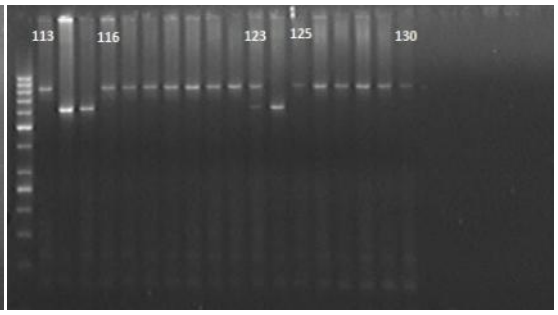
crtRB 5'TE:



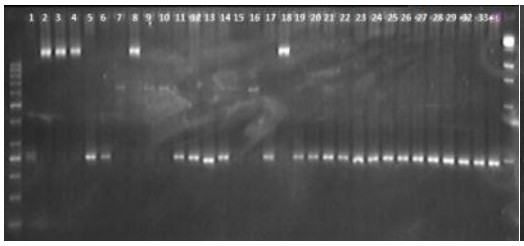


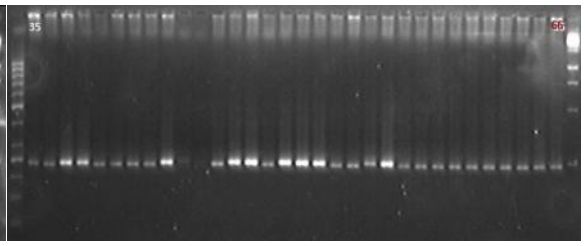


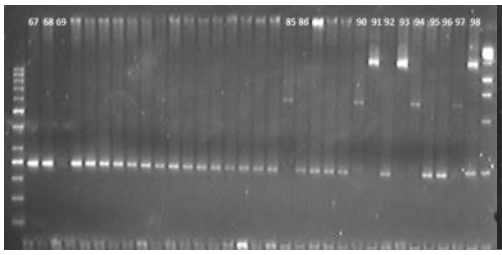


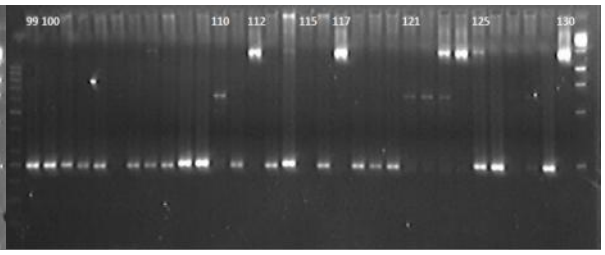


lcyE 5'TE (amplified with one primer pair):
Note: In the thesis, the result of the four primer amplification as suggested by Harjes et al. (2008) was discussed.

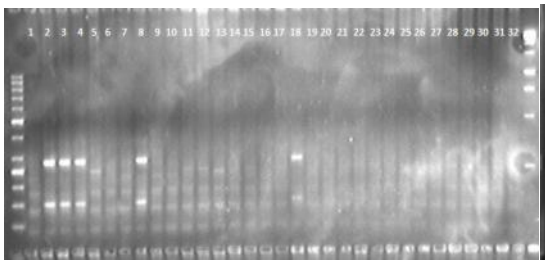


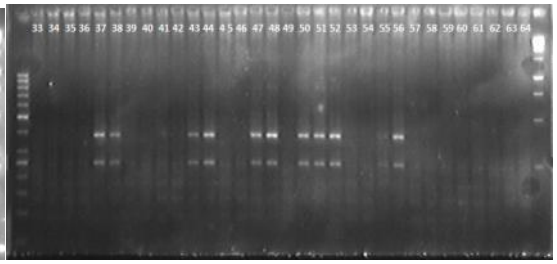


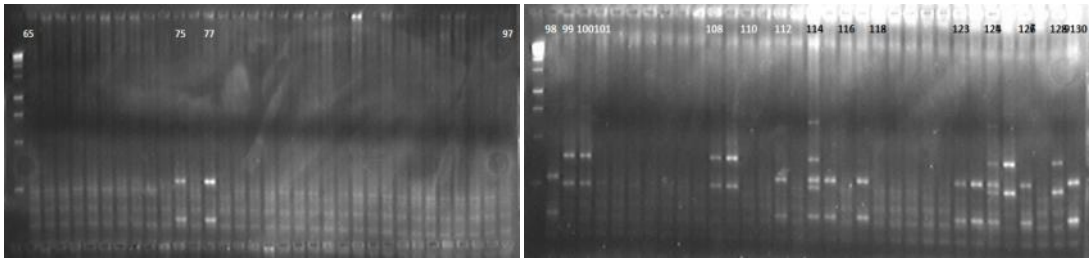




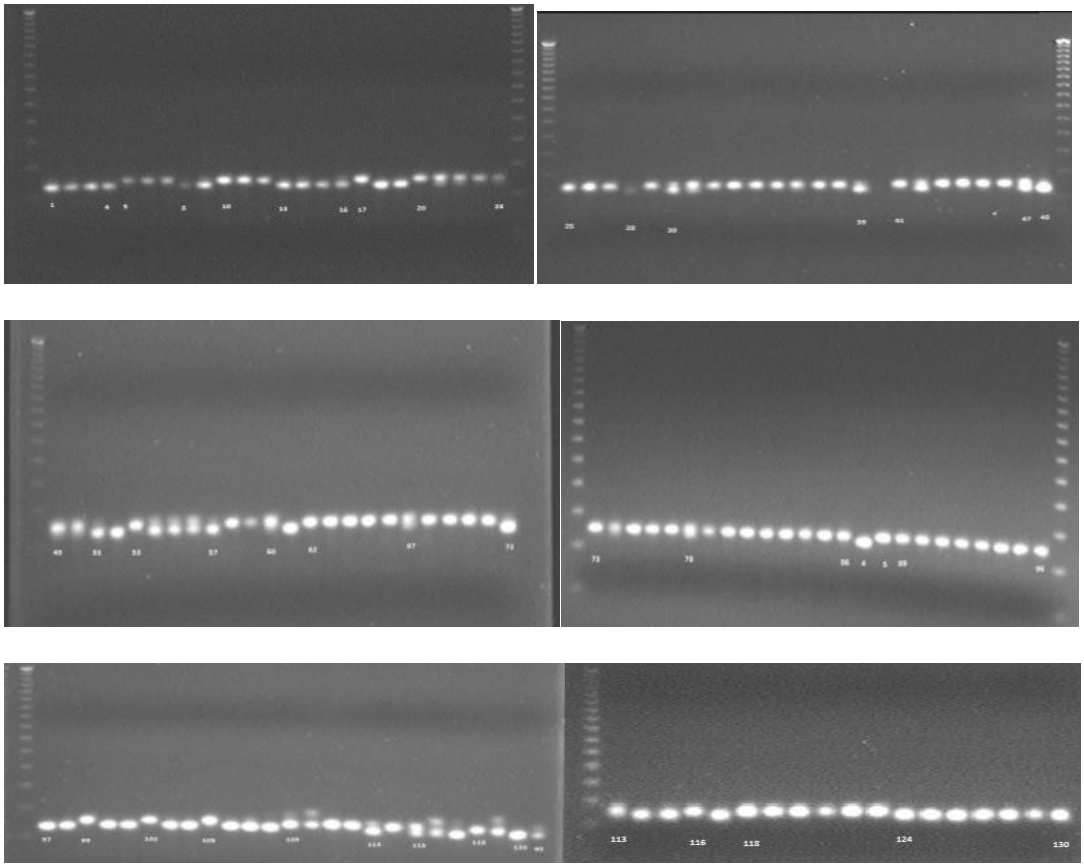
lcyE 5'TE (amplified with two primer pairs):



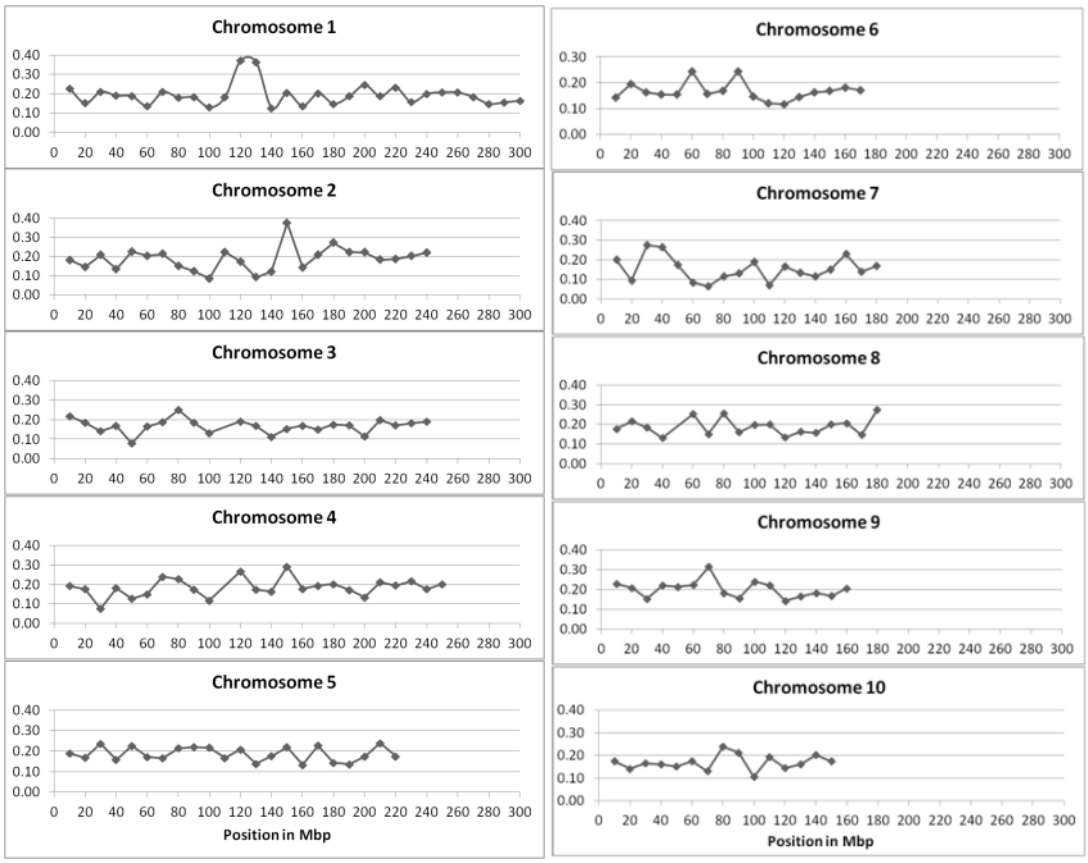




lcyE 3'Indel:



Appendix 2 – Figure 5 Polymorphic information content (PIC) across each maize chromosome averaged per 10 Mbp interval



Appendix 3 –Reports and Protocols/Methods

A) Purification of Total DNA from Plant Tissue (QIAGEN™ DNeasy™ plant mini kitprotocol)

1. 18 to 20 mg freeze dried leaf tissue was ground in 2 ml microcentrifuge tuebe containing metallic balls using Geno/Grinder 2000 (processed ~15 samples at a go).
2. 400 µl Buffer AP1 and 4 µl RNase A stock solution (100 mg/ml) was added to the samples and vortexed vigorously.
3. The mixture was incubated for 10 min at 65°C by mixing 2 or 3 times by inverting tubes.
4. 130 µl Buffer AP2 was added to the lysate, mixed, and incubated for 5 min on ice.
5. The lysate was centrifuged for 5 min at 20,000 x g (14,000 rpm).
6. The lysate was transferred into the QIAshredder Mini spin column (lilac) placed in a 2 ml collection tube using pipette, and centrifuged for 2 min at 20,000 x g (14,000 rpm).
7. The flow-through fraction was transferred from step 11 into a new tube without disturbing the cell-debris pellet.
8. 1.5 volumes of Buffer AP3/E was added to the cleared lysate, and mixed by pipetting.
9. 650 µl of the mixture was pipetted into the DNeasy Mini spin column placed in a 2 ml collection tube. Centrifuged for 1 min at 6000 x g and the flow-through was discarded. This step was repeated once more.
10. The DNeasy Mini spin columns was placed into a new 2 ml collection tube, 500 µl Buffer AW was added, and centrifuged for 1 min at 6000 x g (8000 rpm). Discard the flow-through and reuse the collection tube in step 17.
11. 500 µl Buffer AW was added to the DNeasy Mini spin column, and centrifuged for 2 min at 20,000 x g (14,000 rpm) to dry the membrane.
12. The DNeasy Mini spin column was transferred to a 1.5 ml or 2 ml microcentrifuge tube, and 100 µl Buffer AE was pipetted directly onto the DNeasy membrane, incubated for 5 min at room temperature (15–25°C), and then centrifuged for 1 min at 6000 x g (8000 rpm) to elute.
13. Repeat step 18 once.

B) IITA maize GBS – Reference Pipeline

A report prepared on 10/19/2012 by Katie Hyma, Computational Biology Service Unit, Cornell University, Ithaca, NY (keh233@cornell.edu)

154 uniquely named samples (130 diverse and related inbred lines plus 28 parental lines in 2 to 3 replication, making 190 total) and 2 blank(s) digested with enzyme(s) ApeKI

Library Plate Library Plate ID Flowcell Lane:
IITA_Maize_PvA1 450015763 C12A2ACXX 7
IITA_Maize_PvA2 450015762 C12A2ACXX 8

Analysis Notes:
We used the maize .topm file topm_filtered_042012.topm with md5sum de7e067482076f47f924bb675f471aa3 (provided by the Buckler lab) for this analysis

The GBS analysis pipeline (Tassel Version: 3.0.134 Date: October 4, 2012) was run on these samples
An overview of GBS and the GBS pipeline can be found here: http://www.maizegenetics.net/Table/Genotyping-By-Sequencing/. See the following sections for a list of options used for this analysis.

***Please note,** the GBS pipeline is continually being developed, and as such the results from this run can only be reproduced with the GBS pipeline version noted above, and with the options listed in the following section.

***Please note** that we are providing both hapmap and VCF files. SNP calling for hapmap and VCF files are independent, and as such slight variations between the two files can be expected. The reports and analysis generated from SNPs here are based on genome-wide SNP calls from the VCF file. The VCF file retains information on depth of coverage.

***Please note:** We are providing unfiltered SNP calls, and you may like to filter those SNPs to something that makes sense for your species/dataset. Hapmap files can easily be filtered in Tassel 4.0 (http://www.maizegenetics.net/tassel). VCF files can easily be filtered using VCFtools (http://vcftools.sourceforge.net/).

GBS reference pipeline options that may affect results (not file direction) for reference pipeline:
Plugin option value Description
QseqToTagCountPlugin -s 400000000 Maximum number of good reads per lane. Default: 200,000,000
FastqToTagCountPlugin -s 400000000 Maximum number of good reads per lane. Default: 200,000,000
QseqToTagCountPlugin -c 1 Minimum number of times a tag must be present to be output. Default: 1
FastqToTagCountPlugin -c 1 Minimum number of times a tag must be present to be output. Default: 1
MergeMultipleTagCountPlugin -c 5 Minimum number of times a tag must be present to be output. Default: 1
QseqToTBTPlugin -c 1 Minimum taxa count within a qseq file for a tag to be output. Default: 1
FastqToTBTPlugin -c 1 Minimum taxa count within a fastq file for a tag to be output. Default: 1
MergeTagsByTaxaFilesPlugin -s 400000000 Maximum number of tags the TBT can hold while merging (default:200,000,000). Reduce this only if you run out of memory (omit the commas).
MergeTagsByTaxaFilesPlugin -x Merges tag counts of taxa with identical names if set to -x. Not performed by default
TagsToSNPByAlignmentPlugin -mnMAF 0.01 Minimum minor allele frequency. Defaults to 0.01. SNPs that pass either the specificed minimum minor allele frequency (mnMAF) or count (mnMAC) will be output.
TagsToSNPByAlignmentPlugin -mnLCov 0.1 Minimum locus coverage, i.e., the proportion of taxa with at least one tag at the locus. Default: 0.1
TagsToSNPByAlignmentPlugin -mnF 0 Minimum value of F (inbreeding coefficient). Not tested by default.
TagsToSNPByAlignmentPlugin -mnMAC 10 Minimum minor allele count. Defaults to 10. SNPs that pass either the specificed minimum minor allele count (mnMAC) or frequency (mnMAF) will be output.
TagsToSNPByAlignmentPlugin -inclRare Include the rare alleles at sites (3 or 4th states) (default:false).
TagsToSNPByAlignmentPlugin -mxSites 1000000 The maximum number of SNPs per chromosome for hapmap files (default = 200,000)
MergeDuplicateSNPsPlugin -callHets -callHets When two genotypes at a replicate SNP disagree for a taxon, call it a heterozygote. Defaults to false (= set to missing).
MergeDuplicateSNPsPlugin -misMat 0.05 Threshold mismatch rate above which the duplicate SNPs won't be merged. Default: 0.05.
MergeIdenticalTaxaPlugin -hetFreq 0.8 cutoff frequency between het vs. homozygote calls (default = 0.8)
tbt2vcfPlugin -ak 3 Maximum number of alleles that are kept for each marker across the popuation, default: 3
tbt2vcfPlugin -mnMAF 0.01 Minimum minor allele frequency (default: 0.0)
tbt2vcfPlugin -mnLCov 0.1 Minimum locus coverage (proportion of Taxa with a genotype) (default: 0.0)
MergeDuplicateSNP_vcf_Plugin -ak 3 Maximum number of alleles that are kept for each marker across the population, default: 3

Reads and Tags by lane:
Reading FASTQ file: /workdir/keh233/IITA_maize/internal/GBS/../fastq/C12A2ACXX_7_fastq.gz

Total barcodes found in lane:96

Total barcodes found in lane:96

Total number of reads in lane = 267337805

Total number of good barcoded reads = 231439509

Reading FASTQ file: /workdir/keh233/IITA_maize/internal/GBS/../fastq/C12A2ACXX_8_fastq.gz

Total barcodes found in lane:96

Total barcodes found in lane:96

Total number of reads in lane = 260265941

Total number of good barcoded reads = 241334304

5415553 tags will be output to C12A2ACXX_7.tbt.byte

5439831 tags will be output to C12A2ACXX_8.tbt.byte

Flowcell C12A2ACXX lane 7: failed samples (# of reads les than 0.1 * mean(enzyme,plate)

Flowcell C12A2ACXX lane 8: failed samples (# of reads les than 0.1 * mean(enzyme,plate)

HapMap SNPs: 619,596

VCF SNPs: 835,339

VCF SNP statistics:

	mean	**median**	**standard deviation**
Individual depth	3.496816	3.345645	1.082671
Site depth	3.496827	2.80208	2.367882
Individual missingness	0.487116	0.464239	0.108757
Site missingness	0.487116	0.510417	0.274799

Multi Dimension Scaling (MDS) genome-wide SNPs from the reference genome pipeline:

VCFtools version [v0.1.8] was used to calculate missingness and depth on the unfiltered datasets

VCFtools version [v0.1.8] was used to filter the SNPs and generate an input file of remaining biallelic SNPs for use with Plink version [v1.07]/

• Filtered genotypes to those with sequencing depth between [3 and 1000]

• Filtered individuals with greater than [90] % missing data:

- A1140-108:C12A2ACXX:8:250116903
- A1140-1:C12A2ACXX:7:250117019
- A1140-20:C12A2ACXX:7:250117038
- A1140-31:C12A2ACXX:7:250117049
- A1140-37:C12A2ACXX:7:250117055
- A1140-38:C12A2ACXX:7:250117056
- A1140-43:C12A2ACXX:7:250117061
- A1140-54:C12A2ACXX:7:250117072
- A1140-55:C12A2ACXX:7:250117073
- A1140-61:C12A2ACXX:7:250117079
- A1140-63:C12A2ACXX:7:250117081
- A1140-97:C12A2ACXX:8:250116892
- Blank:C12A2ACXX:7:250117103
- Blank:C12A2ACXX:8:250116896
- P1:C12A2ACXX:8:250116928
- P1:C12A2ACXX:8:250116929

- Filtered sites with more than [20]% missing data, resulting in:
  - [46110] filtered, biallelic SNPs

***Please note:** We are providing unfiltered SNP calls, and you may like to filter those SNPs to something that makes sense for your species/dataset.

Bibliography:

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, et al. (2011) A Robust, Simple Genotyping- by-Sequencing (GBS) Approach for High Diversity Species. PLoS ONE 6(5): e19379. doi:10.1371/ journal.pone.0019379

Bradbury, P. J., Z. Zhang, D. E. Kroon, T. M. Casstevens, Y. Ramdoss, and E. S. Buckler. "TASSEL: Software for Association Mapping of Complex Traits in Diverse Samples." Bioinformatics 23, no. 19 (June 22, 2007): 2633–2635.

Petr Danecek, Adam Auton, Goncalo Abecasis, Cornelis A. Albers, Eric Banks, Mark A. DePristo, Robert Handsaker, Gerton Lunter, Gabor Marth, Stephen T. Sherry, Gilean McVean, Richard Durbin and 1000 Genomes Project Analysis Group. The Variant Call Format and VCFtools. Bioinformatics, 2011

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ & Sham PC (2007) PLINK: a toolset for whole-genome association and population-based linkage analysis. American Journal of Human Genetics, 81, URL:http://pngy.mgh.harvard.edu/purcell/plink

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

End of GBS report sent from Cornell University

END