FRUIT QUALITY IN PAPAYA: USE OF GENOMIC ANALYSIS AND MOLECULAR MARKERS FOR THE CROP BREEDING

DIEIMES BOHRY

UNIVERSIDADE ESTADUAL DO NORTE FLUMINENSE
DARCY RIBEIRO – UENF

CAMPOS DOS GOYTACAZES - RJ
MARCH – 2022

FRUIT Q UALITY IN PAPAYA: USE OF GENOMIC ANALYSIS AND MOLECULAR MARKERS FOR THE CROP BREEDING

DIEIMES BOHRY

"Thesis presented to the Centro de Ciências e Tecnologias Agropecuárias (CCTA) of the Universidade Estadual do Norte Fluminense Darcy Ribeiro as part of the requirements for obtaining the title of Doctor of Science in Genetics and Plant Breeding."

Advisor: Prof.ª Helaine Christine Cancela Ramos

CAMPOS DOS GOYTACAZES - RJ
MARCH – 2022

FICHA CATALOGRÁFICA
UENF - Bibliotecas
Elaborada com os dados fornecidos pelo autor.

B677 Bohry, Dieimes.

FRUIT QUALITY IN PAPAYA : USE OF GENOMIC ANALYSIS AND MOLECULAR MARKERS FOR THE CROP BREEDING / Dieimes Bohry. - Campos dos Goytacazes, RJ, 2022.

133 f. : il.
Bibliografia: 58 - 72.

Tese (Doutorado em Genética e Melhoramento de Plantas) - Universidade Estadual do Norte Fluminense Darcy Ribeiro, Centro de Ciências e Tecnologias Agropecuárias, 2022.
Orientadora: Helaine Christine Cancela Ramos.

1. Papaya breeding. 2. Genomic analysis. 3. Molecular breeding. 4. SNP. 5. PCR markers. I. Universidade Estadual do Norte Fluminense Darcy Ribeiro. II. Título.

CDD - 631.5233

FRUIT QUALITY IN PAPAYA: USE OF GENOMIC ANALYSIS AND MOLECULAR MARKERS FOR THE CROP BREEDING

DIEIMES BOHRY

"Thesis presented to the Centro de Ciências e Tecnologias Agropecuárias (CCTA) of the Universidade Estadual do Norte Fluminense Darcy Ribeiro, as part of the requirements for obtaining the title of Doctor of Science in Genetics and Plant Breeding."

Approved on Mach 24th, 2022.

Examining committee:

Prof. Messias Gonzaga Pereira (Ph.D. Plant Breeding) - UENF

Prof.ª Marcia Flores da Silva Ferreira (D.Sc. Genetics and Breeding) - UFES

Dr. Pedro Henrique Dias dos Santos (D.Sc. Plant Science) - UFV

Prof.ª Helaine Christine Cancela Ramos (D.Sc. Genetics and Plant Breeding) - UENF (Advisor)

DEDICATION

To my parents, Reginaldo and Claudia, my brothers, Lucas and Vinicius, and my girlfriend Erica, with all my love, affection, and gratitude. To my family and all my friends, for the encouragement and love.

I DEDICATE THIS THESIS!

ACKNOWLEDGMENTS

To the Universe, for my existence;

To my parents, Reginaldo and Claudia, for all their love, affection, and dedication during this journey and all my life;

To my brothers, Lucas and Vinicius, for their friendship and companionship;

To my girlfriend Erica for her love, support, encouragement, and for accompanying me during this endeavor;

To the Universidade Estadual do Norte Fluminense Darcy Ribeiro and to the Graduate Program in Genetics and Plant Breeding for this opportunity;

To UENF, FAPERJ, and CAPES for financing the project and the scholarship;

To Professor Helaine for the excellent guidance during this process and for the opportunity to study in this area that has always aroused so much interest in me;

To my advisors, Professor Messias Gonzaga Pereira and Professor Alexandre Pio Viana, for their contributions, suggestions, and support in the development of this work;

To the other professors of the Genetics and Plant Breeding Program and UENF for being part of my training;

To Dr. Pedro Henrique Santos, Dr. Marcela Boechat, and Dr. Fernanda Abreu for the valuable information learned in the laboratory;

To Dr. Pedro Henrique Santos, Professor Marcia Flores, and Professor Messias Pereira for participating in my examining committee;

To my research team and friends, Adriana, Daniel, Grazi, Rafaela, and Lucas for the great discussion and contributions;

To all the friends I have made during my participation in the Study Group “Ganho Genético”;

To all my family members for their support and love, especially my grandparents, Miguel and Maria;

To my friends Gabriel, Luana, Miguel, Samuel, Valter, Divino, Jhean, Ravena, Mara, Renato, Rafael, and all the friends not mentioned here, but who made these years of study more fun and pleasant.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

Thank you so much!

SUMMARY

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 - DISCOVERY OF SNPs AND INDELs IN PAPAYA GENOTYPES AND THE POTENTIAL FOR MARKER ASSISTED SELECTION OF FRUIT QUALITY TRAITS

CHAPTER 2 - DEVELOPMENT OF PCR MARKERS FOR FRUIT FIRMNESS-RELATED GENES IN PAPAYA

ABSTRACT

BOHRY, Dieimes, D.Sc., Universidade Estadual do Norte Fluminense Darcy Ribeiro, March of 2022. Fruit quality in papaya: use of genomic analysis and molecular markers for the crop breeding. Advisor: Prof. Helaine Christine Cancela Ramos. Counselor Professors: Messias Gonzaga Pereira and Alexandre Pio Viana.

Papaya is a tropical and climacteric fruit that is recognized for its nutritional benefits and medicinal applications. Its fruits ripen quickly and undergo dramatic fruit softening, leading to great post-harvest losses. To overcome this scenario, papaya breeding programs must invest in exploring the available genetic variation to continue developing superior cultivars with improved fruit quality traits. The objective of this study was the utilization of genomic analysis and the development of molecular markers for papaya genetic breeding focussed on fruit quality traits. The first chapter consisted of performing whole-genome genotyping (WGG) of papaya, predicting the effects of the identified variants, and developing a list of ripening-related genes (RRGs) with linked variants. The Formosa elite lines of papaya Sekati and JS-12 were submitted to WGG with an Illumina Miseq platform. The effects of variants were predicted using the snpEff program. A total of 28,451 SNPs having a Ts/Tv (Transition/Transversion) ratio of 2.45 and 1,982 small insertions/deletions (InDels) were identified. Most variant effects were predicted in non-coding regions, with only 2,104 and 138 effects placed in exons and splice site regions, respectively. A total of 106 RRGs were found to be associated with

460 variants, which may be converted into PCR markers to facilitate genetic mapping and diversity studies and to apply marker-assisted selection (MAS) for specific traits in papaya breeding programs.

The second chapter involved the development and validation of PCR-based markers to study the genetic variability of fruit firmness-related genes (FFG) in papaya. A total of 13 single nucleotide polymorphisms (SNP) associated with 13 FFG was selected for marker development. The markers were designed through the PRIMER1 web tool. For marker validation, gradient PCR annealing temperature (Ta) was performed. After determining the best Ta, the primers were tested using DNA samples with known genotypes. Primers with similar PCR parameters were tested in combinations to reduce the number of PCRs for genotyping the 13 SNPs. The markers proved efficient at discriminating the 26 SNP alleles in a singleplex reaction. Validation of seven duplex amplifications revealed consistent results when compared with singleplex reactions. The PCR-based markers developed in this study are a valuable resource to study the variability of fruit firmness-related genes in papaya breeding populations and gene banks.

Keywords: Papaya breeding; Genomic analysis; Molecular breeding; SNPs; PCR markers.

RESUMO

BOHRY, Dieimes, D.Sc., Universidade Estadual do Norte Fluminense Darcy Ribeiro, março de 2022. Qualidade de frutos em mamoeiro: uso de análise genômica e marcadores moleculares para o melhoramento genético da cultura. Orientadora: Profª. Helaine Christine Cancela Ramos. Professores Conselheiros: Messias Gonzaga Pereira e Alexandre Pio Viana

O mamão é uma fruta tropical e climatérica reconhecida por seus benefícios nutricionais e aplicações medicinais. Seus frutos amadurecem rapidamente e apresentam um amolecimento drástico dos frutos, levando a grandes perdas pós-colheita. Para superar esse cenário, os programas de melhoramento de mamoeiro devem investir na exploração da variação genética disponível para continuar desenvolvendo cultivares superiores com características de qualidade de frutos melhoradas. O objetivo deste trabalho foi a utilização da análise genômica e desenvolvimento de marcadores moleculares para o melhoramento genético do mamoeiro, com foco em características de qualidade dos frutos. O primeiro capítulo consistiu em realizar a uma genotipagem ampla no genoma (WGG) do mamão, predizer os efeitos das variantes identificadas e desenvolver uma lista de genes relacionados ao amadurecimento (RRGs) com variantes ligadas. As linhagens de mamão do grupo Formosa Sekati e JS-12 foram submetidas a WGG com uma plataforma Illumina Miseq. Os efeitos das variantes foram previstos usando o programa snpEff. Foram identificados 28.451 SNPs com razão Ts/Tv (Transição/Transversão) de 2,45 e 1.982 pequenas inserções/deleções (InDels).

A maioria dos efeitos das variantes foram previstos em regiões não codificantes, com apenas 2.104 e 138 efeitos situados em exons e regiões de splice, respectivamente. Um total de 106 RRGs foram encontrados associados a 460 variantes, que podem ser convertidas em marcadores de PCR para facilitar o mapeamento genético e estudos de diversidade e aplicar a seleção assistida por marcadores (MAS) para características específicas em programas de melhoramento de mamoeiro. O segundo capítulo envolveu o desenvolvimento e validação de marcadores baseados em PCR para estudar a variabilidade genética de genes relacionados à firmeza de frutos (FFG) em mamão. Um total de 13 polimorfismos de nucleotídeo único (SNP) associados a 13 FFG foram selecionados para o desenvolvimento de marcadores. Os marcadores foram desenhados através da ferramenta PRIMER1. Para validação dos marcadores, foi realizado gradiente de temperatura de anelamento (Ta). Após determinar a melhor Ta, os marcadores foram testados usando amostras de DNA com genótipos conhecidos. Os marcadores com parâmetros de PCR semelhantes foram testados em combinações para reduzir o número de PCRs para genotipagem dos 13 SNPs. Os marcadores mostraram-se eficientes em discriminar os 26 alelos dos SNPs em reação singleplex. A validação de sete amplificações duplex revelou resultados consistentes em comparação com reações singleplex. Os marcadores baseados em PCR desenvolvidos neste estudo são um recurso valioso para estudar a variabilidade de genes relacionados à firmeza de frutos em populações de melhoramento e bancos de germoplasma de mamão.

Palavras-chave: Melhoramento do mamoeiro; Análise genômica; Melhoramento molecular; SNPs; Marcadores de PCR.

1. INTRODUCTION

Papaya (*Carica papaya* L.) is the fourth most cultivated tropical fruit crop in the world, behind only bananas, mangoes, and pineapples (FAOSTAT, 2022). Papaya fruits are appreciated for their unique flavor, aroma and texture, excellent nutritional qualities, such as high vitamin A and C content, antioxidants such as β-carotene and lycopene, minerals, and fibers (Chandrika et al., 2003; De Souza et al., 2008).

Brazil is one of the major producers and exporters of fresh fruits worldwide. The significant availability of cultivated land, diverse microclimates, and the use of agricultural technologies are key to Brazilian food production and supply. Regarding papaya, Brazil is world´s third largest producer and exporter. The cultivation of papaya occurs in tropical and subtropical regions in over 60 countries with developing economies (FAOSTAT, 2022). In addition to the great importance of papaya for the economy of these countries and its nutritional benefits for the population, a few breeding programs are working to develop superior papaya cultivars (Pereira et al., 2018), due in part to the fact that the global cultivation of papayas is centered on a small number of cultivars.

Papaya is considered a climacteric fruit crop because during the fruit ripening process there is a burst of ethylene synthesis that coincides with cell respiration (Shen et al., 2017). The ethylene induces several modifications during the ripening process, including pulp, skin color, and texture changes; the synthesis and accumulation of carotenoids and vitamins; the solubilization, synthesis, and

transport of sugars; fruit softening; and other flavor and sensorial attributes (Fabi et al., 2010). Such modifications promote seed dispersal and improve the fruit for human and animal consumption.

Although ethylene blocking compounds proved able to delay fruit ripening in papaya and maintain fruit firmness for a longer period, they affected fruit quality (Fabi et al., 2007). An effective protocol for papaya post-harvest storage has yet to be developed. Another way to improve papaya shelf life is by genetic breeding. Fruit quality in papaya is of great importance throughout the entire supply chain. The major challenge for producers is providing consumers with fruits with good nutritional and sensorial properties, while maintaining sufficient shelf life until consumption (Martín-Pizarro & Posé, 2018). This is especially relevant in breeding programs focussed on papayas because of the fruit´s rapid ripening and loss of fruit firmness, which can lead to significant losses in the post-harvest chain.

Due to the great importance of the fruit ripening process in papaya, studies at the molecular level have been conducted with the aim of identifying the genetic factors that influence this phenomenon. Using gene expression profiling, Fabi et al. (2014) analyzed the correlations and co-expression networks of cell wall-related genes and obtained results suggesting that papaya fruit softening takes place through the interactions of multiple glycoside hydrolases. The study found the transient expression of a polygalacturonase gene resulted in pulp softening and identified this gene as a potential target for biotechnological breeding able to generate non-transgenic plants with improved fruit quality.

In another study, Shen et al. (2017) treated papaya fruit samples with a ripening inducer (exogenous ethylene) and inhibitor (1-methylcyclopropene) and submitted the samples to high-throughput RNA-sequencing. About 50 ripening-related genes with differential expressions among the treatments were isolated from this study, including genes related to the cell-wall metabolism, chlorophyll and carotenoid metabolism, proteinases and their inhibitors, plant hormone signal transduction pathways, and transcription factors. These genes constitute an important source of information regarding the development of strategies for the genetic mapping of fruit quality traits in papaya breeding programs.

However, these studies did not focus on elucidating the genetic foundations of fruit quality traits based on the genetic variability of papaya genotypes, which is essential for its utilization in breeding programs. Thus, since

expanding the Brazilian papaya export market is limited by a lack of cultivars with high firmness fruits able to withstand long transport times there is a clear need for the development of more effective methods for obtaining genetic gains for this fruit crop. Papaya is currently only exported from Brazil by air, and done so at much higher cost compared with conventional fruit shipping by water. In addition, papaya cultivars with an improved shelf life could also reduce losses throughout the entire market chain within the country.

In this thesis we describe the application of genomic analysis and the development of molecular markers to improve fruit quality traits, especially those associated with fruit firmness in papaya. This thesis is divided into two chapters, as follows: CHAPTER 1 consists of performing the whole-genome genotyping (WGG) of papaya, predicting the effects of the identified variants, and developing a list of ripening-related genes (RRGs) with linked variants; CHAPTER 2 covers the development and validation of PCR-based markers to study the genetic variability of fruit firmness-related genes (FFG) in papaya.

2. GENERAL OBJECTIVE

To perform genomic analysis and develop molecular markers for papaya genetic breeding, with a focus on fruit quality traits.

2.1 Specific objectives

1. Performing a genome-wide identification of DNA variants in papaya;
2. Predicting the effects of variants according to genomic location;
3. Developing a list of ripening-related genes with linked variants to facilitate further genotype/phenotype association studies;
4. Designing PCR-based markers to study the genetic variability of fruit firmness-related genes (FFG) in papaya;
5. Validating the PCR-based markers and evaluating their effectiveness.

3. CHAPTERS

3.1 DISCOVERY OF SNPs AND INDELs IN PAPAYA GENOTYPES AND THE POTENTIAL FOR MARKER ASSISTED SELECTION OF FRUIT QUALITY TRAITS

3.1.1 INTRODUCTION

Papaya (*Carica papaya* L.) is a fruit crop cultivated in tropical and subtropical regions of the globe and is listed among the four major fresh tropical fruits. In Brazil, papaya is an important crop with production of around 1.235 million tonnes in 2020, placing the country as the world´s third largest producer and exporter, although with most of the production destined for the domestic market (FAOSTAT, 2022). Papaya fruits are prized for their excellent nutritional and medicinal qualities, high vitamin A and C content, antioxidants such as β-carotene and lycopene, minerals and fibers (Chandrika et al., 2003; De Souza et al., 2008).

Several genetic and genomic resources are available in papaya due to the great advances of sequencing technologies that have contributed to our understanding of the species intriguing sex-determination system (Lee et al., 2018; Liao et al., 2017; Ming et al., 2007; VanBuren et al., 2015, 2016). Besides the sex

determination of papaya, other relevant traits have been investigated through gene expression analysis, such as those related to fruit quality traits (Fabi et al., 2014; Shen et al., 2017, 2019), embryogenesis (Jamaluddin et al., 2017), mild and severe drought resistance (Gamboa-tuz et al., 2018), etc. However, the utilization of sequencing technologies to identify DNA polymorphisms for the genetic mapping of important traits for papaya breeding has been scarce.

The available linkage maps for papaya have varied in coverage, resolution, and type of DNA polymorphisms. The first high-density linkage map was based on 1,498 Amplified Fragment Length Polymorphisms (AFLP) (Ma et al., 2004). The following high-density map was developed with 706 Simple Sequence Repeat (SSR) markers (Chen et al., 2007). The same mapping population was used to improve the map resolution with 277 AFLP and 712 SSR markers and allowed the identification of 14 quantitative trait loci (QTL) related to fruit quality traits (Blas et al., 2012).

More recently, a linkage map based on 219 single nucleotide polymorphisms (SNP) was developed (Nantawan et al., 2019). Although this map was based on SNP markers, which are very abundant and quickly identified through Next Generation Sequencing (NGS), the great distortion of the expected marker segregation observed in F_2 (1:2:1) significantly decreased the map resolution. Still, a total of 21 QTLs for fruit quality traits was detected using this map that will enable candidate gene isolation and the development of marker-assisted selection strategies.

The conventional breeding of papaya for complex traits, such as fruit firmness and sweetness, is time-consuming and renders modest genetic gain per selection cycle. Moreover, Brazilian federal legislation requires that papaya plants, even those in breeding fields, be cut down when showing the first symptoms of viral diseases, mainly the Papaya ringspot virus (PRSV), thus precluding complete measurements in breeding populations. Therefore, the use of molecular markers could speed up the time for selection in papaya breeding programs by allowing the analysis of a higher number of progenies at an early stage of development and increase the genetic gain (Huq et al., 2016).

In Brazil, the papaya breeding program at UENF has had great success in the development of 21 new papaya cultivars (MAPA, 2022), which reduced the need to import hybrid seeds, increased choices for farmers and consumers, and

placed the country as a potential papaya seed exporter. One of these cultivars is the UC10 hybrid, with high yield fruits of around 1.9 kg (Pereira et al., 2019). This hybrid is a relative of the Formosa elite lines Sekati and JS-12, which present contrasting agronomic and fruit quality attributes.

Understanding the genetic and genomic aspects related to fruit quality traits in papaya is essential for the ongoing development of superior cultivars with the unique features valued in both the national and international markets. Numerous studies on climacteric fruit crops have revealed the potential of NGS-based markers for the genetic mapping of fruit quality traits (Liu et al., 2017; Luo et al., 2016; Martínez-García et al., 2013; Nantawan et al., 2019).

In climacteric fruit species such as papaya, peach, tomato, and apple, a rise in cellular respiration, which coincides with ethylene synthesis, is observed during the fruit ripening process. Ethylene is the main phytohormone regulating the ripening of climacteric fruits and its action influences the development of the sensorial and nutritional attributes of such fruits (Lü et al., 2018). One major change in texture during the ripening of such fruits is rapid fruit softening, making it more susceptible to physical injuries and post-harvest diseases. Fruit softening is a complex process characterized by the substantial activity of cell-wall degrading enzymes such as polygalacturonase and beta-galactosidase (Fabi et al., 2014; Shen et al., 2017).

The availability of genomic information related to fruit quality traits will enable the development of tools to aid in the papaya selection process. Thus, in this study, we carried out a genome-wide identification of DNA variants among the Formosa elite lines Sekati and JS-12 using an Illumina MiSeqTM platform. The identified variants were used to predict effects according to genomic location and to develop a list of ripening-related genes with linked variants to facilitate further genotype/phenotype association studies and to apply marker-assisted selection for papaya breeding.

3.1.2 LITERATURE REVIEW

3.1.2.1 Main aspects of papaya

The species *Carica papaya* L., the commercially grown papaya, belongs to the family *Caricaceae*, which is divided into six genera with 35 species distributed as follows: *Carica* (one species), *Jacaratia* (eight species), *Cylicomorpha* (two species), *Jarilla* (three species), *Horovitzia* (one species), and *Vasconcellea* (20 species) (Badillo, 2000).

The papaya has a particular system of sex determination, with three different sexual forms observed: plants with female, male, and hermaphrodite flowers. Several hypotheses have been proposed over recent decades to explain the genetics of trioecious papayas, such as the presence of a single gene with three allelic forms; a group of closely related genes; a genetic balance of the sex chromosome over the autosomal; classic XY chromosomes; and regulatory elements in the biochemical pathway of floral development. However, only studies at the molecular level were able to unravel this system. The papaya has XY chromosomes which determine the sex of the plant, where the Y chromosome has two slightly different forms that determine the male sex (XY) and hermaphrodite (XY^h). Female plants have the chromosomes XX (Ming et al., 2007).

Cultivated papayas are currently divided into two distinct heterotic groups: “Solo” and “Formosa”. "Solo" has the largest number of varieties available on the world market, and most are pure line varieties (Ex.: *Sunrise Solo, Improved Sunrise Solo Line* 72/12, Kapoho, Waimanalo, Higgins, and Baixinho de Santa Amália), thus allowing farmers to use seeds harvested in the own orchard to constitute the next crop generation. The fruits obtained from hermaphrodite plants are pyriform, have an average weight between 300 and 650 grams and red-orange pulp, which makes them more popular in the external market (Dantas et al., 2000). In addition, breeders have recently explored the variability within the heterotic group “Solo”, where the first hybrid cultivars of this group were released (MAPA, 2022). In contrast, “Formosa” consists of hybrids characterized by the production of fruits with an average weight of between 900 and 1,300 grams and reddish pulp. Due to these characteristics, production in the country is destined for domestic consumption. Hybrids of the Tainung series (Ex.: Tainung n.º 1, 2) are the most cultivated. These hybrids have the disadvantage of high-priced seeds imported from Taiwan for US$ 3,000 to US$ 4,000 per kilogram. Another hybrid is the ‘UENF/CALIMAN 01’ (Calimosa), produced by the genetic breeding team of

UENF in partnership with the company Caliman Agrícola S.A. This hybrid is being cultivated in the different producing regions of Brazil, such as the states of Rio Grande do Norte, Bahia, and Espírito Santo (Pereira et al., 2015).

Papaya tree gene banks present wide phenotypic variation for many important agronomic characteristics, such as fruit size and shape, pulp color, flavor, sweetness, and plant height (Kim et al., 2002). The papaya plant develops best at high temperatures, with an ideal range of between 22 and 26 ° C, rainfall of 1,500 mm per year, and relative humidity of 60 to 85%. The papaya tree begins the emission of flowers at three to four months of age and fruiting begins eight months after sowing; thus, while commercially cultivated plants have a production period of 16 months, the total lifespan of the crop is 24 months. Thus, the need to renew the crop field every two years increases the demand for seeds for the maintenance of high yield and quality plantationsh (Ferreguetti, 2003).

3.1.2.2 Economic importance of the crop

Papaya is grown in more than 60 countries, with most of the production taking place in developing economies located in the tropics. The fruit has gained popularity worldwide, appearing in the fourth position (13.9 million tons) in tropical fruit production, behind only bananas (119.8 million tons), mangoes (54.8 million tons), and pineapple (27.8 million tons) (FAOSTAT, 2022).

Due to its nutritional properties, papaya is considered one of the most important fruits for consumption, and is consumed mostly in natura or industrialized in the form of crystallized sweets or raisins and jellies. It is rich in antioxidants (Ex.: carotenoids, vitamin C, and flavonoids), vitamin B (Ex.: folic and pantothenic acid), minerals (Ex.: potassium and magnesium), and fiber. In addition, the fruit is the source of the digestive enzyme papain, which is used as an ingredient in brewing, meat tenderizing, pharmaceuticals, beauty products, and cosmetics (Evans et al., 2012). The carpaine alkaloid, used as a cardiac activator, can also be extracted from leaves, fruits, and seeds (Dantas et al., 2000).

Brazil ranks as the third largest producer of papaya in the world, with a production of 1.235 million tonnes, behind only India (6.011 million tons) and the Dominican Republic (1.271 million tonnes) in production in 2020. In the same year, the country was the third largest exporter of the fruit (43,708 tonnes), with Mexico and Guatemala exporting 167,356 and 55,204 tons, respectively. For their parts,

the United States of America (189,706 tons), Singapore (21,883 tonnes), and Canada (17,899 tonnes) are the largest importers of papaya (FAOSTAT, 2022). The major producing states in Brazil were Espírito Santo (403,278 tons), Bahia (390,075 tons), and Ceará (118,717 tons) in 2019. Together these states represent 78.5% of the Brazilian production, with the states of Ceará and Espírito Santo showing the highest average yield (around 59 t ha -1) (IBGE, 2020).

3.1.2.3 Papaya breeding

In Brazil, the institutions that stand out in the genetic breeding of papaya are the Universidade Estadual do Norte Fluminense Darcy Ribeiro (UENF), in the state of Rio de Janeiro, the Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA Mandioca e Fruticultura), in the state of Bahia, and the Instituto Capixaba de Pesquisa e Extensão Rural (INCAPER), in the state of Espírito Santo.

The objectives of the breeding programs of these institutions have been to carry out research to develop superior cultivars and hybrids that show resistance or tolerance to pests, diseases, and environmental variation. They have sought cultivars and hybrids, associated with good agronomic characteristics in terms of productivity, early fruiting, abundance, and a first fruit height inferior to 90 cm, as well as high fruit quality, high soluble solids content, thin skin, and an absence of spots to supply the internal and external markets. In addition, they work to develop plants with fewer hermaphrodite carpellodic flowers that are sterile and pentandric, storage and transportation resilience, and small internal and star-shaped cavities, among other attributes. Thus, these programs may expand the limited number of varieties and hybrids currently available for cultivation in the country.

The UENF papaya breeding program in partnership with the Caliman Agrícola S.A., located in Linhares-ES, is one of Brazil´s most productive ones in terms of cultivar development. Among the results obtained by this partnership, the registration of the first Brazilian hybrid of papaya, 'UENF / CALIMAN01', popularly known as 'Calimosa', which was recommended to farmers in 2003 and has been cultivated in the main growing regions of Brazil, is noteworthy. Since then, another 20 hybrids have been registered in the *Ministério da Agricultura, Pecuária e Abastecimento* (MAPA, 2022). These advances have allowed the country, a

traditional importer of papaya seeds, mainly of the 'Tainung 01' hybrid, to potentially reduce its dependence on seed imports (Pereira et al., 2015).

In order to attain such results, several studies at UENF used both morpho-agronomic characteristics and papaya crop molecular tools (Pereira et al., 2015) to enhance knowledge of floral behavior, reproduction mode, sex prediction, backcrossing, heterosis, combinatorial ability, topcross, diallel, and resistance to diseases.

While evaluating heterotic effects through partial diallel crosses between 'Solo' and 'Formosa' genotypes, Marin et al. (2006) verified the superiority of the hybrids in relation to their parents and recommended the best hybrid combinations. In another study initiated by Silva et al. (2007), the program sought the sex conversion of the Cariflora genotype from the dioecious (population with female and male plants) to the ginoic-andromonoic (population with female and hermaphrodite plants), to incorporate this material in the process of obtaining inbred lines and superior genotypes. Both classical breeding tools and molecular markers were used to increase the efficiency of the process. The populations RC_1 and RC_2 were obtained from a cross between Cariflora x Sunrise Solo 783 and the subsequent selection of plants with a high degree of similarity to the recurrent parent Cariflora. Several studies were developed for the development, evaluation, and conduction of segregating populations from this crossing (Silva et al., 2007a; Silva et al., 2007b; Silva et al., 2008a; Silva et al., 2008b; Ramos et al., 2011a; Ramos et al., 2011b; Ramos et al., 2011c)

Ramos (2010) has advanced this study of sexual conversion of the genotype 'Cariflora' by conduction of the population RC_3, while Ramos et al. (2012) evaluated the genetic distance between populations derived from three generations of backcrosses (RC_1, RC_2, and RC_3) to select the superior genotypes in order to advance the generations. In addition, combined selection strategies were developed for the main agronomic characteristics of these populations, considering both the phenotypic value and the genetic value obtained by the methodology of the mixed linear models, through the REML / BLUP procedure in order to identify the best families (Ramos et al., 2014).

Several studies have aimed to identify sources of resistance for the main papaya diseases in gene banks and breeding populations, such as black spot, phoma spot, and powdery mildew (Chagas et al., 2021; Moraes et al., 2019;

Pastana et al., 2017; Pirovani et al., 2021; Vivas et al., 2014, 2015, 2016, 2017, 2018).

Cortes et al. (2017) and Santa-Catarina et al. (2018) developed and validated a methodology of analysis and digital image processing focused on the phenotyping of morphological and fruit quality traits of papaya. This methodology has been applied in the phenotyping of the new experimental areas, maintaining accuracy, compared to conventional measurements, and saving time for data collection, at UENF's papaya breeding program.

More recently the UENF/CALIMAN papaya breeding program started a recurrent selection program as an effective strategy for the continuous development of new cultivars. The base population was constituted of superior genotypes and unimproved dioecious varieties with sources of genes for traits of interest and excellent options for the formation of segregating populations (Pereira & Santa-Catarina, 2021). Evaluation of the S1 family revealed significant chance of success in the long, mid, and short-terms. The segregating population has great genetic variability for diseases, fruit yield, and quality, aggregating highly valuable genetic resources that can be explored through different breeding methods, thus contributing to the development of new papaya varieties (Santa-Catarina et al., 2020).

3.1.2.4 Application of OMIC technologies in papaya

Significant advances in OMIC technologies over the last decade have allowed the integration of large datasets for answering important questions related to plant breeding. These include the identification of thousands of DNA polymorphisms for several purposes and the identification on a large scale of transcripts or proteins that are differentially expressed in a given plant tissue or treatment. The main advances regarding genomics and transcriptomics for papaya will be described below.

The species *Carica papaya* is considered a model plant for genomic studies due to a series of characteristics: a short juvenile period; abundant flowering and fruiting; a large seed to fruit ratio; and the fact that efficient in vitro propagation and genetic engineering techniques allow the study of several reproductive and vegetative processes linked to the information from genomic sequences (Yu et al., 2009). In addition, the species is phylogenetically positioned

close to *Arabidopsis thaliana*, the main model plant, and is within one of the main groups of dicotyledons, the order *Brassicales*, making it an optimal candidate for studies of comparative and evolutionary genomics within this group (Chen et al., 2007).

A diploid species with nine pairs of chromosomes, this papaya has a relatively small genome (372 Mpb). About 39% of the genome is composed of the nitrogenous bases Guanine and Cytosine (Ming et al., 2008).

Several genomic studies on papaya have aimed to elucidate the basis of sex determination. Wang et al. (2012) sequenced the hermaphrodite-specific region of the Y^h chromosome (HSY) and its X counterpart, yielding an 8.1 megabase (Mb) HSY pseudomolecule and a 3.5 Mb sequence for the corresponding X region. This difference between the HSY and the X regions is mostly due to duplications and retrotransposon insertions. The authors identified 124 transcriptional units in total, including 50 functional pairs present in both the X and HSY. It was observed that ten HSY genes had functional homologs elsewhere in the papaya autosomal regions, suggesting the movement of genes into the HSY, whereas the X region had none.

Another study aimed to sequence the entire male-specific region of the Y (MSY) chromosome. This study revealed a small 0.4% sequence divergence between the MSY and HSY regions. As the authors sequenced the MSY and resequenced the Y regions of 24 wild males and the Y^h regions of 12 cultivated hermaphrodites, it was possible to identify one haplotype (MSY3), found only in wild dioecious populations from the north Pacific region of Costa Rica, possessing highly similar Y and Y^h sequences. These results support the hypothesis that hermaphrodite papaya is a product of human domestication (VanBuren et al., 2015).

When examining the 0.4% sequence differences between the MSY and HSY regions, Liao et al. (2017) found 1,887 InDels and 21,088 SNPs. A large male-specific retrotransposon insertion of 8,396 bp was also observed and used to develop two papaya male-specific markers. These markers were used to investigate male-to-hermaphrodite reversal mutants, which are crucial materials for validating candidate genes for sex determination in papaya.

Pulp softening during the ripening of papayas is considered one of the most important problems for this crop. In this sense, Fabi et al. (2014) performed a

study to identify and characterize genes related to pulp softening. Because this problem is mainly caused by the degradation of cell walls, the authors used gene expression profiling to analyze the correlations and co-expression networks of cell wall-related genes. The results of this study suggest that papaya pulp softening results from the interactions of multiple glycoside hydrolases. A gene encoding a polygalacturonase (cpPG1) enzyme appeared to play a central role in the network and was further studied. The transient expression of cpPG1 in papaya results in pulp softening and leaf necrosis in the absence of ethylene action and confirms its role in papaya fruit ripening.

The use of large-scale transcriptome sequencing has been used in several crops to identify genes that are differentially expressed under different treatments, plant tissues, stage of development, etc. In papaya, this approach was used to isolate key ripening-related genes and better the understanding of fruit ripening mechanisms. Three papaya samples were used to extract and sequence the total cellular RNA, with one sample treated with Ethylene (ETH), which induces the ripening process of climacteric fruits. The second sample was treated with 1-methylcyclopropene (1-MCP), which is an inhibitor of the ethylene signaling pathway and slows down the ripening process. The third sample was untreated papaya, acting as a control. In this study it was possible to illustrate that papaya fruit softening is a complex process with significant cell wall hydrolases, such as pectinases, cellulases, and hemicellulases involved in the process (Shen et al., 2017).

In another study, Shen et al. (2019) investigated the differential mechanisms of carotenoid biosynthesis in the yellow skin and red flesh of papaya. The skin color changed from green to yellow because of rapid chlorophyll degradation and the appearance of carotenoids such as lutein and β-carotene. Lycopene reached its maximum concentration in the flesh during fruit ripening and together with β-carotene and β-cryptoxanthin, caused the flesh to change from white to the orange-red color of papaya cv. ‘Daqing No.10’. Accumulation changes of carotenoids in the skin and flesh were consistent with the expression of carotenoid biosynthetic pathway genes analyzed using RNA-Seq and RT-qPCR.

A comparative proteomic analysis of the heterosis phenomenon in papaya roots has suggested that the hybrid possesses, compared with its parents, an optimization mechanism for protein synthesis that results in substantial

improvements in cellular energy efficiency and phenotypic performance (Vale et al., 2016). These results may contribute to the understanding of the foundations of the heterosis phenomenon in papaya.

3.1.3 MATERIAL AND METHODS

3.1.3.1 Plant materials

The Formosa elite lines of papaya Sekati and JS-12 were obtained from the UENF/CALIMAN germplasm bank and were cultivated in commercial fields at the Caliman Agrícola S.A. in Linhares-ES, Brazil. The Sekati line (originally from Malaysia) produces large fruits with excellent fruit firmness and moderate total soluble solid content. The JS-12 line (originally from the Embrapa - National Cassava & Fruits Research Center (CNPMF) of Brazil), on the other hand, presents high total soluble solid content and moderate fruit size and firmness[40,41].

3.1.3.2 Sequencing and variant identification

The genomic DNA was extracted from young leaves of the lines using a Plant Genomics DNA Extraction Kit YGP 100 - RBC (BioAmerica), following manufacturer instructions. The DNA concentration and quality were checked using a NanoDrop 2000 spectrophotometer (Thermo Scientific) and superfine resolution agarose gel (1%). The whole-genome libraries of the two lines were constructed using a Nextera library preparation kit (Illumina, Inc.), according to manufacturer instructions. Paired-end (35-251 bp) fragments of the samples were sequenced with a MiSeq platform at the Laboratory of Biotechnology - LBT of the Universidade Estadual do Norte Fluminense. After sequencing, the quality of reads was checked using FastQC (Andrews, 2010). The filtered reads were aligned with the reference genome (Ming et al., 2008) using Bowtie2 (Langmead & Salzberg, 2012). Discovery and filtration of variants were carried out using SAMtools v0.1.18 (H. Li et al., 2009). To facilitate visualization of the overall distribution of variants across the papaya chromosomes, the contigs and scaffolds of the reference genome, which is still a draft version, were associated with 10

papaya linkage groups (LGs) (Chen et al., 2007) and the LGs with a pachytene chromosome-based karyotype of papaya (Zhang et al., 2010).

3.1.3.3 Annotation of Single Nucleotide Polymorphisms and Insertion/Deletion Polymorphism

To predict the putative effects of DNA variants according to genomic location, the snpEff v4.3 program was used (Cingolani et al., 2012). To perform the analysis a *C. papaya* binary database file (.bin) was built in snpEff using the papaya reference genome in Fasta format and an annotation file in gff3 format, both downloaded from the PLAZA: Comparative Genomics In Plants.

A variant call format containing the SNPs and InDels was then annotated with the snpEff program. The variants were classified as genic and intergenic according to their genomic location. The variants in the intergenic regions are classified as Modifier impact and do not affect the coding regions of genes. Variants located in introns are classified as Modifier impact as well.

The variants placed in coding genic regions can generate three types of impacts, such as low, moderate, and high impact. Low impact variants (e.g. synonymous variant) are assumed to be mostly harmless or unlikely to change protein behavior, while a non-disruptive variant that might change protein effectiveness is considered of moderate impact (e.g. missense variant and inframe deletion). The variants with high impact (e.g. stop gained and frameshift variant) probably cause protein truncation or loss of function (Cingolani et al., 2012).

3.1.3.4 Identification of fruit ripening-related genes with linked variants

To identify ripening-related genes, we selected 48 genes isolated from a differential gene expression experiment during the fruit ripening process of papaya fruits[10]. The protein sequences of the 48 differentially expressed genes (DGEs) were used as queries to identify genes with related functions based on sequence similarity within the papaya genome. The Blastp tool available at Phytozome was used, and the ripening-related genes were selected with a minimum of 50% identity and E-value ≤1e-20. We removed from the list of the ripening-related genes those identified by Blastp with no expression during fruit development and

ripening of papaya (Lü et al., 2018) and the genes without variants. We also removed the variants farther than 40 Kb from the gene start/end.

3.1.4 RESULTS AND DISCUSSION

3.1.4.1 SNP and InDel discovery and chromosomal distribution

To obtain new genomic information for studies with papaya, we performed a whole-genome genotyping (WGG) of the elite lines Sekati and JS-12. These lines are both from the heterotic group Formosa, but possess contrasting agronomic and fruit quality traits with those that are widely used in breeding populations as donors of superior alleles at the UENF papaya breeding program. A total of 12,709,090 sequence reads (with lengths ranging from 31 to 251bp) were obtained from the Sekati and JS-12 lines. The Sekati sample generated 1.16 Gb of sequencing data (4,237,292 reads), while the JS-12 sample generated 2.4 Gb (8,471,798 reads).

Mapping of the clean reads, after removing low quality reads, against the papaya reference genome resulted in the identification of 28,451 SNPs and 1,982 InDels (1,061 insertions and 921 deletions). The average coverage of variants was ~3.12x and ~5.02x for the Sekati and JS-12 lines, respectively.

The SNPs were identified in all nine papaya chromosomes (Figure 1). The highest number of SNPs was observed on chromosome 4 (3,375 SNPs) and the lowest on chromosome 5 (1,751 SNPs). A total of 8,079 SNPs (28.4%) were identified in contigs and scaffolds that are not mapped for any papaya linkage group (Chen et al., 2007) and were attributed to an unknown chromosome.

The comparison of SNPs identified in the lines Sekati and the JS-12 revealed that they share about 78% (22,629) and 22% (5,822), respectively, of the genome-wide SNP alleles with the reference genome, which is the SunUp, a transgenic variety of the Solo heterotic group. The lines showed different levels of SNP similarities with the SunUp in all chromosomes. The Ch4 and Ch7 of Sekati share about 94.3% and 82.1% of similarity with the reference genome, respectively. On Ch6, Ch9, and Ch8 the similarity of Sekati with the reference is

the lowest, presenting about 68.3%, 70.66%, and 71.5% similarity, respectively. The remaining chromosomes of Sekati presented SNP similarities close to the genome-wide average. InDels were found in all nine papaya chromosomes (Figure 1). The highest InDel number was found to be 260 on Ch4, while the lowest was 112 on Ch9. A total of 529 InDels were observed in unknown chromosomes.

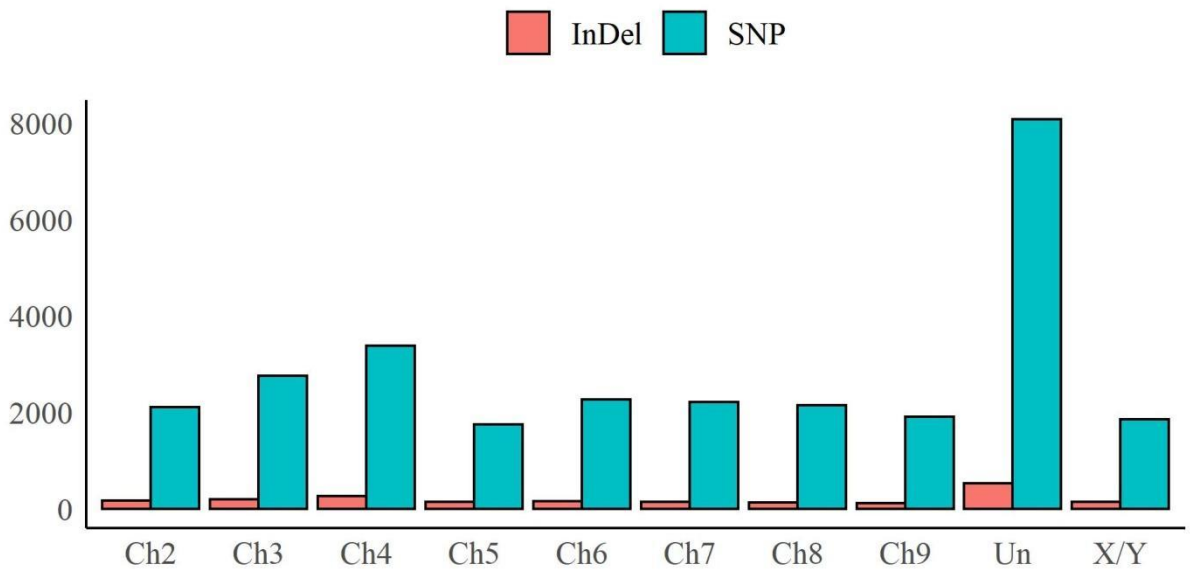

Figure 1. Single nucleotide polymorphism (SNP) and small insertion/deletion (InDel) distribution across papaya chromosomes. Ch: chromosome; Un: unknown chromosome; X/Y: sexual chromosomes; InDel: small insertion/deletion; SNP: single nucleotide polymorphism.

Based on nucleotide substitutions, the SNPs were classified as either transitions (purine-purine and pyrimidine-pyrimidine) or transversions (purine-pyrimidine and pyrimidine-purine). We found 20,199 transitions and 8,252 transversions, with a genome-wide transition to transversion ratio (Ts/Tv) of 2.45. Observation of SNPs in coding regions revealed that the nucleotide substitution frequency and the Ts/Tv ratio were higher at the third codon position (2.40), compared to the second (1.96) and first (1.83) codon positions (Table 1).

3.1.4.2 Functional classification of DNA variants

To annotate and predict the effects of DNA variants in the papaya genome we used the snpEff v4.3 program. Of the 30,433 total variants annotated by the

snpEff, a total of 58,498 effects based on genomic position were predicted. The higher number of effects compared with the number of variants results from the ability of one specific variant to affect multiple genes (e.g. a variant can be downstream from one gene and upstream from another gene).

Table 1. Frequency and type of nucleotide substitutions at codon sites.

Codon Position	Transitions (Ts)	Transversions (Tv)	Ts/Tv Ratio
3rd	755	315	2.40
2nd	350	179	1.96
1st	286	156	1.83
Total	1,391	650	2.14

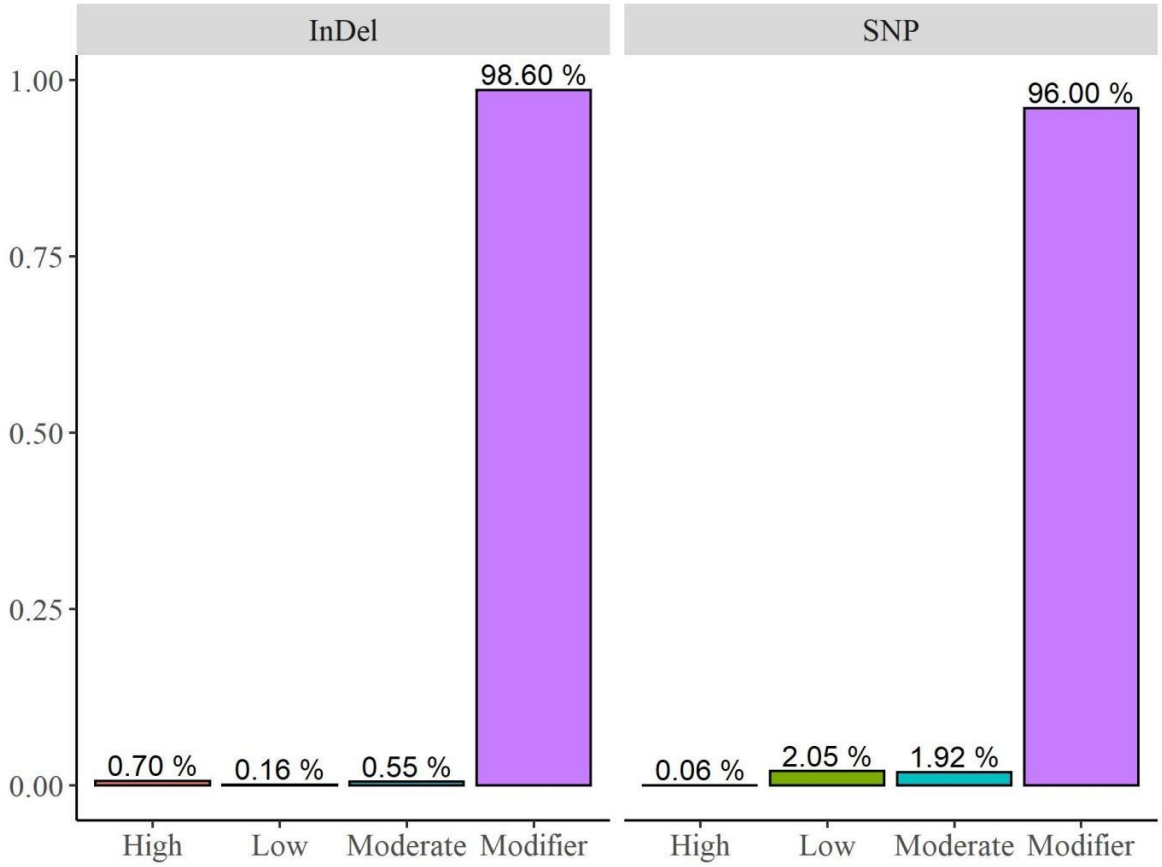

Figure 2. Number of small insertion/deletion (InDel) and single nucleotide polymorphism (SNP) effects by impact.

The SNPs and InDels caused a total of 54,100 and 4,398 effects (Figure 2), respectively. The effects of the variants were classified into four categories: modifier (56,380), low (1117), moderate (1062), and high (63) impact. Only 4% and 1.4% of the SNP and InDel effects, respectively, were placed in coding regions.

High impact variants had a direct impact on gene functionality. A total of 32 and 31 high impact variants were observed for SNPs and InDels, respectively. The most common effects caused by high impact SNPs are stop codon lost and stop codon gain (Figure 3a), which may lead to high levels of functional consequences. Meanwhile, high impact InDels mainly caused disruption of the translational reading frame and may result in abnormal protein products with an incorrect amino acid sequence. Moderate impact SNPs caused a change in one amino acid due to a non-synonymous substitution (Figure 3b). The InDels caused four types of effects in coding regions classified as moderate impact (Figure 3b). Low impact SNPs mainly consisted of synonymous substitutions in which no change of amino acid is observed (Figure 3c). The remaining effects were predicted in non-coding regions and they were classified as modifier impact (Figure 3d).

3.1.4.3 Identification of fruit ripening-related genes with linked variants

We selected 48 DGEs during the papaya fruit ripening process determined by RNAseq, including 20 cell wall-related genes (CW), 13 chlorophyll and carotenoid metabolism-related genes (CCM), four proteinases and their inhibitors (PROT), six plant hormone signal transduction pathway genes (PH), four transcription factors (TF), and one senescence-associated gene (SEN) (Shen et al., 2017). These genes were used as Blastp queries to identify other ripening-related genes within the papaya genome. This search resulted in the identification of 143 other genes that are potentially involved in the fruit ripening process due to sequence similarity.

From 191 selected ripening-related genes (48 DGEs and 143 identified by BLASTp), a total of 106 genes were found to be associated with 460 variants (438 SNPs and 22 InDels) (Table 1A). The 106 ripening-related genes (36 DGEs and 70 identified by BLASTp) with linked variants were classified into five categories: cell wall-related genes (55), chlorophyll and carotenoid metabolism-related genes

(10), proteinases and their inhibitors (13), plant hormone signal transduction pathway genes (11), and transcription factors (17).

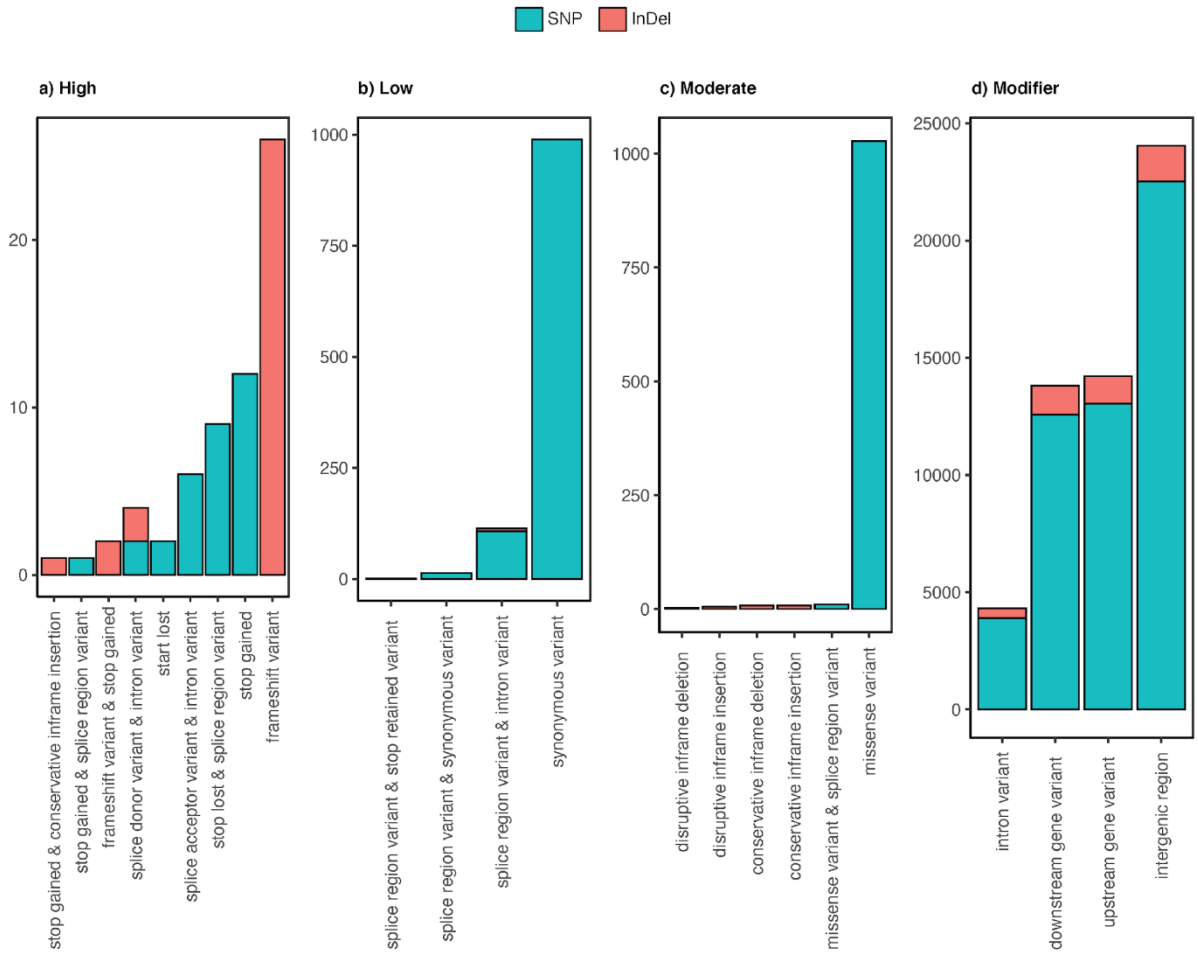

Figure 3. Distribution of single nucleotide polymorphism (SNP) and small insertion/deletion (InDel) effects in the sub-classification of High (Figure 3a), Moderate (Figure 3b), Low (Figure 3b), and Modifier (Figure 3d) impacts.

Most of these variants are located in the flanking regions of the RRGs, including 206 variants in intergenic regions no farther than 40 Kb from the gene and 196 variants downstream/upstream of the genes. Only 58 variants are located inside the genes, including 36 in introns and 22 in exons. The exonic variants are separated as synonymous and missense variants (Table 2).

Table 2. Fruit ripening-related genes with Low and Moderate SNP impacts, gene function and type, chromosome (Chr), variant identification (id), variant effect, and amino acid (AA) change.

Gene id	Gene description	Gene function	Gene type	Chr	SNP id	Variant effect	AA change
evm.model.supercontig_43.43	Hydroxymethylbilane synthase	CCM	DGE	Ch6	SNP_43.43[A/G]569749	SV	Ser384Ser
evm.model.supercontig_92.51	Magnesium-chelatase subunit chlh	CCM	DGE	Ch8	SNP_92.51[G/A]470716	SV	Leu675Leu
evm.model.supercontig_43.43	Hydroxymethylbilane synthase	CCM	DGE	Ch6	SNP_43.43[G/C]569757	MV	Cys387Ser
evm.model.supercontig_151.20	Glycosyl hydrolase 9B13	CW	BLAST	Ch5	SNP_151.20[G/C]175639	SV	Ser435Ser
evm.model.supercontig_106.54	Glycosyl hydrolase family protein	CW	DGE	Ch6	SNP_106.54[C/T]643444 SNP_106.54[T/G]643459 SNP_106.54[T/G]643531	SV SV SV	Pro679Pro Val674Val Leu650Leu
evm.model.supercontig_29.125	Glycosyl hydrolases family 32 protein	CW	BLAST	Ch4	SNP_29.125[G/A]1173309	SV	Asp379Asp
evm.model.supercontig_508.5	Major facilitator superfamily protein	CW	BLAST	Ch3	SNP_508.5[C/T]19647	SV	Thr486Thr

MV: missense variant; SV: synonymous variant; CW: cell wall-related genes; CCM: chlorophyll and carotenoid metabolism-related genes; PH: plant hormone signal transduction pathway genes; TF: transcription factor. SNP id: SNP_250.6_27755[G-C], colors in this example represent: orange: the papaya contig or supercontig; yellow: the gene in that contig or supercontig is linked to the SNP; green: the linear position of SNP in that contig or supercontig; and blue: the Reference and Alternative SNP alleles.

Table 2 – Cont.

Gene id	Gene description	Gene function	Gene type	Chr	SNP id	Variant effect	AA change
evm.model.supercontig_25.194	Pectin methylesterase 3	CW	BLAST	Ch4	SNP_25.194[G/A]2039032 SNP_25.194[G/A]2039146	SV SV	Tyr61Tyr Ile23Ile
evm.model.supercontig_82.65	Sucrose synthase 4	CW	DGE	Un	SNP_82.65[A/C]1098666	SV	Ala803Ala
evm.model.supercontig_93.34	Beta-galactosidase 12	CW	DGE	Ch7	SNP_93.34[T/C]879460	MV	Val624Ala
evm.TU.contig_32583.1	Beta-galactosidase 3	CW	BLAST	Un	SNP_32583.1[C/T]2456	MV	Gly233Ser
evm.model.supercontig_119.14	Expansin A1	CW	BLAST	Ch6	SNP_119.14[G/A]71404	MV	Arg225His
evm.model.supercontig_14.96	Expansin A5	CW	BLAST	X/Y	SNP_14.96[C/T]1096729	MV	Ser17Asn
evm.model.supercontig_3.313	Glycosyl hydrolases family 32 protein	CW	DGE	Ch9	SNP_3.313[G/C]2180909	MV	Glu637Gln
evm.model.supercontig_25.194	Pectin methylesterase 3	CW	BLAST	Ch4	SNP_25.194[G/A]2039051	MV	Ala55Val
evm.model.supercontig_151.32	Ethylene response sensor 1	PH	DGE	Ch5	SNP_151.32[A/T]288630	SV	Val331Val

MV: missense variant; SV: synonymous variant; CW: cell wall-related genes; CCM: chlorophyll and carotenoid metabolism-related genes; PH: plant hormone signal transduction pathway genes; TF: transcription factor. SNP id: SNP_250.6_27755[G-C], colors in this example represents: orange: the papaya contig or supercontig; yellow: the gene in that contig or supercontig is linked to the SNP; green: the linear position of SNP in that contig or supercontig; and blue: the Reference and Alternative SNP alleles.

Table 2 – Cont.

Gene id	Gene description	Gene function	Gene type	Chr	SNP id	Variant effect	AA change
evm.model.supercontig_6.74	Auxin-responsive GH3 family protein	PH	BLAST	Ch4	SNP_6.74[G/A]571741	MV	Ala6Val
evm.model.supercontig_152.36	WRKY family transcription factor	TF	BLAST	X/Y	SNP_152.36[C/T]269630	SV	Ser551Ser
evm.model.supercontig_9.36	WRKY DNA-binding protein 48	TF	BLAST	Ch5	SNP_9.36[A/T]151586	MV	Gln308His

MV: missense variant; SV: synonymous variant; CW: cell wall-related genes; CCM: chlorophyll and carotenoid metabolism-related genes; PH: plant hormone signal transduction pathway genes; TF: transcription factor. SNP id: SNP_250.6_27755[G-C], colors in this example represents: orange: the papaya contig or supercontig; yellow: the gene in that contig or supercontig is linked to the SNP; green: the linear position of SNP in that contig or supercontig; and blue: the Reference and Alternative SNP alleles.

The frequency of SNPs and the Ts/Tv ratio was higher at the third codon position compared with the second and first codon positions (Table 1), revealing a trend of genomic conservation at codon sites during evolution. This trend was also observed in SNPs identified in Expressed Sequence Tags (ESTs) from *Solanum lycopersicum* and *S. habrochiates* (Bhardwaj et al., 2016).

SNPs are known to be associated with many quantitative trait loci in plants (Argyris et al., 2017; Celik et al., 2017; Montero-Pau et al., 2017; Nuñez-Lillo et al., 2019; Pootakham et al., 2015) and an individual SNP can have a large impact on the phenotype (Schreiber et al., 2014; Tzuri et al., 2015). We found 2,180 SNPs located in coding regions and 26,271 in noncoding regions of the papaya genome. Although most SNPs are not located inside genes, their abundance and robustness make them an important and useful source of DNA variation for papaya breeding programs working to develop superior cultivars. InDels also play important roles in the phenotypic variation observed between individuals of a species. In papaya, a dinucleotide insertion mutation in the gene encoding the enzyme lycopene β-cyclase (CpCYC-b) causes the phenotypic variation of red and yellow flesh (Blas et al., 2010). When found in coding regions, InDels generally disrupt the translational reading frame (frameshift variant), except when the mutation is a multiple of three nucleotides. In this study, we identified 62 InDels located in coding regions, 28 of which caused disruptions of the translational reading frame.

Fruit quality is one of the most important features pursued by papaya breeding programs, especially the selection of genotypes that maintain fruit firmness for longer periods, resulting in longer shelf-life and a decrease in post-harvest losses. Studies at the gene expression level were developed to isolate the key genes affecting the fruit ripening process and fruit softening of papaya (Fabi et al., 2012, 2014; Shen et al., 2017). However, these studies analyzed only one genotype at a time and did not consider the variation within DNA sequences among different papaya genotypes.

During the ripening process of climacteric fruits such as papaya and peach, a positive feedback loop regulated by the NAC transcription factor is thought to control the ethylene synthesis. This mechanism is observed in species that lack recent whole-genome duplication (WGD). On the other hand, climacteric fruit species with recent WGD, such as tomato, pear, and apple, appear to have

evolved a MADS-type transcription factor positive feedback loop controlling ripening (Lü et al., 2018). Fruit softening in papaya is mainly caused by the degradation of primary cell wall polymers. Several cell wall-degrading enzymes act cooperatively in a coordinated process to degrade the cellulose-hemicellulose matrix which is embedded in a structurally heterogeneous mixture of pectin (Gapper et al., 2013). While on the one hand ethylene promotes fast fruit softening, on the other it is also thought to improve the rate of sugar synthesis, transport, and degradation during the ripening of papaya. Several genes related to sugar metabolism are upregulated in response to ethylene during the ripening process (Shen et al., 2017). Plant hormones also play important roles in controlling several processes of growth and development in plants. Besides the importance of ethylene for the fruit ripening process, other types of plant hormones can take place synergistically or antagonistically with respect to ethylene action during the ripening of climacteric fruits. In addition, one of the major physiological changes observed during the ripening of papaya is a rapid change in color (Fabi et al., 2012). This results from the rapid degradation of chlorophyll and the appearance of carotenoids such as lutein and β-carotene (Shen et al., 2019). Other genes that are involved in fruit softening include the class of protease enzymes. Studies have shown that some proteases have increased expression during the ripening process of papaya (Shen et al., 2017) and tomatoes (Wang et al., 2017).

The availability of SNPs and InDels strongly associated with ripening-related genes of papaya is essential for developing studies of diversity, genetic mapping, and the application of marker-assisted selection. Thus, we searched for DNA variants that are linked with ripening-related genes that are up or down-regulated in response to exogenous ethylene (Shen et al., 2017) and genes identified using BLASTp. A total of 106 genes with at least one associated variant, either inside or in the flanking region of the gene, were identified (Table 1A). The association between an SNP and InDel with a trait of interest can be accessed through linkage disequilibrium analysis (Flint-Garcia et al., 2003) using, for example, quantitative trait (QTL) analysis. Further analysis will examine the genotype–phenotype association related to fruit ripening traits in a segregant population derived from the cross between the Sekati and JS-12 lines. It is expected that the presence of alleles for these fruit ripening-related genes in the papaya germplasm and breeding populations can contribute to observed

differences in the fruit firmness and TSS content among papaya genotypes. The association of genotypic alleles with a trait of interest points to a genomic region where one or more genes may be affecting the phenotype. The effective application of MAS in breeding programs requires the identification and validation of the candidate genes through functional analysis. After performing all of these identification and validation steps, DNA markers based on PCR, such as the low-cost technique called single nucleotide amplified polymorphism (SNAP) (Drenkard et al., 2000) or the real-time fluorescence-tagged probes technologies, such as TaqMan, Kompetitive allele specific PCR (KASP), or rhAmp (Broccanello et al., 2018), will be developed to apply marker-assisted selection and direct gene editing studies in papaya breeding programs.

3.1.5 CONCLUSION

Comparison of the genomic sequences of two papaya elite lines (Sekati and JS-12) revealed a considerable number of DNA polymorphisms. A total of 30,433 variants (28,451 SNPs and 1,982 InDels) were identified in all the papaya chromosomes. Most variants were enountered in non-coding regions of the papaya genome. A total of 460 variants (438 SNPs and 22 InDels) were found to be associated with 106 ripening-related genes. The availability of SNPs and InDels strongly associated with ripening-related genes of papaya is essential for the development of studies on diversity and genetic mapping. With the effective knowledge of the alleles on the phenotype, DNA markers based on PCR can be developed to apply marker-assisted selection and to direct gene editing studies aiming to improve fruit quality traits in papaya breeding programs.

3.2 DEVELOPMENT OF PCR MARKERS FOR FRUIT FIRMNESS-RELATED GENES IN PAPAYA

3.2.1 INTRODUCTION

Papaya (*Carica Papaya*) belongs to the small Caricaceae family, which includes 35 species divided into six genera mainly native to Central America (Badillo, 2000). *C. papaya* is the only commercially important species within this family and its unique features make papaya highly popular globally. The fruits of papaya have a distinct aroma, flavor, and texture, as well as nutritional benefits (Fabi et al., 2010).

The cultivation of papaya occurs in the tropical and subtropical regions of the world. In 2020, Brazil was the third largest producer and exporter of papaya (FAOSTAT, 2022). Papaya production entails considerable risks, including the high cost of establishing a crop, large numbers of spray applications for pest and disease control, plant losses and decreased yields resulting from viruses, and problems in the post-harvest chain. In terms of this last point, papaya suffers from fast fruit ripening and is easily injured, which reduces fruit shelf life and increases losses (Shen et al., 2017).

To mitigate such losses, papaya breeding programs have focused on developing superior cultivars that are resistant/tolerant to the most common diseases and have improved fruit quality features such as firmness and sweetness. Fruits with higher firmness have a longer shelf life, while sweetness is

a desired feature for consumers. However, increasing fruit firmness and sweetness simultaneously is a formidable task when using classical breeding methods because they are complex inheritance traits, meaning several genes act cooperatively to form the phenotype. In addition, studies have shown that fruit firmness and sweetness are negatively correlated (Quintal, 2009), such that increasing the mean of one trait may decrease that of the other trait. Therefore, during the papaya breeding process, breeders must select genotypes with a balance between the two traits.

Understanding the key genes regulating fruit ripening molecular mechanisms is crucial for improving fruit quality aspects simultaneously in papaya breeding programs. In addition, the availability of DNA variants associated with key genes can enhance understanding of their association with the phenotype. SNPs are the most abundant form of DNA polymorphisms and can be quickly identified through next-generation sequencing (NGS). This powerful source of genetic variability can significantly increase advances in breeding programs. The great availability of SNPs across the genome makes it possible to identify polymorphism tightly associated with specific genes. Such information could be converted into PCR-based markers to study gene variability, genetic diversity, and mapping and to apply marker-assisted selection (MAS) in papaya breeding programs (Bohry et al., 2021). The objective of this study was to develop and validate PCR-based markers to study the genetic variability of fruit firmness-related genes (FFG) in papaya.

3.2.2 LITERATURE REVIEW

3.2.2.1 Use of molecular approaches for plant breeding

Molecular markers allow the analysis of individuals based on their DNA, distinguishing genetically different individuals (Borém, 2009). These markers generally do not represent the genes involved in the expression of a particular trait but act as markers of the observed differences (Collard et al., 2005). However,

due to advances in sequencing technologies, it has been possible to develop markers located within or tightly linked with genes of interest (Kim et al., 2016).

In plant breeding the molecular markers are useful for several fields of study, such as evolution and phylogeny; the investigation of heterosis; genetic diversity assessment; genetic mapping; genome-wide association study (GWAS) and genome-wide selection (GWS); and marker-assisted selection (MAS) (Nadeem et al., 2018). One of the main uses of DNA markers in research in agriculture has been the construction of linkage maps in several plant species. Several markers were previously used for the construction of linkage maps and studies on plants, such as those based on hybridization RFLP (restriction fragment length polymorphism) and those based on PCR, such as RAPD (random amplified polymorphic DNA); SSR (simple sequence repeats or microsatellites); and SCAR (sequence characterized amplified region). Recently the class of SNP markers, which are based on sequencing and DNA chips and characterized as the most abundant form of polymorphism present in both plant and animal genomes (Khlestkina & Salina, 2006), has been used. In addition, the genotyping of thousands of SNPs can be almost fully automated, saving time and effort for obtaining data in breeding programs (Steemers & Gunderson, 2007).

SNP markers have great potential in the construction of highly saturated linkage maps. According to Brookes (1999), an SNP is a single base pair position in the genomic DNA, where alternate sequences (alleles) exist in normal individuals in any population and the frequency of the alternative allele is greater than 1%. SNPs can be classified by the type of altered nucleotide, such as transitions (C / T or G / A) or transversions (A / C, A / T, G / C, G / T).

Linkage maps based on SNPs have been constructed for various plant species, allowing the mapping of genomic regions associated with traits of interest for plant breeding. Sun et al. (2015) developed a high-density genetic map with 3,441 SNP markers from 297 individuals for apples. By using this map, twelve significant QTLs linked to the control of fruit weight, fruit firmness, sugar content, and fruit acidity were mapped to seven linkage groups. Montero-Pau et al. (2017) reported the construction of a high-density linkage map based on 7,718 SNPs identified in a RIL population. This map covers 2,817.6 cM and is the first saturated map for zucchini and was later used for QTL analysis of fruit-related traits. For sacred lotus (*Nelumbo nucifera*), Liu et al. (2016) report the construction

of a high-density genetic map consisting of 8,971 SNP markers sorted into 8 linkage groups (LG) that spanned 581.3 cM. This map will facilitate genetic mapping of important traits in this crop.

In papaya, the linkage map with the highest number of co-segregant markers was developed and published by Chen et al. (2007), who grouped over 700 markers of the microsatellite type in 10 linkage groups. Using SNP marker technology, the UENF breeding team has directed efforts toward the development of a high-density genetic map. However, although this map is already more saturated than the map developed by Chen et al. (2007), further efforts have been made to increase the map saturation and thus the potential for mapping agronomic traits of importance for papaya breeding.

Modern molecular genetic techniques and powerful statistical methods are required to understand the genetic basis of quantitative traits. This approach allows the analysis of both the simply inherited and quantitative traits and identification of the individual genes controlling the traits of interest. Evaluating the contribution of QTL to the phenotype is possible and can accelerate the selection and genetic advances (Dhingani et al., 2015).

The analysis of QTL has been used in several crop species depending on the quantitative trait of interest, such as for seed fiber in *Brassica napus* L (L. Liu et al., 2013), fruit quality traits in peach (Martínez-García et al., 2013) and apple (Sun et al., 2015), yield and resistance against Mediterranean corn borer in maize (Jiménez-Galindo et al., 2017), agronomic traits in adzuki bean (*Vigna angularis*) (Li et al., 2017), and oil palm (*Elaeis guineensis*) (Pootakham et al., 2015).

Despite the power of QTL analysis and the great importance of the papaya, its use in studies aiming to detect QTL for traits of importance in this fruit crop has been scarce. Only two works are observed in the literature for the identification of QTL in papaya. Blas et al. (2012) identified 14 QTLs for the characteristics of mass, length, diameter, and shape of papaya fruits. The low precision in detecting QTL in this study is because of the low density of the genetic map used, which consisted of 11 dominant and 46 co-dominant markers and spanned a total length of 672.1 cM. Thus, using high-density genetic maps could increase the precision for mapping QTL in papaya.

More recently, a linkage map based on 219 single nucleotide polymorphisms (SNP) was developed (Nantawan et al., 2019). Although this map

was based on SNP markers, which are very abundant and quickly identified through Next Generation Sequencing (NGS), the great distortion of the expected marker segregation observed in F_2 (1:2:1) significantly decreased the map resolution. Still, a total of 21 QTLs for fruit quality traits were detected using this map and will enable candidate gene isolation and the development of marker-assisted selection strategies.

Understanding the molecular basis of essential biological phenomena in plants is crucial for effective conservation, management, and efficient utilization of plant genetic resources (Mondini et al., 2009). The adequate knowledge of existing genetic variability is of great importance for breeders, and forms the basis for a successful breeding program. The characterization of plant genetic resources can be done through morphological, biochemical, and molecular markers. Molecular markers have advantages over morphological and biochemical markers because they allow the analysis of the diversity within the DNA of the species. Different markers have different genetic qualities and in the case of the SNP markers, powerful results can be achieved. As the SNPs are the most abundant form of polymorphism, they can more precisely access plant diversity.

When studying the genetic diversity of genes of interest for cassava breeding, Kawuki et al. (2009) verified that the number of SNPs per gene fragment varied from one in the Ethylene response factor and allergenic-related protein Pt2L4 to five in the catalase gene fragment. The SNPs identified in target genes were used to access the diversity of cassava genotypes in this study. In another study, the SNPs were used to evaluate the diversity and genetic structure of the largest cassava germplasm bank in Brazil. The structure and diversity of 1,280 accessions were accessed using 402 SNP markers. The results of this study may benefit cassava germplasm conservation programs and contribute to the maximization of genetic gains in breeding programs (Oliveira et al., 2014).

According to Xu & Crouch (2008), justifications for the development and use of MAS in plant breeding fall within four broad areas that are relevant to almost all crops: a) characteristics that are difficult to conduct through conventional phenotypic selection; b) characteristics whose selection depends on specific environments or stages of development that influence the expression of the target phenotype; c) maintenance of recessive alleles during backcrossing or to accelerate breeding through backcrossing in general; and d) pyramiding of

multiple monogenic characteristics (such as pest and disease resistance or quality characteristics) or multiple QTL for a single target characteristic with complex inheritance (such as drought tolerance or other adaptive characteristics).

According to Kim et al. (2016), the DNA markers associated with characteristics of interest that will be used in marker-assisted selection should have desirable characteristics, such as using simple methods for genotyping, with basic requirements at low cost and high reproducibility.

Thus, the methodology developed by Ye et al. (2001) stands out. This is a robust methodology for the design of allele-specific primers, named tetra-primer amplification refractory mutation system (ARMS) - polymerase chain reaction (PCR). The tetra-primer ARMS-PCR allows the identification of the different alleles of an SNP via conventional polymerase chain reaction (PCR), with ease, speed, and at low cost. The design of the allele-specific primers is limited to the SNP locus and if it is a Guanine/Cytosine rich region, the primers will have a high temperature of melting (Tm), making it hard to design complementary primers with compatible Tm, resulting in primers that may not work well. Primers that form strong secondary structures, such as self-dimer, cross-dimer, and hairpin, are also undesirable because they can drastically reduce PCR efficiency.

Other types of markers are available to genotype SNPs, such as the CAPS markers, which use a common primer to amplify the SNP region followed by cutting the PCR product with a specific restriction enzyme (RE). The original product will be cleaved if the RE site exists at the SNP site, resulting in two small products (allele 1). However, the substitution of one nucleotide at the SNP site must prevent the RE to recognize its specific site, resulting in a non-cleaved product (allele 2).

A more recent and robust SNP genotyping method was developed by Beltz et al. (2018). It makes use of dual enzyme chemistry, is called rhAmp®, and is based on RNase H2-dependent PCR (rhPCR), which provides high signal and specificity for SNP analysis. rhAmp SNP genotyping combines a unique two-enzyme system with 3’ end blocked DNA-RNA hybrid primers to interrogate SNP loci. Activation of the blocked primers occurs upon hybridization to its perfectly matched target. A thermostable hot-start RNase H2 cleaves the primer at exactly the 5’ position of the ribose sugar, releasing the blocking group and allowing primer extension. The rhAmp SNP genotyping is a cost-effective method and

although it was just recently released for the scientific community, this method has demonstrated improved efficiency and a lower cost per reaction compared to the Kompetitive allele-specific PCR (KASP) and TaqMan SNP genotyping assays, which are consolidated methods also based on real-time PCR (Broccanello et al., 2018).

Markers based on other types of polymorphism are useful for marker-assisted selection in crop species as well. In rice, the successful application of marker-assisted breeding to introgress submergence tolerant QTL into the genetic background of BR22, an elite variety, was demonstrated. This strategy was mediated by using SSR and small insertion/deletion (InDel) markers through marker-assisted backcrossing (Amin et al., 2018).

In papaya, studies aiming at the identification of genomic regions controlling both simple traits and quantitative traits remain scarce. Thus, there is a need for the development of new markers associated with the characteristics of interest. In this sense, Vidal et al. (2014) developed and validated an atlas of SSR markers to aid papaya SAM programs. This atlas comprises 160,318 SSRs, of which 21,231 were in gene regions (Ex.: within exons, junctions of exons-introns, or introns). The authors identified 300 genes that are potentially involved in fruit ripening, with about 790 SSRs distributed among them. This study represents a powerful tool to assist and improve MAS programs in papaya breeding and population genetics studies.

3.2.3 MATERIAL AND METHODS

3.2.3.1 Plant materials and DNA extraction

For marker validation, we used the papaya elite lines Sekati and JS-12 and F_2 segregant individuals derived from the cross between the two lines. The genomic DNA was extracted from young leaves taken from one individual of each genotype using a Plant Genomic DNA Extraction Kit YGP 100—RBC (BioAmerica), following the manufacturer's instructions. The DNA concentration

and quality were checked using a NanoDrop 2000 spectrophotometer (Thermo Scientific) and superfine resolution agarose gel (1%).

3.2.3.2 Design of PCR primers for tetra-primer ARMS-PCR

In this study, we developed DNA markers using a simple and economical SNP genotyping method named the tetra-primer amplification refractory mutation system–PCR (ARMS–PCR) (Ye et al. 2001). This technique uses two pairs of primers to genotype one SNP. Each pair of primers has an allele-specific primer (named inner primer) and one reverse primer (named outer primer). Allele specificity is conferred by a mismatch between the 3′ -terminal base of an inner primer and the template. To enhance allelic specificity, a second deliberate mismatch at position –2 from the 3′ -terminus is also incorporated in the inner primers.

We selected 13 SNPs tightly linked with fruit firmness-related genes (FFG) for marker development. Fruit firmness-related genes have been isolated in papaya through RNA sequencing (Shen et al., 2017), while the association of these genes with hundreds of DNA variants was later performed (Bohry et al., 2021).

To design tetra-primer ARMS–PCR markers we downloaded genomic sequences of papaya from the National Center for Biotechnology Information (NCBI). A flanking sequence of 1,000 bases was submitted to PRIMER1 for primer design (http://primer1.soton.ac.uk/primer1.html). The parameters were defined as follows: primer optimal size (20 bases), minimum (18 bases) and maximum (26 bases); optimal (300 bases), maximum (500 bases) and minimum (100 bases) size of PCR amplicon; maximum (3x) and minimum (1.4x) relative size difference between the two PCR amplicons; and optimum annealing temperature (65 °C), maximum (80 °C) and minimum (40 °C). The other parameters were set to defaults. The NetPrimer tool (https://www.premierbiosoft.com/netprimer/) was used to verify the existence of intra and intermolecular structures between the primers, where the best combinations of primers were chosen to form the two sets of specific markers for the two alleles of each SNP.

3.2.3.3 Optimization of singleplex primers

The primer pairs were first tested in a single sample containing bulk DNA (BK) from the lines Sekati and JS-12. Each 13 µl reaction contained the following reagents and concentrations: 1X PCR Buffer (Mg^{2+} free); 0.29 mM each of dNTP; 1.92 mM of Mg^{2+}; 0.08 U/µl of Taq DNA Polymerase; 0.19 µM each of Forward (F) and Reverse (R) primers; 2.31 ng μl^{-1} of DNA template; and the final volume (13 µl) was filled with ultra-pure water.

PCR parameters were one initial cycle of predenaturation at 94°C for 5 min, followed by 30 cycles of denaturation at 94°C for 30 s, annealing for 30 s (Ta according to primer pairs), extension at 72°C for 60 s, and a final extension cycle at 72°C for 10 min. A temperature gradient was performed to find the optimal annealing temperature (Ta). The primers with low-intensity bands after temperature gradient at 30 cycles were tested with 50% more Mg^{2+} (1,5X) and more PCR cycles.

Once we determined the best amplification parameters for each primer pair, we tested the primers using the DNA samples of Sekati, JS-12, and BK. These samples served as controls to check for amplification specificity. Some primers did not present the expected pattern with non-specific amplification, PCR artifacts, and low-intensity amplification. Changes to the PCR parameters Ta, Mg^{2+}, and cycles were made to avoid the above-mentioned problems and have consistent amplification results.

PCR amplicons were stained with six µl of GelRed™ (1:2000) and Blue Juice in the same proportion. Electrophoresis was performed using a 3% agarose gel. A 100bp Ladder, Ready-To-Use (Sinapse Inc) was used to identify amplicon size. The gel image was taken with the MiniBis Pro photo documentation system (Bio-imaging Systems).

3.2.3.4 Optimization of duplex amplification

Duplex PCR is the use of two sets of primers to produce different amplicons in a single PCR reaction. As the reaction involves four primers, it becomes more complex and requires the optimization of various PCR conditions (Pesik et al., 2017). After analyzing the best PCR parameters for singleplex genotyping of 13 SNPs in papaya, we tested primers in a duplex PCR reaction.

The original tetra-primer ARMS–PCR technique only involves a single PCR followed by gel electrophoresis to genotype one SNP. However, the two allele-specific primers of the target SNP may require different conditions to amplify correctly. The conditions for both allele-specific primers must be very similar for efficient amplification in a single reaction.

The 26 primer pairs were divided into two groups, the Sekati specific (s) alleles and the JS-12 specific (s) alleles. Primers within each group with similar amplification conditions were tested in duplex PCR. We used the same three samples (Sekati, JS-12, and BK) for duplex validation. PCR parameters were the best parameters for a singleplex reaction. If two primers had small differences in amplification parameter conditions, both PCR conditions were tested. After the first round of tests, we identified the candidate combination of primer pairs. Then, we adjusted PCR conditions to increase the quality of bands of the two alleles in the gel. Duplex PCR was tested using more PCR cycles, different Ta, and Mg^{2+} and primer concentrations.

3.2.4 RESULTS AND DISCUSSION

3.2.4.1 Marker development for fruit firmness-related genes

To develop PCR-based molecular markers to aid papaya breeding programs we selected 13 SNPs, which are associated with 13 fruit firmness-related genes (FFG) (Table 01). These genes were isolated through gene expression analysis (Shen et al., 2017) and have been previously associated with a great number of DNA variants, such as SNPs and InDels (Bohry et al., 2021). For each FFG, we selected one SNP for marker development. Because in many cases more than one SNP was available for one specific gene, we sided with SNPs inside the target gene or the closest SNP in both up or downstream direction of the gene.

The 13 SNPs were identified in the papaya lines Sekati and JS-12. The lines are homozygous and contrasting for all SNP loci. Four SNPs are located inside the exons of the target genes, with two SNPs causing missense mutation

and the other two causing synonymous mutation (Table 01). The remaining nine SNPs are located in the flanking region of the target genes, being no farther than 10 Kb away from the gene start/end (Bohry et al., 2021).

Based on the selected SNPs, we successfully designed PCR-based markers to amplify all the 26 alleles of the selected SNPs (Tables 2 and 3). Amplicon sizes varied from 108 to 499 bp and primer size from 18 to 26 bases.

3.2.4.2 Optimization of singleplex primers

After performing temperature gradients and adjustments in the amplification parameters, the markers were tested in three samples with known genotypes to check for amplification specificity. Figure 1(a-m) and Figure 2(a-m) show the best amplification for markers specific to the Sekati and JS-12 lines, respectively. The lines Sekati (genotype: ss) and JS-12 (genotype: jj) are homozygous and contrasting for all 13 SNPs and the BK sample has an equal mixture of DNA from both lines, acting as a source of both alleles. The expected result was the amplification of the Sekati-specific (s) primers only in the Sekati and BK samples. In the same way, the JS-12-specific (j) primers were expected to amplify only in the JS-12 and BK sample.

Table 1. Selected fruit firmness-related genes, gene category, SNP identification (id), and SNP effects.

N.	Gene Definition	Category	SNP id	SNP Effects
1	Polygalacturonase	CW	SNP_250.6_27755[G-C]	Inter
2	Beta-D-xylosidase	CW	SNP_106.54_643459[T-G]	SV
3	Beta-galactosidase	CW	SNP_93.34_879460[T-C]	MV
4	Ethylene receptor	PH	SNP_151.32_288630[A-T]	SV
5	UDP -galactose transporter	CW	SNP_183.26_249112[G-C]	Inter
6	NAC domain protein	TF	SNP_165.12_56660[G-A]	Inter
7	Beta glucosidase	CW	SNP_80.123_1055961[G-A]	Inter
8	Subtilisin-like serine endopeptidase family protein	PROT	SNP_12.94_814429[T-C]	Inter
9	Ethylene-insensitive 4-like protein	PH	SNP_2388.2_8708[T-C]	Inter
10	Cysteine proteinases superfamily protein	PROT	SNP_4.28_960396[T-C]	Inter
11	Pectate lyase	CW	SNP_5.270_2394357[T-G]	Inter
12	Endoglucanase 8-like	CW	SNP_2.231_3047903[T-C]	Intron
13	Beta-fructofuranosidase	CW	SNP_3.313_2180909[G-C]	MV

CW: Cell wall-related genes; PROT: proteinase genes; PH: plant hormone and signal transduction pathway genes; TF: transcription factors; MV: missense variant; SV: synonymous variant; Inter: Intergenic region. SNP id: SNP_250.6_27755[G-C], colors in this example represent: orange - the papaya contig or supercontig; yellow - the gene in that contig or supercontig to which the SNP is linked; green - the linear position of SNP in that contig or supercontig; and blue - the Reference and Alternative SNP alleles.

Table 2. Sekati specific alleles, primer sequence, number of bases (Nb), and amplicon size of 13 primer pairs.

Sekati specific alleles					
Marker Name	**Allele**		**Primer Sequence**	**Nb**	**Amplicon Size**
1s	G	F	AGTTTCATATATAAAAAAAAAAAGAG	26	145
		R	ACGAATGAGAAATAATCAAAATA	23	
2s	T	F	CATCGGTCCGTTGTTTCTT	19	266
		R	AGACATGACCATGAGACCAAAC	22	
3s	T	F	TACTTTTAACGCACCCGT	18	278
		R	AAGCAGCTTTGTTAAACACA	20	
4s	A	F	AGCTTATGGAGCAGAATATT	20	398
		R	GTAACAGGTGAGTGATGTGA	20	
5s	G	F	GTATAACGAGAGCATTTTCATG	22	209
		R	GGGCATAGATAAAAAACGAA	20	
6s	G	F	TATAGAATTTTTCATTTGTAGGC	23	447
		R	CTTTTGCAGATTCACATGTT	20	
7s	G	F	CATTTTATTAATTAACCGCCC	21	494
		R	TGTTTGTTTGCTTACCGCTA	20	
8s	T	F	AACTCTATCTTAGGATTTCGGT	22	357
		R	GTTTTGTTTCTTAAGTTTCCC	21	
9s	T	F	CCTTATCCCTTTATGGGTTGTAT	23	145
		R	TTTTAGGAGGAAAGAAAAAGAAA	23	
10s	T	F	TAGAGGTTCCTTATGGGTAGTA	22	217
		R	CCTATATAAAAACACGTTTACTTAG	25	
11s	T	F	AATGTACAAAATCAGGAATCA	21	421
		R	ATCCACCCCTAAATAAATAA	20	
12s	T	F	ATATATGCTACATCACGTGCTT	22	449
		R	ATTTGACCGTTGATGACCT	19	
13s	G	F	TGAATGAAAGCAGAATTCATATG	23	414
		R	TGATCAAATCACGTTTGTATCAT	23	

Table 3. JS-12 specific alleles, primer sequence, number of bases (Nb), and amplicon size of 13 primer pairs.

JS-12 specific alleles					
Marker Name	**Allele**		**Primer Sequence**	**Nb**	**Amplicon Size**
1j	C	F	TTCCTTTTAGATAATATTCTTTTAAG	26	266
		R	CCGAATCACGTTGCTTATA	19	
2j	G	F	TTGACCTCGTTATCGGAATC	20	432
		R	ATTGTCAGTCCATTTTCTGTCTT	23	
3j	C	F	AAAGCTAATGGATCATTACATG	22	135
		R	CATTTGGTACAAGGTGAGG	19	
4j	T	F	ACGAGCCAAATTTAAGACT	19	326
		R	TGTGAGAATAGCGTCTGTAG	20	
5j	C	F	GAAGAAAAAGCACAACCG	18	401
		R	TAACACCACCATCACCACTA	20	
6j	A	F	TTCTTCTCCTTCCCTACCTA	20	315
		R	TGATGGTTTTTAGTGGATGA	20	
7j	A	F	TTGGATGAGGTATGAAGCCA	20	380
		R	GGTCGTCTTGATCTGAAATGAT	22	
8j	C	F	TGTAAAATGTTGAATCATTAGTG	23	183
		R	GCGGGAATACTAGAGTTAATC	21	
9j	C	F	TTCCCCAAATTCCGAGAG	18	499
		R	TGGATGCAGGAGAGAAGGTA	20	
10j	C	F	CCTTATATATTACATTACATATTTGC	26	119
		R	ATATAGGGAAATGAAGATAAGAGA	24	
11j	G	F	AACAAGTAGAACCTGTAAAGAG	22	269
		R	AAATTAGAAGTGAGGTTGGT	20	
12j	C	F	ACAAAATCACAAGTGTCTGAAG	22	315
		R	CTAAAAACTAGGGTCCAACAT	21	
13j	C	F	GCGAAGGTTAAAGTTTCGG	19	108
		R	ACAGTAGGCTCCGTTGTTGT	20	

The PCR parameters varied greatly across all markers (Table 03). The detection sensitivity and accuracy of this genotyping technique are highly influenced by PCR conditions, such as Mg^{2+} concentration, DNA quality and concentration, annealing temperature, and PCR cycles. The best amplification parameters were defined according to band intensity and allele amplification specificity. PCR cycles, Ta, and Mg^{2+} concentration had great importance for marker condition adjustments, although non-specific amplification was observed for some markers when the conditions were highly favorable. Increasing Ta and decreasing the Mg^{2+} concentration and PCR cycles were good strategies to avoid non-specific amplification.

Table 4. Specific PCR conditions for the singleplex genotyping of 13 SNPs.

Sekati specific alleles				**JS-12 specific alleles**			
Marker name	**PCR cycles**	**Ta**	**Mg^{2+} concentration**	**Marker name**	**PCR cycles**	**Ta**	**Mg^{2+} concentration**
1s	32	51°	2.88 mM	1j	35	50°	2.88 mM
2s	32	55°	2.88 mM	2j	30	55°	2.88 mM
3s	30	51°	1.92 mM	3j	30	55°	2.88 mM
4s	30	48°	1.92 mM	4j	32	48°	2.88 mM
5s	30	51°	2.88 mM	5j	35	58°	1.92 mM
6s	30	51°	2.88 mM	6j	30	51°	2.88 mM
7s	30	55°	2.88 mM	7j	30	55°	1.92 mM
8s	30	55°	2.88 mM	8j	30	51°	2.88 mM
9s	35	55°	2.88 mM	9j	35	53°	2.88 mM
10s	35	51°	2.88 mM	10j	30	51°	2.88 mM
11s	32	48°	2.88 mM	11j	32	48°	1.92 mM
12s	30	58°	1.92 mM	12j	32	51°	2.88 mM
13s	30	55°	2.88 mM	13j	32	55°	1.92 mM

Ta: annealing temperature.

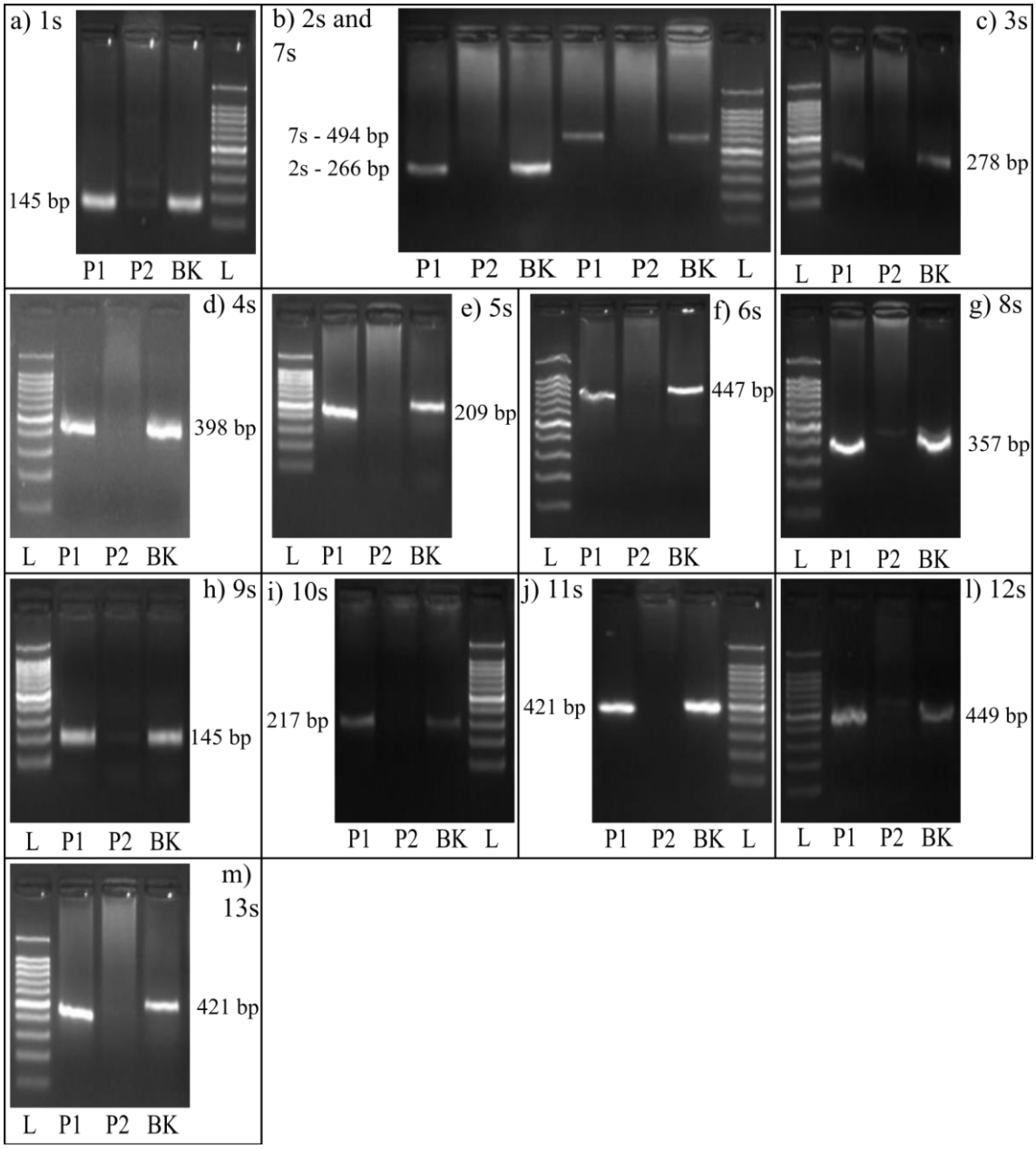

Figure 1. Singleplex genotyping of 13 markers specific to Sekati alleles in three control samples. P1: line Sekati (genotype: s/s); P2: line JS-12 (genotype j/j); BK: control for heterozygote; L: 100 bp ladder.

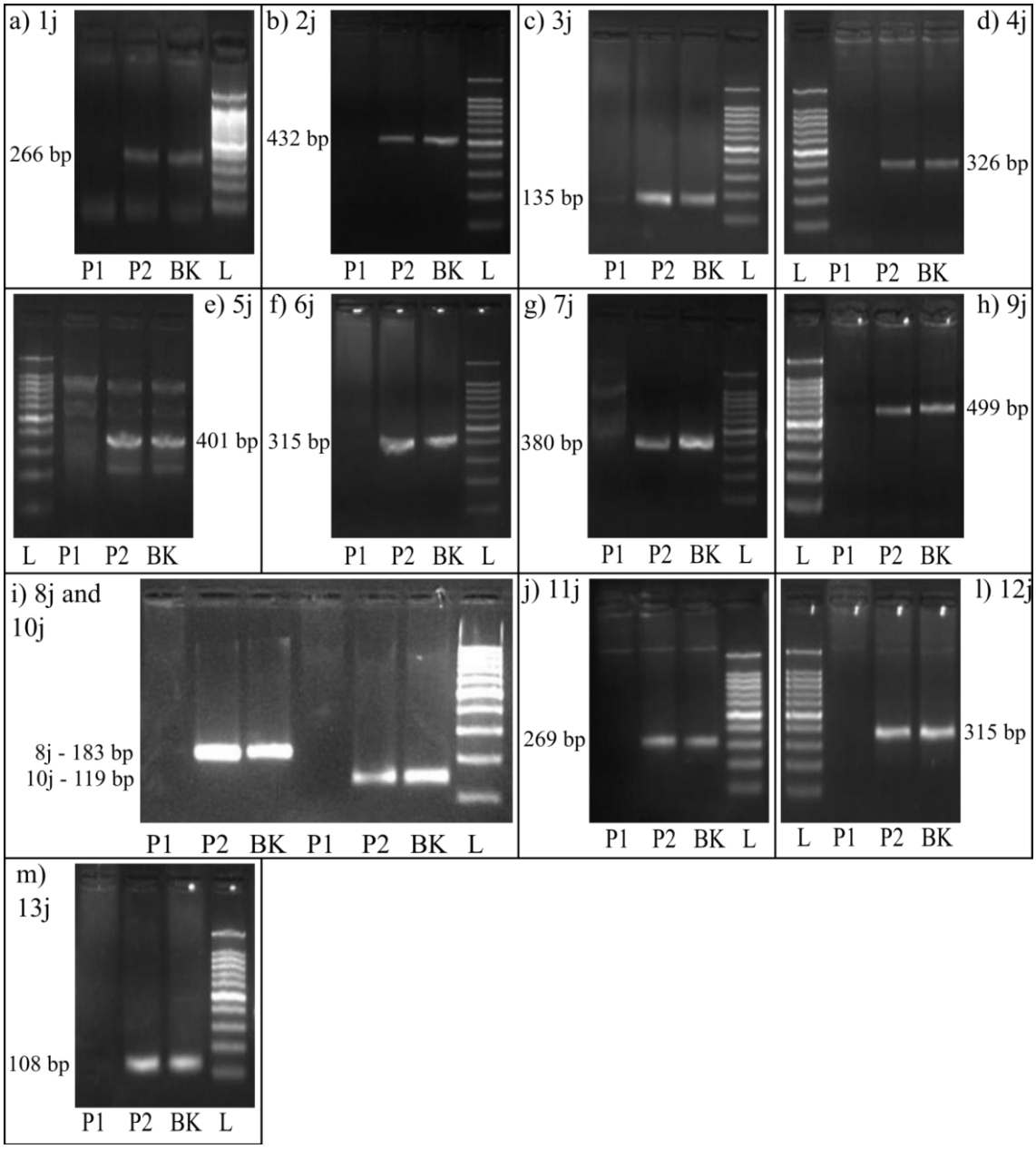

Figure 2. Singleplex genotyping of 13 markers specific to JS-12 alleles in three control samples. P1: line Sekati (genotype: s/s); P2: line JS-12 (genotype j/j); BK: control for heterozygote; L: 100 bp ladder.

The use of saturated genetic maps and QTL analysis is one of the main strategies observed for the genetic mapping of important traits in breeding programs (Liu et al., 2016; Montero-Pa et al., 2017; Sun et al., 2015; Xu et al., 2016). This strategy aims to saturate the genome of the target species with DNA polymorphisms, often identified using NGS and structured mapping populations such as F_2 obtained from contrasting parents. The QTL analysis allows the identification of genomic regions where segregation is statistically associated with the phenotype of interest. The genomic region of the QTL may contain one or more genes influencing the phenotype, allowing the identification of candidate genes for further research. However, sequencing costs remain a major concern for the wide application of this robust strategy of genetic mapping, which could significantly increase breeding program gains.

In this study, an alternative strategy was proposed for the genetic mapping of fruit firmness in papaya. Instead of obtaining saturated genetic maps and performing QTL analysis to search for the candidate genes, a low cost methodology was proposed. A literature review was performed to search for available candidate genes for fruit firmness in papaya. The study developed by Shen et al. (2017) allowed the isolation of about 50 fruit ripening-related genes, including several associated with the fruit softening process. In another study, Bohry et al. (2021) associated these genes with about 460 DNA polymorphisms.

To develop a low cost methodology for genetic mapping in papaya, a total of 13 SNPs tightly linked with 13 fruit firmness genes (Bohry et al., 2021) were selected for marker design. The objective was to make additional use of the SNP information obtained from mapping populations to construct a low resolution genetics map in papaya which may allow the identification of the roles of the selected fruit firmness genes on the phenotype through QTL analysis.

To do so, the markers validated in the three control samples were used to genotype 12 segregating plants derived from the cross between Sekati and JS-12. The three control samples were included in the analysis to help score the alleles and check for amplification specificity. Figures from 3 to 15 show the singleplex amplification patterns of the markers in the mapping population sample.

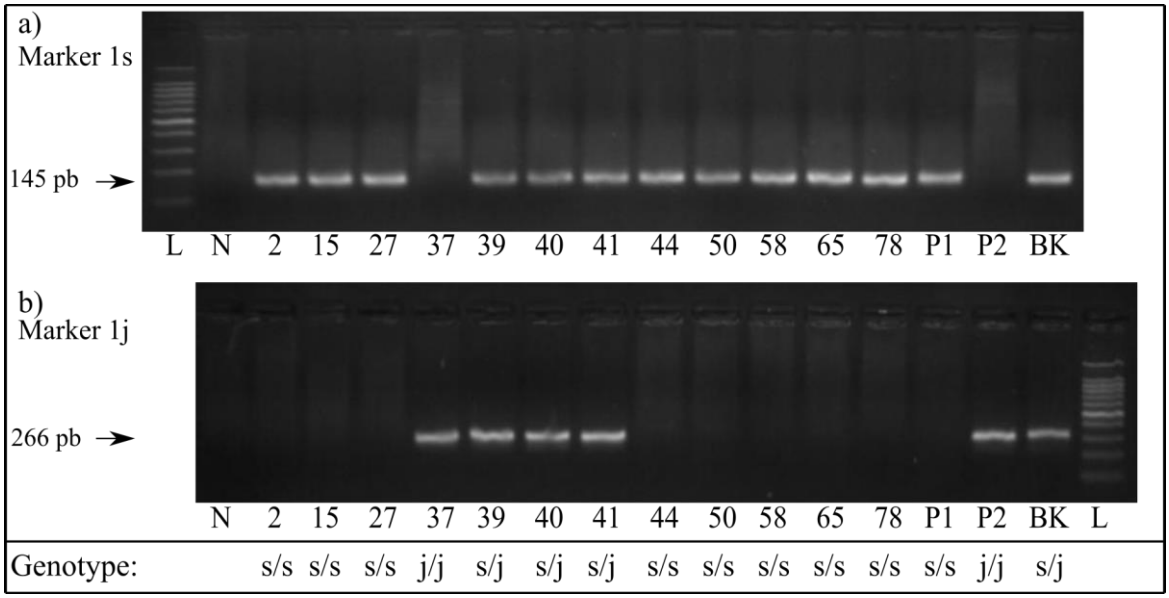

Figure 3. Singleplex genotyping of markers 1s (Figure 3a) and 1j (Figure 3b) in 12 F_2 segregant plants derived from the cross Sekati x JS-12. N: no template control (N); 2-78: F_2 segregant plants; P1: line Sekati; P2: line JS-12; BK: control for heterozygote; L: 100 bp ladder; s/s: homozygote toward Sekati; j/j: homozygote toward JS-12; s/j or j/s: heterozygote.

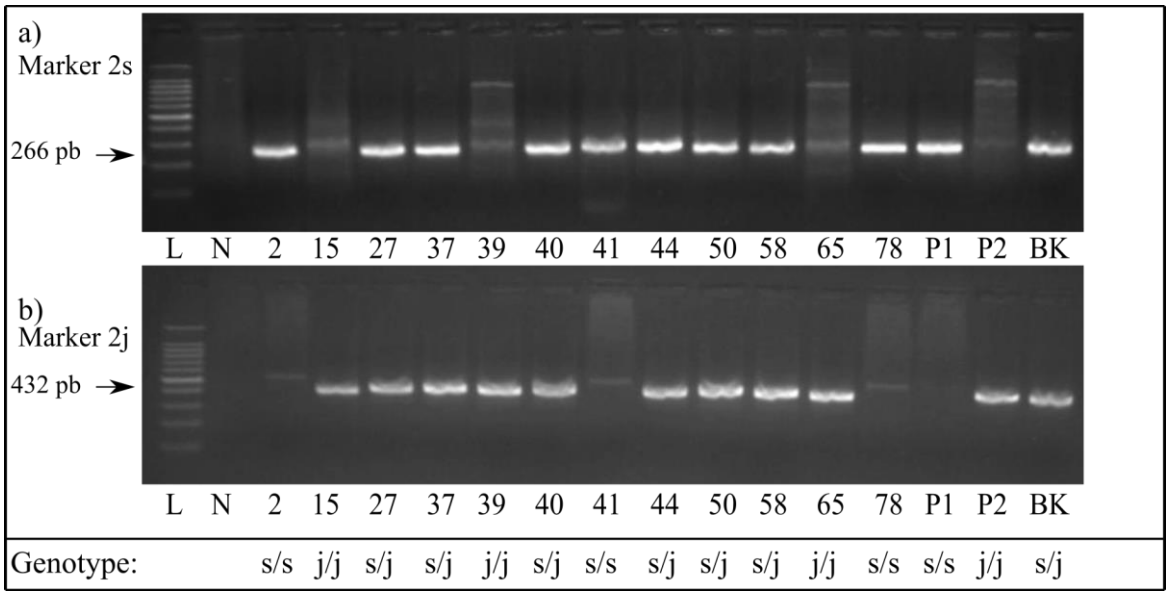

Figure 4. Singleplex genotyping of markers 2s (Figure 2a) and 2j (Figure 2b) in 12 F_2 segregant plants derived from the cross Sekati x JS-12. N: no template control (N); 2-78: F_2 segregant plants; P1: line Sekati; P2: line JS-12; BK: control for heterozygote; L: 100 bp ladder; s/s: homozygote toward Sekati; j/j: homozygote toward JS-12; s/j or j/s: heterozygote.

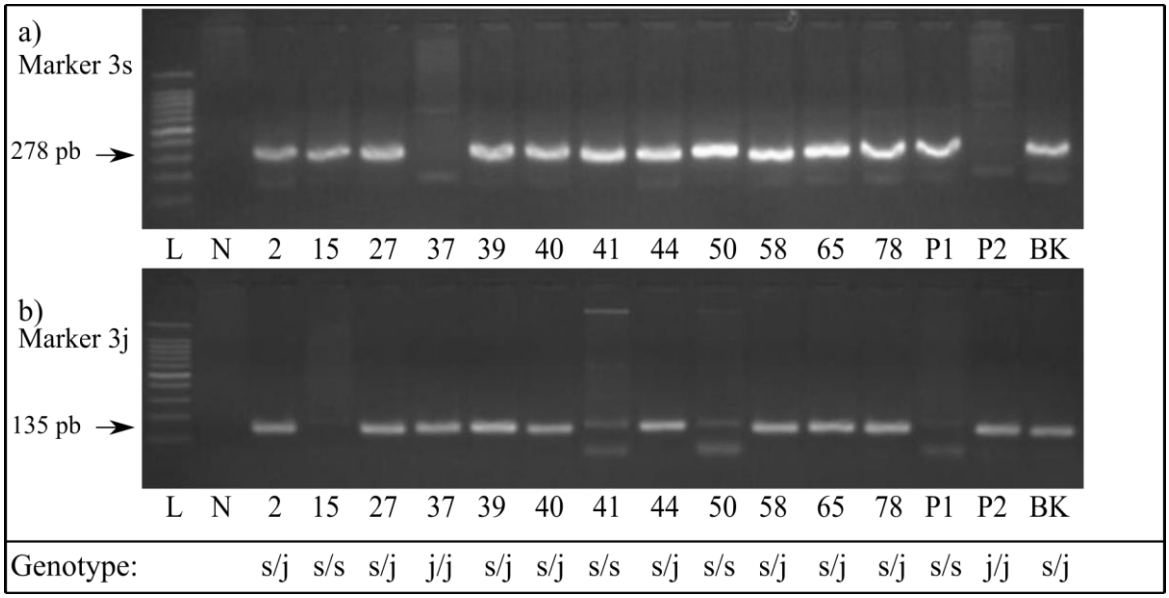

Figure 5. Singleplex genotyping of markers 3s (Figure 5a) and 2j (Figure 5b) in 12 F_2 segregant plants derived from the cross Sekati x JS-12. N: no template control (N); 2-78: F_2 segregant plants; P1: line Sekati; P2: line JS-12; BK: control for heterozygote; L: 100 bp ladder; s/s: homozygote toward Sekati; j/j: homozygote toward JS-12; s/j or j/s: heterozygote.

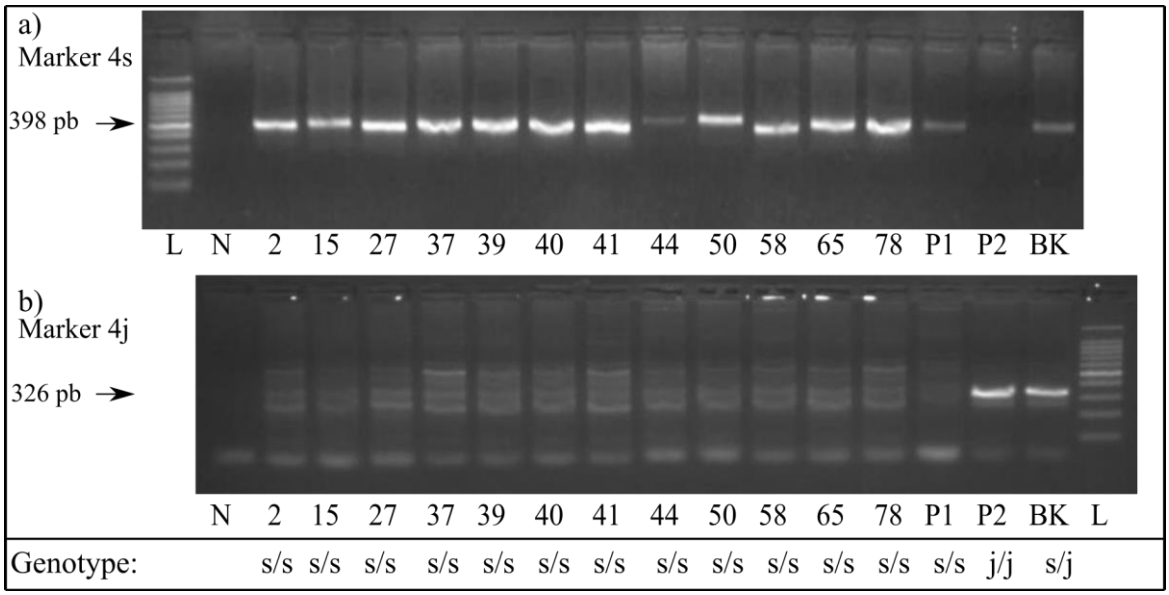

Figure 6. Singleplex genotyping of markers 4s (Figure 6a) and 4j (Figure 6b) in 12 F_2 segregant plants derived from the cross Sekati x JS-12. N: no template control (N); 2-78: F_2 segregant plants; P1: line Sekati; P2: line JS-12; BK: control for heterozygote; L: 100 bp ladder; s/s: homozygote toward Sekati; j/j: homozygote toward JS-12; s/j or j/s: heterozygote.

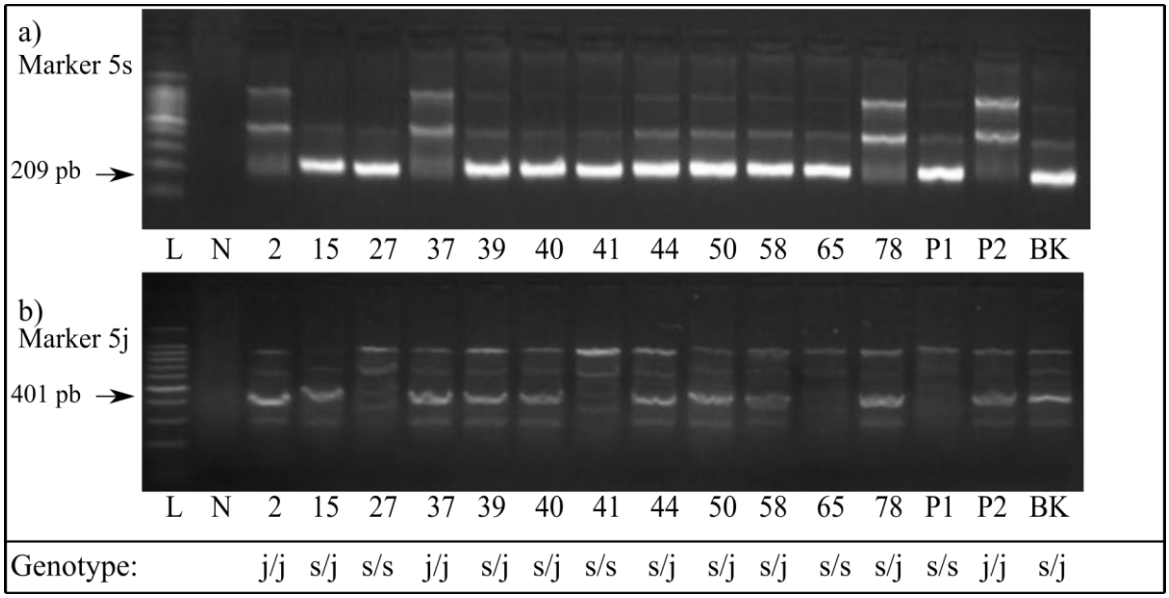

Figure 7. Singleplex genotyping of markers 5s (Figure 7a) and 5j (Figure 7b) in 12 F_2 segregant plants derived from the cross Sekati x JS-12. N: no template control (N); 2-78: F_2 segregant plants; P1: line Sekati; P2: line JS-12; BK: control for heterozygote; L: 100 bp ladder; s/s: homozygote toward Sekati; j/j: homozygote toward JS-12; s/j or j/s: heterozygote.

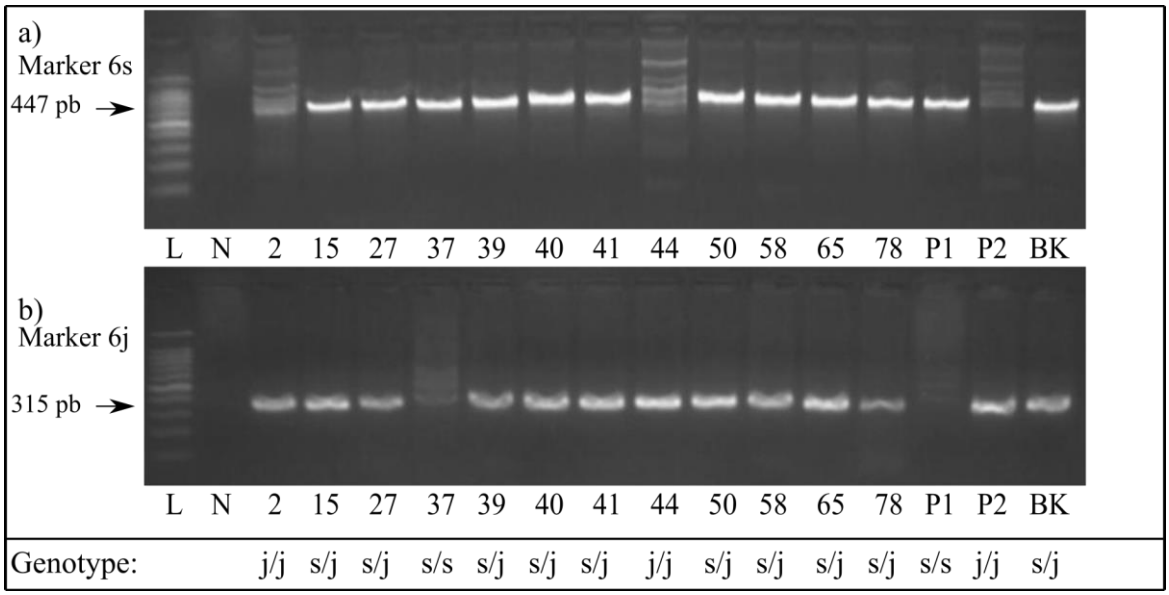

Figure 8. Singleplex genotyping of markers 6s (Figure 8a) and 6j (Figure 8b) in 12 F_2 segregant plants derived from the cross Sekati x JS-12. N: no template control (N); 2-78: F_2 segregant plants; P1: line Sekati; P2: line JS-12; BK: control for heterozygote; L: 100 bp ladder; s/s: homozygote toward Sekati; j/j: homozygote toward JS-12; s/j or j/s: heterozygote.

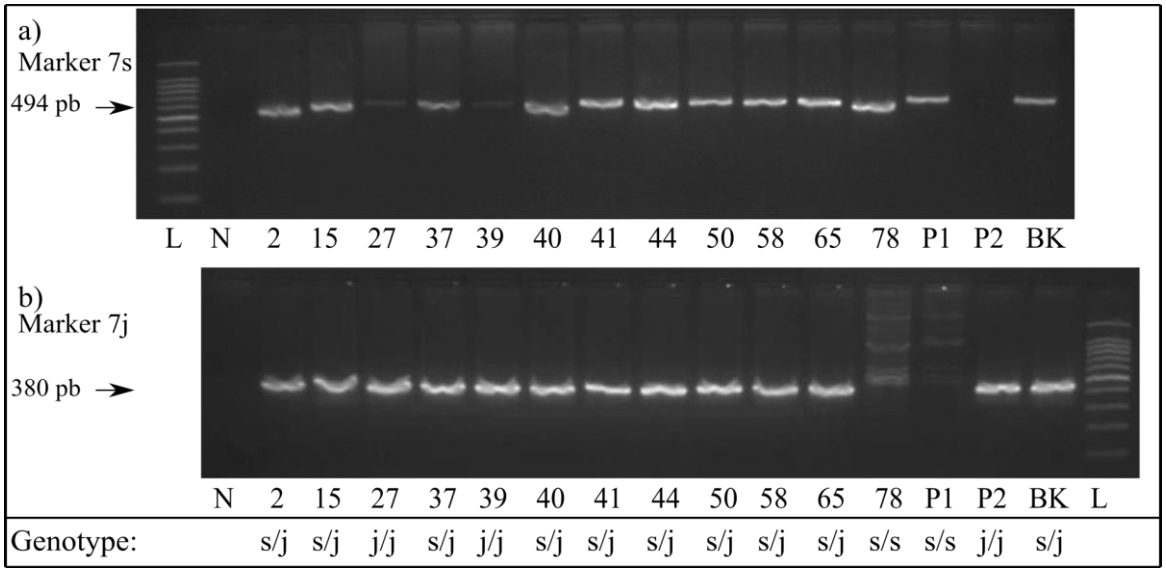

Figure 9. Singleplex genotyping of markers 7s (Figure 9a) and 7j (Figure 9b) in 12 F_2 segregant plants derived from the cross Sekati x JS-12. N: no template control (N); 2-78: F_2 segregant plants; P1: line Sekati; P2: line JS-12; BK: control for heterozygote; L: 100 bp ladder; s/s: homozygote toward Sekati; j/j: homozygote toward JS-12; s/j or j/s: heterozygote.

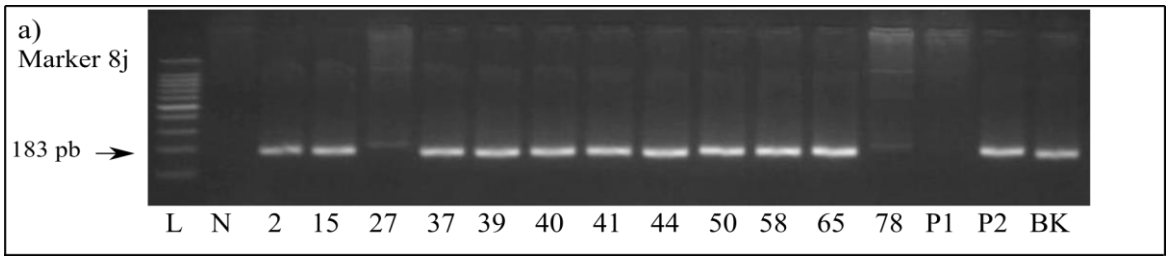

Figure 10. Singleplex genotyping of marker 8j in 12 F_2 segregant plants derived from the cross Sekati x JS-12. N: no template control (N); 2-78: F_2 segregant plants; P1: line Sekati; P2: line JS-12; BK: control for heterozygote; L: 100 bp ladder.

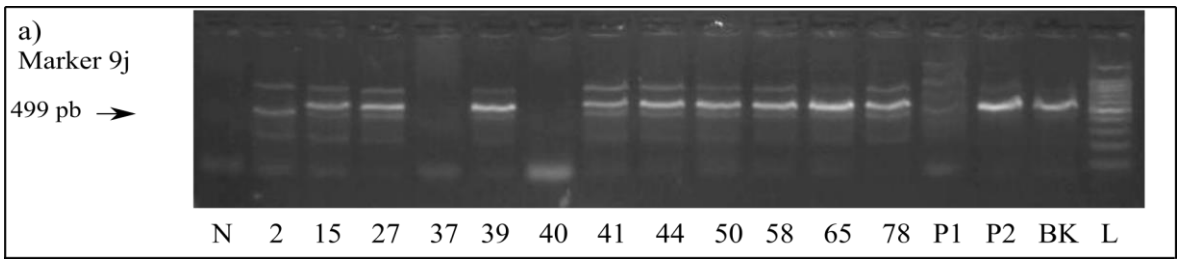

Figure 11. Singleplex genotyping of marker 9j in 12 F_2 segregant plants derived from the cross Sekati x JS-12. N: no template control (N); 2-78: F_2 segregant plants; P1: line Sekati; P2: line JS-12; BK: control for heterozygote; L: 100 bp ladder.

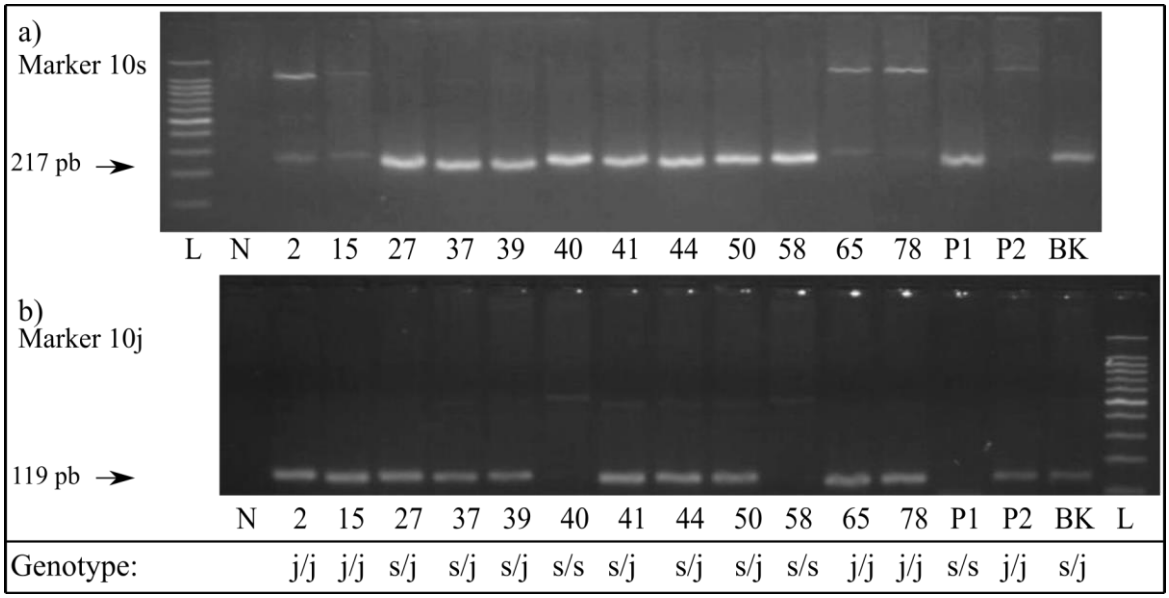

Figure 12. Singleplex genotyping of markers 10s (Figure 12a) and 10j (Figure 12b) in 12 F_2 segregant plants derived from the cross Sekati x JS-12. N: no template control (N); 2-78: F_2 segregant plants; P1: line Sekati; P2: line JS-12; BK: control for heterozygote; L: 100 bp ladder; s/s: homozygote toward Sekati; j/j: homozygote toward JS-12; s/j or j/s: heterozygote.

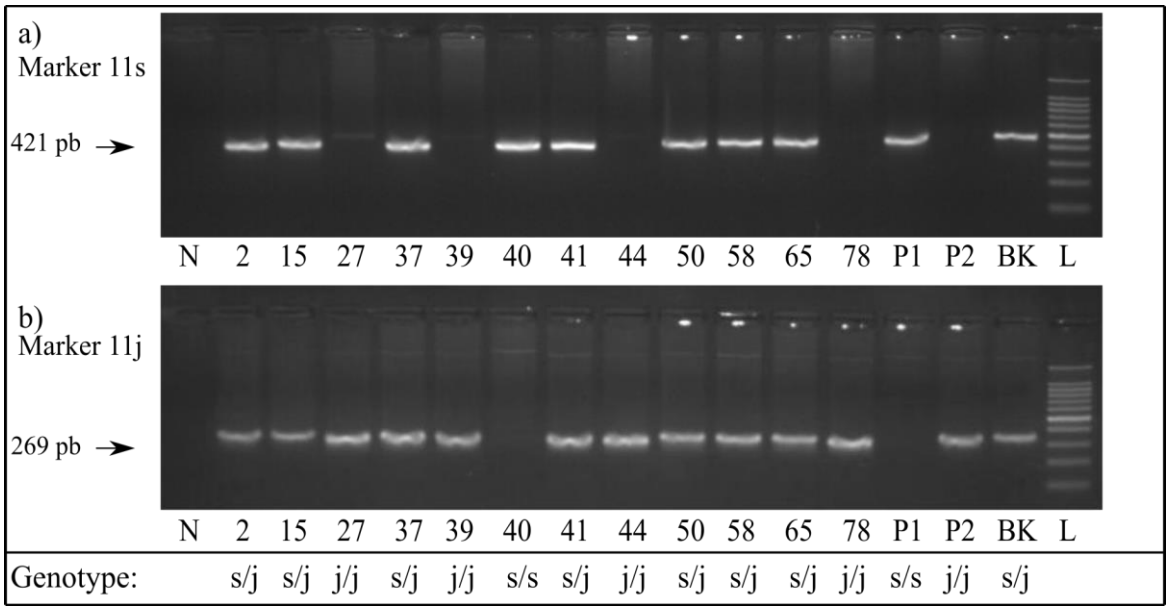

Figure 13. Singleplex genotyping of markers 11s (Figure 13a) and 11j (Figure 13b) in 12 F_2 segregant plants derived from the cross Sekati x JS-12. N: no template control (N); 2-78: F_2 segregant plants; P1: line Sekati; P2: line JS-12; BK: control for heterozygote; L: 100 bp ladder; s/s: homozygote toward Sekati; j/j: homozygote toward JS-12; s/j or j/s: heterozygote.

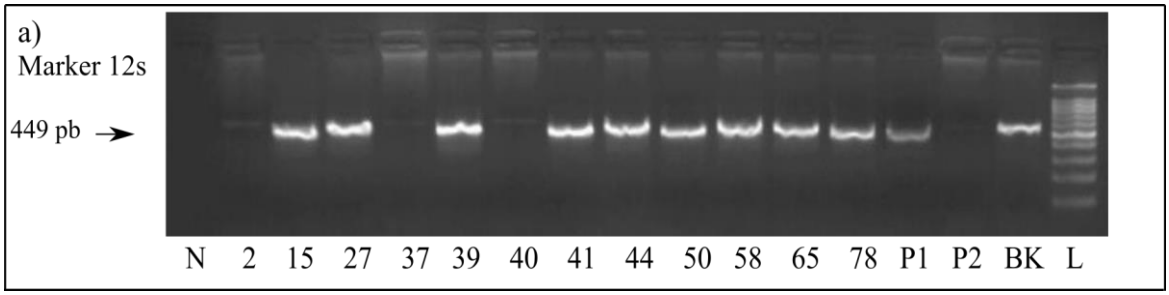

Figure 14. Singleplex genotyping of marker 12s in 12 F_2 segregant plants derived from the cross Sekati x JS-12. N: no template control (N); 2-78: F_2 segregant plants; P1: line Sekati; P2: line JS-12; BK: control for heterozygote; L: 100 bp ladder.

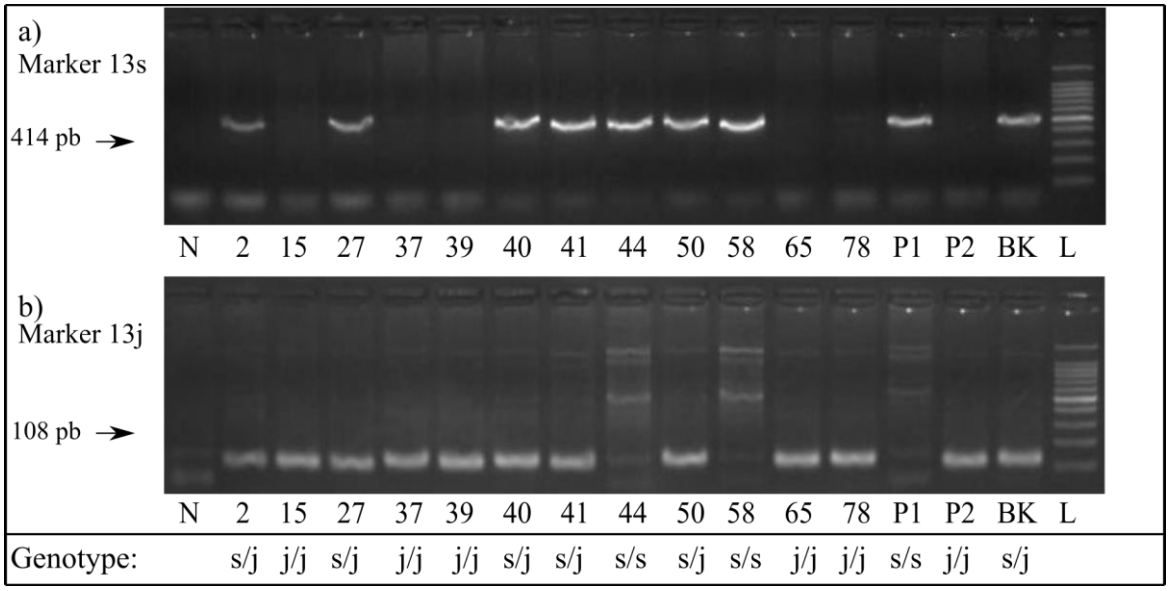

Figure 15. Singleplex genotyping of markers 13s (Figure 15a) and 13j (Figure 15b) in 12 F_2 segregant plants derived from the cross Sekati x JS-12. N: no template control (N); 2-78: F_2 segregant plants; P1: line Sekati; P2: line JS-12; BK: control for heterozygote; L: 100 bp ladder; s/s: homozygote toward Sekati; j/j: homozygote toward JS-12; s/j or j/s: heterozygote.

The markers were efficient in discriminating the alleles of the 13 SNPs. However, not all markers presented clean amplification for observation in the gel electrophoresis. Even after meticulous and time-consuming PCR adjustments, some markers still presented PCR artifacts, such as the markers 5s (Figure 7a), 5j (Figure 7b), 9j (Figure 11a), and10s (Figure 12a).

3.2.4.3 Optimization of duplex primers

To reduce the number of reactions to the genotype a small set of SNPs, duplex PCR reactions were tested. After setting the best PCR parameters for each singleplex marker, the combination of primers with similar parameters was tested. Duplex combination of primer pairs within the Sekati specific alleles group resulted in five successful combinations and two combinations for the JS-12 specific alleles group (Table 4). The primers with no combination had to be analyzed in a singleplex reaction.

Table 5. Specific PCR conditions for the duplex genotyping of 14 SNP alleles.

					Primer Concentration	
Marker duplex	**Amplicon Size**	**PCR cycles**	**Ta**	**Mg^{2+} content**	**1º marker (F+R)**	**2º marker (F+R)**
Sekati specific duplex						
1s-2s	145-266	32	51°	2.88 mM	0.288 µM	0.096 µM
3s-4s	278-398	30	51°	2.88 mM	0.231 µM	0.154 µM
5s-6s	209-447	30	51°	2.88 mM	0.192 µM	0.192 µM
7s-8s	494-357	32	58°	2.88 mM	0.192 µM	0.192 µM
9s-10s	145-217	38	53°	2.88 mM	0.192 µM	0.192 µM
JS-12 specific duplex						
2j-3j	135-432	30	55°	2.88 mM	0.192 µM	0.192 µM
6j-8j	315-183	30	51°	2.88 mM	0.192 µM	0.192 µM

Ta: annealing temperature; F: forward primer; R: reverse primer.

Genotyping of markers in duplex reactions showed consistent results when compared with singleplex ones. Figures 16 to 22 show the amplification patterns of seven duplex combinations of markers. A total of 26 PCR reactions was needed to genotype the set of 13 SNPs of one individual using singleplex reactions. However, when using the duplex primer combinations and singleplex reactions the total number of reactions was reduced to 19. This represents about 27% of PCR reactions, which can save time and reduce genotyping costs.

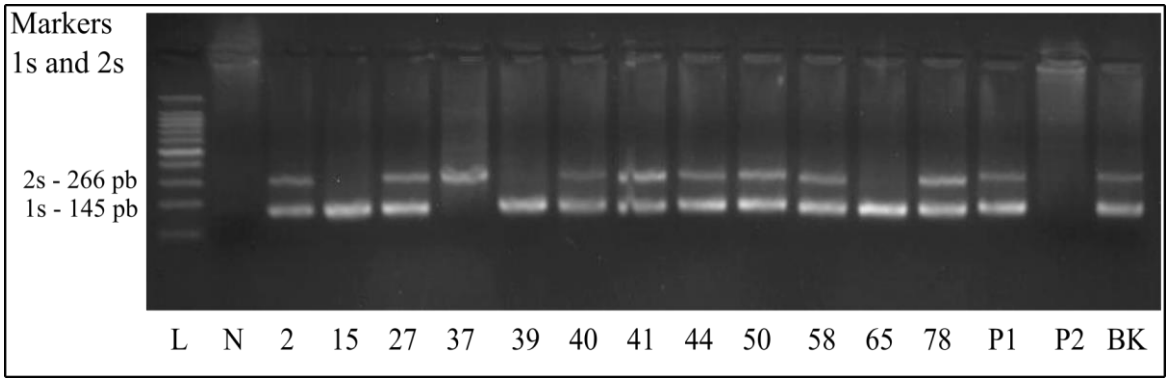

Figure 16. Duplex genotyping of markers 1s and 2s in 12 F_2 segregant plants derived from the cross Sekati x JS-12. N: no template control (N); 2-78: F_2 segregant plants; P1: line Sekati; P2: line JS-12; BK: control for heterozygote; L: 100 bp ladder.

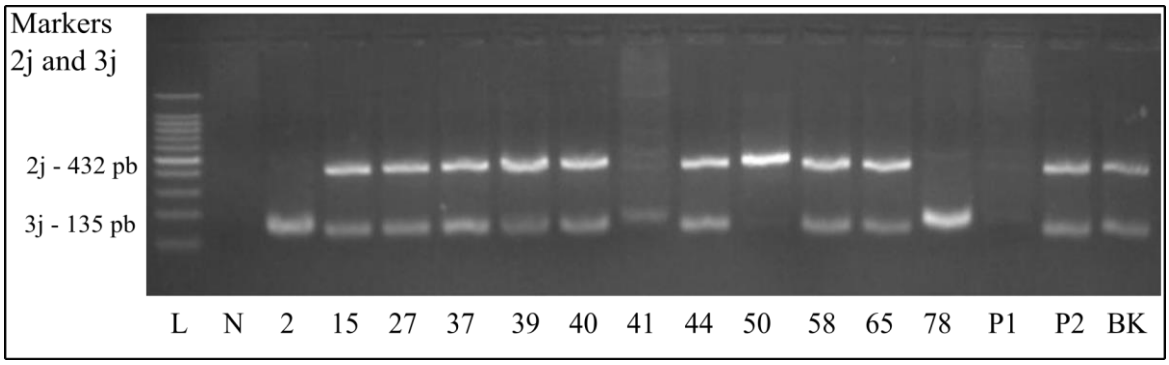

Figure 17. Duplex genotyping of markers 2j and 3j in 12 F_2 segregant plants derived from the cross Sekati x JS-12. N: no template control (N); 2-78: F_2 segregant plants; P1: line Sekati; P2: line JS-12; BK: control for heterozygote; L: 100 bp ladder.

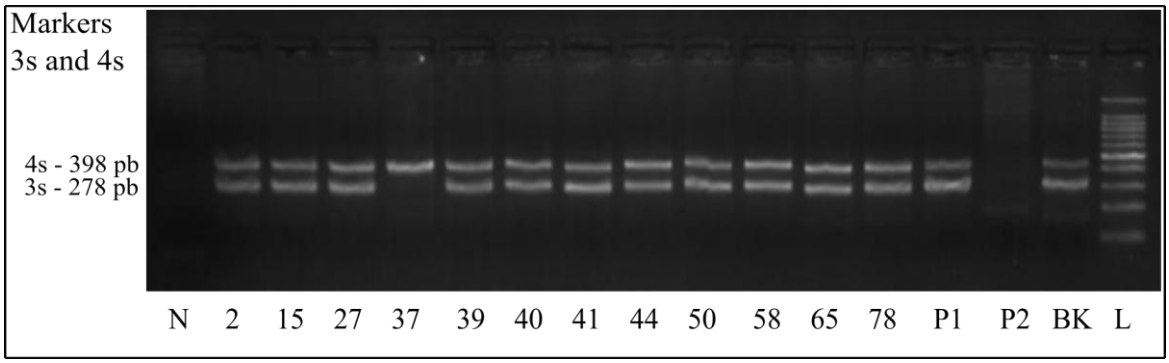

Figure 18. Duplex genotyping of markers 3s and 4s in 12 F_2 segregant plants derived from the cross Sekati x JS-12. N: no template control (N); 2-78: F_2 segregant plants; P1: line Sekati; P2: line JS-12; BK: control for heterozygote; L: 100 bp ladder.

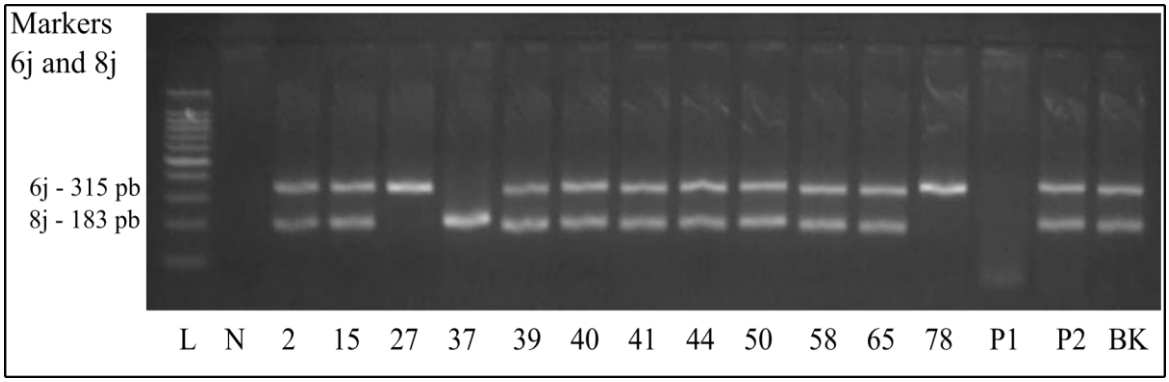

Figure 19. Duplex genotyping of markers 6j and 8j in 12 F_2 segregant plants derived from the cross Sekati x JS-12. N: no template control (N); 2-78: F_2 segregant plants; P1: line Sekati; P2: line JS-12; BK: control for heterozygote; L: 100 bp ladder.

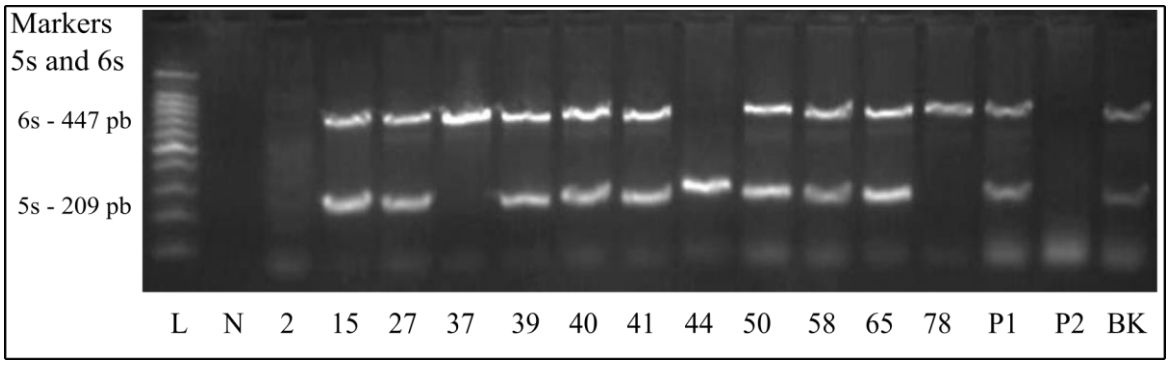

Figure 20. Duplex genotyping of markers 5s and 6s in 12 F_2 segregant plants derived from the cross Sekati x JS-12. N: no template control (N); 2-78: F_2 segregant plants; P1: line Sekati; P2: line JS-12; BK: control for heterozygote; L: 100 bp ladder.

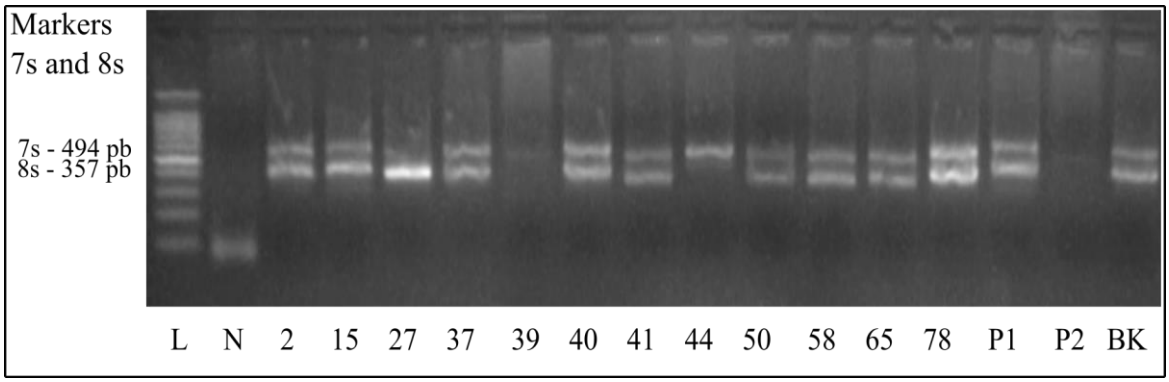

Figure 21. Duplex genotyping of markers 7s and 8s in 12 F_2 segregant plants derived from the cross Sekati x JS-12. N: no template control (N); 2-78: F_2 segregant plants; P1: line Sekati; P2: line JS-12; BK: control for heterozygote; L: 100 bp ladder.

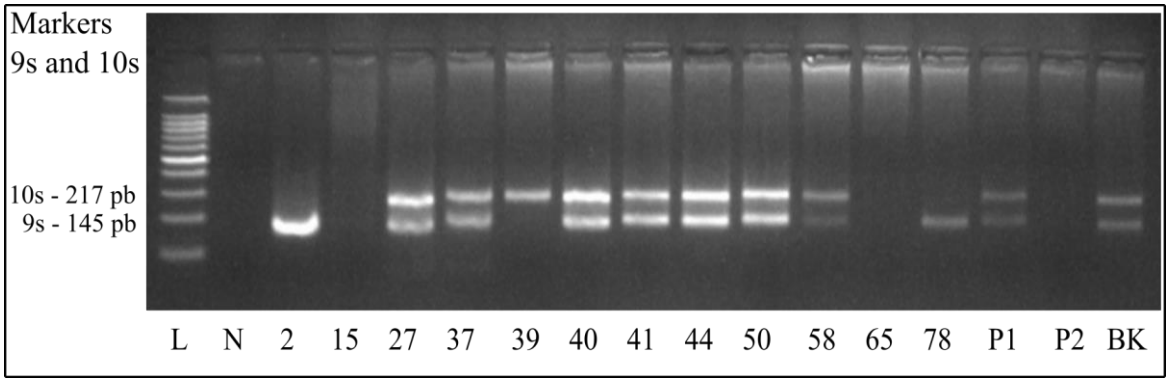

Figure 22. Duplex genotyping of markers 9s and 10s in 12 F_2 segregant plants derived from the cross Sekati x JS-12. N: no template control (N); 2-78: F_2 segregant plants; P1: line Sekati; P2: line JS-12; BK: control for heterozygote; L: 100 bp ladder.

Genotyping of the 26 SNP alleles in 12 F_2 segregating plants revealed apparent 1:2:1 Mendelian segregation of markers (Table 6). However, because of small population size, it was not possible to calculate if the observed segregation fit the expected segregation. SNP4 was the only one with apparent segregation distortion. Marker 4s amplified in all 12 F_2 individuals (Figure 6a), while marker 4j did not amplify in all F_2 samples (Figure 6b). Although only 12 F_2 plants were genotyped, the chances of randomly picking only homozygous samples toward the Sekati (s) allele were very low. Marker segregation distortion has previously been reported in papaya (Chen et al., 2007; Nantawan et al., 2019).

Based on the available literature, this is the first report of the development of PCR markers based on SNPs in papaya. SNPs are the most abundant form of variations in plant genomes and their application in studies involving plants is rapidly expanding. The use of NGS has been largely used to elucidate the intriguing sex-determination system of the papaya species (Lee et al., 2018; Liao et al., 2017; Ming et al., 2007; VanBuren et al., 2015, 2016). However, little effort is observed for the use of NGS for breeding purposes in papaya. A genetic map based on 219 SNPs was obtained in papaya after the use of NGS (Nantawan et al., 2019). In this study, a total of 21 quantitative trait loci (QTL) were identified for seven key fruit quality traits, including flesh sweetness, fruit weight, fruit length, fruit width, skin freckle, flesh thickness, and fruit firmness. Such information reveals candidate SNPs for developing marker-assisted selection strategies in papaya breeding programs.

Table 6. Scores obtained from the genotyping of 13 SNPs associated with fruit firmness-related genes in 12 F_2 plants and three control samples.

SNP	F_2 samples												Control samples			Frequency in F_2		
number	2	15	27	37	39	40	41	44	50	58	65	78	P1	P2	BK	ss	sj	jj
SNP1	ss	ss	ss	jj	sj	sj	sj	ss	ss	ss	ss	ss	ss	jj	sj	8	3	1
SNP2	ss	jj	sj	sj	jj	sj	ss	sj	sj	sj	jj	ss	ss	jj	sj	3	6	3
SNP3	sj	ss	sj	jj	sj	sj	ss	sj	ss	sj	sj	sj	ss	jj	sj	3	8	1
SNP4	ss	ss	ss	ss	ss	ss	ss	ss	ss	ss	ss	ss	ss	jj	sj	12	0	0
SNP5	jj	sj	ss	jj	sj	sj	ss	sj	sj	sj	ss	jj	ss	jj	sj	3	6	3
SNP6	jj	sj	sj	ss	sj	sj	sj	jj	sj	sj	sj	sj	ss	jj	sj	1	9	2
SNP7	sj	sj	jj	sj	jj	sj	sj	sj	sj	sj	sj	ss	ss	jj	sj	1	9	2
SNP8	sj	sj	ss	sj	jj	sj	sj	jj	sj	sj	sj	ss	ss	jj	sj	2	8	2
SNP9	sj	jj	sj	ss	jj	ss	sj	sj	sj	sj	jj	sj	ss	jj	sj	2	7	3
SNP10	jj	jj	sj	sj	sj	ss	sj	sj	sj	ss	jj	jj	ss	jj	sj	2	6	4
SNP11	sj	sj	jj	sj	jj	ss	sj	jj	sj	sj	sj	jj	ss	jj	sj	1	7	4
SNP12	j.	s.	s.	j.	s.	j.	s.	s.	s.	s.	s.	s.	ss	jj	sj			
SNP13	sj	jj	sj	jj	jj	sj	sj	ss	sj	ss	jj	jj	ss	jj	sj	2	5	5

P1: line Sekati; P2: line JS-12; BK: control for heterozygote; ss: homozygote toward Sekati; jj: homozygote toward JS-12; sj: heterozygote.

Fruit firmness is one of the most important traits for papaya breeding. The fruits of papaya ripen very quickly and lose their firmness along the process (Shen et al., 2017), causing great losses in the papaya postharvest chain. The loss of firmness in papaya is under the control of the phytohormone ethylene, as papaya is considered a climacteric fruit. Ethylene regulates the fruit ripening process in climacteric fruits (Liu et al., 2015). As the fruit turns ripe the sensorial attributes of flavor, aroma, color, and texture are developing, making the fruit edible and attractive for seed dispersal. As a result, the fruit naturally loses its firmness during the ripening process. However, significant fruit firmness variability is observed in papaya genotypes, and this seems to be related to fruit size (Barbosa et al., 2011).

Such variability has great importance for developing superior cultivars with improved fruit firmness and other important attributes. However, the complex nature of this trait and its negative relation with fruit sweetness (Quintal, 2009)

make the simultaneous selection of both traits a daunting task in breeding programs. Understanding this genetic variability at the molecular level may significantly contribute to the selection process of both traits.

Loss of firmness in papaya is mainly caused by the disassembling of the fruit cell wall during ripening. Several genes, such as Polygalacturonase, Beta-D-xylosidase, Beta-galactosidase, Beta glucosidase, and Pectate lyase, have upregulation when treated with exogenous ethylene and accelerate the degradation of pectin, cellulose, and hemicellulose polymers (Table 1). Ethylene also induces the expression of genes in the metabolism of sugars such as UDP - galactose transporter and Beta-fructofuranosidase. The genes Ethylene receptor and Ethylene-insensitive 4-like protein also play critical roles during ripening as ethylene sensing genes. Both genes were upregulated when treated with exogenous ethylene and downregulated when treated with 1-MCP, which is an ethylene blocker (Shen et al., 2017). Other genes had higher expressions in exogenous treated papaya, such as the two protease Subtilisin-like serine endopeptidase family protein and Cysteine proteinase superfamily protein. Proteases act as degrading proteins and are related to the senescence process. These genes may contribute to the loss of papaya firmness during ripening. One NAC domain protein was upregulated with ethylene treatments.

3.2.5 CONCLUSION

The markers showed to be efficient in discriminating the 26 SNP alleles in a singleplex reaction. Validation of seven duplex amplifications revealed consistent results compared with singleplex reactions. Genotyping of F_2 segregating plants was carried out with success using the validated markers. The PCR-based markers developed in this study are a valuable resource for studying the variability of fruit firmness-related genes in papaya breeding populations and gene banks.

REFERENCES

Amin, A., Iftekharuddaula, K.M.D., Sarker, A., Ghoshal, S., Aditya, T.L., et al. (2018) Introgression of SUB1 QTL into BR22 Using Marker Assisted Backcrossing. *Int J Plant Biol Res*, *6*(5): 1103.

Andrews, S. *FastQc - A quality control tool for high throughput sequence data* (2010). <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. Accessed April 4, 2019.

Argyris, J.M., Díaz, A., Ruggieri, V., Fernández, M., Jahrmann, T., Gibon, Y., Picó, B., Martín-Hernández, A.M., Monforte, A.J., Garcia-Mas, J. (2017) QTL analyses in multiple populations employed for the fine mapping and identification of candidate genes at a locus affecting sugar accumulation in melon (*Cucumis melo* L.). *Frontiers in Plant Science*, 8:1–20. https://doi.org/10.3389/fpls.2017.01679.

Badillo, V.M. Vasconcella St.-Hil. (Caricaceae) con la rehabilitacion de este ultimo Ernstia, 10 (2000), p. 74-79.

Barbosa, C.D., Viana, A.P., Quintal, S.S.R., Pereira, M.G. (2011) Artificial neural network analysis of genetic diversity in *Carica papaya* L.. *Crop Breeding and Applied Biotechnology* 11(3). https://doi.org/10.1590/S1984-70332011000300004.

Beltz, K., Tsang, D., Wang, J., Rose, S., Bao, Y., Wang, Y., Larkin, K., Rupp, S., Schrepfer, D., Datta, K., Gunderson, K., Sailor, C., Hansen, S., Dobosy, J., Lewis, L., Menezes, A., Walder, J., Behlke, M., Chen, C. (2018) A High-Performing and Cost-Effective SNP Genotyping Method Using rhPCR and Universal Reporters. *Advances in Bioscience and Biotechnology*, 09: 497-512. https://doi.org/10.4236/abb.2018.99034.

Bhardwaj, A., Dhar, Y.V., Asif, M.H., Bag, S.K. (2016) In Silico identification of SNP diversity in cultivated and wild tomato species: Insight from molecular simulations. *Scientific Reports*, 6:1-13. https://doi.org/10.1038/srep38715.

Blas, A.L., Ming, R., Liu, Z., Veatch, O.J., Paull, R.E., Moore, P.H., Yu, Q. (2010) Cloning of the papaya chromoplast-specific lycopene β-cyclase, CpCYC-b, controlling fruit flesh color reveals conserved microsynteny and a recombination hot spot. *Plant Physiology*, *152*(4): 2013-2022. https://doi.org/10.1104/pp.109.152298.

Blas, A.L., Yu, Q., Veatch, O.J., Paull, R.E., Moore, P.H., Ming, R. (2012) Genetic mapping of quantitative trait loci controlling fruit size and shape in papaya. *Molecular Breeding*, 29(2): 457–466. https://doi.org/10.1007/s11032-011-9562-1.

Bohry, D., Ramos, H.C.C., Santos, P.H.D., Boechat, M.S.B., Arêdes, F.A.S., Pirovani, A.A.V., Pereira, M.G. (2021) Discovery of SNPs and InDels in papaya genotypes and its potential for marker assisted selection of fruit quality traits. *Scientific Reports*, 11(1). https://doi.org/10.1038/s41598-020-79401-z.

Broccanello, C., Chiodi, C., Funk, A., McGrath, J.M., Panella, L., Stevanato, P. (2018) Comparison of three PCR-based assays for SNP genotyping in plants. *Plant Methods*, 14(1), 1-8. https://doi.org/10.1186/s13007-018-0295-6.

Brookes, A.J. (1999) The essence of SNPs. *Gene*. 234(2):177-86 https://doi.org/10.1016/S0378-1119(99)00219-X.

Celik, I., Gurbuz, N., Uncu, A.T., Frary, A., Doganlar, S. (2017) Genome-wide SNP discovery and QTL mapping for fruit quality traits in inbred backcross lines

(IBLs) of *Solanum pimpinellifolium* using genotyping by sequencing. *BMC Genomics*, 18(1): 1-10. https://doi.org/10.1186/s12864-016-3406-7.

Chagas, J.T.B., Vivas, M., Moraes, R., Francelino, H.O., Jorge, T.S., Santa-Catarina, R., Amaral Júnior, A.T. Daher, R.F., Pereira, M.G. (2021) Prospection of genotypes resistant to black spot in half-sib families of papaya. *Agronomy Journal*, 113(6): 4661-4671. https://doi.org/10.1002/agj2.20784.

Chandrika, U.G., Jansz, E.R., Wickramasinghe, S.M.D.N., Warnasuriya, N.D. (2003) Carotenoids in yellow- and red-fleshed papaya (*Carica papaya* L). In *Journal of the Science of Food and Agriculture*, 83(12), 1279-1282. https://doi.org/10.1002/jsfa.1533.

Chen, C., Yu, Q., Hou, S., Li, Y., Eustice, M., Skelton, R.L., Veatch, O., Herdes, R.E., Diebold, L., Saw, J., Feng, Y., Qian, W., Bynum, L., Wang, L., Moore, P.H., Paull, R.E., Alam, M., Ming, R. (2007) Construction of a sequence-tagged high-density genetic map of papaya for comparative structural and evolutionary genomics in Brassicales. *Genetics*, 177(4): 2481-2491. https://doi.org/10.1534/genetics.107.081463.

Cingolani, P., Platts, A., Wang, L.L.L., Coon, M., Nguyen, T., Wang, L.L.L., Land, S.J., Lu, X., Ruden, D.M. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118 ; iso-2; iso-3. *Fly*, 5(1): 29-30. https://doi.org/10.1016/S1877-1203(13)70353-7.

Collard, B.C.Y., Jahufer, M.Z.Z., Brouwer, J.B., Pang, E.C.K. (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica*, 142: 169-196. https://doi.org/10.1007/s10681-005-1681-5.

Cortes, D.F.M., Catarina, R.S., Barros, G.B.A., Arêdes, F.A.S., Silveira, S.F., Ferreguetti, G.A., Ramos, H.C.C., Viana, A.P., Pereira, M.G. (2017) Model-assisted phenotyping by digital images in papaya breeding program. *Scientia Agricola*, 74(4): 294-302. https://doi.org/10.1590/1678-992X-2016-0134.

Dhingani, R.M., Umrania, V.V, Tomar, R.S., Parakhia, M.V, Golakiya, B. (2015)

Introduction to QTL mapping in plants. *Annals of Plant Sciences*, 4: 1072-1079.

Drenkard, E., Richter, B.G., Rozen, S., Stutius, L.M., Angell, N.A., Mindrinos, M., Cho, R.J., Oefner, P.J., Davis, R.W., Ausubel, F.M. (2000) A simple procedure for the analysis of single nucleotide polymorphisms facilitates map-based cloning in Arabidopsis. *Plant Physiology*, 124(4): 1483-1492. https://doi.org/10.1104/pp.124.4.1483.

Evans, E.A., Ballen, F.H. (2012) An Overview of Global Papaya Production, Trade , and Consumption. *Florida Cooperative Extension Service EDIS*, 1-8. <http://edis.ifas.ufl.edu/>. Accessed January 19, 2022.

Fabi, J.P., Broetto, S.G., Silva, S.L.G.L., Zhong, S., Lajolo, F.M., Nascimento, J.R.O. (2014) Analysis of papaya cell wall-related genes during fruit ripening indicates a central role of polygalacturonases during pulp softening. *PLoS ONE*, 9(8). https://doi.org/10.1371/journal.pone.0105685.

Fabi, J.P., Cordenunsi, B.R., Barreto, G.P.D.M., Mercadante, A.Z., Lajolo, F.M., Nascimento, J.R.O. (2007) Papaya fruit ripening: Response to ethylene and 1-methylcyclopropene (1-MCP). *Journal of Agricultural and Food Chemistry*, 55(15): 6118-6123. https://doi.org/10.1021/jf070903c.

Fabi, J.P., Peroni, F.H.G., Gomez, M.L.P.A. (2010) Papaya, mango and guava fruit metabolism during ripening: postharvest changes affecting tropical fruit nutritional content and quality. *Fresh Produce*, 4(1): 56-66.

Fabi, J.P., Seymour, G.B., Graham, N.S., Broadley, M.R., May, S.T., Lajolo, F.M., Cordenunsi, B.R., Nascimento, J.R.O. (2012) Analysis of ripening-related gene expression in papaya using an Arabidopsis-based microarray Analysis of ripening-related gene expression in papaya using an Arabidopsis-based microarray. *BMC Plant Biology*, 12(1): 242.

FAOSTAT. (2022) *Food and Agriculture Organization of the United Nations*. <http://www.fao.org/faostat/en/#data/QC/visualize>. Accessed December 11, 2021.

Ferreguetti, G.A. (2003) Caliman 01 - O Primeiro hibrido de Mamão formosa

Brasileiro. *Papaya Brasil*, 211-218.

Flint-Garcia, S.A., Thornsberry, J.M., Buckler, E.S. (2003) Structure of Linkage Disequilibrium in Plants. *Annual Review of Plant Biology*, 54(1): 357-374. https://doi.org/10.1146/annurev.arplant.54.031902.134907.

Gamboa-Tuz, S.D., Pereira-Santana, A., Zamora- Briseño, J.A., Castano, E., Espadas-Gil, F., Ayala-Sumuano, J.T., Keb-Llanes, M.A., Sanchez-Teyer, F., Rodríguez-Zapata, L.C. (2018) Transcriptomics and co-expression networks reveal tissue-specific responses and regulatory hubs under mild and severe drought in papaya *(Carica papaya L .)*. *Scientific Reports*, 8:14539. https://doi.org/10.1038/s41598-018-32904-2.

Gapper, N. E., McQuinn, R.P., Giovannoni, J.J. (2013) Molecular and genetic regulation of fruit ripening. *Plant Molecular Biology*, 82(6): 575-591. https://doi.org/10.1007/s11103-013-0050-3.

Huq, A., Akter, S., Nou, S., Kim, H.T., Jung, Y.J., Kang, K.K. (2016) Identification of functional SNPs in genes and their effects on plant phenotypes. *Journal of Plant Biotechnology*, 43(1): 1-11. https://doi.org/10.5010/JPB.2016.43.1.1.

IBGE. Instituto Brasileiro de Geografia e Estatística. *Produção Agrícola Municipal* (2020) <http://www.cnpmf.embrapa.br/Base_de_Dados/index_pdf/dados/brasil/mamao/b1_mamao.pdf>. Accessed January 19, 2022.

Jamaluddin, N.D., Noor, N.M., Goh, H.H. (2017) Genome-wide transcriptome profiling of *Carica papaya* L. embryogenic callus. *Physiology and Molecular Biology of Plants*, 23(2): 357-368. https://doi.org/10.1007/s12298-017-0429-8.

Jiménez-Galindo, J.C., Ordás, B., Butrón, A., Samayoa, L.F., Malvar, R.A. (2017) QTL Mapping for Yield and Resistance against Mediterranean Corn Borer in Maize. *Frontiers in Plant Science*, 8: 2-11. https://doi.org/10.3389/fpls.2017.00698.

Kawuki, R.S., Ferguson, M., Labuschagne, M., Herselman, L., Kim, D.J. (2009) Identification, characterisation and application of single nucleotide polymorphisms for diversity assessment in cassava (*Manihot esculenta*

Crantz). *Molecular Breeding*, 23(4): 669-684. https://doi.org/10.1007/s11032-009-9264-0.

Khlestkina, E.K., Salina, E.A. (2006) SNP markers: Methods of analysis, ways of development, and comparison on an example of common wheat. *Russian Journal of Genetics*, 42: 585-594 https://doi.org/10.1134/S1022795406060019

Kim, M.S., Moore, P.H., Zee, F., Fitch, M.M.M., Steiger, D.L., Manshardt, R.M., Paull, R.E., Drew, R.A., Sekioka, T., Ming, R. (2002) Genetic diversity of *Carica papaya* as revealed by AFLP markers. *Genome*, 45: 503-512. https://doi.org/10.1139/G02-012.

Kim, S.R., Ramos, J., Ashikari, M., Virk, P.S., Torres, E.A., Nissila, E., Hechanova, S.L., Mauleon, R., Jena, K.K. (2016) Development and validation of allele-specific SNP/indel markers for eight yield-enhancing genes using whole-genome sequencing strategy to increase yield potential of rice, *Oryza sativa* L. *Rice*, 9(1): 12. https://doi.org/10.1186/s12284-016-0084-7.

Langmead, B., Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods*, 9(4): 357-359. https://doi.org/10.1038/nmeth.1923.Fast.

Lee, C.Y., Lin, H.J., Viswanath, K.K., Lin, C.P., Chang, B.C.H., Chiu, P.H., Chiu, C.T., Wang, R.H., Chin, S.W., Chen, F.C. (2018) The development of functional mapping by three sex-related loci on the third whorl of different sex types of *Carica papaya* L. *PLoS ONE*, 13(3): 1-21. https://doi.org/10.1371/journal.pone.0194605.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., Data, G.P., Sam, T. (2009) The Sequence Alignment / Map format and SAMtools. *Bioinformatics*, 25(16): 2078-2079. https://doi.org/10.1093/bioinformatics/btp352.

Li, Y., Yang, K., Yang, W., Chu, L., Chen, C., Zhao, B., Li, Y., Jian, J., Yin, Z., Wang, T., Wan, P. (2017) Identification of QTL and Qualitative Trait Loci for Agronomic Traits Using SNP Markers in the Adzuki Bean. *Frontiers in Plant Science*, 8:840. https://doi.org/10.3389/fpls.2017.00840.

Liao, Z., Yu, Q., Ming, R. (2017) Development of male-specific markers and

identification of sex reversal mutants in papaya. *Euphytica*, 213(2): 1-12. https://doi.org/10.1007/s10681-016-1806-z.

Liu, L., Qu, C., Wittkop, B., Yi, B., Xiao, Y., He, Y., Snowdon, R.J., Li, J. (2013) A high-density snp map for accurate mapping of seed fibre qtl in *Brassica napus* l. *PLoS ONE*, 8(12): 1-9. https://doi.org/10.1371/journal.pone.0083052.

Liu, M., Pirrello, J., Chervin, C., Roustan, J.P., Bouzayen, M. (2015) Ethylene control of fruit ripening: Revisiting the complex network of transcriptional regulation. *Plant Physiology*, 169(4): 2380-2390. https://doi.org/10.1104/pp.15.01361.

Liu, X., Geng, X., Zhang, H., Shen, H., Yang, W. (2017) Association and genetic identification of loci for four fruit traits in tomato using InDel markers. *Frontiers in Plant Science*, 8: 1-14. https://doi.org/10.3389/fpls.2017.01269.

Liu, Z., Zhu, H., Liu, Y., Kuang, J., Zhou, K., Liang, F., Liu, Z., Wang, D., Ke, W. (2016) Construction of a high-density, high-quality genetic map of cultivated lotus (*Nelumbo nucifera*) using next-generation sequencing. *BMC Genomics*, 17:466. https://doi.org/10.1186/s12864-016-2781-4.

Lü, P., Yu, S., Zhu, N., Chen, Y., Zhou, B., Pan, Y., Tzeng, D. (2018) Genome encode analyses reveal the basis of convergent evolution of fleshy fruit ripening. *Nature Plants*, 4, 784-791. https://doi.org/10.1038/s41477-018-0249-z.

Luiz, J., Dantas, L., Oliveira, E. J., Pereira, M. G. (2000) *Melhoramento genético do mamoeiro no Brasil*.

Luo, C., Shu, B., Yao, Q., Wu, H., Xu, W., Wang, S. (2016) Construction of a high-density genetic map based on large-scale marker development in mango using specific-locus amplified fragment sequencing (SLAF-seq). *Frontiers in Plant Science*, 7: 1-8. https://doi.org/10.3389/fpls.2016.01310.

Ma, H., Moore, P.H., Liu, Z., Kim, M.S., Yu, Q., Fitch, M.M.M., Sekioka, T., Paterson, A.H., Ming, R. (2004) High-Density Linkage Mapping Revealed Suppression of Recombination at the Sex Determination Locus in Papaya. *Genetics*, 166(1): 419-436. https://doi.org/10.1534/genetics.166.1.419.

MAPA. (2022) *Ministério da Agricultura, Pecuária e Abastecimento - Registro Nacional de Cultivares (RNC)*. <http://sistemas.agricultura.gov.br/snpc/cultivarweb/cultivares_registradas.php> Accessed January 4, 2022.

Marin, S.L.D., Pereira, M.G., Amaral Júnior, A.T., Martelleto, L.A.P., Ide, C.D. (2006) Partial diallel to evaluate the combining ability for economically important traits of papaya. *Scientia Agricola*, 63(6): 540-546. https://doi.org/10.1590/S0103-90162006000600005.

Martínez-García, P.J., Parfitt, D.E., Ogundiwin, E.A., Fass, J., Chan, H.M., Ahmad, R., Lurie, S., Dandekar, A., Gradziel, T.M., Crisosto, C.H. (2013) High density SNP mapping and QTL analysis for fruit quality characteristics in peach (*Prunus persica* L.). *Tree Genetics & Genomes*, 9(1): 19-36. https://doi.org/10.1007/s11295-012-0522-7.

Martín-Pizarro, C., Posé, D. (2018) Genome editing as a tool for fruit ripening manipulation. *Frontiers in Plant Science*, 9: 1-8. https://doi.org/10.3389/fpls.2018.01415.

Ming, R., Hou, S., Feng, Y., Yu, Q., Dionne-Laporte, A., Saw, J.H., Senin, P., Wang, W., Ly, B.V., Lewis, K.L.T., Salzberg, S.L., Feng, L., Jones, M.R., Skelton, R.L., Murray, J.E., Chen, C., Qian, W., Shen, J., Du, P., Alam, M. et al. (2008) The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). *Nature*, 452(7190): 991-996. https://doi.org/10.1038/nature06856.

Ming, R., Yu, Q., Moore, P.H. (2007) Sex determination in papaya. *Seminars in Cell and Developmental Biology*, 18(3): 401-408. https://doi.org/10.1016/j.semcdb.2006.11.013.

Mondini, L., Noorani, A., Pagnotta, M.A. (2009) Assessing plant genetic diversity by molecular tools. *Diversity*, 1(1): 19-35. https://doi.org/10.3390/d1010019.

Montero-Pau, J., Blanca, J., Esteras, C., Martínez-Pérez, E.M., Gómez, P., Monforte, A.J., 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(1): 1-21. https://doi.org/10.1186/s12864-016-3439-y.

Moraes, R., Vivas, M., Vivas, J.M.S., Daher, R.F., Gravina, G.A., Francelino, H.O., Cruz, D.P., Ramos, G.K.S., Pereira, M.G. (2019) Genetic parameters and performance of papaya genotypes to black spot resistance (*Asperisporium caricae*). *Australian Journal of Crop Science*, 13(5): 649-655. https://doi.org/10.21475/ajcs.19.13.05.p1097.

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

Nantawan, U., Kanchana-udomkan, C., Bar, I., Ford, R. (2019) Linkage mapping and quantitative trait loci analysis of sweetness and other fruit quality traits in papaya. *BMC Plant Biology*, 19(1): 1-11. https://doi.org/10.1186/s12870-019-2043-0.

Nuñez-Lillo, G., Balladares, C., Pavez, C., Urra, C., Sanhueza, D., Vendramin, E., Dettori, M.T., Arús, P., Verde, I., Blanco-Herrera, F., Campos-Vargas, R., Meneses, C. (2019) High-density genetic map and QTL analysis of soluble solid content, maturity date, and mealiness in peach using genotyping by sequencing. *Scientia Horticulturae*, 257: 108734. https://doi.org/10.1016/j.scienta.2019.108734.

Oliveira, E.J., Ferreira, C.F., Santos, V.S., Jesus, O.N., Oliveira, G.A.F., Silva, M.S. (2014) Potential of SNP markers for the characterization of Brazilian cassava germplasm. *Theoretical and Applied Genetics*, 127(6): 1423-1440. https://doi.org/10.1007/s00122-014-2309-8.

Pereira, M. G., Santa-Catarina, R. (2021) Recurrent selection in papaya: An effective strategy for the continuous development of new cultivars Recurrent selection in papaya: An effective strategy for the continuous development of new cultivars. *Crop Breeding and Applied Biotechnology*, 21(S): 381321-

381328. https://doi.org/10.1590/1984.

Pereira, M.G., Ferreguetti, G.A., Pereira, T.N.S., Ramos, H.C.C., Cortes, D.F.M., Santa-Catarina, R., Luz, L.N., Barros, G.B.A., Arêdes, F.A.S., Boechat, M.S.B., Silveira, S.F., Oliveira, J.G. (2015) Melhoramento genético do mamoeiro: programa Uenf/Caliman. *In*: *Anais do VI Simpósio do papaya brasileiro*. Vitória, 6: p. 1-41.

Pereira, M.G., Ferreguetti, G.A., Ramos, H.C.C., Santa-Catarina, R., Vettorazzi, C.J.F., Poltronieri, T.P.S., et al. (2018) *Melhoramento genético do mamoeiro : programa UENF / Caliman*. *In*: *Anais do VII Simpósio do papaya brasileiro*. Vitória, p. 1-59.

Pereira, M.G., Luz, L.N., Santa-Catarina, R., Ramos, H.C.C., Pereira, T.N.S., Barros, G.B., Ferregueti, G.A., Vivas, M., Cortes, D.F.M., Vettorazzi, J.C.F., Azevedo, A.O.N., Silveira, S.F., Oliveira, J.G., Viana, A.P. (2019) UC10: A new early Formosa papaya cultivar. *Crop Breeding and Applied Biotechnology*, 19(1): 131-134. https://doi.org/10.1590/1984-70332019v19n1c18.

Pesik, A., Efendi, D., Novarianto, H., Dinarti, D., Sudarsono, S. (2017) Development of SNAP markers based on nucleotide variability of WRKY genes in coconut and their validation using multiplex PCR. *Biodiversitas*, 18(2): 465-475. https://doi.org/10.13057/biodiv/d180204.

Pirovani, A.A.V., Ramos, H.C.C., Santa-Catarina, R., Poltronieri, T.P.S., Vettorazzi, J.C.F., Santos, P.H.D., Santana, J.G.S., Duarte, R.P., Vivas, M., Pereira, M.G. (2021) New source of alleles for resistance to black spot and phoma spot in papaya (*Carica papaya* L.). *Euphytica*, 217(6). https://doi.org/10.1007/s10681-021-02863-3.

Poltronieri, T.P.S., Silveira, S.F., Santa-Catarina, R., Cortes, D.F.M., Azevedo, A.O.N., Pereira, M.G. (2017) Selecting black-spot resistant papaya genotypes derived from backcrossing and hybrids Selecting black-spot resistant papaya genotypes derived from backcrossing and hybrids. *Genetics and Molecular Research*, 16(1). https://doi.org/10.4238/gmr16019401.

Pootakham, W., Jomchai, N., Ruang-areerate, P., Shearman, J.R., Sonthirod, C., Sangsrakru, D., Tragoonrung, S., Tangphatsornruang, S. (2015) Genome-wide SNP discovery and identification of QTL associated with agronomic traits in oil palm using genotyping-by-sequencing (GBS). *Genomics*, 105(5-6): 288-295. https://doi.org/10.1016/j.ygeno.2015.02.002.

Quintal, S.S.R. (2009) *Caracterização E Avaliação De Um Banco De Germoplasma De Mamoeiro Para Estudo Dos Parâmetros Genéticos E Diversidade Genética*. Thesis (M.Sc. in Genetics and Plant Breeding) – Campos dos Goytacazes – RJ, Universidade Estadual do Norte Fluminense – UENF, 168p.

Ramos, H.C.C. (2010) *Retrocruzamento em mamoeiro: avanço de gerações e seleção de linhagens via procedimentos clássicos e moleculares*. Thesis (D.Sc. in Genetics and Plant Breeding) – Campos dos Goytacazes – RJ, Universidade Estadual do Norte Fluminense – UENF, 134p.

Ramos, H.C.C., Pereira, M.G., Gonçalves, L.S.A., Amaral Júnior, A.T., Scapim, C.A. (2011) Comparison of multiallelic distances for the quantification of genetic diversity in the papaya. *Acta Scientiarum. Agronomy*, 33(1): 59-66. https://doi.org/10.4025/actasciagron.v33i1.11698.

Ramos, H.C.C., Pereira, M.G., Pereira, T.N.S., Barros, G.B.A., Ferreguetti, G.A. (2014) Genetic-molecular characterization of backcross generations for sexual conversion in papaya (*Carica papaya* L.). *Genetics and Molecular Research*, 13(4): 10367-10381. https://doi.org/10.4238/2014.December.4.32.

Ramos, H.C.C., Pereira, M.G., Silva, F.F., Gonçalves, L.S., Pinto, F.O., Souza Filho, G.A., Pereira, T.S. (2011) Genetic characterization of papaya plants (*Carica papaya* L.) derived from the first backcross generation. *Genetics and Molecular Research*, 10(1): 393-403. https://doi.org/10.4238/vol10-1gmr1004.

Ramos, H.C.C., Pereira, M.G., Silva, F.F., Viana, A.P., Ferreguetti, G.A. (2011) Seasonal and genetic influences on sex expression in a backcrossed segregating papaya population. *Crop Breeding and Applied Biotechnology (Online)*, 11(2): 97-105. https://doi.org/10.1590/S1984-70332011000200001.

Santa-Catarina, R., Cortes, D.F.M., Vettorazzi, J.C.F., Ramos, H.C.C., Ferreguetti, G.A., Pereira, M.G. (2018) Image-based phenotyping of morpho-agronomic traits in papaya fruits (*Carica papaya* L. THB var.). *Australian Journal of Crop Science*, 12(11): 1750-1756. https://doi.org/10.21475/ajcs.18.12.11.p1437.

Santa-Catarina, R., Pereira, M.G., Vettorazzi, J.C.F., Cortes, D.F.M., Poltronieri, T.P.S., Azevedo, A.O.N., Moreira, N.F., Miranda, D.P., Moraes, R., Pirovani, A.A.V., Ramos, H.C.C., Vivas, M., Viana, A.P. (2020) Papaya (*Carica papaya* L.) S1 family recurrent selection: Opportunities and selection alternatives from the base population. *Scientia Horticulturae*, 260. https://doi.org/10.1016/j.scienta.2019.108848.

Schreiber, L., Nader-Nieto, A.C., Schönhals, E.M., Walkemeier, B., Gebhardt, C. (2014) SNPs in genes functional in starch-sugar interconversion associate with natural variation of tuber starch and sugar content of potato (*Solanum tuberosum* L.). *G3: Genes, Genomes, Genetics*, 4(10): 1797-1811. https://doi.org/10.1534/g3.114.012377.

Shen, Y.H., Lu, B.G., Feng, L., Yang, F.Y., Geng, J.J., Ming, R., Chen, X.J. (2017) Isolation of ripening-related genes from ethylene/1-MCP treated papaya through RNA-seq. *BMC Genomics*, 18(1): 1-13. https://doi.org/10.1186/s12864-017-4072-0.

Shen, Y.H., Yang, F.Y., Lu, B.G., Zhao, W.W., Jiang, T., Feng, L., Chen, X.J., Ming, R. (2019) Exploring the differential mechanisms of carotenoid biosynthesis in the yellow peel and red flesh of papaya. *BMC Genomics*, 20:49 1-11.

Silva, F.F., Pereira, M.G., Campos, W.F., Damasceno, P.C., Pereira, T.N.S., De Souza Filho, G.A., Ramos, H.C.C., Viana, A.P., Ferreguetti, G.A. (2007) DNA marker-assisted sex conversion in elite papaya genotype (*Carica papaya* L.). *Crop Breeding and Applied Biotechnology*, 7(1): 52-58. https://doi.org/10.12702/1984-7033.v07n01a08.

Souza, L.M., Ferreira, K.S., Chaves, J.B.P., Teixeira, S.L. (2008) L-ascorbic acid, B-carotene and lycopene content in papaya fruits (*Carica papaya*) with or without physiological skin freckles. *Scientia Agricola*, 65(3): 246-250.

https://doi.org/10.1590/S0103-90162008000300004.

Steemers, F.J., Gunderson, K.L. (2007) *Whole* genome genotyping technologies on the BeadArray™ platform. *Biotechnology Journal*, 2(1):41-49. https://doi.org/10.1002/biot.200600213.

Sun, R., Chang, Y., Yang, F., Wang, Y., Li, H., Zhao, Y., Chen, D., Wu, T., Zhang, X., Han, Z. (2015) A dense SNP genetic map constructed using restriction site-associated DNA sequencing enables detection of QTLs controlling apple fruit quality. *BMC Genomics*, 16(1): 1-15. https://doi.org/10.1186/s12864-015-1946-x.

Tzuri, G., Zhou, X., Chayut, N., Yuan, H., Portnoy, V., Meir, A., Sa'Ar, U., Baumkoler, F., Mazourek, M., Lewinsohn, E., Fei, Z., Schaffer, A.A., Li, L., Burger, J., Katzir, N., Tadmor, Y. (2015) A "golden" SNP in CmOr governs the fruit flesh color of melon (*Cucumis melo*). *Plant Journal*, 82(2): 267-279. https://doi.org/10.1111/tpj.12814.

Vale, E.M., Reis, R.S., Santa-Catarina, R., Pereira, M.G., Santa-Catarina, C., Silveira, V. (2016) Comparative proteomic analysis of the heterosis phenomenon in papaya roots. *Scientia Horticulturae*, 209: 178-186. https://doi.org/10.1016/j.scienta.2016.06.031.

VanBuren, R., Wai, C.M., Zhang, J., Han, J., Arro, J., Lin, Z., Liao, Z., Yu, Q., Wang, M.L., Zee, F., Moore, R.C., Charlesworth, D., Ming, R. (2016) Extremely low nucleotide diversity in the X-linked region of papaya caused by a strong selective sweep. *Genome Biology*, 17(1): 1-11. https://doi.org/10.1186/s13059-016-1095-9.

VanBuren, R., Zeng, F., Chen, C., Zhang, J., Wai, C. M., Han, J., Aryal, R., Gschwend, A.R., Wang, J., Na, J.K., Huang, L., Zhang, L., Miao, W., Gou, J., Arro, J., Guyot, R., Moore, R. C., Wang, M. L., Zee, F., Ming, R., et al. (2015) Origin and domestication of papaya Y^h chromosome. *Genome Research*, 25(4): 524-533. https://doi.org/10.1101/gr.183905.114.

Vidal, N.M., Grazziotin, A.L., Ramos, H.C.C., Pereira, M.G., Venancio, T.M. (2014) Development of a gene-centered ssr atlas as a resource for papaya (*Carica*

papaya) marker-assisted selection and population genetic studies. *PLoS ONE*, 9(11). https://doi.org/10.1371/journal.pone.0112654.

Vivas, M., Cardoso, D.L., Ramos, H.C.C., Santos, P.H.D., Silveira, S.F., Moraes, R., Pereira, M.G. (2018) Genetic diversity between papaya lines and their correlation with heterosis in hybrids for disease resistance and morpho-agronomic traits. *Summa Phytopathologica*, 44(2): 110-115. https://doi.org/10.1590/0100-5405/176828.

Vivas, M., Felipe, S., Maganha, J., Vivas, S., Henrique, P., Murizi, B., Daher, R. F., Teixeira, A. (2017) Phenotypic characterization of papaya genotypes to determine powdery mildew resistance. *Crop Breeding and Applied Biotechnology*, 17: 198-205.

Vivas, M., Ramos, H.C.C., Santos, P.H.D., Silveira, S.F., Pereira, T.N.S., Amaral, A.T., Pereira, M.G. (2016) Heterosis and genetic diversity for selection of papaya hybrids for resistance to black spot and phoma spot. *Tropical Plant Pathology*, 41(6): 380-389. https://doi.org/10.1007/s40858-016-0109-1.

Vivas, M., Silveira, S.F., Viana, A.P., Amaral, A.T., Cardoso, D.L., Pereira, M.G. (2014) Efficiency of circulant diallels via mixed models in the selection of papaya genotypes resistant to foliar fungal diseases. *Genetics and Molecular Research*, 13(3): 4797-4804. https://doi.org/10.4238/2014.July.2.9.

Vivas, M., Silveira, S.F., Viana, A.P., Amaral-Júnior, A.T., Ferreguetti, G.A., Pereira, M.G. (2015) Resistance to multiple foliar diseases in papaya genotypes in Brazil. *Crop Protection*, 71: 138-143. https://doi.org/10.1016/j.cropro.2015.02.007.

Wang, J., Na, J.K., Yu, Q., Gschwend, A.R., Han, J., Zeng, F., Aryal, R., VanBuren, R., Murray, J.E., Zhang, W., Navajas-Pérez, R., Feltus, F.A., Lemke, C., Tong, E.J., Chen, C., Wai, C.M., Singh, R., Wang, M.L., Min, X. J., Ming, R., et al. (2012) Sequencing papaya X and Y h chromosomes reveals molecular basis of incipient sex chromosome evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 109(34): 13710-13715. https://doi.org/10.1073/pnas.1207833109.

Wang, W., Cai, J., Wang, P., Tian, S., Qin, G. (2017) Post-transcriptional regulation of fruit ripening and disease resistance in tomato by the vacuolar protease SlVPE3. *Genome Biology*, 18(1): 1-23. https://doi.org/10.1186/s13059-017-1178-2.

Xu, Y., Crouch, J.H. (2008) Marker-assisted selection in plant breeding: From publications to practice. *Crop Science*, 48(2): 391-407. https://doi.org/10.2135/cropsci2007.04.0191.

Xu, Y., Li, P., Yang, Z., Xu, C. (2016) Genetic mapping of quantitative trait loci in crops. *The Crop Journal*, *5*(2): 175-184. https://doi.org/10.1016/j.cj.2016.06.003.

Ye, S., Dhillon, S., Ke, X., Collins, A.R., Day, I.N.M. (2001) An efficient procedure for genotyping single nucleotide polymorphisms. *Nucleic Acids Research*, 29(17): e88-e88. https://doi.org/10.1093/nar/29.17.e88.

Yu, Q., Tong, E., Skelton, R.L., Bowers, J.E., Jones, M.R., Murray, J.E., Hou, S., Guan, P., Acob, R.A., Luo, M.C., Moore, P.H., Alam, M., Paterson, A.H., Ming, R. (2009) A physical map of the papaya genome with integrated genetic map and genome sequence. *BMC Genomics*, 10: 371. https://doi.org/10.1186/1471-2164-10-371.

Zhang, W., Wai, C.M., Ming, R., Yu, Q., Jiang, J. (2010) Integration of Genetic and Cytological Maps and Development of a Pachytene Chromosome-based Karyotype in Papaya. *Tropical Plant Biology*, 3(3): 166-170. https://doi.org/10.1007/s12042-010-9053-2.

APPENDIX

Table 1A. Selected 106 (48 DGEs and 143 identified by BLASTp) ripening-related genes linked with 460 variants (438 SNPs and 22 InDels). Variant alleles of the Formosa elite lines Sekati and JS-12, variant position, variant effects and impact, gene type and function. CCM: chlorophyll and carotenoid metabolism-related genes; PH: plant hormone signal transduction pathway genes; TF: transcription factor. Table 1A – Cont.

Gene.ID	Gene definition	Variant Position	Alleles Sekati	Alleles JS.12	Gene type	Function	Variant Type	EFFECT	IMPACT
evm.TU.supercontig_107.106	beta-hydroxylase 1	564740	G	A	DGE	CCM	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_107.106	beta-hydroxylase 1	568815	A	T	DGE	CCM	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_107.106	beta-hydroxylase 1	573466	G	T	DGE	CCM	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_107.106	beta-hydroxylase 1	574860	A	T	DGE	CCM	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_13.164	magnesium-protoporphyrin IX methyltransferase	1488414	C	G	DGE	CCM	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_130.4	Flavin containing amine oxidoreductase family	17190	G	A	DGE	CCM	SNP	intron_variant	MODIFIER
evm.TU.supercontig_130.4	Flavin containing amine oxidoreductase family	24443	C	T	DGE	CCM	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_130.4	Flavin containing amine oxidoreductase family	25036	T	A	DGE	CCM	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_130.4	Flavin containing amine oxidoreductase family	26002	T	C	DGE	CCM	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_130.4	Flavin containing amine oxidoreductase family	26025	T	C	DGE	CCM	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_130.4	Flavin containing amine oxidoreductase family	28607	A	G	DGE	CCM	SNP	intergenic_region	MODIFIER

Table 1A – Cont.

Gene.ID	Gene definition	Variant Position	Alleles Sekati	Alleles JS.12	Gene type	Function	Variant Type	EFFECT	IMPACT
evm.TU.supercontig_130.4	Flavin containing amine oxidoreductase family	28622	T	C	DGE	CCM	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_132.5	Lycopene beta/epsilon cyclase protein	24283	A	G	DGE	CCM	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_43.43	hydroxymethylbilane synthase	555771	A	G	DGE	CCM	SNP	intron_variant	MODIFIER
evm.TU.supercontig_43.43	hydroxymethylbilane synthase	560154	T	A	DGE	CCM	SNP	intron_variant	MODIFIER
evm.TU.supercontig_43.43	hydroxymethylbilane synthase	561461	T	C	DGE	CCM	SNP	intron_variant	MODIFIER
evm.TU.supercontig_43.43	hydroxymethylbilane synthase	563512	T	C	DGE	CCM	SNP	intron_variant	MODIFIER
evm.TU.supercontig_43.43	hydroxymethylbilane synthase	563532	A	G	DGE	CCM	SNP	intron_variant	MODIFIER
evm.TU.supercontig_43.43	hydroxymethylbilane synthase	564024	C	G	DGE	CCM	SNP	intron_variant	MODIFIER
evm.TU.supercontig_43.43	hydroxymethylbilane synthase	564071	T	C	DGE	CCM	SNP	intron_variant	MODIFIER
evm.TU.supercontig_43.43	hydroxymethylbilane synthase	564087	T	C	DGE	CCM	SNP	intron_variant	MODIFIER
evm.TU.supercontig_43.43	hydroxymethylbilane synthase	565342	T	C	DGE	CCM	SNP	intron_variant	MODIFIER
evm.TU.supercontig_43.43	hydroxymethylbilane synthase	565368	A	G	DGE	CCM	SNP	intron_variant	MODIFIER

Table 1A – Cont.

Gene.ID	Gene definition	Variant Position	Alleles Sekati	Alleles JS.12	Gene type	Function	Variant Type	EFFECT	IMPACT
evm.TU.supercontig_43.43	hydroxymethylbilane synthase	565491	G	T	DGE	CCM	SNP	intron_variant	MODIFIER
evm.TU.supercontig_43.43	hydroxymethylbilane synthase	565752	A	G	DGE	CCM	SNP	intron_variant	MODIFIER
evm.TU.supercontig_43.43	hydroxymethylbilane synthase	565764	C	T	DGE	CCM	SNP	intron_variant	MODIFIER
evm.TU.supercontig_43.43	hydroxymethylbilane synthase	567614	T	C	DGE	CCM	SNP	intron_variant	MODIFIER
evm.TU.supercontig_43.43	hydroxymethylbilane synthase	567707	A	G	DGE	CCM	SNP	intron_variant	MODIFIER
evm.TU.supercontig_43.43	hydroxymethylbilane synthase	567754	T	G	DGE	CCM	SNP	intron_variant	MODIFIER
evm.TU.supercontig_43.43	hydroxymethylbilane synthase	569749	A	G	DGE	CCM	SNP	synonymous_variant	LOW
evm.TU.supercontig_43.43	hydroxymethylbilane synthase	569757	G	C	DGE	CCM	SNP	missense_variant	MODERATE
evm.TU.supercontig_44.117	chlorophyllase 2	1496810	A	C	DGE	CCM	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_44.117	chlorophyllase 2	1503918	A	G	DGE	CCM	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_47.4	Pheophorbide a oxygenase family protein with Rieske [2Fe-2S] domain	21463	A	G	DGE	CCM	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_80.29	protochlorophyllide oxidoreductase A	231119	T	G	DGE	CCM	SNP	downstream_gene_variant	MODIFIER

Table 1A – Cont.

Gene.ID	Gene definition	Variant Position	Alleles Sekati	Alleles JS.12	Gene type	Function	Variant Type	EFFECT	IMPACT
evm.TU.supercontig_80.29	protochlorophyllide oxidoreductase A	232275	T	C	DGE	CCM	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_80.29	protochlorophyllide oxidoreductase A	233024	A	T	DGE	CCM	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_80.29	protochlorophyllide oxidoreductase A	233388	G	C	DGE	CCM	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_80.29	protochlorophyllide oxidoreductase A	234005	G	A	DGE	CCM	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_80.29	protochlorophyllide oxidoreductase A	234551	G	A	DGE	CCM	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_80.29	protochlorophyllide oxidoreductase A	235440	TA	TAACTAA	DGE	CCM	INDEL	downstream_gene_variant	MODIFIER
evm.TU.supercontig_92.51	magnesium-chelatase subunit chlH, chloroplast, putative / Mg-protoporphyrin IX chelatase, putative (CHLH)	466393	G	A	DGE	CCM	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_92.51	magnesium-chelatase subunit chlH, chloroplast, putative / Mg-protoporphyrin IX chelatase, putative (CHLH)	466550	C	T	DGE	CCM	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_92.51	magnesium-chelatase subunit chlH, chloroplast, putative / Mg-protoporphyrin IX chelatase, putative (CHLH)	466569	A	C	DGE	CCM	SNP	downstream_gene_variant	MODIFIER

Table 1A – Cont.

Gene.ID	Gene definition	Variant Position	Alleles Sekati	Alleles JS.12	Gene type	Function	Variant Type	EFFECT	IMPACT
evm.TU.supercontig_92.51	magnesium-chelatase subunit chlH, chloroplast, putative / Mg-protoporphyrin IX chelatase, putative (CHLH)	470716	G	A	DGE	CCM	SNP	synonymous_variant	LOW
evm.TU.supercontig_34.133	protochlorophyllide oxidoreductase C	1241845	G	A	BLAST	CCM	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_1.14	xyloglucan endotransglucosylase/hydrolase 32	343623	G	A	DGE	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_1.14	xyloglucan endotransglucosylase/hydrolase 32	345186	T	C	DGE	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_1.14	xyloglucan endotransglucosylase/hydrolase 32	347097	T	G	DGE	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_1.14	xyloglucan endotransglucosylase/hydrolase 32	350897	G	A	DGE	CW	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_1.14	xyloglucan endotransglucosylase/hydrolase 32	370928	G	A	DGE	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_1.14	xyloglucan endotransglucosylase/hydrolase 32	372105	A	G	DGE	CW	SNP	intergenic_region	MODIFIER

Table 1A – Cont.

Gene.ID	Gene definition	Variant Position	Alleles Sekati	Alleles JS.12	Gene type	Function	Variant Type	EFFECT	IMPACT
evm.TU.supercontig_1.14	xyloglucan endotransglucosylase/hydrolase 32	373711	C	T	DGE	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_1.14	xyloglucan endotransglucosylase/hydrolase 32	375857	G	A	DGE	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_1.14	xyloglucan endotransglucosylase/hydrolase 32	376825	G	A	DGE	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_1.14	xyloglucan endotransglucosylase/hydrolase 32	380721	T	C	DGE	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_1.14	xyloglucan endotransglucosylase/hydrolase 32	380727	C	T	DGE	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_1.14	xyloglucan endotransglucosylase/hydrolase 32	380778	AT	A	DGE	CW	INDEL	intergenic_region	MODIFIER
evm.TU.supercontig_1.419	expansin A4	5164170	T	C	DGE	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_1.419	expansin A4	5165259	T	A	DGE	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_1.419	expansin A4	5166512	A	G	DGE	CW	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_1.419	expansin A4	5166597	T	C	DGE	CW	SNP	downstream_gene_variant	MODIFIER

Table 1A – Cont.

Gene.ID	Gene definition	Variant Position	Alleles Sekati	Alleles JS.12	Gene type	Function	Variant Type	EFFECT	IMPACT
evm.TU.supercontig_1.419	expansin A4	5166611	T	TA	DGE	CW	INDEL	downstream_gene_variant	MODIFIER
evm.TU.supercontig_106.54	Glycosyl hydrolase family protein	639699	C	A	DGE	CW	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_106.54	Glycosyl hydrolase family protein	642393	G	A	DGE	CW	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_106.54	Glycosyl hydrolase family protein	643444	C	T	DGE	CW	SNP	synonymous_variant	LOW
evm.TU.supercontig_106.54	Glycosyl hydrolase family protein	643459	T	G	DGE	CW	SNP	synonymous_variant	LOW
evm.TU.supercontig_106.54	Glycosyl hydrolase family protein	643531	T	G	DGE	CW	SNP	synonymous_variant	LOW
evm.TU.supercontig_106.54	Glycosyl hydrolase family protein	644651	C	T	DGE	CW	SNP	intron_variant	MODIFIER
evm.TU.supercontig_1178.2	sugar transporter 1	8546	T	C	DGE	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_1178.2	sugar transporter 1	9202	T	C	DGE	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_1178.2	sugar transporter 1	9343	C	T	DGE	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_145.20	Pectin lyase-like superfamily protein	472205	T	G	DGE	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_145.20	Pectin lyase-like superfamily protein	479498	T	C	DGE	CW	SNP	downstream_gene_variant	MODIFIER

Table 1A – Cont.

Gene.ID	Gene definition	Variant Position	Alleles Sekati	Alleles JS.12	Gene type	Function	Variant Type	EFFECT	IMPACT
evm.TU.supercontig_145.20	Pectin lyase-like superfamily protein	479501	A	G	DGE	CW	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_1461.3	pectin methylesterase 3	11629	G	A	DGE	CW	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_1461.3	pectin methylesterase 3	11714	G	A	DGE	CW	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_183.26	UDP-galactose transporter 3	249112	G	C	DGE	CW	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_2.231	glycosyl hydrolase 9B1	3044262	G	A	DGE	CW	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_2.231	glycosyl hydrolase 9B1	3047903	T	C	DGE	CW	SNP	intron_variant	MODIFIER
evm.TU.supercontig_208.1	xyloglucan endotransglucosylase/hydrolase 30	31562	C	G	DGE	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_250.6	Pectin lyase-like superfamily protein	27755	G	C	DGE	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_3.313	Glycosyl hydrolases family 32 protein	2176464	G	A	DGE	CW	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_3.313	Glycosyl hydrolases family 32 protein	2180909	G	C	DGE	CW	SNP	missense_variant	MODERATE
evm.TU.supercontig_3.313	Glycosyl hydrolases family 32 protein	2181598	A	G	DGE	CW	SNP	intron_variant	MODIFIER
evm.TU.supercontig_3.313	Glycosyl hydrolases family 32 protein	2188357	T	C	DGE	CW	SNP	upstream_gene_variant	MODIFIER

Table 1A – Cont.

Gene.ID	Gene definition	Variant Position	Alleles Sekati	Alleles JS.12	Gene type	Function	Variant Type	EFFECT	IMPACT
evm.TU.supercontig_3.77	Plant invertase/pectin methylesterase inhibitor superfamily	573253	A	G	DGE	CW	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_40.33	sugar transporter 14	632304	A	T	DGE	CW	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_40.33	sugar transporter 14	632390	GTTT	GTT	DGE	CW	INDEL	upstream_gene_variant	MODIFIER
evm.TU.supercontig_40.33	sugar transporter 14	635407	C	T	DGE	CW	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_5.270	Pectate lyase family protein	2394357	T	G	DGE	CW	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_5.270	Pectate lyase family protein	2398438	C	T	DGE	CW	SNP	intron_variant	MODIFIER
evm.TU.supercontig_80.123	beta glucosidase 41	1050619	CATATATATATATATATATATAT	CATATATATATATATATAT	DGE	CW	INDEL	intron_variant	MODIFIER
evm.TU.supercontig_80.123	beta glucosidase 41	1055961	G	A	DGE	CW	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_82.65	sucrose synthase 4	1083290	G	A	DGE	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_82.65	sucrose synthase 4	1088554	C	T	DGE	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_82.65	sucrose synthase 4	1098666	A	C	DGE	CW	SNP	synonymous_variant	LOW

Table 1A – Cont.

Gene.ID	Gene definition	Variant Position	Alleles Sekati	Alleles JS.12	Gene type	Function	Variant Type	EFFECT	IMPACT
evm.TU.supercontig_93.34	beta-galactosidase 12	871255	G	A	DGE	CW	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_93.34	beta-galactosidase 12	872953	G	C	DGE	CW	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_93.34	beta-galactosidase 12	879460	T	C	DGE	CW	SNP	missense_variant	MODERATE
evm.TU.supercontig_93.34	beta-galactosidase 12	882896	T	C	DGE	CW	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_93.34	beta-galactosidase 12	884504	T	C	DGE	CW	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_93.34	beta-galactosidase 12	884570	C	T	DGE	CW	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_93.34	beta-galactosidase 12	884673	T	C	DGE	CW	SNP	downstream_gene_variant	MODIFIER
evm.TU.contig_29609	sucrose synthase 6	4662	A	G	BLAST	CW	SNP	upstream_gene_variant	MODIFIER
evm.TU.contig_32583	beta-galactosidase 3	2456	T	C	BLAST	CW	SNP	missense_variant	MODERATE
evm.TU.supercontig_102.19	xyloglucan endotransglucosylase/hydrolase 32	657289	G	A	BLAST	CW	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_102.19	xyloglucan endotransglucosylase/hydrolase 32	666038	G	A	BLAST	CW	SNP	intergenic_region	MODIFIER

Table 1A – Cont.

Gene.ID	Gene definition	Variant Position	Alleles Sekati	Alleles JS.12	Gene type	Function	Variant Type	EFFECT	IMPACT
evm.TU.supercontig_102.19	xyloglucan endotransglucosylase/hydrolase 32	666057	T	C	BLAST	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_102.19	xyloglucan endotransglucosylase/hydrolase 32	668023	C	T	BLAST	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_102.19	xyloglucan endotransglucosylase/hydrolase 32	671137	G	A	BLAST	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_102.19	xyloglucan endotransglucosylase/hydrolase 32	671163	T	G	BLAST	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_102.19	xyloglucan endotransglucosylase/hydrolase 32	671971	C	T	BLAST	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_102.19	xyloglucan endotransglucosylase/hydrolase 32	672247	C	A	BLAST	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_102.19	xyloglucan endotransglucosylase/hydrolase 32	672977	A	G	BLAST	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_102.19	xyloglucan endotransglucosylase/hydrolase 32	672979	A	T	BLAST	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_102.19	xyloglucan endotransglucosylase/hydrolase 32	672994	G	T	BLAST	CW	SNP	intergenic_region	MODIFIER

Table 1A – Cont.

Gene.ID	Gene definition	Variant Position	Alleles Sekati	Alleles JS.12	Gene type	Function	Variant Type	EFFECT	IMPACT
evm.TU.supercontig_102.19	xyloglucan endotransglucosylase/hydrolase 32	673020	C	T	BLAST	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_102.19	xyloglucan endotransglucosylase/hydrolase 32	673027	C	A	BLAST	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_102.19	xyloglucan endotransglucosylase/hydrolase 32	673045	C	T	BLAST	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_102.19	xyloglucan endotransglucosylase/hydrolase 32	673111	TTTTCT	TT	BLAST	CW	INDEL	intergenic_region	MODIFIER
evm.TU.supercontig_102.19	xyloglucan endotransglucosylase/hydrolase 32	674341	G	A	BLAST	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_102.19	xyloglucan endotransglucosylase/hydrolase 32	674848	A	G	BLAST	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_102.19	xyloglucan endotransglucosylase/hydrolase 32	674851	C	A	BLAST	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_102.19	xyloglucan endotransglucosylase/hydrolase 32	678511	T	C	BLAST	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_102.19	xyloglucan endotransglucosylase/hydrolase 32	689211	A	G	BLAST	CW	SNP	intergenic_region	MODIFIER

Table 1A – Cont.

Gene.ID	Gene definition	Variant Position	Alleles Sekati	Alleles JS.12	Gene type	Function	Variant Type	EFFECT	IMPACT
evm.TU.supercontig_102.19	xyloglucan endotransglucosylase/hydrolase 32	693278	G	A	BLAST	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_102.19	xyloglucan endotransglucosylase/hydrolase 32	693342	G	T	BLAST	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_102.19	xyloglucan endotransglucosylase/hydrolase 32	693925	C	T	BLAST	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_102.19	xyloglucan endotransglucosylase/hydrolase 32	693953	A	G	BLAST	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_102.19	xyloglucan endotransglucosylase/hydrolase 32	697498	A	G	BLAST	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_1071.2	beta-xylosidase 1	10047	T	G	BLAST	CW	SNP	intron_variant	MODIFIER
evm.TU.supercontig_1071.2	beta-xylosidase 1	12898	G	A	BLAST	CW	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_1071.2	beta-xylosidase 1	13523	AA	AAATTTTTAATGTTTTTCCGACGTTGAAAAACTAAAATACGAAAGGA	BLAST	CW	INDEL	upstream_gene_variant	MODIFIER
evm.TU.supercontig_1071.2	beta-xylosidase 1	15715	A	G	BLAST	CW	SNP	upstream_gene_variant	MODIFIER

Table 1A – Cont.

Gene.ID	Gene definition	Variant Position	Alleles Sekati	Alleles JS.12	Gene type	Function	Variant Type	EFFECT	IMPACT
evm.TU.supercontig_1071.2	beta-xylosidase 1	6208	C	T	BLAST	CW	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_1071.2	beta-xylosidase 1	6221	A	C	BLAST	CW	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_1071.2	beta-xylosidase 1	6250	C	G	BLAST	CW	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_116.16	B-S glucosidase 44	130612	A	C	BLAST	CW	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_116.16	B-S glucosidase 44	130896	A	G	BLAST	CW	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_116.16	B-S glucosidase 44	130949	G	A	BLAST	CW	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_116.16	B-S glucosidase 44	131236	A	T	BLAST	CW	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_116.17	B-S glucosidase 44	133944	G	A	BLAST	CW	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_116.16	B-S glucosidase 44	133944	G	A	BLAST	CW	SNP	intron_variant	MODIFIER
evm.TU.supercontig_116.16	B-S glucosidase 44	137785	G	A	BLAST	CW	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_116.17	B-S glucosidase 44	137785	G	A	BLAST	CW	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_119.14	expansin A1	71404	G	A	BLAST	CW	SNP	missense_variant	MODERATE

Table 1A – Cont.

Gene.ID	Gene definition	Variant Position	Alleles Sekati	Alleles JS.12	Gene type	Function	Variant Type	EFFECT	IMPACT
evm.TU.supercontig_119.14	expansin A1	73725	A	G	BLAST	CW	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_129.18	Glycosyl hydrolase family protein	116569	T	C	BLAST	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_129.18	Glycosyl hydrolase family protein	117006	C	T	BLAST	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_129.18	Glycosyl hydrolase family protein	147134	G	C	BLAST	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_129.18	Glycosyl hydrolase family protein	147142	C	CAACAATAACT	BLAST	CW	INDEL	intergenic_region	MODIFIER
evm.TU.supercontig_14.96	expansin A5	1093376	C	T	BLAST	CW	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_14.96	expansin A5	1096729	C	T	BLAST	CW	SNP	missense_variant	MODERATE
evm.TU.supercontig_14.96	expansin A5	1110409	C	T	BLAST	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_14.219	Pectin lyase-like superfamily protein	2392349	T	C	BLAST	CW	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_14.219	Pectin lyase-like superfamily protein	2396328	G	A	BLAST	CW	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_150.21	glycosyl hydrolase 9B18	213608	A	G	BLAST	CW	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_150.21	glycosyl hydrolase 9B18	218365	T	C	BLAST	CW	SNP	downstream_gene_variant	MODIFIER

Table 1A – Cont.

Gene.ID	Gene definition	Variant Position	Alleles Sekati	Alleles JS.12	Gene type	Function	Variant Type	EFFECT	IMPACT
evm.TU.supercontig_150.21	glycosyl hydrolase 9B18	225486	C	A	BLAST	CW	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_150.21	glycosyl hydrolase 9B18	225531	A	T	BLAST	CW	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_150.21	glycosyl hydrolase 9B18	225810	GTTTTT	GTTTT	BLAST	CW	INDEL	upstream_gene_variant	MODIFIER
evm.TU.supercontig_150.21	glycosyl hydrolase 9B18	225886	G	A	BLAST	CW	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_150.58	beta-galactosidase 3	493964	G	A	BLAST	CW	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_150.58	beta-galactosidase 3	493983	G	C	BLAST	CW	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_150.58	beta-galactosidase 3	494001	T	C	BLAST	CW	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_151.19	Plant invertase/pectin methylesterase inhibitor superfamily	169241	A	G	BLAST	CW	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_151.20	glycosyl hydrolase 9B13	169241	A	G	BLAST	CW	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_151.20	glycosyl hydrolase 9B13	175639	G	C	BLAST	CW	SNP	synonymous_variant	LOW
evm.TU.supercontig_151.21	sucrose synthase 2	175639	G	C	BLAST	CW	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_151.21	sucrose synthase 2	193696	G	A	BLAST	CW	SNP	intergenic_region	MODIFIER

Table 1A – Cont.

Gene.ID	Gene definition	Variant Position	Alleles Sekati	Alleles JS.12	Gene type	Function	Variant Type	EFFECT	IMPACT
evm.TU.supercontig_151.21	sucrose synthase 2	195386	A	C	BLAST	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_151.21	sucrose synthase 2	200498	A	G	BLAST	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_151.21	sucrose synthase 2	200774	C	T	BLAST	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_1540.1	Plant invertase/pectin methylesterase inhibitor superfamily	3869	T	G	BLAST	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_1540.1	Plant invertase/pectin methylesterase inhibitor superfamily	9904	G	A	BLAST	CW	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_178.35	sucrose synthase 5	318309	G	A	BLAST	CW	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_18.25	beta-xylosidase 2	158903	C	T	BLAST	CW	SNP	intron_variant	MODIFIER
evm.TU.supercontig_18.25	beta-xylosidase 2	160286	G	A	BLAST	CW	SNP	intron_variant	MODIFIER
evm.TU.supercontig_18.25	beta-xylosidase 2	161111	AT	ATGT	BLAST	CW	INDEL	upstream_gene_variant	MODIFIER
evm.TU.supercontig_18.25	beta-xylosidase 2	161120	T	A	BLAST	CW	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_2.303	expansin A8	3660694	A	C	BLAST	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_2.303	expansin A8	3663106	C	A	BLAST	CW	SNP	intergenic_region	MODIFIER

Table 1A – Cont.

Gene.ID	Gene definition	Variant Position	Alleles Sekati	Alleles JS.12	Gene type	Function	Variant Type	EFFECT	IMPACT
evm.TU.supercontig_21.97	cellulase 3	1153150	C	T	BLAST	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_21.97	cellulase 3	1154255	C	T	BLAST	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_21.97	cellulase 3	1160041	T	C	BLAST	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_21.97	cellulase 3	1178833	C	T	BLAST	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_21.97	cellulase 3	1180978	C	T	BLAST	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_25.194	pectin methylesterase 3	2039032	G	A	BLAST	CW	SNP	synonymous_variant	LOW
evm.TU.supercontig_25.194	pectin methylesterase 3	2039051	G	A	BLAST	CW	SNP	missense_variant	MODERATE
evm.TU.supercontig_25.194	pectin methylesterase 3	2039146	G	A	BLAST	CW	SNP	synonymous_variant	LOW
evm.TU.supercontig_25.194	pectin methylesterase 3	2041794	T	C	BLAST	CW	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_25.194	pectin methylesterase 3	2044196	T	G	BLAST	CW	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_25.57	beta galactosidase 1	726945	C	G	BLAST	CW	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_260.6	expansin A4	162657	T	C	BLAST	CW	SNP	upstream_gene_variant	MODIFIER

Table 1A – Cont.

Gene.ID	Gene definition	Variant Position	Alleles Sekati	Alleles JS.12	Gene type	Function	Variant Type	EFFECT	IMPACT
evm.TU.supercontig_260.6	expansin A4	173048	A	T	BLAST	CW	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_260.6	expansin A4	173048	A	T	BLAST	CW	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_260.6	expansin A4	173072	C	T	BLAST	CW	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_260.6	expansin A4	173072	C	T	BLAST	CW	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_260.6	expansin A4	173097	T	C	BLAST	CW	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_260.6	expansin A4	173097	T	C	BLAST	CW	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_260.6	expansin A4	175172	T	C	BLAST	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_260.6	expansin A4	175172	T	C	BLAST	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_260.6	expansin A4	180614	T	C	BLAST	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_260.6	expansin A4	180614	T	C	BLAST	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_260.6	expansin A4	184787	A	G	BLAST	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_29.123	Glycosyl hydrolases family 32 protein	1153515	C	T	BLAST	CW	SNP	intergenic_region	MODIFIER

Table 1A – Cont.

Gene.ID	Gene definition	Variant Position	Alleles Sekati	Alleles JS.12	Gene type	Function	Variant Type	EFFECT	IMPACT
evm.TU.supercontig_29.123	Glycosyl hydrolases family 32 protein	1167173	A	G	BLAST	CW	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_29.125	Glycosyl hydrolases family 32 protein	1173309	G	A	BLAST	CW	SNP	synonymous_variant	LOW
evm.TU.supercontig_37.20	beta-galactosidase 10	153220	GTGCATAT	GT	BLAST	CW	INDEL	upstream_gene_variant	MODIFIER
evm.TU.supercontig_444.5	beta glucosidase 32	22994	A	G	BLAST	CW	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_444.5	beta glucosidase 32	25195	C	T	BLAST	CW	SNP	intron_variant	MODIFIER
evm.TU.supercontig_444.5	beta glucosidase 32	25497	T	C	BLAST	CW	SNP	intron_variant	MODIFIER
evm.TU.supercontig_444.5	beta glucosidase 32	33868	C	T	BLAST	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_488.2	sugar transporter 1	15440	A	T	BLAST	CW	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_488.2	sugar transporter 1	15445	C	G	BLAST	CW	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_488.2	sugar transporter 1	18218	A	T	BLAST	CW	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_488.2	sugar transporter 1	7381	T	C	BLAST	CW	SNP	intergenic_region	MODIFIER

Table 1A – Cont.

Gene.ID	Gene definition	Variant Position	Alleles Sekati	Alleles JS.12	Gene type	Function	Variant Type	EFFECT	IMPACT
evm.TU.supercontig_5.156	expansin A16	1239391	CTATATATATATATATATATATATATATAT	CTATATATATATATATATATAT	BLAST	CW	INDEL	downstream_gene_variant	MODIFIER
evm.TU.supercontig_50.160	sucrose synthase 2	1536970	A	G	BLAST	CW	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_508.3	Major facilitator superfamily protein	13209	G	A	BLAST	CW	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_508.3	Major facilitator superfamily protein	13239	C	A	BLAST	CW	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_508.3	Major facilitator superfamily protein	19647	C	T	BLAST	CW	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_508.5	Major facilitator superfamily protein	19647	C	T	BLAST	CW	SNP	synonymous_variant	LOW
evm.TU.supercontig_508.5	Major facilitator superfamily protein	32669	G	T	BLAST	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_561.2	beta glucosidase 32	21302	G	A	BLAST	CW	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_561.2	beta glucosidase 32	22754	A	G	BLAST	CW	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_561.2	beta glucosidase 32	23686	G	A	BLAST	CW	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_561.2	beta glucosidase 32	23724	C	A	BLAST	CW	SNP	upstream_gene_variant	MODIFIER

Table 1A – Cont.

Gene.ID	Gene definition	Variant Position	Alleles Sekati	Alleles JS.12	Gene type	Function	Variant Type	EFFECT	IMPACT
evm.TU.supercontig_561.2	beta glucosidase 32	24208	G	A	BLAST	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_561.2	beta glucosidase 32	24256	C	T	BLAST	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_561.2	beta glucosidase 32	24845	T	C	BLAST	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_6.189	beta-galactosidase 3	1532239	G	A	BLAST	CW	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_70.9	sugar transporter 1	100499	G	A	BLAST	CW	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_860.2	Pectin lyase-like superfamily protein	21321	G	A	BLAST	CW	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_860.2	Pectin lyase-like superfamily protein	21422	T	C	BLAST	CW	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_860.2	Pectin lyase-like superfamily protein	21562	G	A	BLAST	CW	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_860.2	Pectin lyase-like superfamily protein	22105	A	G	BLAST	CW	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_99.11	beta-galactosidase 2	132154	T	C	BLAST	CW	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_99.11	beta-galactosidase 2	132172	CGTCATGTCAGTAG	CG	BLAST	CW	INDEL	intergenic_region	MODIFIER
evm.TU.supercontig_99.11	beta-galactosidase 2	133342	C	A	BLAST	CW	SNP	intergenic_region	MODIFIER

Table 1A – Cont.

Gene.ID	Gene definition	Variant Position	Alleles Sekati	Alleles JS.12	Gene type	Function	Variant Type	EFFECT	IMPACT
evm.TU.supercontig_99.11	beta-galactosidase 2	136227	G	A	BLAST	CW	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_99.11	beta-galactosidase 2	146890	C	T	BLAST	CW	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_151.32	ethylene response sensor 1	288630	A	T	DGE	PH	SNP	synonymous_variant	LOW
evm.TU.supercontig_151.32	ethylene response sensor 1	290435	G	A	DGE	PH	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_151.32	ethylene response sensor 1	293123	G	A	DGE	PH	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_151.32	ethylene response sensor 1	293194	A	T	DGE	PH	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_151.32	ethylene response sensor 1	293516	G	A	DGE	PH	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_2388.2	Signal transduction histidine kinase, hybrid-type, ethylene sensor	8708	T	C	DGE	PH	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_25.85	homology to ABI1	926286	G	T	DGE	PH	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_25.85	homology to ABI1	926308	T	G	DGE	PH	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_25.85	homology to ABI1	927597	C	T	DGE	PH	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_292.1	Auxin-responsive GH3 family protein	1980	C	T	DGE	PH	SNP	intergenic_region	MODIFIER

Table 1A – Cont.

Gene.ID	Gene definition	Variant Position	Alleles Sekati	Alleles JS.12	Gene type	Function	Variant Type	EFFECT	IMPACT
evm.TU.supercontig_292.1	Auxin-responsive GH3 family protein	6339	T	A	DGE	PH	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_292.1	Auxin-responsive GH3 family protein	9243	A	T	DGE	PH	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_2.295	response regulator 17	3571203	GCC	GCCC	BLAST	PH	INDEL	downstream_gene_variant	MODIFIER
evm.TU.supercontig_2.295	response regulator 17	3591547	A	G	BLAST	PH	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_2.295	response regulator 17	3591578	G	A	BLAST	PH	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_2.295	response regulator 17	3591609	A	G	BLAST	PH	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_2.295	response regulator 17	3592814	C	T	BLAST	PH	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_2.295	response regulator 17	3594140	C	T	BLAST	PH	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_2.295	response regulator 17	3594970	T	C	BLAST	PH	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_2.295	response regulator 17	3597711	G	C	BLAST	PH	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_2.295	response regulator 17	3597793	G	A	BLAST	PH	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_2.295	response regulator 17	3600491	A	T	BLAST	PH	SNP	intergenic_region	MODIFIER

Table 1A – Cont.

Gene.ID	Gene definition	Variant Position	Alleles Sekati	Alleles JS.12	Gene type	Function	Variant Type	EFFECT	IMPACT
evm.TU.supercontig_2.295	response regulator 17	3600513	T	C	BLAST	PH	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_2.295	response regulator 17	3603386	G	A	BLAST	PH	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_2.295	response regulator 17	3609057	T	C	BLAST	PH	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_2.295	response regulator 17	3609198	A	G	BLAST	PH	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_2.295	response regulator 17	3610564	A	G	BLAST	PH	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_44.97	Signal transduction histidine kinase, hybrid-type, ethylene sensor	1215070	G	A	BLAST	PH	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_5.212	response regulator 9	1835282	T	A	BLAST	PH	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_5.328	response regulator 4	3022912	A	G	BLAST	PH	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_5.328	response regulator 4	3024141	G	C	BLAST	PH	SNP	intron_variant	MODIFIER
evm.TU.supercontig_5.328	response regulator 4	3027314	A	C	BLAST	PH	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_5.328	response regulator 4	3028498	C	T	BLAST	PH	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_5.328	response regulator 4	3028529	G	A	BLAST	PH	SNP	upstream_gene_variant	MODIFIER

Table 1A – Cont.

Gene.ID	Gene definition	Variant Position	Alleles Sekati	Alleles JS.12	Gene type	Function	Variant Type	EFFECT	IMPACT
evm.TU.supercontig_5.328	response regulator 4	3028549	T	C	BLAST	PH	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_5.328	response regulator 4	3028578	A	C	BLAST	PH	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_5.328	response regulator 4	3028610	A	G	BLAST	PH	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_6.74	Auxin-responsive GH3 family protein	571741	G	A	BLAST	PH	SNP	missense_variant	MODERATE
evm.TU.supercontig_6.74	Auxin-responsive GH3 family protein	577509	G	A	BLAST	PH	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_64.99	response regulator 9	693605	AATATATATATATATAT	AATATATATATATATAT	BLAST	PH	INDEL	upstream_gene_variant	MODIFIER
evm.TU.supercontig_9.204	putative indole-3-acetic acid-amido synthetase GH3.9	1382431	G	A	BLAST	PH	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_9.204	putative indole-3-acetic acid-amido synthetase GH3.9	1382456	G	A	BLAST	PH	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_9.204	putative indole-3-acetic acid-amido synthetase GH3.9	1386086	C	T	BLAST	PH	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_9.204	putative indole-3-acetic acid-amido synthetase GH3.9	1386116	G	C	BLAST	PH	SNP	upstream_gene_variant	MODIFIER

Table 1A – Cont.

Gene.ID	Gene definition	Variant Position	Alleles Sekati	Alleles JS.12	Gene type	Function	Variant Type	EFFECT	IMPACT
evm.TU.supercontig_9.204	putative indole-3-acetic acid-amido synthetase GH3.9	1386442	T	C	BLAST	PH	SNP	upstream_gene_variant	MODIFIER
evm.TU.contig_38010	Serine protease inhibitor (SERPIN) family protein	1000	A	G	DGE	PROT	SNP	downstream_gene_variant	MODIFIER
evm.TU.contig_38010	Serine protease inhibitor (SERPIN) family protein	250	G	A	DGE	PROT	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_12.94	Subtilisin-like serine endopeptidase family protein	814429	T	C	DGE	PROT	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_12.94	Subtilisin-like serine endopeptidase family protein	819635	CTTT	CTTTT	DGE	PROT	INDEL	intron_variant	MODIFIER
evm.TU.supercontig_12.94	Subtilisin-like serine endopeptidase family protein	833228	G	T	DGE	PROT	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_12.94	Subtilisin-like serine endopeptidase family protein	833237	C	T	DGE	PROT	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_12.94	Subtilisin-like serine endopeptidase family protein	833296	A	C	DGE	PROT	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_12.94	Subtilisin-like serine endopeptidase family protein	833353	T	G	DGE	PROT	SNP	intergenic_region	MODIFIER

Table 1A – Cont.

Gene.ID	Gene definition	Variant Position	Alleles Sekati	Alleles JS.12	Gene type	Function	Variant Type	EFFECT	IMPACT
evm.TU.supercontig_12.94	Subtilisin-like serine endopeptidase family protein	833370	C	A	DGE	PROT	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_4.28	Cysteine proteinases superfamily protein	956211	G	C	DGE	PROT	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_4.28	Cysteine proteinases superfamily protein	960396	T	C	DGE	PROT	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_136.72	Subtilisin-like serine endopeptidase family protein	744077	AATTTATATATATATATAT	AAT	BLAST	PROT	INDEL	upstream_gene_variant	MODIFIER
evm.TU.supercontig_136.72	Subtilisin-like serine endopeptidase family protein	745405	C	G	BLAST	PROT	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_136.72	Subtilisin-like serine endopeptidase family protein	745629	C	G	BLAST	PROT	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_136.72	Subtilisin-like serine endopeptidase family protein	752360	G	T	BLAST	PROT	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_136.72	Subtilisin-like serine endopeptidase family protein	752489	G	T	BLAST	PROT	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_136.72	Subtilisin-like serine endopeptidase family protein	752545	T	C	BLAST	PROT	SNP	downstream_gene_variant	MODIFIER

Table 1A – Cont.

Gene.ID	Gene definition	Variant Position	Alleles Sekati	Alleles JS.12	Gene type	Function	Variant Type	EFFECT	IMPACT
evm.TU.supercontig_136.72	Subtilisin-like serine endopeptidase family protein	752900	G	A	BLAST	PROT	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_21.158	Granulin repeat cysteine protease family protein	1595350	G	T	BLAST	PROT	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_21.158	Granulin repeat cysteine protease family protein	1598591	T	C	BLAST	PROT	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_21.158	Granulin repeat cysteine protease family protein	1602400	A	G	BLAST	PROT	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_21.158	Granulin repeat cysteine protease family protein	1602701	A	G	BLAST	PROT	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_21.158	Granulin repeat cysteine protease family protein	1608292	A	G	BLAST	PROT	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_21.158	Granulin repeat cysteine protease family protein	1608304	C	G	BLAST	PROT	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_21.158	Granulin repeat cysteine protease family protein	1608945	T	C	BLAST	PROT	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_21.158	Granulin repeat cysteine protease family protein	1608959	T	C	BLAST	PROT	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_21.158	Granulin repeat cysteine protease family protein	1609496	T	C	BLAST	PROT	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_21.158	Granulin repeat cysteine protease family protein	1609570	A	C	BLAST	PROT	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_21.158	Granulin repeat cysteine protease family protein	1609926	A	G	BLAST	PROT	SNP	intergenic_region	MODIFIER

Table 1A – Cont.

Gene.ID	Gene definition	Variant Position	Alleles Sekati	Alleles JS.12	Gene type	Function	Variant Type	EFFECT	IMPACT
evm.TU.supercontig_21.158	Granulin repeat cysteine protease family protein	1609972	A	G	BLAST	PROT	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_21.158	Granulin repeat cysteine protease family protein	1610308	C	G	BLAST	PROT	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_21.158	Granulin repeat cysteine protease family protein	1610317	C	T	BLAST	PROT	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_21.158	Granulin repeat cysteine protease family protein	1610326	T	G	BLAST	PROT	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_21.158	Granulin repeat cysteine protease family protein	1610845	A	G	BLAST	PROT	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_21.158	Granulin repeat cysteine protease family protein	1610876	T	C	BLAST	PROT	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_21.158	Granulin repeat cysteine protease family protein	1611367	A	G	BLAST	PROT	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_21.158	Granulin repeat cysteine protease family protein	1612003	A	G	BLAST	PROT	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_21.158	Granulin repeat cysteine protease family protein	1612636	T	C	BLAST	PROT	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_21.158	Granulin repeat cysteine protease family protein	1612675	A	C	BLAST	PROT	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_21.158	Granulin repeat cysteine protease family protein	1613340	G	T	BLAST	PROT	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_21.158	Granulin repeat cysteine protease family protein	1613348	G	A	BLAST	PROT	SNP	intergenic_region	MODIFIER

Table 1A – Cont.

Gene.ID	Gene definition	Variant Position	Alleles Sekati	Alleles JS.12	Gene type	Function	Variant Type	EFFECT	IMPACT
evm.TU.supercontig_21.158	Granulin repeat cysteine protease family protein	1613820	C	A	BLAST	PROT	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_21.158	Granulin repeat cysteine protease family protein	1614278	G	A	BLAST	PROT	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_21.158	Granulin repeat cysteine protease family protein	1614307	A	G	BLAST	PROT	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_21.158	Granulin repeat cysteine protease family protein	1614399	C	T	BLAST	PROT	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_21.158	Granulin repeat cysteine protease family protein	1614686	T	C	BLAST	PROT	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_21.158	Granulin repeat cysteine protease family protein	1614758	G	A	BLAST	PROT	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_21.158	Granulin repeat cysteine protease family protein	1614830	T	C	BLAST	PROT	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_21.158	Granulin repeat cysteine protease family protein	1614864	T	C	BLAST	PROT	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_21.158	Granulin repeat cysteine protease family protein	1616835	C	T	BLAST	PROT	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_21.158	Granulin repeat cysteine protease family protein	1616876	T	C	BLAST	PROT	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_232.4	xylem cysteine peptidase 1	115128	A	G	BLAST	PROT	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_27.197	Subtilase family protein	1967781	C	T	BLAST	PROT	SNP	downstream_gene_variant	MODIFIER

Table 1A – Cont.

Gene.ID	Gene definition	Variant Position	Alleles Sekati	Alleles JS.12	Gene type	Function	Variant Type	EFFECT	IMPACT
evm.TU.supercontig_27.197	Subtilase family protein	1967798	T	G	BLAST	PROT	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_286.15	xylem cysteine peptidase 1	103806	G	A	BLAST	PROT	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_34.197	cystatin B	1799430	C	T	BLAST	PROT	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_34.197	cystatin B	1802990	A	T	BLAST	PROT	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_34.197	cystatin B	1807651	A	G	BLAST	PROT	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_34.197	cystatin B	1807906	A	G	BLAST	PROT	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_34.197	cystatin B	1807955	C	T	BLAST	PROT	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_34.197	cystatin B	1807959	C	G	BLAST	PROT	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_34.199	cystatin B	1827936	A	G	BLAST	PROT	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_34.200	cystatin B	1827936	A	G	BLAST	PROT	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_34.199	cystatin B	1828554	G	A	BLAST	PROT	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_34.201	cystatin-1	1854552	G	A	BLAST	PROT	SNP	intergenic_region	MODIFIER

Table 1A – Cont.

Gene.ID	Gene definition	Variant Position	Alleles Sekati	Alleles JS.12	Gene type	Function	Variant Type	EFFECT	IMPACT
evm.TU.supercontig_34.201	cystatin-1	1857426	T	C	BLAST	PROT	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_34.201	cystatin-1	1857476	G	A	BLAST	PROT	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_34.201	cystatin-1	1861948	G	A	BLAST	PROT	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_34.201	cystatin-1	1861982	G	A	BLAST	PROT	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_34.201	cystatin-1	1861987	A	G	BLAST	PROT	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_34.201	cystatin-1	1862207	A	G	BLAST	PROT	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_34.201	cystatin-1	1862233	T	TC	BLAST	PROT	INDEL	intergenic_region	MODIFIER
evm.TU.supercontig_34.201	cystatin-1	1862873	G	A	BLAST	PROT	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_34.201	cystatin-1	1863357	C	T	BLAST	PROT	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_34.201	cystatin-1	1863523	T	C	BLAST	PROT	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_34.201	cystatin-1	1863525	G	A	BLAST	PROT	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_34.201	cystatin-1	1864242	A	G	BLAST	PROT	SNP	intergenic_region	MODIFIER

Table 1A – Cont.

Gene.ID	Gene definition	Variant Position	Alleles Sekati	Alleles JS.12	Gene type	Function	Variant Type	EFFECT	IMPACT
evm.TU.supercontig_34.201	cystatin-1	1864635	A	G	BLAST	PROT	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_34.201	cystatin-1	1866945	G	A	BLAST	PROT	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_34.201	cystatin-1	1866951	C	T	BLAST	PROT	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_34.201	cystatin-1	1867857	G	T	BLAST	PROT	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_34.201	cystatin-1	1867940	G	A	BLAST	PROT	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_55.102	cysteine proteinase1	613665	G	A	BLAST	PROT	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_55.102	cysteine proteinase1	614450	T	G	BLAST	PROT	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_165.12	NAC domain containing protein 74	65284	GAGAAGAAGAAGAAGAAGAA	GAGAAGAAGAAGAAGAAGAAGAA	DGE	TF	INDEL	intron_variant	MODIFIER
evm.TU.supercontig_18.82	WRKY DNA-binding protein 23	538479	A	T	DGE	TF	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_18.82	WRKY DNA-binding protein 23	543259	G	A	DGE	TF	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_18.82	WRKY DNA-binding protein 23	543265	C	G	DGE	TF	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_18.82	WRKY DNA-binding protein 23	543387	A	G	DGE	TF	SNP	intergenic_region	MODIFIER

Table 1A – Cont.

Gene.ID	Gene definition	Variant Position	Alleles Sekati	Alleles JS.12	Gene type	Function	Variant Type	EFFECT	IMPACT
evm.TU.supercontig_18.82	WRKY DNA-binding protein 23	543392	C	T	DGE	TF	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_18.82	WRKY DNA-binding protein 23	543728	C	T	DGE	TF	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_18.82	WRKY DNA-binding protein 23	543751	C	T	DGE	TF	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_18.82	WRKY DNA-binding protein 23	543766	C	T	DGE	TF	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_18.82	WRKY DNA-binding protein 23	552968	A	T	DGE	TF	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_18.82	WRKY DNA-binding protein 23	553694	C	T	DGE	TF	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_18.82	WRKY DNA-binding protein 23	554792	G	C	DGE	TF	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_19.44	WRKY DNA-binding protein 70	376310	A	G	DGE	TF	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_10.75	WRKY DNA-binding protein 40	1253829	C	T	BLAST	TF	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_10.75	WRKY DNA-binding protein 40	1271163	C	G	BLAST	TF	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_10.75	WRKY DNA-binding protein 40	1275382	A	T	BLAST	TF	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_126.49	WRKY DNA-binding protein 50	483744	T	C	BLAST	TF	SNP	intergenic_region	MODIFIER

Table 1A – Cont.

Gene.ID	Gene definition	Variant Position	Alleles Sekati	Alleles JS.12	Gene type	Function	Variant Type	EFFECT	IMPACT
evm.TU.supercontig_126.49	WRKY DNA-binding protein 50	492480	G	A	BLAST	TF	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_126.49	WRKY DNA-binding protein 50	493045	A	C	BLAST	TF	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_126.49	WRKY DNA-binding protein 50	497264	TCC	TC	BLAST	TF	INDEL	intergenic_region	MODIFIER
evm.TU.supercontig_126.49	WRKY DNA-binding protein 50	504145	C	T	BLAST	TF	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_126.49	WRKY DNA-binding protein 50	504195	A	G	BLAST	TF	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_126.49	WRKY DNA-binding protein 50	509720	G	A	BLAST	TF	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_126.49	WRKY DNA-binding protein 50	515640	C	T	BLAST	TF	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_126.49	WRKY DNA-binding protein 50	519192	T	C	BLAST	TF	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_126.49	WRKY DNA-binding protein 50	521289	C	A	BLAST	TF	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_127.25	WRKY DNA-binding protein 18	398323	T	C	BLAST	TF	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_127.25	WRKY DNA-binding protein 18	402923	G	C	BLAST	TF	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_127.25	WRKY DNA-binding protein 18	404587	T	G	BLAST	TF	SNP	intergenic_region	MODIFIER

Table 1A – Cont.

Gene.ID	Gene definition	Variant Position	Alleles Sekati	Alleles JS.12	Gene type	Function	Variant Type	EFFECT	IMPACT
evm.TU.supercontig_136.16	zinc-dependent activator protein-1	201792	A	G	BLAST	TF	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_14.101	NAC domain containing protein 100	1191890	T	A	BLAST	TF	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_14.101	NAC domain containing protein 100	1193299	A	T	BLAST	TF	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_14.101	NAC domain containing protein 100	1193314	G	A	BLAST	TF	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_14.101	NAC domain containing protein 100	1193331	T	C	BLAST	TF	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_14.101	NAC domain containing protein 100	1193337	T	C	BLAST	TF	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_14.101	NAC domain containing protein 100	1198819	G	A	BLAST	TF	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_14.101	NAC domain containing protein 100	1202898	C	A	BLAST	TF	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_14.101	NAC domain containing protein 100	1202934	A	C	BLAST	TF	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_14.101	NAC domain containing protein 100	1202997	TTGTG	TTGTGTG	BLAST	TF	INDEL	upstream_gene_variant	MODIFIER
evm.TU.supercontig_14.101	NAC domain containing protein 100	1213086	G	A	BLAST	TF	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_14.101	NAC domain containing protein 100	1216513	C	T	BLAST	TF	SNP	intergenic_region	MODIFIER

Table 1A – Cont.

Gene.ID	Gene definition	Variant Position	Alleles Sekati	Alleles JS.12	Gene type	Function	Variant Type	EFFECT	IMPACT
evm.TU.supercontig_152.36	WRKY family transcription factor	265757	G	C	BLAST	TF	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_152.36	WRKY family transcription factor	265992	G	A	BLAST	TF	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_152.36	WRKY family transcription factor	266005	G	A	BLAST	TF	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_152.36	WRKY family transcription factor	266016	A	G	BLAST	TF	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_152.36	WRKY family transcription factor	267526	G	A	BLAST	TF	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_152.36	WRKY family transcription factor	269630	C	T	BLAST	TF	SNP	synonymous_variant	LOW
evm.TU.supercontig_3.311	NAC 007	2172526	T	A	BLAST	TF	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_43.86	WRKY DNA-binding protein 75	1117116	C	T	BLAST	TF	SNP	upstream_gene_variant	MODIFIER
evm.TU.supercontig_43.86	WRKY DNA-binding protein 75	1120473	G	A	BLAST	TF	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_43.86	WRKY DNA-binding protein 75	1120526	A	G	BLAST	TF	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_43.86	WRKY DNA-binding protein 75	1120559	T	A	BLAST	TF	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_684.1	WRKY family transcription factor	828	C	T	BLAST	TF	SNP	upstream_gene_variant	MODIFIER

Table 1A – Cont.

Gene.ID	Gene definition	Variant Position	Alleles Sekati	Alleles JS.12	Gene type	Function	Variant Type	EFFECT	IMPACT
evm.TU.supercontig_72.18	WRKY DNA-binding protein 13	374487	A	C	BLAST	TF	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_72.18	WRKY DNA-binding protein 13	392211	T	C	BLAST	TF	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_72.18	WRKY DNA-binding protein 13	392672	G	A	BLAST	TF	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_72.18	WRKY DNA-binding protein 13	395810	T	C	BLAST	TF	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_72.18	WRKY DNA-binding protein 13	404383	C	T	BLAST	TF	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_72.18	WRKY DNA-binding protein 13	407326	G	A	BLAST	TF	SNP	intron_variant	MODIFIER
evm.TU.supercontig_72.18	WRKY DNA-binding protein 13	407781	G	C	BLAST	TF	SNP	intron_variant	MODIFIER
evm.TU.supercontig_72.18	WRKY DNA-binding protein 13	411420	C	A	BLAST	TF	SNP	intron_variant	MODIFIER
evm.TU.supercontig_72.18	WRKY DNA-binding protein 13	435653	G	A	BLAST	TF	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_72.18	WRKY DNA-binding protein 13	435689	C	A	BLAST	TF	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_72.18	WRKY DNA-binding protein 13	435707	A	G	BLAST	TF	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_72.18	WRKY DNA-binding protein 13	435710	G	A	BLAST	TF	SNP	intergenic_region	MODIFIER

Table 1A – Cont.

Gene.ID	Gene definition	Variant Position	Alleles Sekati	Alleles JS.12	Gene type	Function	Variant Type	EFFECT	IMPACT
evm.TU.supercontig_80.94	NAC domain containing protein 2	725276	A	G	BLAST	TF	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_80.94	NAC domain containing protein 2	725306	C	T	BLAST	TF	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_80.94	NAC domain containing protein 2	725330	C	T	BLAST	TF	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_80.94	NAC domain containing protein 2	725733	C	A	BLAST	TF	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_80.94	NAC domain containing protein 2	725740	C	T	BLAST	TF	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_80.94	NAC domain containing protein 2	725854	G	A	BLAST	TF	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_80.94	NAC domain containing protein 2	726152	A	G	BLAST	TF	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_80.94	NAC domain containing protein 2	727174	A	G	BLAST	TF	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_80.94	NAC domain containing protein 2	730539	T	A	BLAST	TF	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_80.94	NAC domain containing protein 2	730555	GTATATATATATATATAT	GTATATATATATATATATAT	BLAST	TF	INDEL	intergenic_region	MODIFIER
evm.TU.supercontig_80.94	NAC domain containing protein 2	737242	C	T	BLAST	TF	SNP	intron_variant	MODIFIER
evm.TU.supercontig_86.65	WRKY DNA-binding protein 38	839141	A	G	BLAST	TF	SNP	upstream_gene_variant	MODIFIER

Table 1A – Cont.

Gene.ID	Gene definition	Variant Position	Alleles Sekati	Alleles JS.12	Gene type	Function	Variant Type	EFFECT	IMPACT
evm.TU.supercontig_9.36	WRKY DNA-binding protein 48	147417	C	A	BLAST	TF	SNP	downstream_gene_variant	MODIFIER
evm.TU.supercontig_9.36	WRKY DNA-binding protein 48	151586	A	T	BLAST	TF	SNP	missense_variant	MODERATE
evm.TU.supercontig_919.2	WRKY DNA-binding protein 51	10778	G	A	BLAST	TF	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_919.2	WRKY DNA-binding protein 51	10814	A	T	BLAST	TF	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_919.2	WRKY DNA-binding protein 51	12032	C	T	BLAST	TF	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_919.2	WRKY DNA-binding protein 51	14351	T	C	BLAST	TF	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_919.2	WRKY DNA-binding protein 51	14383	A	C	BLAST	TF	SNP	intergenic_region	MODIFIER
evm.TU.supercontig_919.2	WRKY DNA-binding protein 51	15245	T	C	BLAST	TF	SNP	intergenic_region	MODIFIER