REFINING OF THE EGUSI LOCUS AND MAPPING OF SEED COAT COLOR GENES IN WATERMELON

by

LUCKY PAUDEL

(Under the Direction of Cecilia McGregor)

ABSTRACT

Seed coat type (SCT) and seed coat color (SCC) are economically important traits in edible seed watermelons. Current efforts to develop edible seed watermelon varieties with different seed coats are faced with difficulty due to limited knowledge of these traits. This study aimed to refine the egusi locus for identification of candidate genes and enabling marker assisted selection (MAS) for the egusi trait and to map loci conferring SCC in watermelon. Using QTL-seq and KASP™ assays, the egusi locus was reduced from a 4.28 Mb region to a 398.25 Kb region, extending from 6.95 Mb to 7.35 Mb on chromosome 6. KASP™ assay UGA6_7026576 was identified as associated with egusi trait and useful for MAS. The inheritance of SCC was studied in three F_2 populations. The phenotypic segregation in these populations indicated that the stipple SCC is dominant to green, red and clump SCC, the *R* locus is dominantly epistatic to T^1, and the *D* locus is recessively epistatic to *W*. The SCC loci, *R*, T^1, *W* and *D*, were mapped on chromosome 3, 5, 6 and 8, respectively.

INDEX WORDS: *Citrullus lanatus, C. mucosospermus*, seed coat color (SCC), edible seed watermelon, egusi locus, QTL-seq, KASP™ assay

REFINING OF THE EGUSI LOCUS AND MAPPING OF SEED COAT COLOR GENES IN WATERMELON

by

LUCKY PAUDEL

B.Sc., CCS Haryana Agricultural University, India, 2011

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2018

© 2018

Lucky Paudel

All Rights Reserved

REFINING OF THE EGUSI LOCUS AND MAPPING OF SEED COAT COLOR GENES IN WATERMELON

by

LUCKY PAUDEL

Major Professor: Cecilia McGregor
Committee: Anish Malladi
Joann Conner

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
August 2018

DEDICATION

I dedicate this dissertation to my family. This would have never been possible without their unconditional love and support.

ACKNOWLEDGEMENTS

I am grateful to my family who made my goals their own. I am also thankful to Dr. Cecilia McGregor for her guidance and mentorship and members of McGregor's Lab, Dr. Yihua Chen, Reeve Legendre, Winnie Gimode, Leigh Ann Fall, Jorge Reyes and Jesse Daniel Kuzy for their help with my projects.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS v

LIST OF TABLES viii

LIST OF FIGURES ix

CHAPTER

1 INTRODUCTION 1

2 LITERATURE REVIEW 4

Origin and Taxonomy 4

Watermelon breeding 5

Genomic tools in watermelon breeding 12

References 16

3 REFINING THE EGUSI LOCUS IN WATERMELON USING KASP ASSAYS 26

Abstract 27

Introduction 28

Materials and methods 30

Results 34

Discussion 36

References 39

4 CHROMOSOMAL LOCATIONS AND INTERACTIONS OF FOUR LOCI ASSOCIATED WITH SEED COAT COLOR IN WATERMELON 55

Abstract 56

Introduction 57

Materials and methods 59

Results and discussion 63

Conclusion 70

References 71

5 CONCLUSION 88

References 91

APPENDIX

A CUSTOM-MADE PYTHON SCRIPT 93

LIST OF TABLES

Page

Table 2.1: Development of various seed coat colors in watermelon due to interaction of four genes 25

Table 3.1: Watermelon cultivars and Plant Introductions used to develop mapping populations (MP), validation populations (VP) and the diversity panel (DP) for mapping and marker validation of the egusi seed type locus. 43

Table 3.2: Primer sequence of KASP^{TM} assays used for identification and narrowing down of the egusi (*eg*) locus on in watermelon. 45

Table 3.3: Candidate genes present in the refine egusi locus from 6.95 Mb to 7.35 Mb based on the 97103 watermelon genome (Guo et al. 2013) along with their annotated function.....48

Table 4.1: KASP^{TM} assays used to test association between significant genomic regions, identified from QTL-seq, and the seed coat color phenotype in watermelon...................77

LIST OF FIGURES

Page

Figure 3.1: Seeds of the parents used to develop mapping and validation populations. (a) Seeds of Strain II (PI 279461), (b) Dried and fresh (lower left corner insert) seeds of PI 560023 (Egusi), (c) Seeds of UGA147 (PI 169233)..50

Figure 3.2: Absolute ΔSNP-index graph plot of all chromosomes obtained from QTL-seq analysis..51

Figure 3.3: Bar graphs showing the marker phenotype association between seed coat phenotype and KASP™ markers UGA6_7026576 (solid) and UGA8_17929262 (dotted) in the (a) Strain II (PI 279461) x Egusi (PI 560023) and (b) UGA147 (PI 169233) x Egusi population 52

Figure 3.4: Partial genetic map of chromosome 6 of the Strain II (PI 279461) x Egusi (PI 560023) F_2 watermelon population showing the refined position of the egusi locus. Map positions are in cM. 53

Figure 3.5: (a) Genotypic and phenotypic data for KASP™ markers UGA6_6903757 (n = 136), UGA6_6958189 (n = 139) and UGA6_7026576 (n = 137), in the mapping population, Strain II (PI 279461) x Egusi (PI 560023), and (b) KASP™ marker UGA6_7026576 in the validation population UGA147 (PI 169233) x Egusi, (n = 145). 54

Figure 4.1: Seed coat color of parents, F_1, and F_2 progenies in (a) the stipple x green population, (b) the stipple x red population and (c) the stipple x clump population.........78

Figure 4.2: (a) Absolute ΔSNP-index of all chromosomes in the inter-specific stipple x green F_2 population developed from a cross between Sugar Baby (*C. lanatus*) and PI 482379 (*C. amarus*) plotted along with statistical confidence intervals under the null hypothesis of no QTL (P = 0.01) (red line). (b) Magnified view of *R* locus, a significant ΔSNP-index peak (yellow), along with absolute ΔSNP-index of SNPs (black circles) plotted against the SNP position. (c) Association of KASP™ marker UGA3_5820134 with seed coat color phenotype in the stipple x green F_2 population (n = 126) 79

Figure 4.3: (a) Mapping of the T^1 locus as a significant peak on chromosome 5 using QTL-seq. (b) Magnified view of significant ΔSNP-index peak (yellow) associated with T^1 locus along with absolute ΔSNP-index of SNPs (black circles) plotted against SNP position based on 97103 watermelon genome (Guo et al., 2013). (c) Association of KASP™ marker UGA5_4591722 with the khaki and red seed coat phenotype in the stipple x red F_2 population. (d) Bar graph indicating the phenotypic prediction accuracy of KASP™ markers UGA3_5820134 and UGA5_4591722 in the stipple x red population (n = 96). 81

Figure 4.4: A genetic linkage map of the Stipple x Clump F_2 population developed from a cross between stipple seeded Charleston Gray and a clump seeded UGA147, selection from PI 169233 83

Figure 4.5: Genetic maps of chromosome 6 (a) and 8 (b) showing the position of the *W* and *D* loci in the stipple x clump F_2 population developed from a cross between stipple seeded Charleston Gray and clump seeded UGA147 (selection from PI 169233). (c) Bar graph indicating phenotypic prediction accuracy of SNP markers

UGA6_7076766 and UGA8_22729513 in the stipple x clump population (n = 174). 87

CHAPTER 1

INTRODUCTION

Watermelon [*Citrullus lanatus* (Thumb.) Matsum. & Nakai)] is an economically important crop of the *Cucurbitaceae* family. It is a member of the xerophytic genus *Citrullus* and is grown in warmer climates and sandy soils (Wehner et al. 2001; Wehner 2008; Paris 2015). It is a trailing, annual crop with lobed leaves and large, edible fruit. Plants are monoecious or andromonoecious with many staminate flowers and a pistilllate or a hermaphrodite flower at every seventh or eighth leaf axis (Rosa 1928; Porter 1933; Paris 2015). Watermelon is an outcrossing species grown for the fruit, which has a thick, outer rind and an inner sweet, juicy flesh.

Watermelon has 2n = 2x = 22 chromosomes and a relatively small genome size of 425 Mb (Arumuganathan and Earle 1991; Ren et al. 2012; Guo et al. 2013). Naturally occurring watermelon is diploid, however, tetraploid watermelon can be developed artificially by chromosome doubling, and triploid watermelons for seedless production can be produced by crossing tetraploid and diploid plants (Kihara 1951).

Watermelon is one of the most important vegetable crops cultivated in the warm regions of the world. In most parts of the world, watermelon is cultivated for the sweet, juicy, red fleshed fruit, but in some countries, watermelon is also produced for the oleaginous, proteinous seed. Such watermelon are called edible seed watermelon (Zhang and Jiang 1990; Zhang 1996). In 2013, approximately 3.35 million hectares was under watermelon cultivation worldwide, and a total of 108.08 million metric tons of watermelon was produced (FAO 2013). In the same year,

China produced 73.15 million metric tons of watermelon on 1.84 million hectares which accounts for two thirds of both total production and land under cultivation worldwide. After China, Turkey, Iran, Brazil and the United States of America are the other top producers of watermelon (FAO 2013). In the US, a total of 1.59 million metric tons of watermelon was harvested from 46,300 hectares in 2015, with an estimated market worth of $ 483 million. Approximately 44 percent of total watermelon produced in the U.S. is cultivated in California, Florida, Georgia and Texas and 91 percent of the watermelon produced was seedless watermelon (USDA-NASS 2016).

Although seedless watermelon is popular in the U.S., there are many parts of the world where watermelon is cultivated particularly for the seeds. Watermelon cultivation for seeds is common in several Central and Western African nations, China, India, and other Middle-Eastern countries (National Research Council 2006; Mahla et al. 2014) and are called edible seed watermelon (Zhang and Jiang 1990). Egusi watermelon, an edible seed watermelon with unique mucilage around the seed, is the predominant type of watermelon used for seed cultivation in west-African countries, while normal seed type, landrace watermelon are cultivated for seed, in India and China (Mahla et al. 2014). The land under watermelon for edible seed production in China in 1996 was 95,000 ha, about double the total amount of land under current watermelon production in the U.S (Zhang 1996; USDA-NASS 2016).

The land dedicated to production of edible watermelon seed is increasing globally every year (National Research Council 2006; Mahla et al. 2014). Despite this, edible seed watermelons have received limited attention from plant breeders. Even economically important traits, yield, seed color and seed quality have not been studied in detail for decades and most of the breeding for new variety development is still done using conventional approaches which is inefficient,

time consuming and resource intensive (Wehner et al. 2001; Mahla et al. 2014). Routine utilization of modern tools like marker assisted selection (MAS) is crucial to accelerate watermelon breeding programs. However, MAS requires adequate knowledge about the genetics of traits and the development of molecular markers tightly linked to those traits. This study aimed to understand the (i) inheritance pattern, (ii) identify associated genomic regions, and (iii) develop high-throughput markers for MAS of traits conferring seed coat color and seed coat type in watermelon.

CHAPTER 2

LITERATURE REVIEW

Origin and Taxonomy

The taxonomy and nomenclature of watermelon and its closest relatives have long been debated. The first scientific name for watermelon, *Cucurbita Citrullus,* was provided by Carl Von Linniè in his book *Species Plantarum,* which was published in 1753 (Paris 2015). However, since then several different names have been assigned (Renner et al. 2014). Moreover, several botanists and breeders have often made mistakes in the taxonomy of citron, egusi and dessert watermelons (Chomicki and Renner 2015) and have misplaced citron, egusi and dessert watermelons as a subspecies of *Citrullus lanatus* (Paris 2015). In the most recent literature, watermelon and its relatives have been classified into seven distinct species (Chomicki and Renner 2015; Paris 2015). Sweet watermelon, also known as dessert watermelon, has been designated *C. lanatus*. Its closest relative is egusi watermelon, *C. mucosospermus,* which usually has seed with a unique mucilaginous seed coat. The preserving melon, also known as citron, tsamma, or cow melon, has been designated *C. amarus.* These melons are cultivated in Southern Africa and are used in jam preparation and as a source of water. Other species related to watermelon are *C. colocynthis, C. rehmii* and *C. ecirrhosus* (Jarret et al. 1997; Robinson and Decker-Walters 1997; Chomicki and Renner 2015; Paris 2015)

Like its nomenclature and taxonomy, the origin of watermelon is also a highly-debated topic. Watermelon has been cultivated since ancient times. The earliest remains of watermelon seed was dated as 5000 years old and was discovered in southwest Libya (Wasylikowa and Van

der Veen 2004). This discovery has reopened the already complex discussion about the origin of the crop. There are several hypotheses claiming different parts of Africa as a centre of origin. One of the hypotheses is that watermelon originated in Northern Africa from *Citrullus colocynthis* (Singh 1978; Sain et al. 2002; McCreight et al. 2013; Paris 2015). But, the veracity of this hypothesis is highly dubious since hybrid plants from a cross between watermelon and colocynth have reduced fertility, and often have chromosomal irregularities (Whitaker and Davis 1962; Shimotsuma 1963; Sain et al. 2002; Paris 2015). Another hypothesis that was widely accepted until recently, states that watermelon originated in Southern Africa from *C. amarus* (Robinson and Decker-Walters 1997; Maynard and Maynard 2000; Rubatsky 2001). However, findings of a molecular phylogenetic study of watermelon relatives by Chomicki and Renner (2015) and archaeological findings in southwest Libya contradict this hypothesis. Lately, it has been proposed that watermelon originated in Western Africa, and the progenitor of the sweet watermelon is in fact *C. mucosospermus* (Guo et al. 2013; Paris 2015). Chomicki and Renner (2015) conducted a molecular phylogenetic study and discovered that *C. mucosospermus* and *C. lanatus* are sister species which diverged 3.1 million years ago, which supports this theory. West Africa as the center of origin for *C. lanatus* has again been disputed in a recent study that showed that wild germplasm from northeast Africa (historically classified as *C. lanatus* subsp. *cordophanus)* is more closely related to cultivated watermelon than any other *Citrullus* species (Fursa and Gavrilyuk 1990; Renner et al. 2017).

Watermelon breeding

The current global watermelon breeding trend is mostly focused on breeding seeded or seedless varieties for sweet flesh (Kumar and Wehne 2010). The normal seeded watermelon cultivars are diploid and are developed either as open pollinated or F_1 hybrids. Older, seeded

watermelon cultivars like “Sugar Baby”, “Crimson Sweet” and “Calhoun Gray” are still popular among home gardeners, whereas new F_1 hybrids like “Topgun”, “Sangria”, “Estrella”, “Walker” and “Jade Star” have been developed by several private seed companies for commercial cultivation.

Seedless watermelon cultivars are more popular among consumers in the U.S. and some parts of Europe, hence, most of the watermelon breeding effort in these regions is focused on developing new seedless varieties. Seedless cultivars are triploid and developed by crossing a diploid plant with a tetraploid plant which is developed through artificial chromosomal doubling. Triploid plants have reduced fertility as they fail to have normal meiotic pairing, and hence fail to produce seeds. This natural inability of triploid watermelon to produce seeds has been commercially exploited to develop seedless watermelons. Some of the seedless cultivars are “Fascination”, “Sweet Dawn”, “Exclamation”, “Captivation”, “Summer Breeze”, “Joy Ride”, “Embasy” and “Warrior”. New seeded and seedless varieties are selected for economically important traits related to fruit quality, fruit yield and resistance against common diseases and pests.

Breeding of edible seed watermelons is done on a much smaller scale compared to seedless watermelon breeding. New edible seed watermelon varieties are developed with an emphasis on increasing total yield of watermelon seeds by increasing seed size, number of seeds per fruits and number of fruits per plant. Most of the edible seed watermelon belonging to *C. lanatus* species are cultivated in Asian countries, whereas edible seed landraces belonging to *C. mucosospermus*, locally known as egusi watermelon, are cultivated in Africa. Some of the edible seed watermelon varieties developed in China are “Hetaopi”, “Lahzhou Daban”, “GN-1”,

"Jingyuan Daban" and "Ningxia Red" (Zhang and Jiang 1990; Zhang 1996). Unlike in China, most of the egusi watermelon cultivated in African countries are landraces.

Economically important traits in edible seed watermelon

Watermelon seeds are rich in proteins (25-35%) and lipids (15-43%) and serve as an important part of the human diet (Akobundu et al. 1982; El-Adawy and Taha 2001; Jarret and Levy 2012; Prothro et al. 2012). They are eaten as snacks or used in preparing soups, stews, dumplings, and seasonings (National Research Council 2006; Achigan-Dako et al. 2008; Ntui et al. 2010; Prothro et al. 2012b). Flour of egusi watermelon seed is used in preparing meals and the oil is used in cooking and salad dressing (National Research Council 2006). Seed size, seed coat type and seed coat color are economically important traits since they affect yield, quality and market value of edible seeds.

Seed size

Seed size is an important consideration in watermelon breeding. In the case of watermelon cultivation for edible seed, breeders mostly select for large seed, since the seed size is an important yield component and an increase in seed size increases total yield. However, when breeding for flesh production, the breeder's objective is to develop large sized seed for direct seeding that will produce fruits with small sized seeds. (Tanaka et al. 1995; Prothro et al. 2012b).

Watermelon has a wide genetic diversity for seed size, seed length and seed width. Poole et al. (1941) grouped watermelon seed into three groups on the basis of seed size: long, medium and short with average seed lengths of 13 mm, 10 mm and 6 mm, respectively. They determined that two genes "*s*' and "*l*", and the interaction between them (*s* epistatic to *l*), is responsible for the short, medium, and long seed length. The genotypes *LLSS*, *llSS*, and *llSS* or *LLss* give seed

with medium, long and short length respectively (Poole et al. 1941; Guner and Wehner 2004; Prothro et al. 2012b). Similarly, Tanaka et al. (1995) found that seed size was controlled by dominant short seed gene *Ti* and recessive medium seed gene *ti*. The short seed gene was epistatic to the medium gene. However, they could not confirm whether these new genes were the same genes as described by Poole et al. (1941). Zhang et al. (1994) reported tomato seed size (length by width: 4.2×2.6 mm) in watermelon which is smaller than the small seed size described by Poole et al. (1941). This phenotype is controlled by a single recessive gene *ts* and is expressed when the genotype is *LLsststs* (Wehner et al. 2001).

The first high quality single nucleotide polymorphic (SNP) genetic map was produced from two populations segregating for seed size by Sandlin et al. (2012), and Prothro et al. (2012b) used them to identify quantitative trait loci associated with seed size. They observed a high correlation between seed length, seed width and 100 seed weight and identified six quantitative trait loci (QTL) associated with these traits. A major QTL which explained 26.9% to 73.6% of phenotypic variation was discovered for all three traits ($Qsl6^{M}$, $Qswd6^{M}$, $Q100swt6^{M}$) between 4.58 Mb and 6.44 Mb on chromosome 6 and an intermediate QTL ($Q100swt2\text{-}1^{I}$, $Qfl2\text{-}2^{I}$, $Qfwd2\text{-}2^{M}$) associated with all traits between 28.24 Mb and 31.96 Mb on chromosome 2. Such colocalization of QTL implies that either a single gene with a pleiotropic effect controls the three seed traits or different genes controlling seed traits are tightly linked. In this case, it is probably a pleiotropic effect. There was also a significant interaction between the loci on chromosomes 2 and 6. Additional minor QTL were also identified on chromosome 8 for 100 seed weight (*Q100swt8*) and seed length (*Qsl8*). These QTL are protected by patent US20140041078 A1 (Methods and compositions for producing watermelon plants with selected seed sizes), owned by Seminis Vegetable Seeds, Inc.

Seed size in watermelon is important because it is positively or negatively correlated with other important traits. Larger seeds have higher germination percentage than smaller seeds (Kim et al. 2009). Fruit weight and total soluble solid of flesh are negatively correlated with seed surface, seed length and seed width (Achigan-Dako et al. 2008).

Seed coat type

Watermelon seeds can be classified into two groups based on seed coat types: normal seed and egusi seed. A normal watermelon seed has a glossy texture, a pyriform shape, and a hard seed coat with varied color, whereas egusi seed types have a soft, fleshy, swollen pericarp that develops around the seed. Such tissues are not present in normal watermelon seeds. The fleshy tissues covering egusi seeds are believed to be the remnants of nucellar tissues. These tissues develop during the second and third week of fruit development (Gusmini et al. 2004) and are present even when the fruit is mature. However, during drying the fleshy pericarp dries and becomes completely unnoticeable (Guner and Wehner 2004). The dried egusi seeds are large, flat and have a rough texture, ovoid shape, light yellow to cream-white color with a typical groove around the circumference of the seed. Among seeds of different watermelon species, egusi seeds usually have the largest seed length, seed width, seed surface and seed weight (Meru and McGregor 2013; Achigan-Dako et al. 2015). However, this is not true for all egusi seeds. Oyolu (1977) referred to egusi watermelon as *Colocynthis citrullus* and has grouped different egusi seeds into five groups based upon the seed size, texture of the seed coat and thickness of the seed edge. The five seed types that Oyolu (1977) described are as follows:

Type 1: Miniature seeds with a thin seed coat and flat edges

Type 2: Small seeds with a thin seed coat and flat edges

Type 3: Small seeds with an encrusted thick seed coat and flat edges

Type 4: Large seeds with thick seed coat and moulded edges

Type 5: Large seeds with a thick seed coat and flat edges

Besides physical traits, the bio-chemical composition of seed also varies significantly between the two seed types. Egusi seeds usually have a higher percentage of seed oil and protein (Akobundu et al. 1982; Jarret and Levy 2012) as compared to normal seeds. In the study conducted by Prothro et al. (2012), seed oil percentage in egusi seed ranged from 30.30 to 40.60% whereas seed oil percentage in normal type seed ranged from 20.14 to 26.55%.

The egusi phenotype is genetically controlled by a single recessive gene "*eg*" (Gusmini et al. 2004). The inheritance pattern of the egusi trait was validated by Prothro et al. (2012) and the *eg* locus was mapped at 57.8 cM on linkage group 2 between markers NW0248325 and NW0250248 of the Strain II (PI 279461) x Egusi (PI 560023) genetic map. On the physical map, the *eg* region extends from 6.75 Mb to 11.03 Mb region on chromosome 6. The closest marker to the egusi locus is NW0248325 but this marker is still 5.1 cM away from the egusi locus. The *eg* region is 4.28 Mb wide and contains 241 candidate genes.

Seed coat color

Watermelon seeds have a wide range of colors. The common watermelon seed coat colors are white, cream, yellow, tan, red, green, black and stippled. The seed color that is visible to our eye is in fact the color of the testa and it develops in two stages (UPOV 2013). First, the ground color appears on the testa of the developing seed, and later, the over color develops upon the ground color (UPOV 2013). The premature seeds which only have ground color are white, while the matured seeds have a distinct color which is the result of the superimposed final color over the ground color. However, some seeds are white even when they are mature. In some white seeds, a black spot at the hilum is present (UPOV 2013).

The inheritance of seed color is very complex, and little is known about the nature and location of genes, and gene interactions responsible for the development of color. The early studies of inheritance of watermelon seed coat color were done by Kanda (1931), McKay (1936), and Poole et al. (1941). Through a cross between flat black and stippled black seeded watermelon, Kanda (1931) demonstrated that a single dominant gene is responsible for the flat black color. Similarly, McKay (1936) conducted a cross between tan and red seeded watermelons and another cross between green and red seeded watermelons and in both crosses, the former color was monogenically dominant over red. Poole et al. (1941) developed 40 different populations to understand inheritance of seed color in watermelon. They found that three main genes (*R, T* and *W)* and a fourth, modifier gene (*D),* control the development of various seed colors (Table 2.1). The most dominant seed coat color to the least dominant can be ordered as black, stippled, tan, green, and red. (Kanda 1931; McKay 1936; Poole et al. 1941).

A good understanding of seed color development has become an important topic in watermelon breeding, partly because the consumer preference for specific color seed has increased. In China, people prefer large, red colored edible watermelon seeds, but breeders are finding it difficult to produce enough seeds of high yielding varieties for farmers because of complex interaction among genes controlling seed coat color (Zhang and Jiang 1990). Moreover, seed coat color is often associated with other traits. Poole et al. (1941) found that the *W* gene governing seed color was linked with the *L* gene controlling seed length. This phenomenon is also common in other plant species. In sesame, the seed coat color is associated with different biochemical properties and disease resistance (Zhang et al. 2013). In soybean, genes controlling seed colors effect the cell wall protein and carotenoid components (Lindstrom and Vodkin 1991;

Monma et al. 1994). Because of lack of adequate research on watermelon seed coat color, we still do not know much about the association of seed color with other important agronomic traits.

Genomic tools in watermelon breeding

The selection of plants with desirable traits is the fundamental principle of plant breeding. Until recently, most of the crop improvements were done exclusively through conventional breeding approaches. However, several recent breakthroughs in genomic technologies have opened opportunities in employing genomic technologies as a tool in plant breeding.

Reference genome

In the past few decades, scientists have been able to produce reference genomes for several crops like maize (Schnable et al. 2009), soybean (Schmutz et al. 2010), tomato (The Tomato Genome Consortium 2012), rice (Goff et al. 2002), chickpea (Varshney et al. 2013), and watermelon (Guo et al. 2013). These reference genomes play a crucial role in discovering new genes and regulatory sequences, QTL mapping, allele mining and developing molecular markers (Huq et al. 2016). The first draft genome sequence of watermelon was developed by Guo et al. (2013) using Illumina sequencing technology. They sequenced an elite inbred Chinese watermelon cultivar, 97103. The final assembled genome was 353.5 Mb, and it represents 83.2 % of the total genome. The draft genome was assembled using 1973 scaffolds with N50 length of 2.38 Mb and 26.38 kb for the scaffolds and contigs, respectively. The remaining uncovered regions were suspected to be transposable elements. Importantly, this draft sequence is publicly available. The American cultivar, Charleston Gray, has also recently been sequenced and is publicly available (www.icugi.org/). The draft genome is 396.5 Mb representing around 93.3% of the watermelon genome. The *de novo* assembly was done using scaffolds with N50 value of 7.47 Mb.

Marker-assisted selection

Developing markers that can be used for MAS is a high priority in most plant breeding programs (Collard and Mackill 2008). Widespread application of MAS in watermelon breeding has only very recently become common practice in selection for horticulturally important traits like disease resistance (Kim et al. 2013). Having a high-density genetic map is highly valuable in both mapping genes and QTL as well as identifying markers for MAS. Several types of DNA markers like random-amplified polymorphic DNA (RAPD), restriction fragment length polymorphism (RFLP), simple sequence repeat (SSR), and single nucleotide polymorphism (SNP) have been used over the last few decades to create genetic maps in different inter sub-specific and intra specific populations of watermelon, but only in the past 5 to 6 years that truly useful MAS tools have become available to impact watermelon breeding.

In watermelon, the first genetic maps using SNP markers were constructed by Sandlin et al. (2012). They developed three mapping populations: a recombinant inbred line (RIL) population (n = 164) from a cross between two elite cultivars, Klondike Black Seed × New Hampshire Midget (KBS × NHM); an F_2 population resulting (n = 187) from a cross between an elite cultivar and a wild egusi accession, Strain II × PI 560023 (SII × Egusi); and an F_2 population (n = 182) developed from a cross between an elite cultivar and a wild *C. amarus* accession, ZWRM50 × PI 244019 (ZWRM × Citroides). The three genetic maps were 1,438 cM, 1,514 cM, and 1,144 cM with 378, 357 and 338 markers with an average marker distance of 3.8, 4.2 and 3.4 cM. Since then, many high-density genetic maps have been developed and different QTL related to watermelon fruit, seed, flowering and disease resistance traits have been mapped (Reddy et al. 2014; Ren et al. 2015; Meru and McGregor 2016; Shang et al. 2016). These genetic maps have been very helpful in developing different molecular markers for MAS of many

agronomically important traits. However, the number of functional markers and other markers capable of discerning traits in different genetic background are still very limited.

QTL-seq

QTL-seq is a new, powerful and cost effective genetic approach which combines the concept of bulk segregant analysis (Michelmore et al. 1991) and whole genome sequencing for easy and rapid identification of QTL associated with a trait of interest (Takagi et al. 2013; Illa-Berenguer et al. 2015). The principle of QTL-seq involves preparing two DNA bulks from an equal number of plants with contrasting phenotypes. Whole genome sequencing of the two DNA bulks is then carried out and a comparative SNP-index is used to identify QTL associated with the traits (Takagi et al. 2013; Illa-Berenguer et al. 2015). The first use of QTL-seq in identifying QTL in plants was by Takagi et al. (2013) to identify QTL associated with several important agronomic and economic traits like disease resistance and seedling vigor in rice. In watermelon, QTL-seq was used to identify a dwarfism gene *Cla010726* (*ClaGA20ox*) on chromosome 7 (Dong et al. 2018). In tomato, Illa-Berenguer et al. (2015) utilized QTL-seq to identify and map a new QTL controlling fruit weight, *fw11.2,* in the distal part of chromosome 11. Similarly, in chickpea, Das et al. (2015) used QTL-seq and traditional QTL mapping to narrow down a 1.37 Mb region for a major seed weight QTL to a 35 kb genomic region containing only six genes. In cucurbits, QTL-seq has been used in cucumber by Lu et al. (2014) and watermelon (Dong et al. 2018). Lu et al. (2014) used QTL-seq along with classical QTL analysis to narrow down the region containing a major QTL, *Ef1.1*, controlling flowering time in cucumber cultivar ‘Muromskij’, to an 890 kb genomic region and Dong et al. (2018) used QTL-seq to identify *Cla010726* on chromosome 7 as candidate dwarfism gene in watermelon.

Kompetitive allele-specific PCR (KASP^{TM}) genotyping assay

In order to use MAS routinely on a large scale, a high throughput method for genotyping linked or functional markers for the trait of interest must be developed. One of the most popular high throughput methods currently used for MAS is the Kompetitive Allele Specific PCR (KASP^{TM}; LGC Genomics service lab, Teddington, UK) assay. KASP^{TM} uses competitive allele-specific PCR to detect variation at a nucleotide level and reports using a homogenous fluorescence-based reporting system (He et al. 2014; Semagn et al. 2014). The fluorometric dyes HEX and FAM are added at the 5' end of the allele specific primers. The fluorometric dye and quencher gets separated when the primer is hybridized with DNA, and the fluorescence emitted during separation can be used to identify the genotype. In the case of a homozygous genotype, one of the two possible fluorescence signal will be emitted, while in case of heterozygous genotype, a mixed fluorescent signal will be emitted (He et al. 2014; Semagn et al. 2014).

References

Achigan-Dako EG, Avohou ES, Linsoussi C, Ahanchede A, Vodouhe RS, Blattner FR (2015) Phenetic characterization of *Citrullus* spp. (*Cucurbitaceae*) and differentiation of egusi-type (*C. mucosospermus*). Genet Resour Crop Evol 62:1159-1179

Achigan-Dako EG, Fagbemissi R, Avohou HT, Vodouhe RS, Coulibaly O, Ahanchede A (2008) Importance and practices of Egusi crops (*Citrullus lanatus* (Thunb.) Matsum. & Nakai, *Cucumeropsis mannii* Naudin and *Lagenaria siceraria* (Molina) Standl. cv.'Aklamkpa') in sociolinguistic areas in Benin. Biotechnologie, Agronomie, Société et Environnement 12:393-403

Akobundu E, Cherry J, Simmons J (1982) Chemical, functional, and nutritional properties of egusi (*Colocynthis citrullus* L.) seed protein products. J Food Sci 47:829-835

Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Molecular Biology Reporter 9:208-218

Chomicki G, Renner SS (2015) Watermelon origin solved with molecular phylogenetics including Linnaean material: another example of museomics. New Phytol 205:526-532

Collard BC, Mackill DJ (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philosophical Transactions of the Royal Society B: Biological Sciences 363:557-572

Das S, Upadhyaya HD, Bajaj D, Kujur A, Badoni S, Kumar V, Tripathi S, Gowda CL, Sharma S, Singh S (2015) Deploying QTL-seq for rapid delineation of a potential candidate gene underlying major trait-associated QTL in chickpea. DNA Res:193-203

Dong W, Wu D, Li G, Wu D, Wang Z (2018) Next-generation sequencing from bulked segregant analysis identifies a dwarfism gene in watermelon. Scientific reports 8:1-7

El-Adawy TA, Taha KM (2001) Characteristics and composition of watermelon, pumpkin, and paprika seed oils and flours. J Agric Food Chem 49:1253-1259

FAO (2013) Statistical yearbook 2013. World Food and Agriculture, FAO, Food and Agriculture Organization, UN Rome, Italy

Fursa T, Gavrilyuk I (1990) Phylogenetic relations in the genus *Citrullus* Schrad. according to data from immunochemical analysis of the seed proteins. Sbornik Nauchnykh Trudov po Prikladnoĭ Botanike, Genetike i Selektsii 133:19-26

Goff SA, Ricke D, Lan T-H, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science 296:92-100

Guner N, Wehner TC (2004) The genes of watermelon. HortScience 39:1175-1182

Guo S, Zhang J, Sun H, Salse J, Lucas WJ, Zhang H, Zheng Y, Mao L, Ren Y, Wang Z (2013) The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. Nat Genet 45:51-58

Gusmini G, Wehner T, Jarret R (2004) Inheritance of egusi seed type in watermelon. J Hered 95:268-270

He C, Holme J, Anthony J (2014) SNP Genotyping: The KASP Assay. In: Fleury D, Whitford R (eds) Crop Breeding: Methods and Protocols. Springer New York, New York, NY, pp 75-86

Huq MA, Akter S, Jung Y-J, Nou IS, Cho Y-G, Kang K-K (2016) Genome sequencing, a milestone for genomic research and plant breeding. Plant Breeding and Biotechnology 4:29-39

Illa-Berenguer E, Van Houten J, Huang Z, van der Knaap E (2015) Rapid and reliable identification of tomato fruit weight and locule number loci by QTL-seq. Theor Appl Genet 128:1329-1342

Jarret R, Merrick L, Holms T, Evans J, Aradhya M (1997) Simple sequence repeats in watermelon (*Citrullus lanatus* (Thunb.) Matsum & Nakai). Genome 40:433-441

Jarret RL, Levy IJ (2012) Oil and fatty acid contents in seed of *Citrullus lanatus* Schrad. J Agric Food Chem 60:5199-5204

Kanda T (1931) The inheritance of seed-coat colouring in the watermelon. Jap J Genet 7:30-48

Kihara H (1951) Triploid watermelons. Proc Amer Soc Hort Sci 58:217-230

Kim K-H, Ahn S-G, Hwang J-H, Choi Y-M, Moon H-S, Park Y-H (2013) Inheritance of resistance to powdery mildew in the watermelon and development of a molecular marker for selecting resistant plants. Horticulture, environment, and biotechnology 54:134-140

Kim Y-J, Yang T-J, Park Y-H, Lee Y-J, Kang S-C, Kim Y-K, Cho J-L (2009) Development of near isogenic lines with various seed sizes and study on seed size-related characteristics in watermelon. Korean journal of breeding science 41:403-411

Kumar R, Wehne T (2010) Breeding for yield in Watermelon-A review. HortScience 42:1332-1336

Lindstrom JT, Vodkin LO (1991) A soybean cell wall protein is affected by seed color genotype. The Plant Cell 3:561-571

Lu H, Lin T, Klein J, Wang S, Qi J, Zhou Q, Sun J, Zhang Z, Weng Y, Huang S (2014) QTL-seq identifies an early flowering QTL located near flowering locus T in cucumber. Theor Appl Genet 127:1491-1499

Mahla HR, Singh JP, Roy MM (2014) Seed purpose watermelon in arid zone. Central arid zone research institute, Jodhpur, Raj. India

Maynard D, Maynard DN (2000) Cucumber, melon and watermelon. In: Kniple KF, Orneles KC (eds) The cambridge world history of food. Cambridge University Press, Cambridge, UK, pp 298-313

McCreight JD, Staub JE, Wehner TC, Dhillon NP (2013) Gone global: familiar and exotic cucurbits have Asian origins. HortScience 48:1078-1089

McKay J (1936) Factor interaction in *Citrullus* seed-coat color, fruit shape and markings show evidence of mendelian inheritance in watermelon crosses. J Hered 27:110-112

Meru G, McGregor C (2013) Genetic mapping of seed traits correlated with seed oil percentage in watermelon. HortScience 48:955-959

Meru G, McGregor C (2016) Genotyping by sequencing for SNP discovery and genetic mapping of resistance to race 1 of *Fusarium oxysporum* in watermelon. Scientia Horticulturae 209:31-40

Michelmore RW, Paran I, Kesseli R (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. Proceedings of the national academy of sciences 88:9828-9832

Monma M, Ito M, Saito M, Chikuni K (1994) Carotenoid components in soybean seeds varying with seed color and maturation stage. Biosci, Biotechnol, Biochem 58:926-930

National Research Council (2006) Lost crop of Africa: volume II: vegetables. The national academies press, Washington, D.C

Ntui VO, Chin DP, Nakamura I, Mii M (2010) An efficient *Agrobacterium tumefaciens*-mediated genetic transformation of "Egusi" melon (*Colocynthis citrullus* L.). Plant Cell, Tissue and Organ Culture (PCTOC) 103:15-22

Oyolu C (1977) Extraction rates and chemical composition of seed types in egusi (*Colocynthis citrullus* L.). Acta Hort 53:287-290

Paris HS (2015) Origin and emergence of the sweet dessert watermelon, *Citrullus lanatus*. Ann Bot 116:133-148

Poole CF, Grimball PC, Porter DR (1941) Inheritance of seed characters in watermelon. Journal of agricultural research 63:433-456

Porter DR (1933) Watermelon breeding. Hilgarda:585-624

Prothro J, Sandlin K, Abdel-Haleem H, Bachlava E, White V, Knapp S, McGregor C (2012b) Main and epistatic quantitative trait loci associated with seed size in watermelon. J Am Soc Hort Sci 137:452-457

Prothro J, Sandlin K, Gill R, Bachlava E, White V, Knapp SJ, McGregor C (2012) Mapping of the egusi seed trait locus (*eg*) and quantitative trait loci associated with seed oil percentage in watermelon. J Am Soc Hort Sci 137:311-315

Reddy UK, Nimmakayala P, Levi A, Abburi VL, Saminathan T, Tomason YR, Vajja G, Reddy R, Abburi L, Wehner TC (2014) High-resolution genetic map for understanding the effect of genome-wide recombination rate on nucleotide diversity in watermelon. G3: Genes, Genomes, Genetics 4:2219-2230

Ren R, Ray R, Li P, Xu J, Zhang M, Liu G, Yao X, Kilian A, Yang X (2015) Construction of a high-density DArTseq SNP-based genetic map and identification of genomic regions with

segregation distortion in a genetic population derived from a cross between feral and cultivated-type watermelon. Mol Genet Genomics 290:1457-1470

Ren Y, Zhao H, Kou Q, Jiang J, Guo S, Zhang H, Hou W, Zou X, Sun H, Gong G (2012) A high resolution genetic map anchoring scaffolds of the sequenced watermelon genome. PLoS One 7:e29453-29410.21371/journal.pone.0029453.

Renner SS, Chomicki G, Greuter W (2014) Proposal to conserve the name *Momordica lanata* (*Citrullus lanatus*)(watermelon, *Cucurbitaceae*), with a conserved type, against *Citrullus battich*. Taxon 63:941-942

Renner SS, Sousa A, Chomicki G (2017) Chromosome numbers, Sudanese wild forms, and classification of the watermelon genus *Citrullus*, with 50 names allocated to seven biological species. Taxon 66:1393-1405

Robinson RW, Decker-Walters D (1997) Cucurbits. CAB International, Wallingford

Rosa J (1928) The inheritance of flower types in *Cucumis* and *Citrullus*. Hilgardia 3:233-250

Rubatsky V (2001) Origin, distribution, and uses. In: Maynard D (ed) Watermelons, characteristics, production, and marketing. ASHS Press, Alexandria, VA, pp 21-26

Sain R, Joshi P, Sastry ED (2002) Cytogenetic analysis of interspecific hybrids in genus *Citrullus* (*Cucurbitaceae*). Euphytica 128:205-210

Sandlin K, Prothro J, Heesacker A, Khalilian N, Okashah R, Xiang W, Bachlava E, Caldwell DG, Taylor CA, Seymour DK (2012) Comparative mapping in watermelon [*Citrullus lanatus* (Thunb.) Matsum. et Nakai]. Theor Appl Genet 125:1603-1618

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA (2009) The B73 maize genome: complexity, diversity, and dynamics. science 326:1112-1115

Semagn K, Babu R, Hearne S, Olsen M (2014) Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement. Mol Breed 33:1-14

Shang J, Li N, Li N, Xu Y, Ma S, Wang J (2016) Construction of a high-density genetic map for watermelon (*Citrullus lanatus* L.) based on large-scale SNP discovery by specific length amplified fragment sequencing (SLAF-seq). Scientia Horticulturae 203:38-46

Shimotsuma M (1963) Cytogenetic and evolutionary studies in the genus *Citrullus*. Seiken Zihô:24-34

Singh A (1978) Cytogenetics of semi-arid plants. III. a natural interspecific hybrid of *Cucurbitaceae* (*Citrullus colocynthis* Schrad* *C. vulgaris* Schrad). Cytologia 43:569-574

Takagi H, Abe A, Yoshida K, Kosugi S, Natsume S, Mitsuoka C, Uemura A, Utsushi H, Tamiru M, Takuno S (2013) QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. The Plant Journal 74:174-183

Tanaka T, Wimol S, Mizutani T (1995) Inheritance of fruit shape and seed size of watermelon. Journal of the Japanese society for horticultural science 64:543-548

The Tomato Genome Consortium (2012) The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485:635

UPOV (2013) Watermelon. (*Citrullus lanatus* (Thunb.) Matsum. et Nakai). Guideline for the conduct of tests for distinctness, univormity and stability. International union for the protection of new varieties of plants Switzerland, Geneva

USDA-NASS (2016) United states department of agriculture, National agriculture statistics service.

Varshney RK, Song C, Saxena RK, Azam S, Yu S, Sharpe AG, Cannon S, Baek J, Rosen BD, Tar'an B (2013) Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. Nat Biotechnol 31:240

Wasylikowa K, Van der Veen M (2004) An archaeobotanical contribution to the history of watermelon, *Citrullus lanatus* (Thunb.) Matsum. & Nakai (syn. *C. vulgaris* Schrad.). Vegetation History and Archaeobotany 13:213-217

Wehner T, Shetty N, Elmstrom G (2001) Breeding and seed production. Watermelons: characteristics, production, and marketing. ASHS press, alexandria, Virginia, USA, pp 27-73

Wehner TC (2008) Watermelon. In: Prohens J, Nuez F (eds) Vegetables I *Asteraceae*, *Brassicaceae*, *Chenopodiaceae*, and *Cucurbitaceae*. Springer, New York., pp 381-418

Whitaker TW, Davis GN (1962) Cucurbits: botany, cultivation, and utilization. Leonard Hill Ltd., London

Zhang H, Miao H, Wei L, Li C, Zhao R, Wang C (2013) Genetic analysis and QTL mapping of seed coat color in sesame (*Sesamum indicum* L.). PLoS One 8:e63898

Zhang J (1996) Breeding and production of watermelon for edible seed in China. Cucurbit Genet Coop Rpt 19:66-67

Zhang X, Jiang Y (1990) Edible seed watermelons (*Citrullus lanatus* (Thunb.) Matsum. & Nakai) in Northwest China. Report-Cucurbit Genetics Cooperative 13:40-42

Zhang X, Rhodes B, Wang M (1994) Genes controlling watermelon seed size. In: Lester G, Dunlap J (eds) Cucurbitaceae ′94: Evaluation and enhancement of cucurbit germplasm. ASHS Press, Alexandria, VA, pp 144-147

Table 2.1 Development of various seed coat colors in watermelon due to interaction of four SCC genes (Poole et al. 1941; Poole 1944)

Gene	Seed color
RTWD	Complete Black
RTWd	Stippled surface of black dots; tannish or reddish undercoat in segregating phenotype
RtW	Tan, ranging from dark Tuscany brown to cacao color
RTw	Clump pattern, pigment of seed clumps about the margin or in the centre of seed
Rtw	Whitish or white-tan-tip seed, shading is present at the hilum
rTW	Green seed color
rtW	Red or reddish-orange colored seed, color ranges from fez red to powder pink
rtw	White seed with light pink tip at the hilum often called as white-pink-tip seed

CHAPTER 3

REFINING THE EGUSI LOCUS IN WATERMELON USING KASP ASSAYS[1]

[1]Lucky Paudel, J. Clevenger, and C. McGregor. To be submitted to Sci. Hort.

Abstract

Egusi watermelon (*Citrullus mucosospermus*), the closest relative of sweet watermelon, is an economically important crop grown in many West African countries for its protein and lipid rich edible seeds. Egusi watermelon seeds have a thick, fleshy mucilaginous seed coat layer surrounding the seed coat which is unique to egusi watermelon. The egusi seed trait is controlled by a single recessive Mendelian locus, *eg*, located on chromosome 6 from 6.75 Mb to 11.03 Mb. This region is 4.28 Mb wide and contains 241 candidate genes. The region lacks adequate markers for fine mapping and for marker-assisted selection (MAS) of the egusi trait. In this study, we used QTL-seq to validate the position of the *eg* locus and to identify SNP markers in the region to refine the *eg* locus. A genomic region associated with the egusi trait was confirmed on chromosome 6 from 5.25 Mb to 7.85 Mb partially overlapping the previously mapped *eg* locus. SNPs identified from QTL-seq were used to design Kompetitive Allele Specific PCR (KASP^{TM}) assays for refining the *eg* locus. The *eg* locus was remapped at 86.3 cM on the genetic map, co-segregating with marker UGA6_7026576. The refined *eg* locus is 398.25 Kb long, extending from 6.95 Mb to 7.35 Mb, containing 30 candidate genes. Additional validation of the markers in the validation population and a diversity panel identified marker UGA6_7026576 as associated with the egusi phenotype and useful for MAS.

Introduction

Watermelon (*Citrullus lanatus* (Thumb.) Matsum. & Nakai) is an annual, vining, herbaceous crop which is cultivated throughout the world, predominantly for its sweet flesh. However, in several Asian and African countries different types of watermelons are cultivated for edible seeds and are collectively called edible seed watermelon (Zhang and Jiang 1990). In Asia, most edible seed watermelon that are cultivated are *C. lanatus*, the same species grown for edible flesh, whereas in Africa, especially in West African countries, *C. mucosospermus* is cultivated. *C. mucosospermus* is the closest relative of sweet watermelon (Guo et al. 2013) and is locally known as egusi watermelon. The term 'egusi' comes from the igbo and yoruba language spoken in Nigeria, meaning 'melon' (Adebayo and Yusuf 2015). Egusi watermelons have round fruits, with light green rind and white, bland flesh. The large, flat seeds present within the fruit are used for human consumption and are a rich source of oil (approx. 35%), protein (approx. 28%), carbohydrate, vitamins and minerals (Oyolu 1977a; Akobundu et al. 1982; Jarret and Levy 2012; Prothro et al. 2012). The composition of egusi seed oil is similar to sunflower, soybean and safflower oil which makes it a potential feedstock for biodiesel production (Giwa et al. 2010; Bello and Makanju 2011; Jarret and Levy 2012).

The seed coat enveloping the seed in egusi watermelon has a typical fleshy outer mucilaginous layer when fresh. The mucilaginous tissue on the seed coat is considered a remnant of nucellar tissues and it appears during the second and the third week of seed development (Gusmini et al. 2004). Upon drying seeds, the mucilaginous layer becomes desiccated and seeds look like normal seeds however, rehydration causes the layer to reappear. This seed coat layer is unique to *C. mucosospermus*. However, not all accessions classified as *C. mucosospermus* have the egusi type seed coat. Accessions like PI 189317 do not possess the egusi type seed coat but

are still classified as *C. mucosospermus* (*C. lanatus* subsp. *mucosospermus* in Guo et al. (2013)). Therefore, it is important to note that the egusi phenotype cannot be used to identify *C. mucosospermus*.

Oyolu (1977b) classified egusi seeds into five different classes based upon the morphological traits of seeds, including seed thickness, seed size and the thickness of seed edges, but does not discuss if all classes possess egusi type seed coat. The type one class of miniature seeds with thin seed coat and flat edge looks morphologically similar to *C. lanatus* 'neri-type' in Achigan-Dako et al. (2015). This confusion is also complicated by the fact that in West Africa, where egusi watermelon is predominantly cultivated, several other cucurbits that do not have egusi type seed coat are collectively referred as egusi. In the current study, only seeds with a thick fleshy mucilaginous seed coat layer will be considered egusi type seed coat and watermelons with egusi type seed coat will be called egusi watermelon for the sake of simplicity.

The egusi seed coat is genetically controlled by a single recessive locus, *eg* (Gusmini et al. 2004; Prothro et al. 2012). The *eg* locus has been mapped to the region between markers NW0248325 and NW0250248 of the Strain II (PI 279461) x Egusi (PI 560023) genetic map (Prothro et al. 2012), which corresponds to the 6.75 Mb to 11.03 Mb region on chromosome 6 of the 97103 reference genome (Guo et al. 2013). The region is 4.28 Mb long, contains 241 candidate genes and the closest marker, NW0248325, is 5.1 cM from the *eg* locus. This number of genes are still too large for candidate gene identification.

Recent advances in next generation sequencing (NGS) technology and the availability of reference genomes have facilitated the use of different tools to conduct a genome wide comparison to map loci and to rapidly detect large numbers of molecular markers throughout the genome. QTL-seq combines the principles of bulk segregant analysis (BSA) (Michelmore et al.

1991) and whole genome sequencing to map QTL and to identify markers simultaneously. QTL-seq has been used in mapping several major QTL in a wide variety of cereals (Takagi et al. 2013; Masumoto et al. 2016) and vegetables (Lu et al. 2014; Illa-Berenguer et al. 2015; Shu et al. 2018), including watermelon (Dong et al. 2018).

The objective of this study is to refine the *eg* locus using the QTL-seq approach to generate abundant SNPs and to determine the utility of the linked SNPs for the MAS in different genetic backgrounds.

Materials and Methods

Plant materials

An interspecific Strain II (*C. lanatus,* PI 279461) x Egusi (*C. mucosospermus*, PI 560023) F_2 mapping population (hereafter SII x Egusi) previously described by Sandlin et al. (2012) and used by Prothro et al. (2012) to map the *eg* locus was used for QTL-seq, fine mapping and identification of markers associated with the egusi phenotype. An additional F_2 population UGA147 (normal seed type, selection from PI 169233) x Egusi (hereafter 147 x Egusi) with 156 individuals was developed as a validation population to test marker-phenotype association (Fig. 3.1). Parental, F_1 and F_2 plants of the latter population were grown in the summer of 2017 at the Durham horticulture farm (Watkinsville, GA) and were visually phenotyped in the field. An additional diversity panel consisting of 12 egusi seed coat type and 17 normal seed coat type PIs and cultivars was compiled to verify marker-phenotype association (Table 3.1).

DNA extraction for QTL-seq

Leaf tissue of F_3 plants originating from the open-pollinated F_2 progenies of the SII x Egusi population was used to extract DNA for preparation of bulks. Based on the phenotype of

F_4 seeds obtained after selfing of F_3 plants, 7 F_3 plants, with normal type F_4 seeds, originating from 6 F_2 plants were selected to prepare the N-bulk and 7 F_3 plants, with egusi type F_4 seeds, originating from 6 F_2 plants were selected to prepare the E-bulk. DNA from all 14 samples was extracted using E. Z. N. A. Plant DNA kit (Omega Bio-Tek Inc., Norcross, GA) using the manufacturer's protocol. DNA concentrations were measured using an Infinite M200Pro plate reader (Tecan, Group Ltd., Mannerdorf, Switzerland). An equal amount of DNA was pooled from each individual to create the bulks. Both normal type and egusi type bulks were sent to the HudsonAlpha Institute for Biotechnology (Huntsville, AL) for library preparation and 151 base pair paired-end whole genome sequencing on the Illumina HiSeq X (Illumina, San Diego, CA).

DNA extraction for refining the egusi locus and marker validation

DNA from freeze dried samples of the SII x Egusi population (Prothro et al. 2012; Sandlin et al. 2012) was extracted using the modified Dellaporta et al. (1983) protocol with a few modifications (CJ Tsai, personal communications). A TissueLyser II (QIAGEN, Hilden, Germany) was used to grind 35 mg of leaf tissue. For each sample, 525 µl of extraction buffer (50 mM of Tris-HCL pH 8, 10 mM EDTA pH 8, 100 mM NaCl, 1% SDS, and 10 mM β-mercaptoethanol) and 140 µg of proteinase K were added, followed by vortexing. Samples were incubated at 60°C for 20 min, after which 140 µl of ice-cold 5M potassium acetate (pH 4.8) was added and samples were incubated on ice for another 20 min. Samples were centrifuged for 25 min at 4000 rpm and 560 µl of supernatant was transferred into a new plate and mixed with 336 µl of isopropanol. The mixture was centrifuged at 4000 rpm for 10 min to precipitate the DNA. DNA pellets were washed twice with 70% ethanol, dried and resuspended in 100 µl diH_2O.

DNA from leaf tissue of the validation population and diversity panel was extracted using the protocol described by King et al. (2014) with some modifications. Approximately 50 mg of

frozen leaf tissue was ground using a TissueLyser II (QIAGEN) and 500 µl of extraction buffer mixture containing 40 % (v/v) 5 M NaCl and 60 % (v/v) Edward's extraction buffer (200 mM Tris/HCl pH 7.5, 250 mM NaCl, 25 mM EDTA, 0.5 % SDS) was added to the ground samples. Samples were incubated at 65°C for 20 min and centrifuged for 10 min at 3600 rpm. The supernatant was mixed with isopropanol (0.6 times the volume of supernatant) to precipitate the DNA. DNA pellets obtained after centrifuging at 3600 rpm for 10 min were washed with 70 % ethanol, dried and resuspended in TE buffer.

Analysis of reads

The quality of raw reads obtained from whole genome sequencing of bulks was analyzed using FastQC (Andrews 2010). The first two and last seven bases of the forward reads and the first two and last twenty bases of the reverse reads were trimmed to make sure that the average phred score for all base positions of all reads was higher than 20. The trimmed reads from both bulks were indexed and aligned against the 97103 watermelon genome (Guo et al. 2013) using BWA and BWA MEM (Li and Durbin 2009). The SAM files were converted to BAM files, sorted and indexed using SAM tools (Li et al. 2009). SAM tools was also used to calculate the genotype likelihood. SNP calling and filtering with a minimum depth of 10 were done using BCF tools and python tool. The SNP-index for all positions of the genome was calculated by counting the number of reads harboring the SNP as compared to the reference genome sequence and dividing it by the total number of reads. The SNP-indices between two bulks was subtracted to obtain the ΔSNP-index and a custom-made python script (Appendix A) was used to generate a marker-specific threshold for $p < 0.01$ and $p < 0.001$ as described by Takagi et al. (2013). ΔSNP-index was plotted along with threshold $p < 0.001$ to identify the genomic region containing the egusi locus.

Kompetitive Allele Specific PCR (KASP™) assay design

Primers for KASP™ assays (LGC Genomics LLC, Teddington, UK) were designed (Table 3.2) for SNPs in the egusi region identified by QTL-seq and Prothro et al. (2012) using Primer3Plus (Untergasser et al. 2007). PCR amplification was done using an S1000™ thermo cycler (Bio-Rad Laboratories, Inc., Hercules, CA) in a 4 µl reaction volume, comprised of 1.96 µl 2x low rox KASP™ master mix (LGC Genomics LLC), 0.06 µl of 0.81mM primer mix, and 2 µl of 50-100 ng/µl genomic DNA. PCR conditions were set as 95°C for 15 min, followed by 10 cycles of touch down PCR with 20s of 95°C, 25s of primer annealing temperature + 9°C, with 1°C decrease each cycle, and 15s of 72°C, then 35 cycles of 10s at 95°C, 1 min at primer annealing temp, and 15s at 72°C. KASP™ florescent end-point readings and genotyping calls were done using an Infinite M200Pro plate reader (Tecan, Group Ltd.) and KlusterCaller™ (LGC Genomics LLC), respectively.

Mapping of the egusi locus and identification of candidate genes

One hundred and thirty-nine F_2 individuals from the SII x Egusi population were genotyped with 15 KASP™ assays. A genetic map was developed using JoinMap 5 software (Van Ooijen 2006). The maximum likelihood mapping function using a LOD score of 5 was used for mapping. The final genetic map with the refined position of the egusi locus was drawn in MapChart (Voorrips 2002). KASP™ assays which mapped close to the egusi locus in the genetic map were tested on the diversity panel and markers polymorphic between the parents of the validation population were tested on the latter population. For identification of candidate genes the physical position of the markers flanking the refined egusi locus were identified on the 97103 reference genome (Guo et al. 2013) and genes present in the region were examined for

their annotated functions, homologs and orthologs in the Cucurbit Genomics Database (http://cucurbitgenomics.org/).

Results

Validation of inheritance of the egusi phenotype

In this experiment, an F_2 population, 147 x Egusi, was developed as a validation population and phenotyped for seed coat type. The F_1 plants had a normal seed coat, confirming that the normal type is dominant over the egusi type and the F_2 individuals segregated at the ratio of 108 normal: 48 egusi ($\chi^2_{(0.05,1)}$, $P = 0.96$). These results confirm the conclusion made by Gusmini et al. (2004) and Prothro et al. (2012) that the trait is controlled by a single locus, with the egusi type being recessive.

QTL-Seq

We obtained a total of 79,551,582 and 155,774,971 reads for the N-bulk and the E-bulk respectively from the next generation sequencing. 99.02 % of trimmed reads from the N-bulk and 98.90% of trimmed reads from the E-bulk were successfully aligned against the 97103 watermelon genome (Guo et al. 2013). 699,060 SNPs were detected between the two bulks. When plotting the ΔSNP-index against the genome position, we found two statistically significant ($p < 0.001$) peaks. The first peak was located from 5.25 Mb to 7.85 Mb on chromosome 6 and the second was located at 17.65 Mb to 19.63 Mb on chromosome 8 (Fig. 3.2). The peak on chromosome 6 partially overlaps the previously mapped egusi locus (Prothro et al. 2012). SNP markers within both peak regions (UGA6_7026576 and UGA8_17929262) were utilized to design KASP™ assays to test association of the peaks with the phenotype in both mapping and validation populations. KASP™ assay UGA6_7026576 identified that the genomic region on chromosome 6 was highly associated to the egusi phenotype in both mapping and

validation populations. However, the peak on the chromosome 8 did not associate with the phenotype. The recombination frequency between KASP^{TM} marker UGA8_17929262 and the egusi phenotype was 62.06 and 37.52 mapping units in the mapping SII x Egusi population and validation 147 x Egusi population, respectively, confirming that genomic region identified on chromosome 8 was a false discovery (Fig. 3.3a, b).

Refining the egusi locus

We designed 21 KASP^{TM} assays in the egusi region identified by QTL-seq and Prothro et al. (2012). Fifteen assays that gave the expected genotypes in the parental and F_1 plants were utilized to genotype the population for fine mapping the egusi locus (Table 3.2). The egusi locus was remapped to the 86.3 cM position, co-segregating with marker UGA6_7026576 on chromosome 6 of the SII x Egusi genetic map (Fig. 3.4). Marker UGA6_6958189, the closest flanking marker, was 1.3 cM away from the *eg* locus, whereas the other flanking marker UGA6_7356440 was 3.1 cM away from the *eg* locus. In the 97103 genome physical map (Guo et al. 2013), the refined egusi locus is 398.25 Kb and extends from 6.95 Mb to 7.35 Mb and contains 30 candidate genes (Table 3.3).

Marker-phenotype association

In the SII x Egusi population, three KASP^{TM} markers: UGA6_7026576, UGA6_6958189 and UGA6_6903757, were able to accurately predict 100% (137/137), 100% (139/139) and 99.26% (135/136) of the F_2 individuals (Fig. 3.5a). We further tested these markers on the diversity panel and validation population. Only KASP^{TM} marker UGA6_7026576 was polymorphic in the validation, 147 x Egusi population, and it was able to accurately predict 90.34% (131/145) of phenotypes (Fig. 3.5b). In the diversity panel, UGA6_7026576 was fixed in the egusi genotypes, but segregating in the normal type watermelon (Table 3.1). Cultivars

“Allsweet”, “Blacktail Mountain”, “Charleston Gray”, “Crimson Sweet”, “Georgia Rattlesnake”, “Stars and Stripes”, “Sugarlee” and genotype ZWRM (PI 593359) had the same genotype as the egusi seed type watermelons. Among these cultivars, “Allsweet”, “Charleston Gray”, “Crimson Sweet” and “Sugarlee” share ancestry and are genetically related.

Discussion

The QTL-seq approach is a powerful tool which has been successfully utilized in QTL mapping and SNP discovery. The power of QTL-seq to reliably identify genomic regions associated with a trait depends on the nature and size of the population, the number of individuals in each bulk, the heritability of the QTL and the phenotypic variation explained by the QTL. Studies deploying QTL-seq have shown that an F_2 population with 200-300 individuals utilizing 15-20% of the individuals for each bulk is sufficient to identify both major and minor QTL (Magwene et al. 2011; Takagi et al. 2013; Tiwari et al. 2016). An F_2 population with 100 individuals has been shown to be sufficient in identifying a single qualitative locus or a QTL (Takagi et al. 2013; Illa-Berenguer et al. 2015).

In our study, we used QTL-seq to map a qualitative trait, *eg*, in watermelon. Each bulk was comprised of DNA from seven F_3 individuals originating from six open pollinated F_2 individuals. Lack of good quality leaf tissue or selfed seeds from the F_2 individuals prevented us from increasing bulk size. Nonetheless, we were able to map a significant peak partially overlapping the previously mapped egusi locus. We believe that the qualitative nature of the egusi trait made it possible to map the *eg* locus even with such non-ideal bulks. However, the use of smaller sized bulks increases the possibility of having bulks that are different, not only for alleles linked to the trait of interest but also for other alleles. This increases the probability of mapping false positives (Giovannoni et al. 1991; Michelmore et al. 1991; Tiwari et al. 2016). We

also mapped a significant peak on chromosome 8 that was not associated with the egusi trait. However, the chi-square test and an association test between the phenotype and the genotype of UGA8_17929262 in the peak region in both mapping and validation populations confirmed that the egusi phenotype is controlled by a single gene and that the peak on chromosome 8 is a false positive (Fig. 3.3a, b).

QTL-seq identified 3,112 SNPs in the egusi region, providing plenty of resources for developing KASP™ assays in the region. Fifteen KASP™ markers were added in the region of the egusi locus identified by QTL-seq and Prothro et al. (2012). The egusi locus co-segregated with marker UGA6_7026576 and was near the flanking marker UGA6_6958189. The *eg* locus was narrowed from 4.29 Mb to 398.25 Kb and the number of candidate genes between the flanking markers was decreased from 249 to 30. We examined the annotated function of each gene present in the region of interest and their homologs and orthologs in *Arabidopsis* and other cucurbits, however, none were previously associated with seed coat development. Three uncharacterized genes (Cla007525, Cla007540 and Cla007541) were also found in this region. It is possible that none of these 30 genes are candidate genes. The 97103 reference genome we used to align reads and to identify candidate genes may lack the gene conferring the egusi phenotype since cultivar 97103 belongs to *C. lanatus*. The candidate gene may be present only in the genomic region specific to the *C. mucosospermus* species because the egusi phenotype is present only in *C. mucosospermus* species. However, the unavailability of a public reference genome of *C. mucosospermus* species hinders confirming this hypothesis.

Recently, breeders have started breeding egusi watermelons with red sweet edible flesh (Orji et al. 2016). The flesh of egusi watermelons is white, hard, bland and inedible (Oyolu 1977b). Breeding egusi watermelon with red sweet flesh will make both the seed and flesh

edible, and it will increase food security in many West African countries. Harvesting seeds from the hard flesh is a very time consuming and unsanitary process since fruit is cracked and then left to rot for several days (Oyolu 1977b) before extracting seeds. Breeding watermelon with soft red flesh will save farmer's time and protect their health. However, both favorable traits, red flesh and the egusi trait are recessive to white flesh and normal seed phenotype, respectively. Therefore, breeding for watermelon with these traits without MAS will be difficult and time consuming. The KASP^{TM} marker UGA6_7026576 we developed can be used for MAS for the egusi trait in populations segregating for this marker.

References

Achigan-Dako EG, Avohou ES, Linsoussi C, Ahanchede A, Vodouhe RS, Blattner FR (2015) Phenetic characterization of *Citrullus* spp. (*Cucurbitaceae*) and differentiation of egusi-type (*C. mucosospermus*). Genet Resour Crop Evol 62:1159-1179

Adebayo A, Yusuf K (2015) Performance Evaluation of an Egusi-melon Washing Machine. Analele Universitatii' Eftimie Murgu' 22:11-22

Akobundu E, Cherry J, Simmons J (1982) Chemical, functional, and nutritional properties of egusi (*Colocynthis citrullus* L.) seed protein products. J Food Sci 47:829-835

Andrews S (2010) FastQC: a quality control tool for high throughput sequence data.

Bello E, Makanju A (2011) Performance evaluation of Egusi melon (*Citrullus colocynthis* L.) seeds oil biodiesel. Journal of Emerging Trends in Engineering and Applied Sciences 2:741-745

Dellaporta SL, Wood J, Hicks JB (1983) A plant DNA minipreparation: version II. Plant molecular biology reporter 1:19-21

Dong W, Wu D, Li G, Wu D, Wang Z (2018) Next-generation sequencing from bulked segregant analysis identifies a dwarfism gene in watermelon. Scientific reports 8:1-7

Giovannoni JJ, Wing RA, Ganal MW, Tanksley SD (1991) Isolation of molecular markers from specific chromosomal intervals using DNA pools from existing mapping populations. Nucleic Acids Res 19:6553-6568

Giwa S, Abdullah LC, Adam NM (2010) Investigating “Egusi”(*Citrullus colocynthis* L.) seed oil as potential biodiesel feedstock. Energies 3:607-618

Guo S, Zhang J, Sun H, Salse J, Lucas WJ, Zhang H, Zheng Y, Mao L, Ren Y, Wang Z (2013) The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. Nat Genet 45:51-58

Gusmini G, Wehner T, Jarret R (2004) Inheritance of egusi seed type in watermelon. J Hered 95:268-270

Illa-Berenguer E, Van Houten J, Huang Z, van der Knaap E (2015) Rapid and reliable identification of tomato fruit weight and locule number loci by QTL-seq. Theor Appl Genet 128:1329-1342

Jarret RL, Levy IJ (2012) Oil and fatty acid contents in seed of *Citrullus lanatus* Schrad. J Agric Food Chem 60:5199-5204

King Z, Serrano J, Boerma HR, Li Z (2014) Non-toxic and efficient DNA extractions for soybean leaf and seed chips for high-throughput and large-scale genotyping. Biotechnol Lett 36:1875-1879

Li H, Durbin R (2009) Fast and accurate short read alignment with burrows–wheeler transform. Bioinformatics 25:1754-1760

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078-2079

Lu H, Lin T, Klein J, Wang S, Qi J, Zhou Q, Sun J, Zhang Z, Weng Y, Huang S (2014) QTL-seq identifies an early flowering QTL located near flowering locus T in cucumber. Theor Appl Genet 127:1491-1499

Magwene PM, Willis JH, Kelly JK (2011) The statistics of bulk segregant analysis using next generation sequencing. PLoS Comp Biol 7:e1002255

Masumoto H, Takagi H, Mukainari Y, Terauchi R, Fukunaga K (2016) Genetic analysis of NEKODE1 gene involved in panicle branching of foxtail millet, Setaria italica (L.) P. Beauv., and mapping by using QTL-seq. Mol Breed 36:59

Michelmore RW, Paran I, Kesseli R (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. Proceedings of the national academy of sciences 88:9828-9832

Orji VO, Ogbonna PE, Onyia VN, Okechukwu EC (2016) Improvement of fruit taste quality of 'egusi'melon (*Colocynthis citrullus* L.) through hybridization with watermelon (*Citrullus lanatus* Thumb.). Journal of Tropical Agriculture 54:1-6

Oyolu C (1977a) Extraction rates and chemical composition of seed types in egusi (*Colocynthis citrullus* L.). Acta Hort 53:287-290

Oyolu C (1977b) A quantitative and qualitative study of seed types in egusi (*Colocynthis citrullus* L.). Tropical Sci 19:287-290

Prothro J, Sandlin K, Gill R, Bachlava E, White V, Knapp SJ, McGregor C (2012) Mapping of the egusi seed trait locus (*eg*) and quantitative trait loci associated with seed oil percentage in watermelon. J Am Soc Hort Sci 137:311-315

Sandlin K, Prothro J, Heesacker A, Khalilian N, Okashah R, Xiang W, Bachlava E, Caldwell DG, Taylor CA, Seymour DK (2012) Comparative mapping in watermelon [*Citrullus lanatus* (Thunb.) Matsum. et Nakai]. Theor Appl Genet 125:1603-1618

Shu J, Liu Y, Zhang L, Li Z, Fang Z, Yang L, Zhuang M, Zhang Y, Lv H (2018) QTL-seq for rapid identification of candidate genes for flowering time in broccoli× cabbage. Theor Appl Genet 131:917-928

Takagi H, Abe A, Yoshida K, Kosugi S, Natsume S, Mitsuoka C, Uemura A, Utsushi H, Tamiru M, Takuno S (2013) QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. The Plant Journal 74:174-183

Tiwari S, Krishnamurthy S, Kumar V, Singh B, Rao A, SV AM, Rai V, Singh AK, Singh NK (2016) Mapping QTLs for salt tolerance in rice (*Oryza sativa* L.) by bulked segregant analysis of recombinant inbred lines using 50K SNP chip. PLoS One 11:e0153610

Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JA (2007) Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Res 35:W71-W74

Van Ooijen J (2006) JoinMap 4: Software for the calculation of genetic linkage maps in experimental populations. Kyazma BV, Wageningen, the Netherlands

Voorrips R (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. J Hered 93:77-78

Zhang X, Jiang Y (1990) Edible seed watermelons (*Citrullus lanatus* (Thunb.) Matsum. & Nakai) in Northwest China. Report-Cucurbit Genetics Cooperative 13:40-42

Table 3.1 Watermelon cultivars and Plant Introductions used to develop mapping population (MP), validation population (VP) and the diversity panel (DP) for mapping and marker validation of the egusi seed type locus

Accession names	Species	Seed coat type	Origin	Uses	Genotype for KASP UGA6_7026576
Allsweet	*C. lanatus*[b]	Normal	U.S.	DP	C:C
Black Diamond	*C. lanatus*[b]	Normal	U.S.	DP	T:T
Blacktail Mountain	*C. lanatus*[b]	Normal	U.S.	DP	C:C
Charleston Gray	*C. lanatus*[b]	Normal	U.S.	DP	C:C
Cream of Saskatchewan	*C. lanatus*[b]	Normal	U.S.	DP	T:T
Crimson Sweet	*C. lanatus*[b]	Normal	U.S.	DP	C:C
Georgia Rattlesnake	*C. lanatus*[b]	Normal	U.S.	DP	C:C
Mickylee	*C. lanatus*[b]	Normal	U.S.	DP	C:T
New Hampshire Midget	*C. lanatus*[b]	Normal	U.S.	DP	T:T
Klondite Black seeded	*C. lanatus*[b]	Normal	U.S.	DP	T:T
Orangeglo	*C. lanatus*[b]	Normal	U.S.	DP	T:T
PI 296341-FR	*C. amarus*[b]	Normal	South Africa	DP	T:T
ZWRM (PI 593359)	*C. lanatus*[b]	Normal	China	DP	C:C
Stars and Stripe	*C. lanatus*[b]	Normal	U.S.	DP	C:C
Strain II (PI 279461)	*C. lanatus*[b]	Normal	Japan	MP	T:T
Sugar Baby	*C. lanatus*[b]	Normal	U.S.	DP	T:T
Sugarlee	*C. lanatus*[b]	Normal	U.S.	DP	C:C

Accession names	Species	Seed coat type	Origin	Uses	Genotype for KASP UGA6_7026576
UGA147 (PI 169233)	*C. lanatus*[b]	Normal	Turkey	VP	T:T
PI 189317[a]	*C. mucosospermus*[c]	Normal	Nigeria	DP	T:T
PI 494532[a]	*C. mucosospermus*[b]	Egusi	Nigeria	DP	C:C
PI 559994[a]	*C. mucosospermus*[b]	Egusi	Nigeria	DP	C:C
PI 559997[a]	*C. mucosospermus*[b]	Egusi	Nigeria	DP	C:C
PI 560006[a]	*C. mucosospermus*[b]	Egusi	Nigeria	DP	C:C
PI 560010[a]	*C. mucosospermus*[b]	Egusi	Nigeria	DP	C:C
PI 560011[a]	*C. mucosospermus*[b]	Egusi	Nigeria	DP	C:C
PI 560014[a]	*C. mucosospermus*[b]	Egusi	Nigeria	DP	C:C
PI 560017[a]	*C. mucosospermus*[b]	Egusi	Nigeria	DP	C:C
PI 560018[a]	*C. mucosospermus*[b]	Egusi	Nigeria	DP	C:C
PI 560020[a]	*C. mucosospermus*[b]	Egusi	Nigeria	DP	C:C
PI 560023 (Egusi)[a]	*C. mucosospermus*[b]	Egusi	Nigeria	MP, VP	C:C
PI 560024[a]	*C. mucosospermus*[b]	Egusi	Nigeria	DP	C:C
PI 595203[a]	*C. mucosospermus*[b]	Egusi	U.S.	DP	C:C

[a]Plant materials obtained from USDA-ARS Plant Genetic Resources Unit (Griffin, GA)

[b]Species classification based on USDA-ARS Plant Genetic Resources Unit (Griffin, GA)

[c]Species classification based on (Guo et al. 2013)

Table 3.2 Primer sequence of KASP^{TM} assays used for identification and narrowing down of the egusi (*eg*) locus on in watermelon. Numbers after "UGA" indicate chromosome number and physical position of the SNP on the 97103 watermelon genome (Guo et al. 2013)

KASP assay	Primer type	Primer Sequence (5'-3')	Tm (°C)
UGA6_4756466	FAM	GAAGGTGACCAAGTTCATGCTGTGTTTTGACCGCAGTGCATC	63.9
	VIC	GAAGGTCGGAGTCAACGGATTGTGTTTTGACCGCAGTGCATT	63.9
	Reverse	GCAGCTTCAAGGCATCTTGT	60.5
UGA6_5333406	FAM	GAAGGTGACCAAGTTCATGCTTGCCACTAGCCAACCTTTAAAACC	64.5
	VIC	GAAGGTCGGAGTCAACGGATTTGCCACTAGCCAACCTTTAAAACT	64.5
	Reverse	AATGCATTTGACAACTCCTTCC	60.4
UGA6_5596144	FAM	GAAGGTGACCAAGTTCATGCTTTTTACATCTTGGGATAATTACGGATA	61.2
	VIC	GAAGGTCGGAGTCAACGGATTTTTTACATCTTGGGATAATTACGGATT	61.2
	Reverse	CACTCGATTGTTAAGGGGCATT	62.4
UGA6_5706068	FAM	GAAGGTGACCAAGTTCATGCTCCCATGACCTCCCTATTTTCAC	61.9
	VIC	GAAGGTCGGAGTCAACGGATTCCCATGACCTCCCTATTTTCAG	61.9
	Reverse	TCTTCGATGTCGAATGAATGGA	62.8
UGA6_5912358	FAM	GAAGGTGACCAAGTTCATGCTAGGAAAATCAAAATATACAGGTCGAGA	61.6
	VIC	GAAGGTCGGAGTCAACGGATTAGGAAAATCAAAATATACAGGTCGAGT	61.6

KASP assay	Primer type	Primer Sequence (5'-3')	Tm (°C)
	Reverse	AACACATCTACCCCGGAGCTT	62.1
UGA6_6396525	FAM	GAAGGTGACCAAGTTCATGCTGGGATCAAAGAGCAACCAGTG	62.0
	VIC	GAAGGTCGGAGTCAACGGATTGGGATCAAAGAGCAACCAGTA	62.0
	Reverse	TTGTATCGTAACATTTGTTGTGTGC	61.0
UGA6_6501374	FAM	GAAGGTGACCAAGTTCATGCTGCCATGTTGTTTTTGCAAGTC	60.5
	VIC	GAAGGTCGGAGTCAACGGATTGCCATGTTGTTTTTGCAAGTT	60.5
	Reverse	TGATCAATGGGCCAAGTTTTG	63
UGA6_6737954	FAM	GAAGGTGACCAAGTTCATGCTTGGTCGTGGCTTACACATAAAAG	61.3
	VIC	GAAGGTCGGAGTCAACGGATTTGGTCGTGGCTTACACATAAAAC	61.3
	Reverse	AGTGGAGCTAAGGATTCCAACA	60.1
UGA6_6829416	FAM	GAAGGTGACCAAGTTCATGCTACCCATCCTTGTTCCTTCCTTT	62.2
	VIC	GAAGGTCGGAGTCAACGGATTACCCATCCTTGTTCCTTCCTTC	62.2
	Reverse	TGCTAACTGGCGTTTCAAGATAGA	62.4
UGA6_6903757	FAM	GAAGGTGACCAAGTTCATGCTTCAACTGACATTAAGTTCATACAATCG	60.6
	VIC	GAAGGTCGGAGTCAACGGATTTCAACTGACATTAAGTTCATACAATCT	60.6
	Reverse	TGTGGGTGGAAGAATCAAACC	62.1
UGA6_6958189	FAM	GAAGGTGACCAAGTTCATGCTCGCTTCAGTCGGCAGCTAT	61.6

KASP assay	Primer type	Primer Sequence (5'-3')	Tm (°C)
	VIC	GAAGGTCGGAGTCAACGGATTCGCTTCAGTCGGCAGCTAC	61.6
	Reverse	TTTGCTCACCTATACTCAGACCCATC	63.8
UGA6_7026576	FAM	GAAGGTGACCAAGTTCATGCTAATGTAAATGAGGTTCAAAGATGTGAT	60.4
	VIC	GAAGGTCGGAGTCAACGGATTAATGTAAATGAGGTTCAAAGATGTGAC	60.4
	Reverse	CCTTTTTGGGACCTCCAAATGT	63.4
UGA6_7356440	FAM	GAAGGTGACCAAGTTCATGCTTACAACAATGCCTAAATCCAACC	60.1
	VIC	GAAGGTCGGAGTCAACGGATTTACAACAATGCCTAAATCCAACA	60.1
	Reverse	TGGAAACCAAGCCCCTTATTG	63.2
UGA6_7533594	FAM	GAAGGTGACCAAGTTCATGCTGGTGCATGGAATTCAAACTGAC	61.7
	VIC	GAAGGTCGGAGTCAACGGATTGGTGCATGGAATTCAAACTGAT	61.7
	Reverse	GATGGATGTAACGACGGTCAA	60.8
UGA6_8289797	FAM	GAAGGTGACCAAGTTCATGCTTGTCATGTGATGTGTGCTAAACTAAAC	61.6
	VIC	GAAGGTCGGAGTCAACGGATTTGTCATGTGATGTGTGCTAAACTAAAG	61.6
	Reverse	CCATGATGGTTGGTTCATAATTCA	62.8
UGA8_17929262	FAM	GAAGGTGACCAAGTTCATGCTCACCAATAGTGCATGTAACCCTCA	63.2
	VIC	GAAGGTCGGAGTCAACGGATTCACCAATAGTGCATGTAACCCTCG	63.2
	Reverse	GGTTGTCGAAGGTGGTCGTC	62.9

Table 3.3 Candidate genes present in the refine egusi locus from 6.95 Mb to 7.35 Mb based on the 97103 watermelon genome (Guo et al. 2013) along with their annotated function

Gene ID	Annotated function
Cla007521	Homocysteine s-methyltransferase
Cla007522	Inositol 1 4 5-trisphosphate 5-phosphatase
Cla007523	Stress-induced-phosphoprotein 1
Cla007524	IQ calmodulin-binding motif family protein
Cla007525	Uncharacterized protein
Cla007526	Protein forked1
Cla007527	Protein forked1
Cla007528	Transmembrane BAX inhibitor motif-containing protein 4
Cla007529	DNA-directed RNA polymerase III subunit RPC9, transcription of DNA
Cla007530	Epidermal patterning factor-like protein 5, Stomata patterning
Cla007531	Protein TIFY 4A
Cla007532	Acireductone dioxygenase
Cla007533	Aldo/keto reductase
Cla007534	Aldo/keto reductase
Cla007535	Sulfate adenylyltransferase
Cla007536	Calmodulin
Cla007537	Os11g0586300 protein
Cla007538	G-protein gamma-subunit 1
Cla007539	Pyruvate kinase
Cla007540	Uncharacterized protein
Cla007541	Plant-specific domain TIGR01615 family protein, Uncharacterized protein
Cla007542	CBL-interacting protein kinase 20
Cla007543	Acylamino-acid-releasing enzyme, prolyl oligopeptidase active site region

Gene ID	Annotated function
Cla007544	Auxin responsive protein
Cla007545	Auxin responsive protein
Cla007546	Transport protein sec23
Cla007547	Magnesium transporter MRS2-2
Cla007548	Voltage-gated potassium channel beta subunit
Cla007549	Sieve element occlusion
Cla007550	Sieve element occlusion

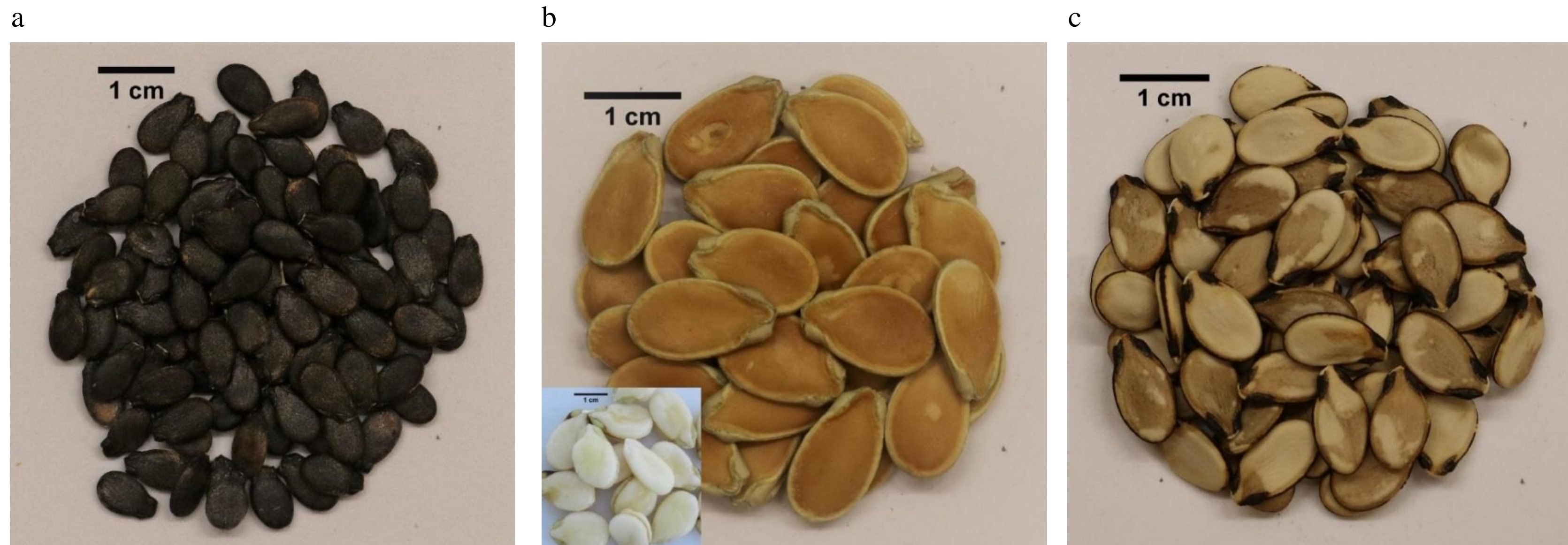


Fig. 3.1 Seeds of the parents used to develop mapping and validation populations. (a) Seeds of Strain II (PI 279461), (b) Dried and fresh (lower left corner insert) seeds of PI 560023 (Egusi), (c) Seeds of UGA147 (PI 169233)

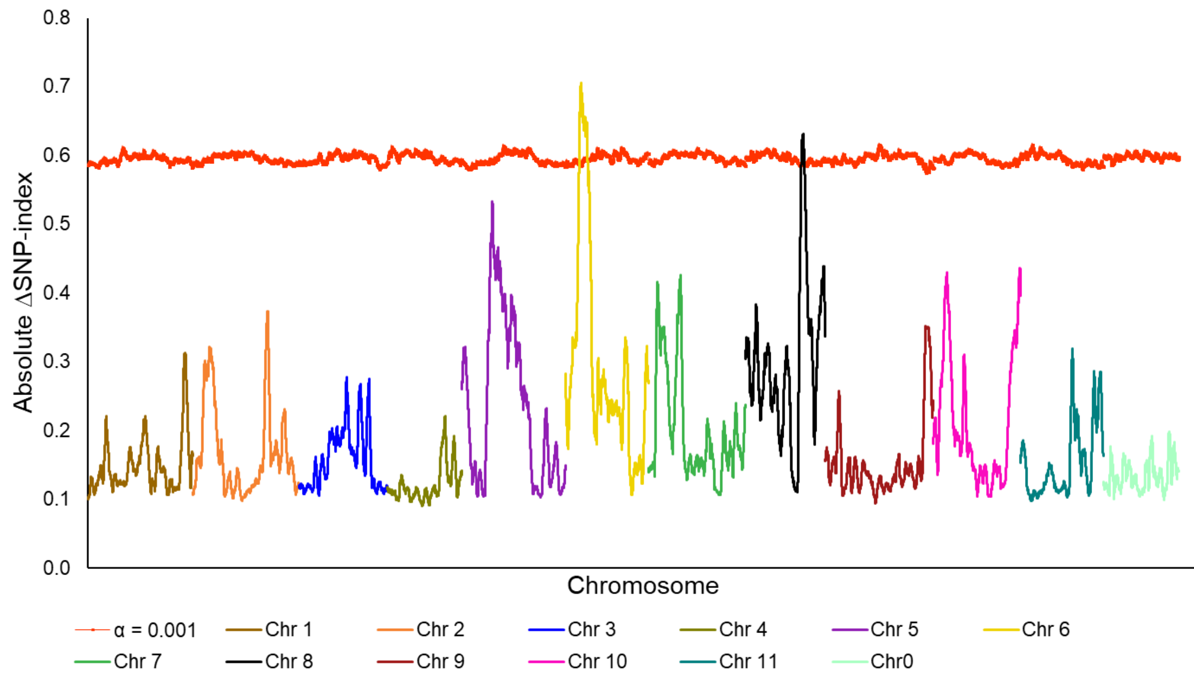
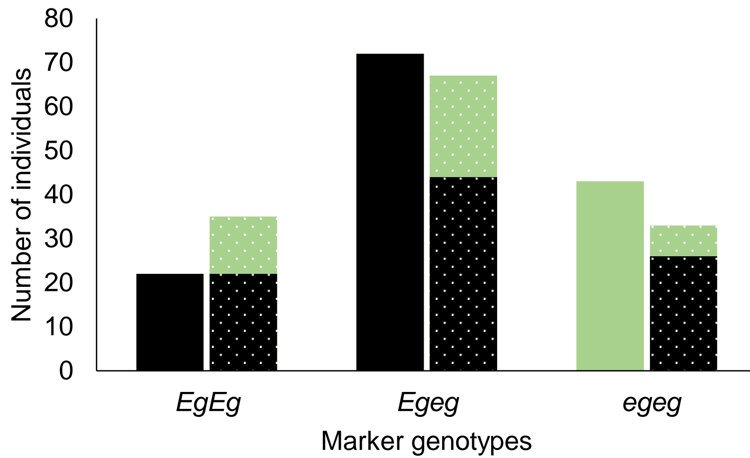


Fig. 3.2 Absolute ΔSNP-index graph plot of all chromosomes obtained from QTL-seq analysis. X-axis indicates chromosomes of watermelon placed in sequential order from 1 to 11 and 0 denoted by different colors. Y-axis indicates absolute ΔSNP-index. Absolute ΔSNP-index for each chromosome were calculated using 1Mb sliding window with a 10Kb increment increase and plotted with statistical threshold ($P < 0.001$) (horizontal red line) with criteria that no QTL below threshold. Two significant peaks on chromosome 6 (yellow) and chromosome 8 (black) were identified

a



b

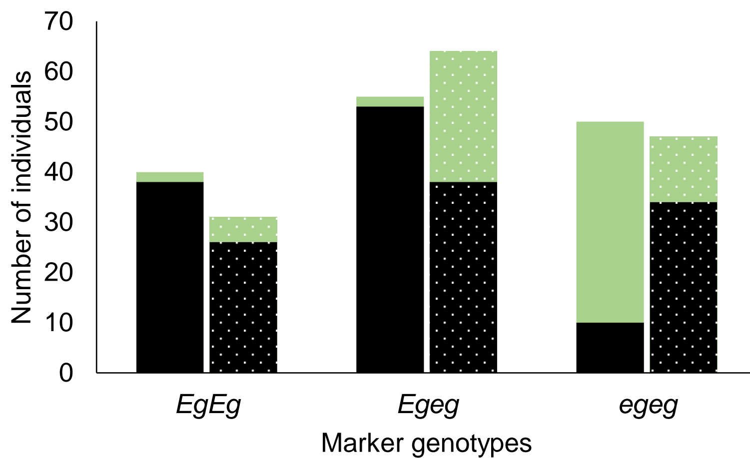


Fig 3.3 Bar graphs showing the marker phenotype association between seed coat phenotype and KASP^{TM} markers UGA6_7026576 (solid) and UGA8_17929262 (dotted) in the (a) Strain II (PI 279461) x Egusi (PI 560023) and (b) UGA147 (PI 169233) x Egusi population. Black and green sections denote the number of individuals with normal and egusi type seed coat, respectively. The x-axis genotypes represent the marker genotypes coded as normal (*Eg_*) or egusi (*egeg*) type allele

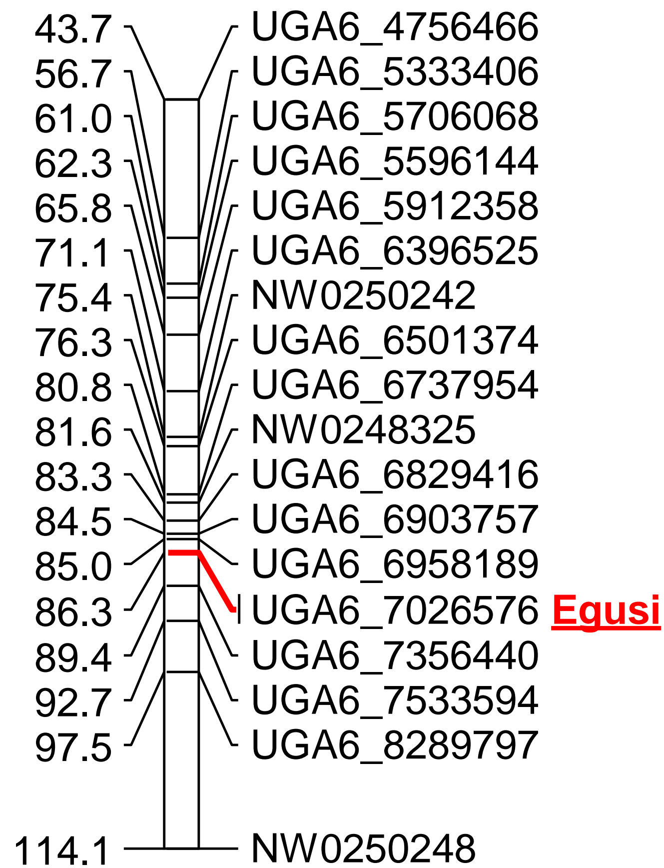


Fig. 3.4 Partial genetic map of chromosome 6 of the Strain II (PI 279461) x Egusi (PI 560023) F_2 watermelon population showing the refined position of the egusi locus. Map positions are in cM

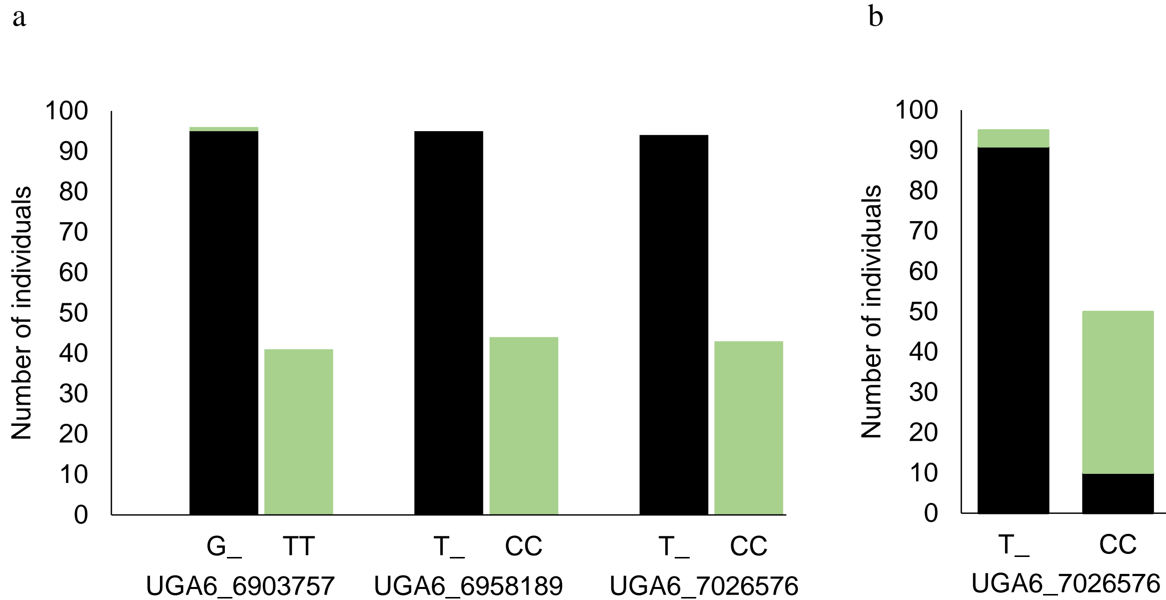


Fig. 3.5 (a) Genotypic and phenotypic data for KASP™ markers UGA6_6903757 (n = 136), UGA6_6958189 (n = 139) and UGA6_7026576 (n = 137), in the mapping population, Strain II (PI 279461) x Egusi (PI 560023), and (b) KASP™ marker UGA6_7026576 in the validation population UGA147 (PI 169233) x Egusi, (n = 145). X-axis indicates genotypes of respective KASP™ markers and y-axis indicates number of F_2 individuals. Black and green sections denote the number of individuals with normal and egusi type seed coat, respectively

CHAPTER 4

CHROMOSOMAL LOCATIONS AND INTERACTIONS OF FOUR LOCI ASSOCIATED WITH SEED COAT COLOR IN WATERMELON[1]

[1]Lucky Paudel, J. Clevenger, and C. McGregor. To be submitted to Theor. Appl. Genet

Abstract

Different species of watermelons (*Citrullus* spp.) are cultivated throughout Asia and Africa for their lipid and protein rich edible seeds. These watermelon seeds have diverse seed coat colors including black, stipple, red, green and white. Recent changes in consumer preference for specific seed coat colors have made this trait economically important for watermelon breeders. In this study, we developed three F_2 populations segregating for seed coat color: Sugar Baby (stippled seed, *RRTTWW*) x PI 482379 (green seed, *rrTTWW*), Charleston Gray (stippled seed, *RRTTWW*) x PI 189225 (red seed, *rrttWW*) and Charleston Gray (stipple seeded, *RRTTWWdd*) x UGA147 (clump seed, *RRTTwwDD*, a selection from PI 169233) to map four genes *R*, *T*, *W* and *D* associated with seed coat color development in watermelon. The observed segregation ratios of the stipple x green population indicated that the dominant stipple seed coat color and the recessive green seed coat color segregate for a single gene, *R.* In the stipple x red population, the dominant stipple seed coat color and the recessive red seed coat color segregate for *R* and *T*, where the *R* gene is dominantly epistatic to the *T* gene. However, based on the phenotypic segregation and mapping information, the presumed *T* locus in the stipple x red population is either a novel locus or a different allele of the *T* locus previously described. We propose to name this locus T^1. In the stipple x clump population, the clump seed coat color and the stipple seed coat color segregate for *W* and *D*, where *D* is recessively epistatic to *W*. We identified the genomic regions associated with the *R* and T^1 on chromosomes 3 and 5, respectively using QTL-seq while genotyping by sequencing was used to identify the *W* and *D* loci on chromosome 6 and 8, respectively.

Keywords

Citrullus lanatus, *Citrullus amarus*, edible seed watermelon, seed coat color, QTL-seq

Introduction

Watermelon (*Citrullus lanatus*) is an annual, warm season vegetable crop which is grown throughout the tropical and sub-tropical regions of the world, predominantly for consumption of the sweet flesh. However, in many Asian and African countries, watermelons are instead cultivated for edible seeds. In China and India, most of the edible seed watermelon are from *C. lanatus* (Zhang 1996; Mahla et al. 2014), whereas in West Africa , egusi watermelon, from the indigenous *C. mucosospermus* are extensively cultivated for edible seed (Oyolu 1977; Gusmini et al. 2004). *C. colocynthis* is also cultivated as an edible seed watermelon in the Arabian peninsula and in India (Schafferman et al. 1998; Mahla et al. 2014). The land under edible seed watermelon production is increasing and the market has expanded from China, India and Africa to Europe and the Americas (Zhang 1996; National Research Council 2006; Mahla et al. 2014).

Seed coat color is an economically important trait because consumers prefer watermelon seeds with a specific color of seed. In China, seeds with red seed coat color or seed with a black margin and white center are preferred (Zhang 1996). Watermelon have a wide variety of seed coat colors ranging from black, stipple (dotted), tan (brown), green, red, clump (white seeds with black rim on margin and/or two black spots on the hilum end) to white which is genetically controlled by a number of genes involving complex genetic interactions (Poole et al. 1941). Naming of seed phenotypes has often been inconsistent among authors (Weetman 1937; Poole et al. 1941; Sachan and Nath 1976; Nath and Khandelwal 1978) and for the sake of simplicity, we will use the phenotypic descriptors used by Poole et al. (1941).

The earliest attempt to study the inheritance of seed coat color was by Kanda (1931). He crossed black seeded watermelon with stipple seeded watermelon and demonstrated that the black seed coat color is monogenically dominant to the stipple seed coat color. Later, McKay

(1936) developed two crosses: tan x red and green x red and showed that in each cross the former phenotype was monogenically dominant to the later phenotype. Weetman (1937) developed populations from three different crosses and showed that (i) the stipple seed coat is monogenically dominant to the clump seed coat, and (ii) different combinations of two genes produce clump and tan seed coat color. Poole et al. (1941) developed 40 different segregating populations and from the results they proposed a four-gene model controlling seed coat color in watermelon. According to this model, different combinations of three genes: *R*, *T* and *W* with a modifier gene *D* (which only acts when the other three genes are in the dominant state) produce different seed coat colors, like black (*RTWD*), stipple (*RTWd*), green (*rTW*), tan (*RtW*), clump (*RTw*), red (*rtW*), white tan-tip (*Rtw*) and white pink-tiped (*rtw*). This 1941 study was the last in-depth, large scale study on the genetics of watermelon seed coat color.

Next generation sequencing (NGS) technologies has made high-throughput sequencing, less error-prone and very cost effective. As a result, NGS has become popular for the discovery of molecular markers throughout the genome (Varshney et al. 2009). Genotyping by sequencing (GBS) is a simple but highly scalable NGS based genotyping model that can be used to genotype large populations and to identify thousands of genomic markers throughout the genome simultaneously (Elshire et al. 2011). GBS has been widely used to develop linkage maps and map QTL in several crops including watermelon (Lambel et al. 2014; Meru and McGregor 2016a; Meru and McGregor 2016b; Branham et al. 2017), zucchini (Montero-Pau et al. 2017), cucumber (Wang et al. 2018), pumpkin (Zhang et al. 2015), barley (Liu et al. 2014), pea (Boutet et al. 2016), and rice (Bhatia et al. 2018). Another relatively recent NGS based technology is QTL-seq (Takagi et al. 2013). It combines bulk segregant analysis (Michelmore et al. 1991) with whole genome sequencing to identify quantitative trait loci (QTL) and to discover genetic

markers necessary for marker assisted selection (MAS). One of the advantages of QTL-seq is that it does not require genotyping of entire populations (Takagi et al. 2013). The first use of the QTL-seq approach in watermelon was utilized to map a dwarfism locus on chromosome 7 (Dong et al. 2018). QTL-seq has been employed in several other crops like rice (Takagi et al. 2013), tomato (Illa-Berenguer et al. 2015), cucumber (Lu et al. 2014), chickpea (Das et al. 2015; Singh et al. 2016), and peanut (Clevenger et al. 2018).

Identification of the genomic regions associated with seed coat color is a crucial step in identifying candidate genes and in developing molecular markers for MAS. In this study, we used two interspecific and one intraspecific F_2 populations segregating for different seed coat colors to (i) determine the location of the *R*, *T*, *W* and *D* loci responsible for seed coat color development in watermelon and (ii) determine the interaction among these loci.

Materials and Methods

Plant materials and phenotyping

Three segregating F_2 populations were used to identify the loci responsible for seed coat color development in watermelon. The stipple x green F_2 population (n = 128) was produced by crossing stipple seeded Sugar Baby (*C. lanatus*) with green seeded PI 482379 (*C. amarus*) (Fig. 4.1a). A stipple x red F_2 population (n = 96) was developed by crossing stipple seeded Charleston Gray (*C. lanatus*) with red seeded PI 189225 (*C. amarus*) (Fig. 4.1b). The stipple x clump population (n = 178) used in this study was developed by Meru and McGregor (2016b) to map *Fusarium oxysporum* f. sp. *niveum* race 2 in sweet watermelon. This F_2 population was produced by crossing stipple seeded Charleston Gray (*C. lanatus*) with clump seeded UGA147 (*C. lanatus*), a selection from PI 169233 (Fig. 4.1c).

The stipple x green and stipple x red parental cultivar/accession along with F_1 plants, and both F_2 populations were phenotyped in the field at the Durham horticulture farm (Watkinsville, GA) in the summer of 2017. Mature fruits from each plant were cut open and seeds were visually phenotyped. Dry seeds from the parental, F_1 and F_2 plants from the stipple x clump population, grown in the greenhouse in 2012 and 2013, were visually phenotyped under daylight conditions.

Bulk construction and DNA isolation for QTL-seq

For QTL-seq of the stipple x green population, a stipple bulk (S-bulk) and a green bulk (G-bulk) were constructed by pooling equal amounts of DNA from 18 individuals of each phenotype. Similarly, for the QTL-seq of the stipple x red population, a khaki bulk (K-bulk) was developed by pooling equal amounts of DNA from 20 individuals with khaki seed coat color and a red bulk (R-bulk) was developed by pooling equal amounts of DNA from 7 individuals with red seed coat color. Genomic DNA was extracted using the E. Z. N. A. Plant DNA kit (Omega Bio-Tek Inc., Norcross, GA) following the manufacturer's protocol. DNA concentrations were measured using an Infinite M200Pro plate reader (Tecan, Group Ltd., Switzerland) and bulks were comprised from equal amounts of DNA from the selected individuals and sent to the HudsonAlpha Institute for Biotechnology (Huntsville, AL) for library preparation and 151 bp paired-end whole genome sequencing on the Illumina HiSeq X (Illumina, San Diego, CA).

Analysis of NGS data

The quality of the reads obtained from NGS was analyzed using FastQC (Andrews 2010). To ensure that the average phred score for all of the base positions in all the reads was higher than 28, bases with a low phred score were trimmed on both ends for all the bulks as follows: the first seven bases of forward and reverse reads of all bulks, the last two bases of all forward reads, the last 41 bases of reverse reads of the S-bulk, the last 27 bases of reverse reads of the G-bulk,

the last 31 bases of reverse reads of the K-bulk and the last 27 bases of reverse reads of the R-bulk. The downstream analysis for all the bulks was the same. The trimmed reads were aligned against the 97103 watermelon genome (Guo et al. 2013) using default BWA and BWA MEM options (Li and Durbin 2009). SAM tools (Li et al. 2009) was used to sort, index and calculate the genotype likelihood. BCF tools and a python script was used for SNP calling and filtering. The SNP-index, which is the proportion of reads harboring SNPs divided by the total number of reads for a genomic position, was calculated for every base in the genome for all bulks. The SNP-index in the G-bulk was subtracted from the SNP-index in the S-bulk to obtain a ΔSNP-index for the stipple x green population, and the SNP-index of the R-bulk was subtracted from the SNP-index of the K-bulk to obtain a ΔSNP-index for the stipple x red population. A custom python script (Appendix A) was used to conduct sliding window analysis by averaging the ΔSNP-index within a 1 Mb window region with a 10 kb stepwise incremental. A permutation test was conducted to develop a null model assuming no QTL as explained by Takagi et al. (2013) and Clevenger et al. (2018). Thresholds for $P < 0.05$ and $P < 0.01$ were calculated for both population taking population size, number of individuals in each bulk and read depth into account.

Genotyping of stipple x clump and construction of a genetic linkage map

GBS of the stipple x clump population is described in Meru and McGregor (2016b). The original 501 SNPs for the population was filtered using Joinmap 5.0 (Van Ooijen 2006) for missing data (with up to 20% missing data) and segregation distortion ($P < 0.0001$). The remaining 230 SNPs were ordered using the regression mapping algorithm and grouped into linkage groups at LOD 5. A linkage map was generated using the Kosambi mapping function by converting recombination frequencies into map distances in centimorgan (cM).

DNA extraction of F_2 populations and Kompetitive Allele Specific PCR (KASP^{TM}) genotyping

Approximately 50 mg of leaf material from each individual of the stipple x green and the stipple x red parental cultivar/line, F_1, and F_2 populations were frozen in liquid nitrogen and ground using a TissueLyser II (QIAGEN, Hilden, Germany). DNA was extracted from leaf material using the King et al. (2014) extraction method with the following modifications. 500 µl of extraction buffer mix [40% (v/v) 5 M NaCl and 60% (v/v) extraction buffer (200 mM Tris/HCl pH 7.5, 250 mM NaCl, 25 mM EDTA, 0.5% SDS)] was added on ground leaf material. Samples were vortexed, incubated for 30 min at 60°C and centrifuged for 10 min at 3600 rpm. An equal amount of supernatant and isopropanol were mixed and centrifuged to obtain DNA pellets. The DNA pellets were washed with 70% alcohol, dried and resuspended in 200 µl TE buffer.

To validate the association of significant peaks with seed coat color, SNPs identified through QTL-seq were converted into KASP^{TM} assays (Table 4.1). Primers were designed using Primer3Plus (Untergasser et al. 2007) and PCR amplification was done using a S1000^{TM} Thermo Cycler (Bio-Rad Laboratories, Inc., Hercules, CA). The 4 µl PCR reaction included 2 µl of 50-100 ng/µl genomic DNA, 1.96 µl 2x low rox KASP^{TM} master mix (LGC Genomics LLC, Teddington, UK) and 0.06 µl primer mix for a final concentration of 0.81 µM. The PCR conditions for the KASP^{TM} assays were set as follows: 95°C for 15 min, followed by 10 cycles of touch down PCR (95°C for 20s, primer annealing temperature + 9°C for 25s with 1°C decrease each cycle and 72°C for 15s), then followed by 35 additional cycles (95°C for 10s, primer annealing temp for 1 min, and 72°C for 15s). PCR florescent end-point readings was done using an Infinite M200Pro plate reader (Tecan, Group Ltd., Mannerdorf, Switzerland) and genotyping calls were carried out using KlusterCaller^{TM} (LGC Genomics LLC).

Results and Discussion

Phenotypic segregation in the stipple x green population and mapping of the *R* locus

In the stipple x green population, F_1 plants had seeds with stipple seed coat color indicating that stipple is dominant over the green. The F_2 progenies had either the stipple or the green seed coat color (Fig. 4.1a). The number and size of black dots on stipple coated seeds varied among individuals ranging from a few fine dots to many large dots. The individuals with green seed color also had different shades of green ranging from light brown to dark green to brownish green. The shade of green color depended on the maturity of the fruit. Within the same individual, more mature fruit had brownish green seed color, while less mature fruit had light brown to light green seeds. The F_2 progenies segregated at a ratio of 88 stipple to 40 green seeded individuals. A chi-square goodness of fit test shows that the observed segregation ratio fits a 3:1 ratio ($\chi2_{(0.05,1)}$, $P = 0.10$). This result confirms the conclusion made by Poole (1944) that the stipple (*R_*) seed coat color is monogenically dominant to the green seed coat color (*rr*).

QTL-seq was used to determine the location of the *R* locus in the watermelon genome. A total of 168,613,320 and 172,686,615 reads for the S-bulk and the G-bulk were generated from NGS, respectively. 165,489,026 (98.15%) reads from the S-bulk and 170,855,441(98.94%) reads from the G-bulk were aligned to the 97103 watermelon genome (Guo et al. 2013) with an average read depth > 96x. A total of 4,953,800 SNPs were identified between the two bulks and a significant ΔSNP-index peak ($P < 0.01$) was identified from 4.48 Mb to 12.98 Mb on Chromosome 3 of the *C*. *lanatus* genome (Fig. 4.2a, b).

A KASP^{TM} assay, UGA3_5820134 (Table 4.1), was designed for a SNP located near the highest peak (5820134 bp on chromosome 3 of 97103 reference genome (Guo et al. 2013)) to test the association between the significant peak and the phenotype. The assay was able to

accurately predict the phenotype of 85.7% (n = 126) of individuals (Fig. 4.2c), confirming the association of this region with the *R* locus.

Phenotypic segregation in the stipple x red population and mapping of the T^1 locus

The F_1 plants in the stipple x red population have seeds with stippled seed coat denoting that the stippled seed coat is dominant over the red seed coat color (Fig. 4.1b). The segregating F_2 progenies had either stipple, red or khaki like seed coat color. According to the four-gene model, F_2 individuals in a stipple x red population is expected to have either stipple, tan, green or red seed coat color at a 9 stipple (*R_T_*): 3 tan (*R_tt*): 3 green (*rrT_*): 1 red (*rrtt*) ratio. In the current study, no individuals with green or tan, described as "dark Tuscany brown through cacao" by Poole et al. (1941), seed color were observed in the stipple x red population. The khaki like seed coat color did not fit any phenotypic descriptions by Poole et al. (1941) and Poole (1944). Therefore, a new phenotype descriptor, khaki, was created. The F_2 progenies segregated at the ratio of 67 stipple: 22 khaki: 7 red which statistically corresponds to 12:3:1 ratio ($\chi^2_{(0.05,2)} = 1.40$, $P = 0.49$) and indicates dominant epistasis.

One of the possible reasons for inconsistency between the current study and previous observations could be because the stipple x red population might have different alleles or genes conferring seed coat color compared to the populations described in Poole et al. (1941). This hypothesis is supported by several independent studies that have reported that the inheritance of watermelon seed coat color contradicts the four-gene model. Nath and Khandelwal (1978) made a cross between black seeded and red seeded genotypes of *C. lanatus* and observed that F_2 progenies segregated for a single gene instead of two genes as predicted by the four-gene model. In a similar study conducted by Sharma and Choudhury (1982), they found that black seed coat

color and white seed color segregate only for one gene instead of two or three genes (depending on whether white seed is white-tan tip or white-pink tip) as predicted by the four-gene model.

Based on the 12:3:1 ratio associated with dominant epistasis, we inferred that the khaki type seed coat color and the red seed coat color were segregating for a single gene. Therefore, we pooled DNA from individuals with the khaki seed coat color and the red seed coat color to develop the K-bulk and the R-bulk for QTL-seq. From the whole genome sequencing, we obtained a total of 124,764,246 and 154,206,455 reads for the K-bulk and R-bulk, respectively. 123,169,517 (98.72%) reads from the K-bulk and 152,346,070 (98.79%) reads from the R-bulk were aligned against to the reference 97103 watermelon genome with an average read depth >83x. A total of 3,401,764 SNPs was identified between the two bulks and a significant ΔSNP-index peak ($P < 0.01$) was mapped from 1.89 Mb to 6.46 Mb on chromosome 5 of the *C. lanatus* genome (Fig. 4.3a, b).

A SNP present within a significant peak and positioned at 4,591,722 bp on chromosome 5 of the 97103 reference genome (Guo et al. 2013) was utilized to develop the UGA5_4591722 KASP™ assay (Table 4.1) to test the association of the peak and the phenotype. The marker was able to predict khaki (genotype: A:A or G:A) or red seed color (genotype: G:G) with 96.55% accuracy (n = 29) validating that the peak is related to the seed coat color (Fig. 4.3c). Since the region mapped in this population was different from the region mapped in the stipple x green population, we concluded that this region is not the *R* locus and based on the nature of inheritance, it can be inferred that this region is either a novel locus or a different allele of the *T* locus described by Poole et al. (1941). Therefore, we propose to name this locus T^1, in conformance with gene nomenclature rules for Cucurbitaceae (Cucurbit Gene List Committee 1982).

We also tested the KASP™ assay UGA3_5820134 associated with the *R* locus and found that the stipple x red population was segregating for the *R* locus, as predicted by the four-gene model. Approximately 97.01% of individuals with stipple seed color had the genotype G:G or T:G and 72.41% of individuals with khaki or red seed color had the genotype T:T (data not shown). In addition, the genotypic data from the combination of KASP™ markers UGA3_5820134 and UGA5_4591722 were analyzed to understand the interaction between the two loci. Out of 71 F_2 individuals that had the genotype G:G or G:T for marker UGA3_5820134, 65 individuals (91.54%) had stipple seed coat color, independent of the UGA5_4591722 genotype (Fig. 4.3d). Among 16 F_2 individuals that had the genotype T:T for marker UGA3_5820134 and A_ for marker UGA5_4591722, 15 individuals (93.75%) had khaki seed coat color. Similarly out of 9 F_2 individuals that had the genotype T:T for marker UGA3_5820134 and G:G for marker UGA5_4591722, 7 individuals (77.77%) had red seed coat color. This confirms our hypothesis that the *R* locus is dominantly epistatic to T^1 locus.

<u>Phenotypic segregation in the stipple x clump population and mapping of the W and D loci</u>

Based on the four-gene model, the stipple genotype (*RTWd*) and the clump genotype (*RTwD* or *RTwd*) segregate either for the *W* gene or for both *W* and *D* genes. In the stipple x clump population, the F_1 had black seed coat color (*W_D_*) meaning that the genotype of the clump parent, UGA147, is expected to be *RRTTwwDD* and the population is segregating for two genes (Fig. 4.1c). The F_2 progenies segregated as black (*W_D_*, n = 94), stipple (*W_dd*, n = 35), and clump (*wwD_* or *wwdd*, n = 49) which statistically fits a 9:3:4 ratio ($\chi^2_{(0.05,2)} = 0.91$, P = 0.63), confirming the conclusion by Poole et al. (1941) that the *D* gene is recessively epistatic to *W*.

For mapping of the *W* and *D* genes, the seed phenotypes were translated into the *"abhcd*" genotype code format as described in the Joinmap® 4 manual (Van Ooijen 2006). For the *W* locus, all individuals with non-clump seed coat color (black and stipple, genotype: *W_*) were coded *d* (not clump parent genotype) and individuals with clump seed coat color (genotype: *ww*) were coded *b* (clump parent genotype). Similarly, for the *D* locus, all the individuals with black seed color (genotype: *D_)* were coded *c* (not stipple parent genotype), and individuals with stipple seed coat color (genotype: *dd*) were coded *a* (stipple parent genotype). Individuals with the clump phenotype (genotype: *D_* or *dd*) were coded as missing data since the genotype of clump seeded individuals could not be predicted at the F_2 generation. The two phenotypic markers along with 230 SNP markers were used to construct a genetic map. Thirteen linkage groups with a total length of 1226 cM and an average marker distance of 5.3 cM were developed for the 11 watermelon chromosomes (Fig. 4.4). The *W* locus was mapped at 14.5 cM on chromosome 6 between markers UGA6_5820584 and UGA6_7076766 (Fig. 4.4). The closest marker UGA6_7076766 is 9.8 cM away from the *W* locus. The genomic region associated with *W* locus partially overlapped with the major seed length QTL, $Qsl6^M$ (Prothro et al. 2012b; Meru and McGregor 2013). This is in accordance with the conclusion by Poole et al. (1941) that the *W* locus is linked with the *L* locus associated with seed length. The *D* locus was mapped between markers UGA8_21660128 and UGA8_22729513 at position 77.7 cM on chromosome 8 on the stipple x clump genetic map (Fig. 4.4). The closest marker, UGA8_22729513 is 3.4 cM away from the *D* locus.

We analyzed the genotypic data of SNP markers UGA6_7076766 and UGA8_22729513 to examine if the combination of *W* and *D* loci could predict seed coat color. Whenever F_2 individuals were homozygous dominant or heterozygous for both *W* (A:A or A:G genotype for

marker UGA6_7076766) and *D* locus (A:A or A:T genotype for marker UGA8_22729513), 90.69%) of individuals had black seed color (Fig. 4.5). Similarly, when F_2 individuals were homozygous dominant or heterozygous for *W* locus but recessive for *D* locus (T:T genotype for marker UGA8_22729513), 71.42% individuals had stipple seed coat color. However, when F_2 individuals were homozygous recessive for the *W* locus, (G:G genotype for marker UGA6_7076766), 82.60% of individuals had clump seed color irrespective of *D* locus. In total, the phenotype prediction accuracy of markers UGA6_7076766 and UGA8_22729513 when used as a proxy for *W* and *D* loci was 82.60% (n = 174). The percentage of inaccurate phenotype prediction (17.40%) is similar to the value of total recombination between SNP markers and respective loci (13.2 cM). Our result confirms that the genomic regions identified on chromosome 6 and 8 are associated with the *W* and *D* locus respectively, and that the *D* locus is recessively epistatic to the *W* locus.

<u>Re-evaluation of the four-gene model</u>

The four-gene model was developed by Poole et al. (1941) based on the inheritance of seed coat color in their populations and the populations developed by McKay (1936). Since the development of the four-gene model, only a few studies have been conducted to study inheritance of seed coat color. Nath and Dutta (1973) crossed a tan seeded individual with a red seeded individual and found that tan is monogenically dominant to red as predicted by the model. Similar studies by Sachan and Nath (1976) and Nath and Khandelwal (1978), crossing black seeded and brown (tan) seeded individuals, also fit the four-gene model. However, the model was contradicted in a cross made by Nath and Khandelwal (1978) where black seed coat color was monogenically dominant to red seed coat color. Similarly, the findings of Sharma and Choudhury (1982) that black seed coat color is monogenically dominant to white seed coat color

also challenged the validity of the four-gene model. In the current study the inheritance of the *R*, *W* and *D* loci fit the four-gene model, however, inheritance of the T^1 locus did not. The T^1 locus mapped in the stipple x red population could be a different allele of the *T* locus or even a novel gene. Further testing of allelism is hampered by the lack of information about the identity of the red parental genotype used in the study by McKay (1936), which was used by Poole et al. (1941) in developing the four-gene model. The genotype is simply described as “citron”, which is equivalent to *C. amarus*, but no further information is provided. The parental genotypes, “Peerless” and “Baby Delight”, which produced red phenotype when crossed in Poole et al. (1941), are not currently available to replicate the cross. Nevertheless, findings from the current study and several others demonstrate that the four-gene model is incomplete and requires amendment.

One of the major obstacles in improving the understanding of genetic control of watermelon seed coat color is to the lack of standard phenotypic descriptors. In each study, the authors develop their own phenotyping methodology which makes it difficult to compare results among experiments and to derive a consensus conclusion (Weetman 1937; Poole et al. 1941; Sachan and Nath 1976; Nath and Khandelwal 1978). This has been exacerbated by the fact that some of the lines/cultivars used in previous studies are no longer available to replicate the crosses. Since the phenotypic description developed by Poole et al. (1941) is the most detailed among any studies previously conducted, we propose that future studies related to seed coat color in watermelon should use the phenotypic description developed by Poole et al. (1941). Any new phenotypic class like “khaki” should only be used if it is distinct from the previously developed class or has different inheritance pattern.

Conclusion

Most of the research related to the genetics of watermelon seed color was carried out before the advent of molecular tools. This is the first study to map seed color gene loci in watermelon and to report SNP markers associated with these loci. In this study, we mapped the *R*, T^1, *W* and *D* loci on chromosomes 3, 5, 6 and 8, respectively. Further research is necessary to verify if T^1 is a different allele or different locus than the previously described *T* locus. Moreover, identification of the T^1 locus indicates that there are additional genes/alleles that confer seed coat color in watermelon. This opens future research opportunities to fine map genomic regions to pinpoint the genes conferring seed coat color and to identify DNA markers for MAS of seed coat color in watermelon.

References

Andrews S (2010) FastQC: a quality control tool for high throughput sequence data.

Bhatia D, Wing RA, Yu Y, Chougule K, Kudrna D, Lee S, Rang A, Singh K (2018) Genotyping by sequencing of rice interspecific backcross inbred lines identifies QTLs for grain weight and grain length. Euphytica 214:41

Boutet G, Carvalho SA, Falque M, Peterlongo P, Lhuillier E, Bouchez O, Lavaud C, Pilet-Nayel M-L, Rivière N, Baranger A (2016) SNP discovery and genetic mapping using genotyping by sequencing of whole genome genomic DNA from a pea RIL population. BMC Genomics 17:121

Branham S, Vexler L, Meir A, Tzuri G, Frieman Z, Levi A, Wechter WP, Tadmor Y, Gur A (2017) Genetic mapping of a major codominant QTL associated with β-carotene accumulation in watermelon. Mol Breed 37:146

Clevenger JP, Chu Y, Chavarro C, Botton S, Culbreath A, Isleib TG, Holbrook CC, Ozias-Akins P (2018) Mapping late leaf spot resistance in peanut (*Arachis hypogaea*) using QTL-seq reveals markers for marker assisted selection. Frontiers in Plant Science 9:83

Cucurbit Gene List Committee (1982) Update of cucurbit gene list and nomenclature rules. Cucurbit Genet Coop Rpt 5:62-66

Das S, Upadhyaya HD, Bajaj D, Kujur A, Badoni S, Kumar V, Tripathi S, Gowda CL, Sharma S, Singh S (2015) Deploying QTL-seq for rapid delineation of a potential candidate gene underlying major trait-associated QTL in chickpea. DNA Res:193-203

Dong W, Wu D, Li G, Wu D, Wang Z (2018) Next-generation sequencing from bulked segregant analysis identifies a dwarfism gene in watermelon. Scientific reports 8:1-7

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PloS one 6:e19379

Guo S, Zhang J, Sun H, Salse J, Lucas WJ, Zhang H, Zheng Y, Mao L, Ren Y, Wang Z (2013) The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. Nat Genet 45:51-58

Gusmini G, Wehner T, Jarret R (2004) Inheritance of egusi seed type in watermelon. J Hered 95:268-270

Illa-Berenguer E, Van Houten J, Huang Z, van der Knaap E (2015) Rapid and reliable identification of tomato fruit weight and locule number loci by QTL-seq. Theor Appl Genet 128:1329-1342

Kanda T (1931) The inheritance of seed-coat colouring in the watermelon. Jap J Genet 7:30-48

King Z, Serrano J, Boerma HR, Li Z (2014) Non-toxic and efficient DNA extractions for soybean leaf and seed chips for high-throughput and large-scale genotyping. Biotechnol Lett 36:1875-1879

Lambel S, Lanini B, Vivoda E, Fauve J, Patrick Wechter W, Harris-Shultz KR, Massey L, Levi A (2014) A major QTL associated with *Fusarium oxysporum* race 1 resistance identified in genetic populations derived from closely related watermelon lines using selective genotyping and genotyping-by-sequencing for SNP discovery. Theor Appl Genet 127:2105-2115

Li H, Durbin R (2009) Fast and accurate short read alignment with burrows–wheeler transform. Bioinformatics 25:1754-1760

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078-2079

Liu H, Bayer M, Druka A, Russell JR, Hackett CA, Poland J, Ramsay L, Hedley PE, Waugh R (2014) An evaluation of genotyping by sequencing (GBS) to map the Breviaristatum-e (ari-e) locus in cultivated barley. BMC Genomics 15:104

Lu H, Lin T, Klein J, Wang S, Qi J, Zhou Q, Sun J, Zhang Z, Weng Y, Huang S (2014) QTL-seq identifies an early flowering QTL located near flowering locus T in cucumber. Theor Appl Genet 127:1491-1499

Mahla HR, Singh JP, Roy MM (2014) Seed purpose watermelon in arid zone. Central arid zone research institute, Jodhpur, Raj. India

McKay J (1936) Factor interaction in *Citrullus* seed-coat color, fruit shape and markings show evidence of mendelian inheritance in watermelon crosses. J Hered 27:110-112

Meru G, McGregor C (2013) Genetic mapping of seed traits correlated with seed oil percentage in watermelon. HortScience 48:955-959

Meru G, McGregor C (2016a) Genotyping by sequencing for SNP discovery and genetic mapping of resistance to race 1 of *Fusarium oxysporum* in watermelon. Scientia Horticulturae 209:31-40

Meru G, McGregor CE (2016b) A Genetic Locus Associated with Resistance to *Fusarium oxysporum* f. sp. *niveum* Race 2 in *Citrullus lanatus*-type Watermelon. J Am Soc Hort Sci 141:617-622

Michelmore RW, Paran I, Kesseli R (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic

regions by using segregating populations. Proceedings of the national academy of sciences 88:9828-9832

Montero-Pau J, Blanca J, Esteras C, Martínez-Pérez EM, Gómez P, Monforte AJ, Cañizares J, Picó B (2017) An SNP-based saturated genetic map and QTL analysis of fruit-related traits in Zucchini using Genotyping-by-sequencing. BMC Genomics 18:94

Nath P, Dutta O (1973) Inheritance of Some Seed Characters in *Citrullus lanatus* Thunb. Mansf. Indian Journal of Horticulture 30:388-390

Nath P, Khandelwal R (1978) Genetics of seed-coat colour in *Citrullus lanatus* (Thunb.) Mansf. Theor Appl Genet 53:29-31

National Research Council (2006) Lost crop of Africa: volume II: vegetables. The national academies press, Washington, D.C

Oyolu C (1977) A quantitative and qualitative study of seed types in egusi (*Colocynthis citrullus* L.). Tropical Sci 19:287-290

Poole C (1944) Genetics of cultivated cucurbits. J Hered 35:122-128

Poole CF, Grimball PC, Porter DR (1941) Inheritance of seed characters in watermelon. Journal of agricultural research 63:433-456

Prothro J, Sandlin K, Abdel-Haleem H, Bachlava E, White V, Knapp S, McGregor C (2012b) Main and epistatic quantitative trait loci associated with seed size in watermelon. J Am Soc Hort Sci 137:452-457

Sachan S, Nath P (1976) Inheritance of some qualitative characters in watermelon, *Citrullus lanatus* (Thunb.) Mansf. Proc Third Intn Symp Sub-tropic Tropic Hort, Bangalore, pp 99-105

Schafferman D, Beharav A, Shabelsky E, Yaniv Z (1998) Evaluation of *Citrullus colocynthis*, a desert plant native in Israel, as a potential source of edible oil. J Arid Environ 40:431-439

Sharma R, Choudhury B (1982) Studies on inheritance of fruit and seed characters in watermelon *Citrullus lanatus*. Vegetable science 9:89-95

Singh VK, Khan AW, Jaganathan D, Thudi M, Roorkiwal M, Takagi H, Garg V, Kumar V, Chitikineni A, Gaur PM (2016) QTL-seq for rapid identification of candidate genes for 100-seed weight and root/total plant dry weight ratio under rainfed conditions in chickpea. Plant Biotechnol J 14:2110-2119

Takagi H, Abe A, Yoshida K, Kosugi S, Natsume S, Mitsuoka C, Uemura A, Utsushi H, Tamiru M, Takuno S (2013) QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. The Plant Journal 74:174-183

Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JA (2007) Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Res 35:W71-W74

Van Ooijen J (2006) JoinMap 4: Software for the calculation of genetic linkage maps in experimental populations. Kyazma BV, Wageningen, the Netherlands

Varshney RK, Nayak SN, May GD, Jackson SA (2009) Next-generation sequencing technologies and their implications for crop genetics and breeding. Trends Biotechnol 27:522-530

Wang Y, VandenLangenberg K, Wen C, Wehner TC, Weng Y (2018) QTL mapping of downy and powdery mildew resistances in PI 197088 cucumber with genotyping-by-sequencing in RIL population. Theor Appl Genet 131:597-611

Weetman LM (1937) Inheritance and correlation of shape, size and color in the watermelon, *Citrullus vulgaris* Schrad. Research Bulletin (Iowa Agriculture and Home Economics Experiment Station) 228:224-256

Zhang G, Ren Y, Sun H, Guo S, Zhang F, Zhang J, Zhang H, Jia Z, Fei Z, Xu Y (2015) A high-density genetic map for anchoring genome sequences and identifying QTLs associated with dwarf vine in pumpkin (*Cucurbita maxima* Duch.). BMC Genomics 16:1101

Zhang J (1996) Breeding and production of watermelon for edible seed in China. Cucurbit Genet Coop Rpt 19:66-67

Table 4.1 KASP^{TM} assays used to test association between significant genomic regions, identified from QTL-seq, and the seed coat color phenotype in watermelon. The marker names indicate the chromosome number and physical position of the marker based on 97103 watermelon genome (Guo et al., 2013)

KASP assay	Primer type	Primer Sequence (5'-3')	Tm (°C)
UGA3_5820134	FAM	GAAGGTGACCAAGTTCATGCTTAGAGACACAAGAAAGTTGCAAAGG	61.1
	VIC	GAAGGTCGGAGTCAACGGATTTAGAGACACAAGAAAGTTGCAAAGT	61.1
	Reverse	TCATTTATTTCCCTCCTTAGCTTTCA	62.5
UGA5_4591722	FAM	GAAGGTGACCAAGTTCATGCTTTGTGAAATCAAAGATATGGACCAA	61.7
	VIC	GAAGGTCGGAGTCAACGGATTTTGTGAAATCAAAGATATGGACCAG	61.7
	Reverse	GAGTTACTTGAATTTGGAAAGGAAAGG	62.6

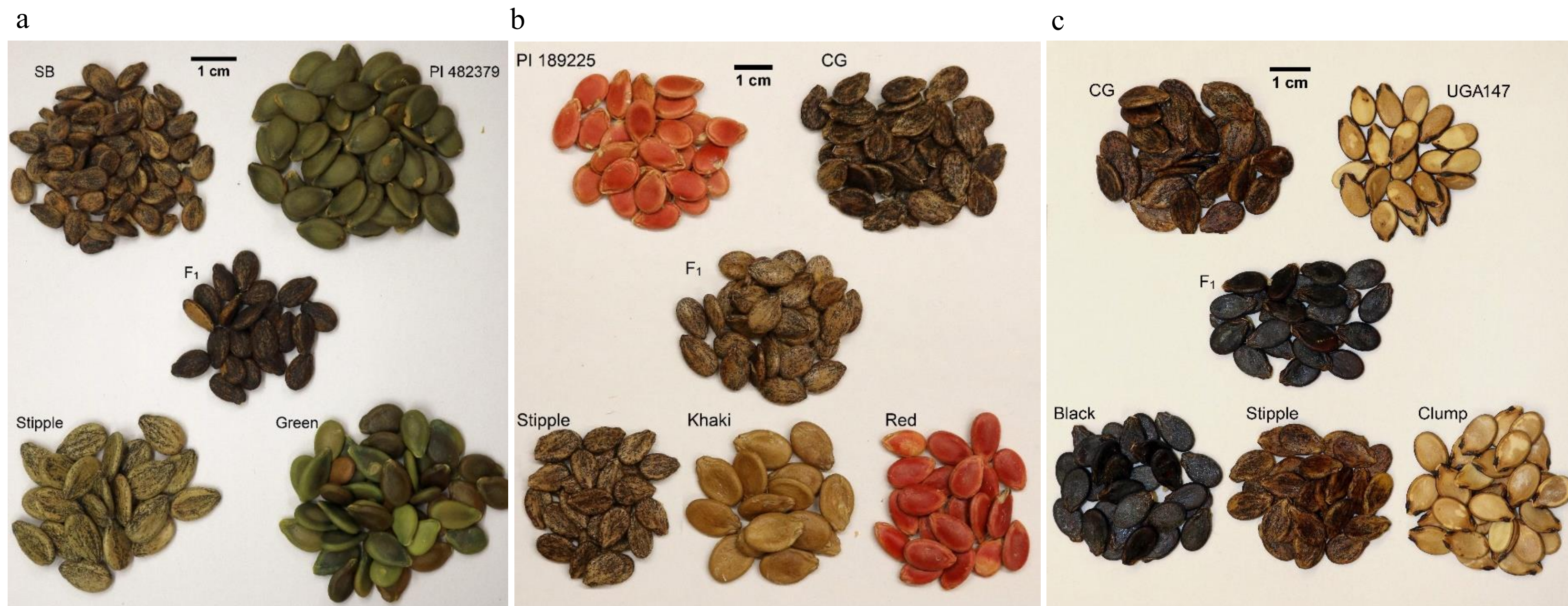
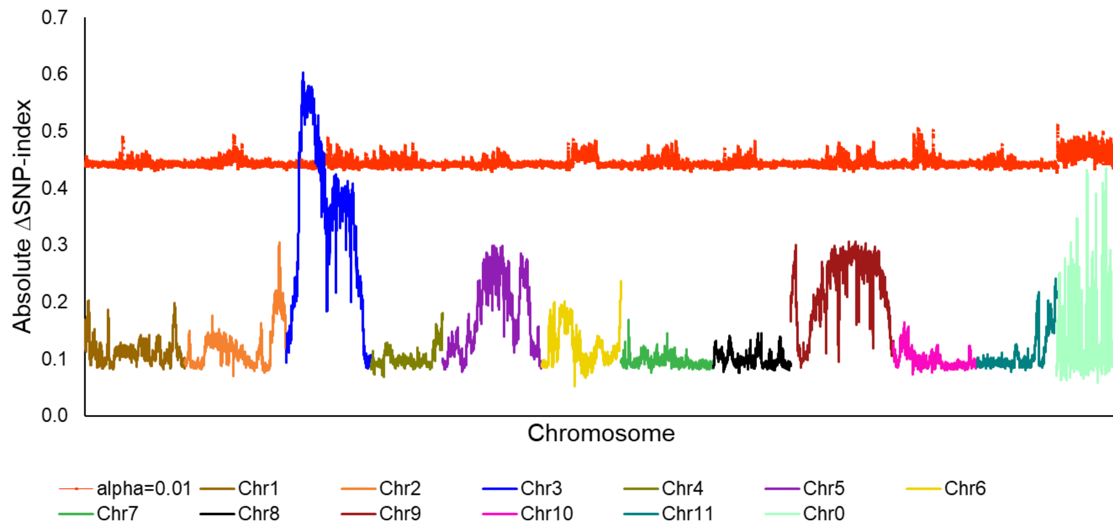
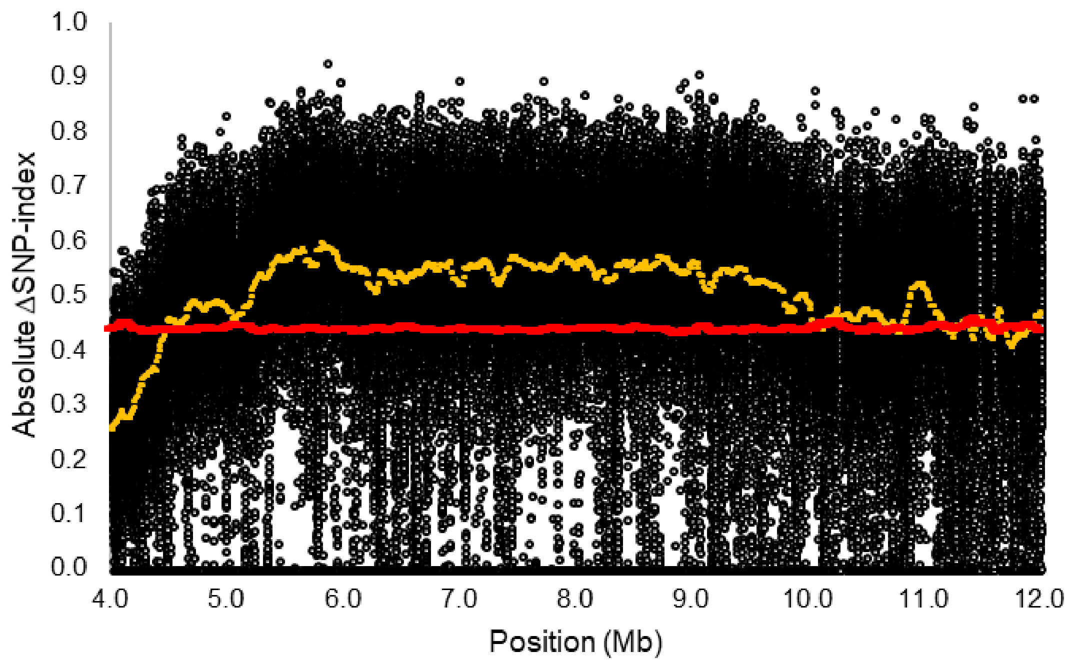


Fig. 4.1 Seed coat color of parents, F_1, and F_2 progenies in (a) the stipple x green population, (b) the stipple x red population and (c) the stipple x clump population. In the stipple x green population (a), seed of stipple seeded Sugar Baby (SB), female parent, green seeded PI 482379, male parent, F_1 and F_2 individuals with stipple and green phenotype. In the stipple x red population (b), seed of red seeded PI 189225, female parent, stipple seeded Charleston Gray (CG), male parent, F_1 and F_2 individuals with stipple, khaki and red phenotype. In stipple x clump population (c), seed of stipple seeded Charleston Gray (CG), female parent, clump seeded UGA147, selection from PI 169233, male parent, F_1 and F_2 progenies with black, stipple and clump phenotype

a



b



c

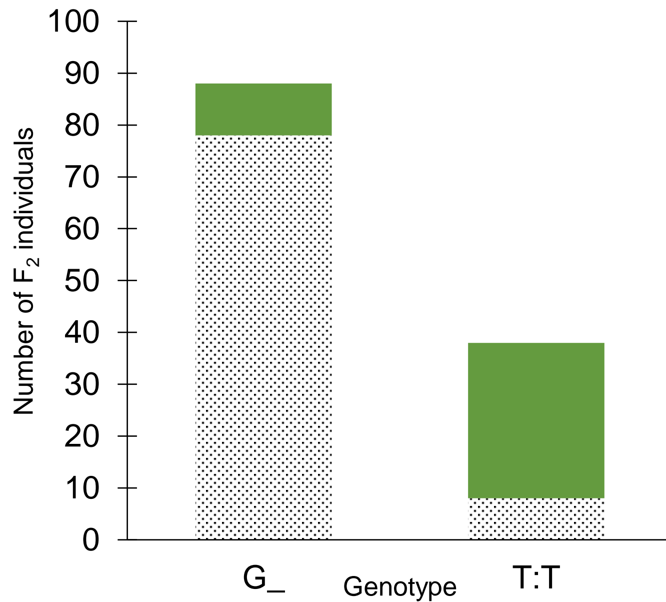
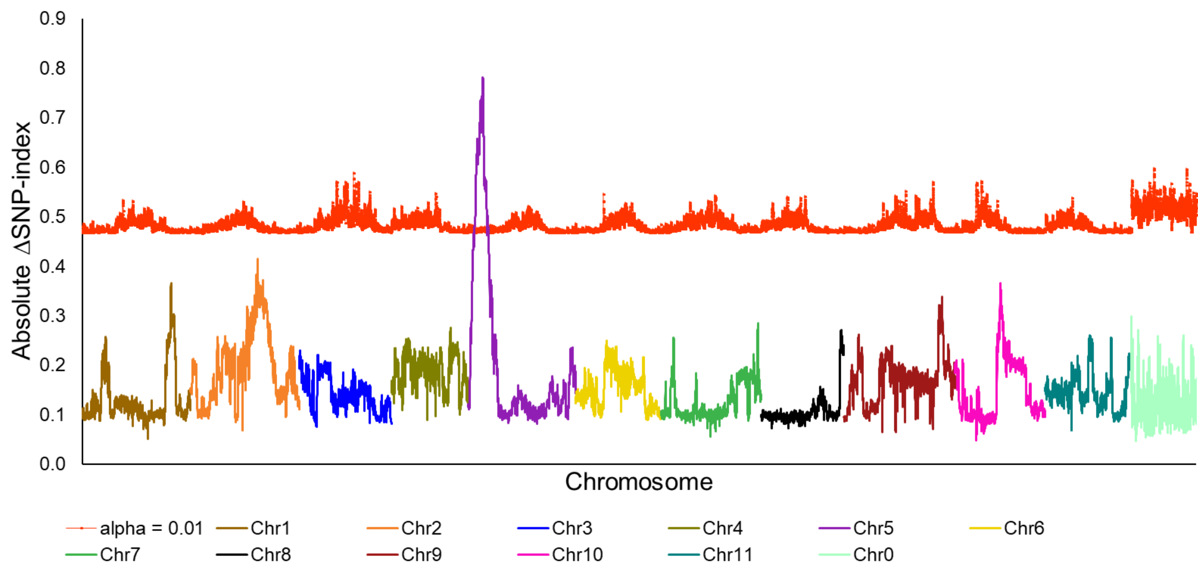
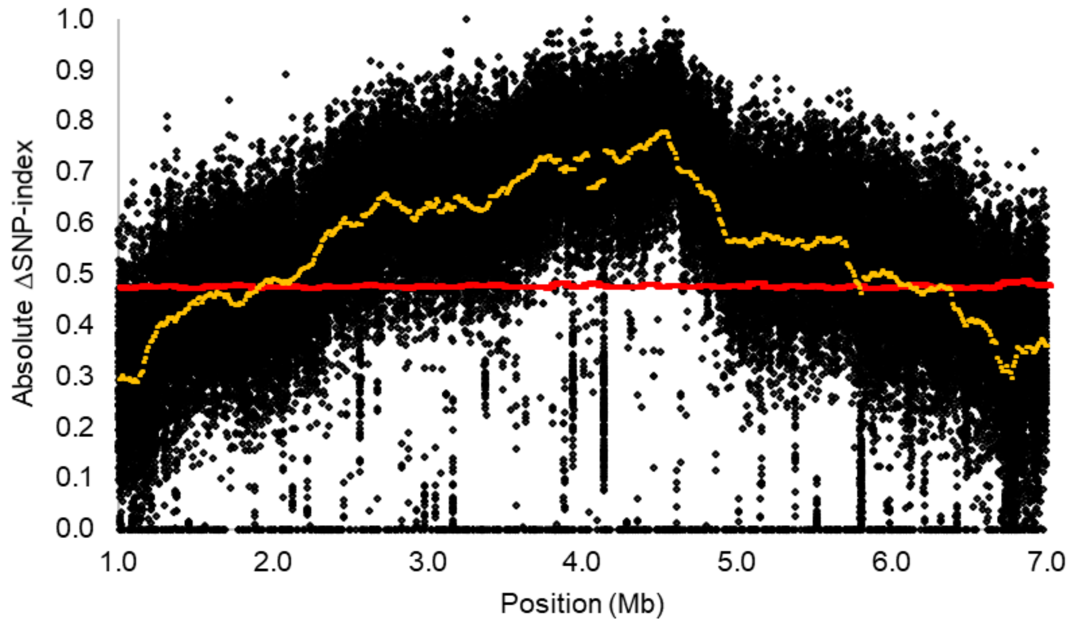


Fig. 4.2 (a) Absolute ΔSNP-index of all chromosomes in the inter-specific stipple x green F_2 population developed from a cross between Sugar Baby (*C. lanatus*) and PI 482379 (*C. amarus*) plotted along with statistical confidence intervals under the null hypothesis of no QTL ($P = 0.01$) (red line). The *R* locus is mapped on chromosome 3. (b) Magnified view of *R* locus, a significant ΔSNP-index peak (yellow), along with absolute ΔSNP-index of SNPs (black circles) plotted against the SNP position. SNP positions are based on the 97103 watermelon genome (Guo et al., 2013). (c) Association of KASP™ marker UGA3_5820134 with seed coat color phenotype in the stipple x green F_2 population (n = 126). The stipple and green sections in the graph indicate the number of F_2 individuals with stipple and green seed coat color, respectively

a



b



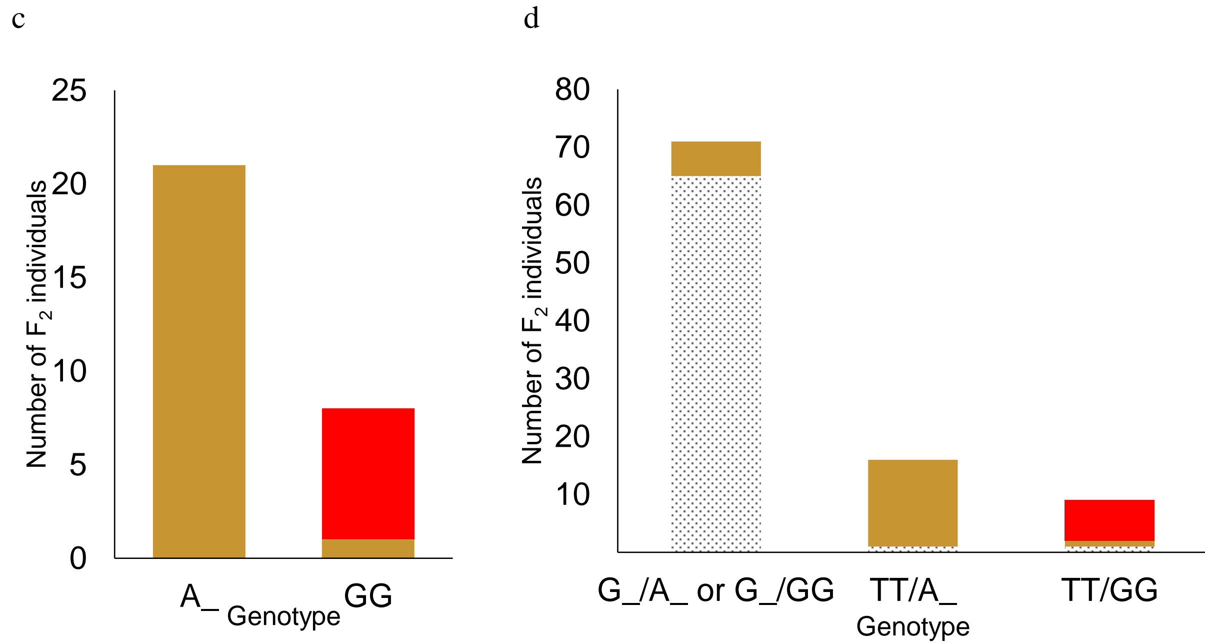
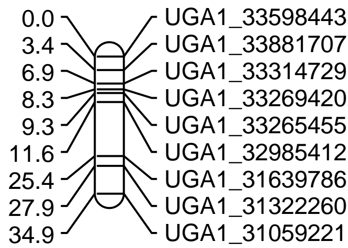
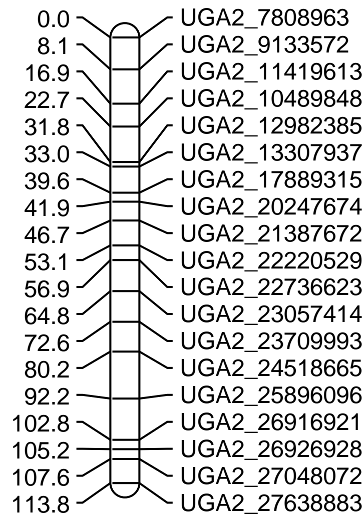


Fig. 4.3 (a) Mapping of the T^1 locus as a significant peak on chromosome 5 using QTL-seq. Absolute ΔSNP-index of all chromosomes, obtained by subtracting SNP-index of the red bulks from the khaki bulk in the stipple x red population, is plotted along with statistical confidence intervals under the null hypothesis of no QTL (P = 0.01) (red line). (b) Magnified view of significant ΔSNP-index peak (yellow) associated with T^1 locus along with absolute ΔSNP-index of SNPs (black circles) plotted against SNP position based on 97103 watermelon genome (Guo et al., 2013). (c) Association of KASP™ marker UGA5_4591722 with the khaki and red seed coat phenotype in the stipple x red F_2 population (n = 29). The x-axis denotes the genotype of F_2 individuals for KASP™ marker UGA5_4591722 and the y-axis denotes the number of F_2 individuals with khaki (khaki bar) and red seed coat color (red bar). (d) Bar graph indicating the phenotypic prediction accuracy of KASP™ markers UGA3_5820134 and UGA5_4591722 in the stipple x red population (n = 96). The genotypes on the x-axis represents the alleles of the UGA3_5820134/UGA5_4591722 markers. The stipple, khaki and red sections in the graph indicate the number of F_2 individuals with respective seed coat color

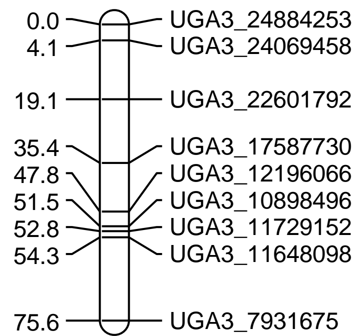
Chr_1



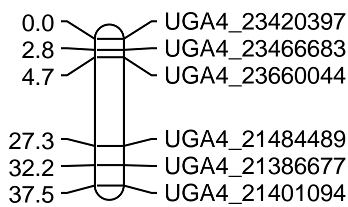
Chr_2



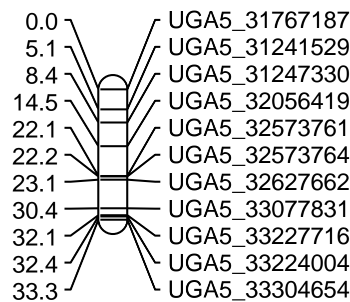
Chr_3



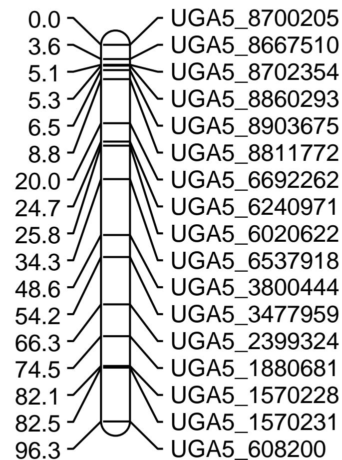
Chr_4

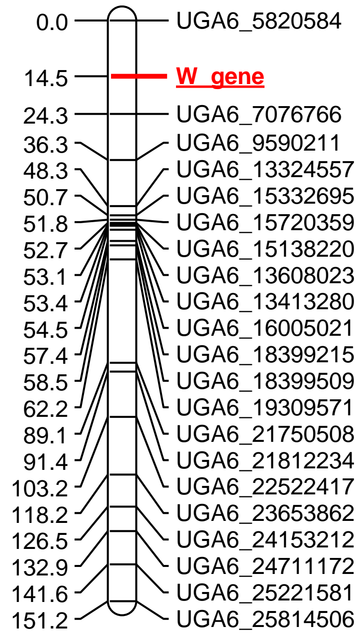


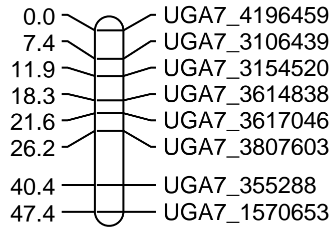
Chr_5A

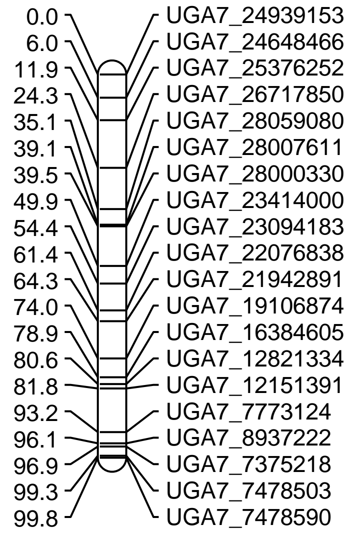


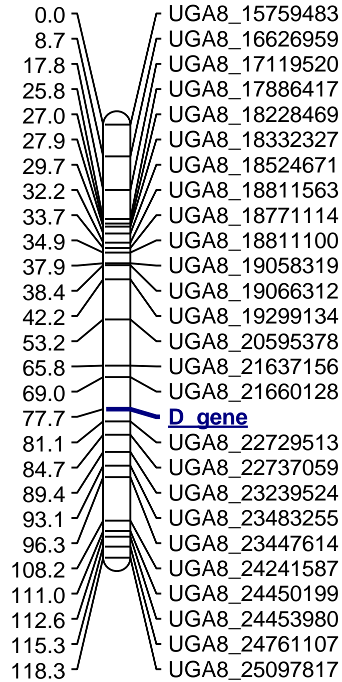
Chr_5B

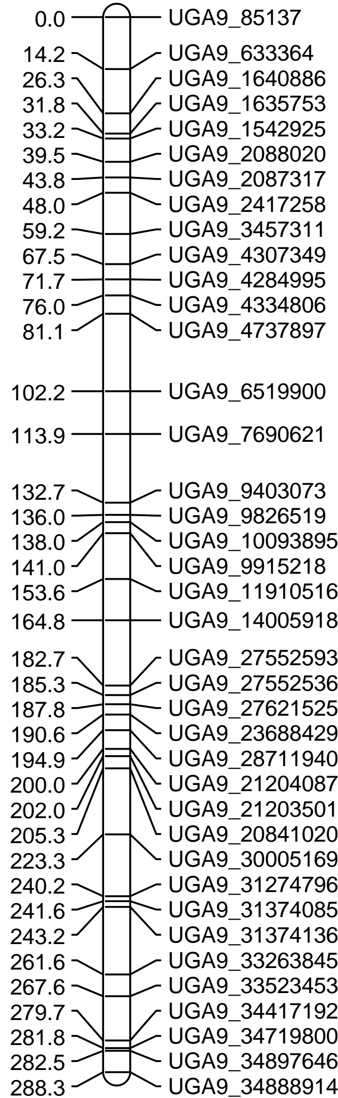


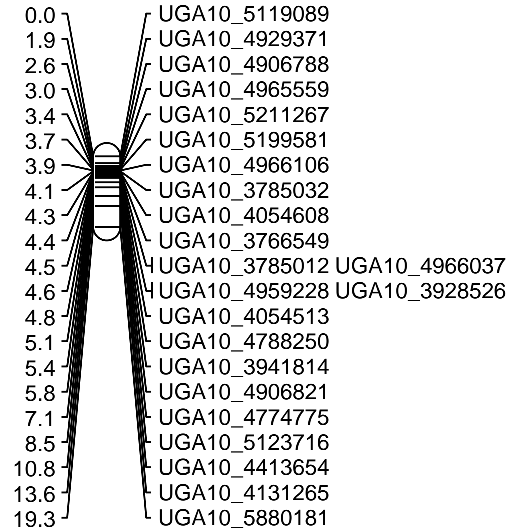

Chr_6
0.0 UGA6_5820584
14.5 W_gene
24.3 UGA6_7076766
36.3 UGA6_9590211
48.3 UGA6_13324557
50.7 UGA6_15332695
51.8 UGA6_15720359
52.7 UGA6_15138220
53.1 UGA6_13608023
53.4 UGA6_13413280
54.5 UGA6_16005021
57.4 UGA6_18399215
58.5 UGA6_18399509
62.2 UGA6_19309571
89.1 UGA6_21750508
91.4 UGA6_21812234
103.2 UGA6_22522417
118.2 UGA6_23653862
126.5 UGA6_24153212
132.9 UGA6_24711172
141.6 UGA6_25221581
151.2 UGA6_25814506


Chr_7A
0.0 UGA7_4196459
7.4 UGA7_3106439
11.9 UGA7_3154520
18.3 UGA7_3614838
21.6 UGA7_3617046
26.2 UGA7_3807603
40.4 UGA7_355288
47.4 UGA7_1570653


Chr_7B
0.0 UGA7_24939153
6.0 UGA7_24648466
11.9 UGA7_25376252
24.3 UGA7_26717850
35.1 UGA7_28059080
39.1 UGA7_28007611
39.5 UGA7_28000330
49.9 UGA7_23414000
54.4 UGA7_23094183
61.4 UGA7_22076838
64.3 UGA7_21942891
74.0 UGA7_19106874
78.9 UGA7_16384605
80.6 UGA7_12821334
81.8 UGA7_12151391
93.2 UGA7_7773124
96.1 UGA7_8937222
96.9 UGA7_7375218
99.3 UGA7_7478503
99.8 UGA7_7478590


Chr_8
0.0 UGA8_15759483
8.7 UGA8_16626959
17.8 UGA8_17119520
25.8 UGA8_17886417
27.0 UGA8_18228469
27.9 UGA8_18332327
29.7 UGA8_18524671
32.2 UGA8_18811563
33.7 UGA8_18771114
34.9 UGA8_18811100
37.9 UGA8_19058319
38.4 UGA8_19066312
42.2 UGA8_19299134
53.2 UGA8_20595378
65.8 UGA8_21637156
69.0 UGA8_21660128
77.7 D_gene
81.1 UGA8_22729513
84.7 UGA8_22737059
89.4 UGA8_23239524
93.1 UGA8_23483255
96.3 UGA8_23447614
108.2 UGA8_24241587
111.0 UGA8_24450199
112.6 UGA8_24453980
115.3 UGA8_24761107
118.3 UGA8_25097817


Chr_9
0.0 UGA9_85137
14.2 UGA9_633364
26.3 UGA9_1640886
31.8 UGA9_1635753
33.2 UGA9_1542925
39.5 UGA9_2088020
43.8 UGA9_2087317
48.0 UGA9_2417258
59.2 UGA9_3457311
67.5 UGA9_4307349
71.7 UGA9_4284995
76.0 UGA9_4334806
81.1 UGA9_4737897
102.2 UGA9_6519900
113.9 UGA9_7690621
132.7 UGA9_9403073
136.0 UGA9_9826519
138.0 UGA9_10093895
141.0 UGA9_9915218
153.6 UGA9_11910516
164.8 UGA9_14005918
182.7 UGA9_27552593
185.3 UGA9_27552536
187.8 UGA9_27621525
190.6 UGA9_23688429
194.9 UGA9_28711940
200.0 UGA9_21204087
202.0 UGA9_21203501
205.3 UGA9_20841020
223.3 UGA9_30005169
240.2 UGA9_31274796
241.6 UGA9_31374085
243.2 UGA9_31374136
261.6 UGA9_33263845
267.6 UGA9_33523453
279.7 UGA9_34417192
281.8 UGA9_34719800
282.5 UGA9_34897646
288.3 UGA9_34888914


Chr_10
0.0 UGA10_5119089
1.9 UGA10_4929371
2.6 UGA10_4906788
3.0 UGA10_4965559
3.4 UGA10_5211267
3.7 UGA10_5199581
3.9 UGA10_4966106
4.1 UGA10_3785032
4.3 UGA10_4054608
4.4 UGA10_3766549
4.5 UGA10_3785012 UGA10_4966037
4.6 UGA10_4959228 UGA10_3928526
4.8 UGA10_4054513
5.1 UGA10_4788250
5.4 UGA10_3941814
5.8 UGA10_4906821
7.1 UGA10_4774775
8.5 UGA10_5123716
10.8 UGA10_4413654
13.6 UGA10_4131265
19.3 UGA10_5880181

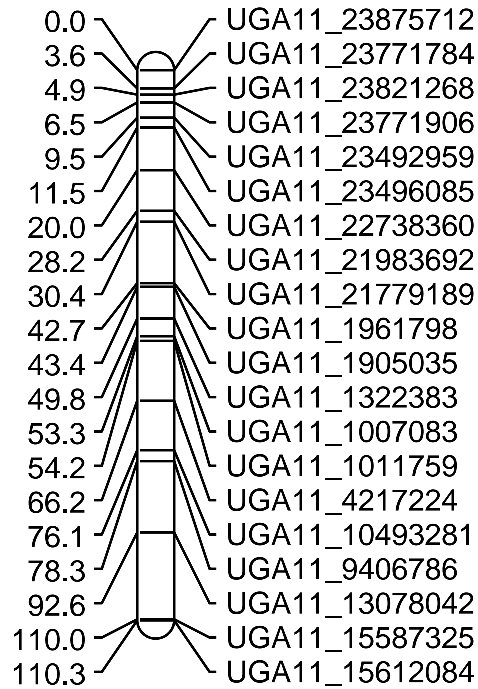


Fig. 4.4 A genetic linkage map of the stipple x clump F_2 population developed from a cross between stipple seeded Charleston Gray and a clump seeded UGA147, selection from PI 169233. 230 SNP markers and 2 phenotypic markers present on 11 watermelon chromosomes have been grouped into 13 linkage groups which spans 1226 cM. The *W* and *D* loci are mapped on chromosome 6 and 8, respectively. The marker names are indicated in the left side of the bars and genetic distance in cM is indicated in the right side of the bars

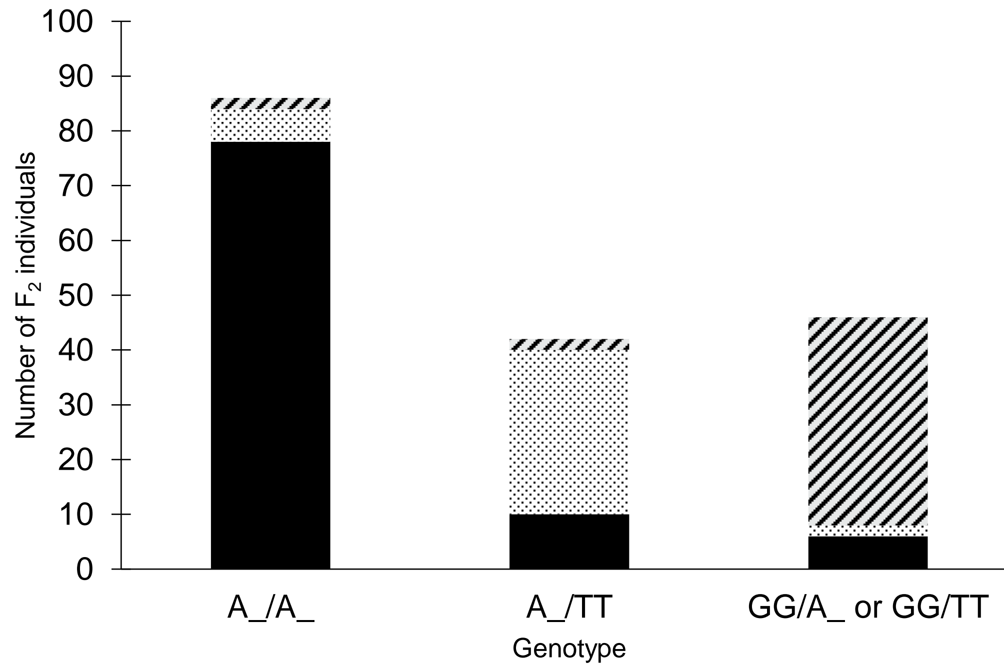


Fig. 4.5 Bar graph indicating phenotypic prediction accuracy of SNP markers UGA6_7076766 and UGA8_22729513 in the stipple x clump population (n = 174). The genotype on the x-axis represents the alleles of markers UGA6_7076766/UGA8_22729513 in the F_2 population. The sections with black, stipple (dotted) and diagonal black lines in the graph indicate number of F_2 individuals with black, stipple and clump seed coat color, respectively

CHAPTER 5

CONCLUSION

Watermelon is cultivated worldwide for its sweet juicy flesh. Even though seedless watermelons are predominantly cultivated in the U.S. and European countries, most of the watermelon produced globally are normal seeded watermelons (FAO 2013) produced for their edible flesh, while seeds are discarded. A lesser known vegetable crop, edible seed watermelon, is cultivated only in a few Asian and west African countries for its lipid and protein rich edible seeds (National Research Council 2006; Jarret and Levy 2012; Mahla et al. 2014). In these countries, edible seed watermelon is an important part of the human diet. It is cultivated in a significant proportion of land and it is of high social importance (Achigan-Dako et al. 2008). Edible seed watermelon has big edible seeds and bland/bitter, inedible flesh. Recently, breeders have tried to hybridize normal watermelon with egusi watermelon to develop varieties which have sweet flesh from normal watermelon and edible seeds from egusi watermelon (Orji et al. 2016). However, they have been facing severe challenges since the desirable traits are largely due to recessive alleles and they lack molecular markers for marker-assisted selection (MAS) of these traits.

The egusi trait is conferred by a single recessive locus, *eg*, mapped on chromosome 6 between markers NW0248325 and NW0250248 (Gusmini et al. 2004; Prothro et al. 2012). In the physical map, this locus is 4.28 Mb wide extending from 6.75 Mb to 11.03 Mb. Since this region is devoid of markers, further narrowing has not been possible. In this study, we used an independent validation population, UGA147 (selection from PI 169233) x Egusi (PI 560023) and

confirmed that the egusi trait is conferred by a single recessive locus. In the mapping population, Strain II (PI 279461) x Egusi, the *eg* locus was remapped using QTL-seq on chromosome 6 from 5.25 Mb to 7.85 Mb, partially overlapping the previously mapped locus. Fifteen Kompetitive Allele Specific PCR (KASP^{TM}) assays were developed using the SNPs identified by QTL-seq to narrow down the *eg* locus. The refined *eg* locus was mapped to the 86.3 cM position co-segregating with marker UGA6_7026576 on chromosome 6 of the SII x Egusi genetic map. In the 97103 genome physical map (Guo et al. 2013), the refined *eg* locus is 398.25 Kb wide, extends from 6.95 Mb to 7.35 Mb and contains 30 candidate genes. Further testing of KASP^{TM} markers on the validation population identified marker UGA6_7026576 as associated with the egusi phenotype. The marker will be useful for MAS of the egusi trait when used with select normal type watermelon cultivars.

Watermelon exhibits a wide natural variation in seed coat color ranging from black, stipple (dotted), red, green to white. Recent increase in consumer preference for specific seed coat colors has made seed coat color an economically important trait in edible seed watermelon. However, inadequate knowledge on the genetics of the trait has made breeding for watermelon seed coat color difficult. In this study, we used three F_2 populations segregating for seed coat color: Sugar Baby (stippled seed) x PI 482379 (green seed), Charleston Gray (stippled seed) x PI 189225 (red seed) and Charleston Gray (stippled seed) x UGA147 (clump seed) to understand the inheritance of seed coat color and to map four genes, *R*,*T*, *W* and *D*, described in the four-gene model (Poole et al. 1941; Poole 1944). Phenotypic segregation of F_2 individuals in the stipple x green and the stipple x clump population indicate that (i) the stipple is dominant to the green and clump seed coat color, (ii) stipple and green seed coat color segregate for one gene, *R* and (iii) stipple and clump seed coat color segregate for two genes, *W* and *D*, with recessive

epistatic interaction. These results fit the four-gene model. QTL-seq mapped the *R* gene on chromosome 3 from 4.48 Mb to 12.98 Mb in the stipple x green population. In the stipple x clump population, the *W* and *D* loci were mapped on chromosome 6 from 5.82 Mb to 7.07 Mb and on chromosome 8 from 21.66 Mb to 22.72 Mb, respectively, using GBS and traditional genetic mapping. In the stipple x red population, the F_2 individuals had stipple, khaki and red color seed segregating at 9:3:4 ratio indicating dominant epistasis. This phenotypic segregation does not fit the four-gene model which predicted this population to segregate at 9 stipple (*R_T_*): 3 tan (*R_tt*): 3 green (*rrT_*): 1 red (*rrtt*). The new seed coat color, khaki, is distinct from both tan and green, and it has not been described before in any literature. QTL-seq identified a significant genomic region when comparing khaki and red bulk on chromosome 5 from 1.89 Mb to 6.46 Mb in the stipple x red population. Based on the nature of inheritance and position of mapped region, we inferred that the genomic region identified from QTL-seq is either a novel locus or different allele of the T locus described by four-gene model. We name this locus T^1. Phenotype prediction accuracy of KASP™ markers UGA6_7076766 and UGA8_22729513 and SNP markers UGA6_7076766 and UGA8_22729513 were used to confirm that genomic regions in chromosomes 3, 5, 6 and 8 were associated to *R*, T^1, *W* and *D* genes. Genomic regions identified in this study are important steps in the development of molecular markers for MAS of seed coat color in watermelon breeding programs.

References

Achigan-Dako EG, Fagbemissi R, Avohou HT, Vodouhe RS, Coulibaly O, Ahanchede A (2008) Importance and practices of Egusi crops (Citrullus lanatus (Thunb.) Matsum. & Nakai, Cucumeropsis mannii Naudin and Lagenaria siceraria (Molina) Standl. cv.'Aklamkpa') in sociolinguistic areas in Benin. Biotechnologie, Agronomie, Société et Environnement 12:393-403

FAO (2013) Statistical yearbook 2013. World Food and Agriculture, FAO, Food and Agriculture Organization, UN Rome, Italy

Guo S, Zhang J, Sun H, Salse J, Lucas WJ, Zhang H, Zheng Y, Mao L, Ren Y, Wang Z (2013) The draft genome of watermelon (Citrullus lanatus) and resequencing of 20 diverse accessions. Nat Genet 45:51-58

Gusmini G, Wehner T, Jarret R (2004) Inheritance of egusi seed type in watermelon. J Hered 95:268-270

Jarret RL, Levy IJ (2012) Oil and fatty acid contents in seed of Citrullus lanatus Schrad. J Agric Food Chem 60:5199-5204

Mahla HR, Singh JP, Roy MM (2014) Seed purpose watermelon in arid zone. Central arid zone research institute, Jodhpur, Raj. India

National Research Council (2006) Lost crop of Africa: volume II: vegetables. The national academies press, Washington, D.C

Orji VO, Ogbonna PE, Onyia VN, Okechukwu EC (2016) Improvement of fruit taste quality of ‘egusi’melon (Colocynthis citrullus L.) through hybridization with watermelon (Citrullus lanatus Thumb.). Journal of Tropical Agriculture 54:1-6

Poole C (1944) Genetics of cultivated cucurbits. J Hered 35:122-128

Poole CF, Grimball PC, Porter DR (1941) Inheritance of seed characters in watermelon. Journal of agricultural research 63:433-456

Prothro J, Sandlin K, Gill R, Bachlava E, White V, Knapp SJ, McGregor C (2012) Mapping of the egusi seed trait locus (eg) and quantitative trait loci associated with seed oil percentage in watermelon. J Am Soc Hort Sci 137:311-315

APPENDIX

APPENDIX A. CUSTOM-MADE PYTHON SCRIPT

```
import sys
infile=open(sys.argv[1],"r")
chr=[]
pos=[]
index=[]
smooth=[]
tall=[]
short=[]
start = 0
w=int(sys.argv[2])
s=int(sys.argv[3])
for line in infile:
    line=line.strip()
    if line.startswith('#'):
        continue
    line=line.split('\t')
    chr.append(line[0])
```

```
    pos.append(int(line[1]))
    tall.append(float(line[4]))
    short.append(float(line[5]))
#    index.append(abs(float(line[6])))
    index.append(float(line[6]))
    smooth.append(0.0)
infile.close()
start_chr = chr[0]
start_pos = pos[0]
sum = 0.0
low = 0.0
low_1 = 0.0
ave = 0.0
total = 0.0
i = 0
back = 0
while i < len(chr):
    if chr[i] == start_chr:
        if pos[i] > start and pos[i] < start + w:
            sum += index[i]
            low += tall[i]
```

```
        low_1 += short[i]
        total += 1.0
        i+=1
        continue
    if pos[i] > start and pos[i] >= start + w:
        if total == 0.0:
            print("%s\t%i\t%i\t%f" % (start_chr,start,start+w,0.0))
            start_chr = chr[i]
            sum = 0.0
            total = 0.0
            start += s
            continue
        else:
            if total > 3:
                print("%s\t%i\t%i\t%f" % (start_chr,start,start+w,sum/total))
                while (pos[i] > start + s and start_chr == chr[i]):
                    i -= 1
                i+=1
            if total < 3:
                print("%s\t%i\t%i\t%f" % (start_chr,start,start+w,0.0))
                while (pos[i] > start + s and start_chr == chr[i]):
```

```
                i -= 1
            i+=1
        start_chr = chr[i]
        sum = index[i]
        low = tall[i]
        low_1 = short[i]
        total = 1.0
        start += s
        continue
if chr[i] != start_chr:
    print("%s\t%i\t%i\t%f" % (start_chr,start,start+w,sum/total))
    start_chr = chr[i]
    start = 0
    sum = index[i]
    total = 1.0
    i+=1
    continue
```