CANDIDATE GENE ANALYSIS ON SELF-INCOMPATIBILITY IN COCOA

(*Theobroma cacao* L.)

By

SHARAT PRABHAKARAN

(2018-11- 165)

CENTRE FOR PLANT BIOTECHNOLOGY AND MOLECULAR BIOLOGY

COLLEGE OF HORTICULTURE

VELLANIKKARA, THRISSUR – 680656

KERALA, INDIA

2020

CANDIDATE GENE ANALYSIS ON SELF-INCOMPATIBILITY IN COCOA (*Theobroma cacao* L.)

By

SHARAT PRABHAKARAN

(2018 11 165)

THESIS

Submitted in partial fulfilment of the requirement for the degree of

Master of Science in Agriculture

(PLANT BIOTECHNOLOGY)

Faculty of Agriculture

Kerala Agricultural University, Thrissur

CENTRE FOR PLANT BIOTECHNOLOGY AND MOLECULAR BIOLOGY

COLLEGE OF HORTICULTURE

VELLANIKKARA, THRISSUR – 680656

KERALA, INDIA

2020

DECLARATION

I, hereby declare that this thesis entitled "**Candidate gene analysis on self-incompatibility in cocoa (*Theobroma cacao* L.)**" is a bonafide record of research work done by me during the course of research and the thesis has not previously formed the basis for the award to me of any degree, diploma, associateship, fellowship or other similar title, of any other University or Society.

Vellanikkara

Date: 15/12/2020

Sharat Prabhakaran

(2018-11-165)

CERTIFICATE

Certified that the thesis entitled **"Candidate gene analysis on self-incompatibility in cocoa (*Theobroma cacao* L.)"** is a record of research work done independently by **Sharat Prabhakaran (2018-11-165)** under my guidance and supervision and that it has not previously formed the basis for the award of any degree, diploma, fellowship or associateship to him.

Vellanikkara

Date: 15/12/2020

Dr. Minimol J. S.

(Major Advisor)
Associate Professor
Plant Breeding and Genetics
Cocoa Research Centre
Vellanikkara, Thrissur

CERTIFICATE

We, the undersigned members of the advisory committee of **Mr. Sharat Prabhakaran** (2018-11-165) a candidate for the degree of **Master of Science in Agriculture** with major field in **Plant Biotechnology**, agree that this thesis entitled **"Candidate gene analysis on self-incompatibility in cocoa (*Theobroma cacao* L.)"** may be submitted by **Mr. Sharat Prabhakaran** in partial fulfilment of the requirement for the degree.

Dr. Minimol J. S.
(Chairman, Advisory Committee)
Associate Professor,
Plant Breeding and Genetics,
Cocoa Research Centre,
Kerala Agricultural University.

Dr. Deepu Mathew
(Member, Advisory Committee)
Assistant Professor,
CPBMB,
College of Horticulture, Vellanikkara.

14/12/2023

Dr. Abida P. S.
(Member, Advisory Committee)
Professor and Head,
CPBMB,
College of Horticulture, Vellanikkara.

Dr. Suma B
(Member, Advisory Committee)
Professor and Head,
Cocoa Research Centre,
Kerala Agricultural University

ACKNOWLEDGEMENT

It is very difficult to express one's feelings in words but formality demands to do so to the extent possible. I fell the inadequacy of diction in expressing my sincere heartfelt gratitude to ***Dr. Minimol J. S.,*** *Associate Professor (Plant Breeding and Genetics) Cocoa Research Centre, KAU, Vellanikkara and Chairman of my advisory committee. Her level of guidance, lively encouragement and generous assistance at every stage of my research work. Her keen observations in detecting errors and correcting manuscript which consumed her valuable time and efforts is far beyond the call of duty. Her kind and understanding nature have been overwhelming and unforgettable throughout my life.*

I am highly indebted to ***Dr. Deepu Mathew****, Assistant Professor, CPBMB, for his constant encouragement, affectionate advice, endless support and generous guidance rendered during the whole venture.*

I express my sincere thanks and gratitude to ***Dr. Abida P. S*** *Head CPBMB and* ***Dr. M. R. Shylaja,*** *former Head CPBMB, College of Horticulture, Vellanikkara who were always there to help us students during the time of need. I have immense pleasure to express my sincere thanks to* ***Dr. Suma B.,*** *Professor and Head, Cocoa Research Center for the avail of facilities at the center during the course of my research.*

Dr. Rehena Agustine, Dr. Smitha Nair, Dr. Binu, Dr. Donald, Dr. Marriet *who have introduced me to this beautiful subject, been really amazing teachers and have managed to inculcate passion and interest for this subject in me. I also thank* ***Ravishankar sir*** *who helped me with bioinformatics during this study.*

This thesis would be incomplete if I do not reckon the sacrifices, love, affection and support of my mother ***Smt. Meena*** *for her selfless persuasions, sacrifices, warm blessing and love.*

It was a smooth ride on a bumpy road thanks to all my beloved friends ***Alfiya, Ashwathi, Akhil, Ashwini, Athira, Dida, Haritha, Laya, Pavi, Pranali, Rasha, Ribin, Sreekutty, Sunny, and Vaibhav.***

My sincere thanks ***Nimmi chechi*** *and* ***Asha chechi*** *who introduced the laboratory to and were always there to lend a hand whenever we students needed any kind of support, guidance and advice in the laboratory.*

I am thankful to my loveable buddies ***Sanjuttan, Koduram, Abin*** *and* ***Ashok Madala*** *for always being just a call away to extend their help. A sincere thanks from the bottom of my heart to my seniors* ***Shilpa, Shivaji, Vipul, Feba*** *and* ***Architha*** *who provided critical suggestions and motivations throughout this journey.*

My special thanks to, ***Panjami Chechi, Chithira Chechi*** *and all the other Research Associates, Cocoa Research Centre, whose helping hands were always with me throughout the research work.*

I take this opportunity to extend my gratitude to all non- teaching staff of CPBMB and Cocoa Research Centre.

I wish to thank all the others who had helped me in one way or the other in completing the study.

Errors and omissions are entirely unintentional.

Sharat Prabhakaran

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF PLATES

ABBREVIATIONS

%	Percentage
>	Greater than
B	Beta
µg	Microgram
µl	Microliter
CAPS	Cleaved Amplified Polymorphic Sequence
CCRP	KAU- Cadbury Co-operative Cocoa Research Project
CRC	Cocoa Research Centre
Cm	Centimeter
CPBMB	Centre for Plant Biotechnology and Molecular Biology
CTPP	Confronted Two Primer Polymerase Chain Reaction
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
dNTPs	Deoxyribo Nucleoside Triphosphate
EBI	European Bioinformatics Institute
EDTA	Ethylene Diamine Tetra Acetic acid
Gm	Gram
ICTG	International Cocoa Gene bank
ISSR	Inter Simple Sequence Repeat
KAU	Kerala Agricultural University
Kb	Kilo base pairs

KOAC	Potassium acetate
L	Litre
M	Molar
mg	Milligram
ml	Millilitre
NaOAC	Sodium acetate
mM	Milli Molar
NCBI	National Center for Biotechnology Information
ng	Nano gram
^{o}C	Degree Celsius
OD	Optical Density
PBR	Plant Breeders Right
PCR	Polymerase Chain Reaction
pH	Hydrogen ion concentration
PIC	Polymorphic Information Content
PVP	Poly Vinyl Pyrrolidone
RAPD	Random Amplified Polymorphic DNA
RFLP	Restriction Fragment Length Polymorphism
RNA	Ribonucleic acid
RNase	Ribonuclease
rpm	Revolutions per minute
SDS	Sodium dodecyl sulfate
SNP	Single Nucleotide Polymorphism
TAE	Tris Acetate EDTA
TE	Tris EDTA
UV	Ultra Violet
V	Volts

Introduction

1. Introduction

Theobroma is a genus of diploid plants which belong to the Malvaceae family and are native to South America. The genus comprises of 22 species of tropical trees of which *Theobroma cacao* (2n=20) commonly known as cocoa, is the species with utmost economic importance. The species of second most importance is *T. grandiflorum*, which belongs to Brazillian Amazon.

Physiological and genetic investigations had unveiled that the yield potential of cocoa is not yet fully exploited (Bertus, 2004). Demand for chocolate is ever increasing at a rate of 15-20 percent every year and to meet this demand more area has to be brought under cocoa cultivation using improved genetic stock. Development of superior hybrids had significantly contributed to improve cocoa productivity in many countries (Kennedy *et al*., 1987; Dias *et al*., 2003). When cocoa hybrids were evaluated for their performance it was observed that they showed wide adaptability, low environmental interaction, improved yield when compared to traditional local cultivars (Dias *et al*., 2003).

However artificial pollination in cocoa is very tedious due to the small and fragile nature of the flower. The success rate is only one percent even with minimum disturbance (Mallika *et al*., 2000). Self-incompatibility is a pollination control mechanism which prevents self-fertilization. Hence this can be exploited in hybrid production by avoiding emasculation which is a cumbersome process (Minimol and Amma, 2013). Moreover, emasculation will damage the flower leading to reduction in success rate.

Conventional method to assess self-incompatibility is by selfing 100 flowers per tree. If no fruit set was observed then the plant is classified as self-incompatible (Mallika *et al.*, 2006). This is a time-consuming tedious process which reduces the pace of breeding programme.

In, cocoa the first evidence of self-incompatibility was reported by Pound (1932) and he could characterize the cacao trees into self-compatible and self-incompatible.

Mechanism of self-incompatibility acting in cocoa was reported to be a late acting ovarian self-incompatibility (OSI). It has also been evident that the genetic control of self-incompatibility in cacao has both aspects of sporophytic and gametophytic control (Cope, 1958).

After 24 hours of pollination, in case of the compatible reactions the pollen penetrates into the ovary through the micropyle, the fusion of the plasma membrane of the sperm nuclei and the egg nucellus takes place (Cope, 1939 and 1940). This is followed by the union of the male and female nuclei, followed by normal double fertilization and the flower is retained. In case of the self-incompatible reaction, the plasma membranes fail to fuse and syngamy doesn't occur and finally abscission of the flower was observed (Knight and Rodger, 1955). Cocoa is reported to be the first model of angiosperms to have an incompatibility mechanism, which is based on the success or failure of syngamy (Cope, 1962).

From the fine mapping, GWAS (Genome Wide Association Studies) and cytological studies conducted by Lanaud *et al.* (2017), the existence of two loci independent of each other, one in Chromosome 1 and the other in Chromosome 4 was confirmed.

Many candidate genes have been reported for self-incompatibility in various other crops (McCormick, 1998). However, the actual sequence variations in these genes are not yet studied in cocoa. Identification of appropriate gene resulting in self-incompatibility will help to identify the mechanism in early stage and definitely help to quicken the breeding programme.

In this context, the present study was taken up with the following specific objective:

- Study the sequence variation in ten putative candidate genes contributing to self-incompatibility in Forastero cocoa.

Review of literature

2. Review of literature

2.1 Introduction and general background

Theobroma is a genus of diploid plants which belong to the Malvaceae family and are native to South America. The genus comprises of 22 species of tropical trees of which *Theobroma cacao* (2n=20) commonly known as cocoa, is the species with utmost economic importance. The species of second most importance is *T. grandiflorum*, which belongs to Brazillian Amazon.

In India, the area under cultivation and the production of cocoa is increasing continuously. The area under cultivation escalated from 57,000 ha in 2010 to 89,000 ha in 2018. The annual production has leaped, from 14,000 metric tonnes in 2010 to 20,000 metric tonnes in 2018. Within a span of 5 years the export value of the crop has drastically risen from 573.22 crores in 2013 to 1144.35 crores in 2018. The global and domestic demand for cocoa produce is ever increasing (GOI, 2016 and 2019). The southern states of India, *viz*. Kerala, Karnataka, Tamil Nadu, Telangana and Andhra Pradhesh are the major producers of the crop. Cocoa cultivation in Kerala is under an area of 16594, and an annual production of 7507 (second highest producer) and a productivity of 750 kg per hectare (second best productivity) (DCCD, 2018).

Cocoa is a perennial tree with typical plant habit, specific fruit characteristics, highly influenced by climate change and growing environment, which make it necessary to have long term and dynamic breeding programme (Malhotra and Hubali, 2016).

2.2 Self-incompatibility

In nature, angiosperms generally produce perfect flowers which possess both anther and pistil very close to each other on the same flower; as a result, they would have a high disposition to self-fertilize (Kao and McCubbin, 1996). As it seems, this trait should have promoted self-fertilization and inbreeding, resulting in a reduction in genetic variability and hence be unfavorable for the evolution of angiosperms. Quite obviously that has not

been the case; angiosperms are pretty much the most successful group of terrestrial plants on this planet (Silva and Goring, 2001).

This reproductive triumph is the result of the evolution of various mechanisms that control and hinder self-fertilization. Certain common methods that hermaphrodites have developed to reduce the chances of self-pollination are herkogamy- which is the spatial isolation of reproductive parts and dichogamy – which is the asynchronous maturation of the reproductive parts (Barrett, 2002).

Self-incompatibility is the most common and efficient approach followed by plants to promote out breeding. By this mechanism a pistil is able to recognize a self-pollen or pollen from genetically identical individuals and reject it, and hence lead to out crossing rather than inbreeding. This results in an increase in fitness of the progeny and genetic diversity within a species (McCubbin and Kao, 2000).

Massive diversification of angiosperms has occurred during the Cretaceous period. This is very much attributed to the fact that a much earlier evolution of the mechanism of self-incompatibility has occurred in this group of plants (Whitehouse, 1950).

Self-incompatibility or self-sterility systems are present throughout angiosperms. They have been known to be present in 60 percent of species, present in about 19 orders, 71 families and 250 genera (East and Mangelsdorf, 1925; Brewbaker, 1957; de Nettancourt, 1997; Charlesworth, 1985; Hiscock et al., 1995).

Many angiosperms have evolved to possess an inherent ability by which they can detect and reject its own pollen during the pollen pistil interaction. This ability which prevents self-fertilization and inbreeding is called self-incompatibility. By definition self-incompatibility is “the inability of a fertile hermaphrodite seed plant to produce zygotes after self-pollination” (de Nettancourt, 1997).

In flowering plants, based on morphological distinguishability of the flowers, self-incompatibility is classified into two systems - heteromorphic and homomorphic (Barrett,

1992). Heteromorphic are those plants that produce two or more morphologically different flowers within the same species. Homomorphic are plants that show no difference in the floral morphology within the same species (Gibbs, 1986).

2.3 Types of homomorphic self-incompatibility

Even though a majority of the angiosperms utilize self-incompatibility in order to maintain genetic variability within the species, not every species follows the common mechanism for self-incompatibility. Through various genetic studies conducted in the early 20^{th} century it has been well established that two types of homomorphic SI systems exist based on the genetic control, *i.e.*, gametophytic self-incompatibility system and sporophytic self-incompatibility system (de Nettancourt, 1997).

A highly polymorphic single locus, referred to as the S locus, is known to be responsible for self-incompatibility in most of the well characterized self-incompatibility systems (de Nettancourt, 1997, 2001; Hiscock and McInnis, 2003). Apart from this, in a few species multi loci systems are also noted to be found (Lundqvist, 1990 and 1991), the popular one being the two loci system (S and Z system) present in Poaceae family (Li *et al.*, 1997). Two very tightly linked polymorphic genes are known to be present in the S locus, one for the pollen identity and the other for the pistil identity.

Pioneering studies conducted independently by East and Mangelsdorf (1925), whose research revealed that a single locus consisting of multiple alleles controls self-incompatibility in some plants. In this system, the incompatibility status of the pollen was determined by the haploid S allele it possesses in its haploid gametophytic genome. This system is referred to as the gametophytic self-incompatibility system. This S locus is reported to be very highly polymorphic within the species (Emerson, 1938).

Later on, another homomorphic self-incompatibility was discovered which did not resemble the gametophytic self-incompatibility model. The initially discovered exceptions were in Brassicaceae (Kakizaki, 1930), *Crepis foetida* (Asteraceae) (Hughes and Babcock,

1950), *Parthenium argentatum* (Asteraceae) (Gestrel, 1950). Hughes and Babcock, 1950 and Gestrel, 1950 independently proposed another model with a multi allelic single locus. The pollen compatibility was determined by both the alleles of the parent on which the pollen was produced. This came to be known as sporophytic self-incompatibility.

Compatibility and incompatibility patterns resulting from diallele reciprocal crossing experiment in species can be used to distinguish between gametophytic self-incompatibility and sporophytic self-incompatibility. In a gametophytic self-incompatibility system, crosses between individuals sharing one *S* allele, in common tend to display half compatibility. In case of sporophytic self-incompatibility system, there can be only one of either two; either a cross is fully compatible or fully incompatible. Dominance between the alleles was also observed since the reciprocal crosses of certain compatible crosses were found to be incompatible (de Nettancourt, 2001; Hiscock and McInnis, 2003).

Significant crucial correlations have been found to exist between the possession of a type of self-incompatibility system and certain floral attributes (Brewbaker, 1957; Zavada, 1984; Heslop-Harrison and Shivanna, 1997). Gametophytic self-incompatibility systems are attributed to the possession of a moist stigma, a bicellular (non-divided) pollen and the occurrence of inhibition at the stylar region, where as a dry stigma, a tricellular (divided) pollen and the inhibition at the stigmatic surface is found to be common in sprophytic self-incompatibility system. Correlations of this sort can be quite helpful for developing hypotheses on the kind of self-incompatibility present for those species of plants for which genetic data is limited (Gibbs, 1986).

Besides the above mentioned gametophytic and sporophytic self-incompatibility systems, a late acting self-incompatibility system has also been observed to be present. In this system, it is inside the ovary the inhibition of pollen tube occurs, just before fertilization or at some point of time soon after fertilization (de Nettancourt 1997, 2001; Seavey and Bawa 1986; Barrett, 1988; Gibbs and Bianchi, 1999).

The existence of late-acting ovarian self-incompatibility has been found in all major clades throughout angiosperms. In a lot of woody perennial species of different families, such as, Malvaceae (*Theobroma cacao, Chorisia sp*.); Ericaceae (Rhododendron sp.); Fabaceae (*Acacia retinodes*); Winteraceae (*Pseudowintera colorata*); Myrtaceae (*Eucalyptus sp.*); the entry of pollen tube into the ovary with no stylar hindrance, followed by a very poor seed set has been observed (Kenrick *et al.,* 1986; Gibbs and Bianchi, 1999; Sage and Sampson, 2003).

Late acting ovarian self-incompatibility (OSI) has shown significant variation in the timing and site of pollen rejection. The OSI systems are divided into two; pre-zygotic and post zygotic (Sage *et al.,* 1994; Sage and Sampson, 2003).

In *Theobroma cacao* a prezygotic mechanism is observed, where the gametic fusion is inhibited at the ovule (Knight and Rogers, 1955). In certain species, the growth of pollen tube is arrested at the nucellus, for example *Acacia retinoides* (Kenrick, *et al.* 1986). In case of post zygotic OSI, the fertilization of the ovule takes place, yet the zygote or the endosperm so formed is malformed and aborted. This type is reported in *Rhododendron* species (Williams *et al.,* 1986). In *Gasteria* spp, Sears (1937) reported that self-fertilization does not stimulate the development of the ovule.

Though the genetic basis of OSI is not definitely known, it is most often hypothesized to be gametophytic (Sage *et al.,* 1994). A few efforts have been taken via pollination experiments to characterize the genetic mechanism governing OSI. Controlled crossing experiments conducted in *Acacia retinoides* indicated that it was under gametophytic regulation. Experiments involving extensive crossing in cocoa has revealed a combination of both sporophytic as well as gametophytic regulation was taking place (Cope, 1958; Knight and Rogers, 1955). It was hypothesized to be controlled by three unlinked loci.

2.4 Molecular mechanisms underlying self-incompatibility

Even though both the systems carry out the same purpose, the genetics that govern them and the timing of inhibition are significantly dissimilar. In the same way the molecular mechanisms or pathways undertaken to achieve this feat, the various genes and proteins involved are significantly contrasting.

2.4.1 Gametophytic self-incompatibility

In angiosperms, the most abundant mechanism controlling self-incompatibility is GSI and is observed in about 80-90 families. This is quite evident from the phylogenetic study conducted by Allen and Hiscock (2008) of the GSI. In all the major clades in angiosperms, the GSI is found to be quite widespread.

Within the GSI, at the molecular level, two drastically different mechanisms both of which involve a single *S* locus have been identified and studied in detail. First GSI to be identified and characterized in detail was an S-RNase mechanism in Solanaceae. This was subsequently found to be present in Rosaceae, Scrophulariaceae and Campanulaceae (Sassa *et al.,* 1993). The second is a complex mechanism that till date been has been only reported in *Papaver rhoeas* L. (Lawrence *et al.,* 1978). The mechanisms set off a number of signal transduction cascades, which leads to immediate inhibition of the pollen tube leading to cell death.

2.4.1.1 S-RNase based SI - Solanceae

The stylar protein S-RNase, a ribonuclease encoded by the S locus, plays the key role of selective inhibition of the growth of self pollen. All the S-RNase proteins studied till date possess seven domains out of which 5 domains are highly conserved and two of them are hyper variable. Two of the conserved domains C3 and C4 domains possess the active site for ribonuclease activity and the two hyper variable regions are responsible for the selectivity and specificity for the self pollen (Matton *et al.,* 1999).

The male determinant of self-incompatibility was discovered to be an *S* locus F box protein (SLF) through various gain and loss of function experiments. The SLF protein takes the role of an E3 ubiquitin ligase, which takes part in proteasomal degradation by ubiquitination (Hua and Kao, 2007).

The mechanism works on the basis of non self-recognition based model. As the pollen tube grows through the stylar canal, it takes up the S-RNases from the pistil. Each of the SFB proteins acts as a detoxification agent. It ubiquinates the non self S-RNases and directs them for proteasomal degradation in case of cross compatible cases (Kubo *et al.*, 2010).

2.4.1.2 Self-incompatibility in Papaver

Papaver rhoeas displays a gametophytic genetic control but the underlying molecular mechanism is totally different from the S-Rnase based GSI. The specific interaction between a protein secreted by the pistil *PrsS* (pistil determinant) and a transmemberane protein *PrpS* (male determinant) on the pollen leads to the self-recognition and programmed cell death (PCD) or apoptosis (Poulter *et al.*, 2010).

In case of an incompatible pollen, the interaction between *PrsS* and *PrpS* sets off a Ca^{2+} dependent signaling system. This leads to a multitude of downstream processes like the modification of the cellular cytoskeleton, phosphorylation of Mitogen Activated Protein Kinases (MAPK) and soluble pyrophosphates (sPPases), rise in Reactive oxygen species (ROS) and nitric oxide levels, activation of several enzymes that displays caspase like activity. Towards the end, fragmentation of chromosomes has also been reported (Jordan *et al.*, 2000). This eventually leads to PCD or apoptosis.

2.4.2 Sporophytic Self-incompatibility

The S- haplotype or the S – locus comprising of three tightly linked genes *SLG, SRK* and *SCR* is known to be present in all the species that have been studied for self-incompatibility in Brassicaceae (Kusaba *et al*, 2001).

The male determinant is the product of *S* locus *Cysteine Rich* gene (SCR or SP - 11) which was secreted on the tapetum wall of the pollen parent, this is incorporated on to the outer wall of the pollen (Schopfer *et al.*, 1999, Takayama and Isogai, 2005). The *S* locus glycoprotein (SLG), expressed in the stylar canal, acts as an enhancer of self-incompatibility. It is hypothesized that the self-recognizing interaction leads to the phosphorylation reactions with the arm repeats containing protein (*ARC1*) leading to a signaling cascade (Stone *et al*., 1999).

In, Brassicaceae, the dominance expressed between the S alleles in the pollen was reported to be much more common compared to the level of dominance observed in the stigma (Hatakeyama *et al.,* 2001). Kusaba *et al.* (2002) through expression studies reported that the dominance of the *SCR/SP11* alleles is at the transcription level. Shiba *et al.* (2002) hypothesized that the dominance was as the result of some kind of epigenetic silencing mechanism. Unlike *SCR/ SP11* in *SRK*, the expression rates of both the alleles were found to be similar, as a result not governed by epigenetic regulations (Hatakeyama *et al.,* 2001).

2.5 Self-incompatibility in cocoa

The first evidence of self-incompatibility in cocoa was reported by Pound (1932), he characterized the cacao trees into self-compatible and self-incompatible. Self-incompatibility in cocoa was found to be a late acting ovarian self-incompatibility (OSI) mechanism.

The SI mechanism in cocoa is eccentric for two reasons: 1) unlike other mechanisms, the recognition, the S- gene expression occurred only after the ovule and pollen had made contact. 2) The result of rejection is the abscission of the entire flower and not just the pollen tube (de Nettancourt, 1997).

Mather (1944) has stated that the incompatibility system in cocoa is comparatively a primitive and least developed one. The reason being there is no inhibition of the pollen at the stylar region.

Knight and Rogers (1953 and 1955) conducted cross pollination experiment between the progenies of three self-incompatible genotypes and came to the conclusion that self-incompatibility in cocoa was under the sporophytic genetic mechanism.

Cytological studies conducted by Cope (1962) proved that apart from sprorohytic mechanism, a gametophytic mechanism was also found to regulate self-incompatibility in cocoa. In his study, the most common manifestation of incompatibility in self and cross pollination was in the ovary, 25 per cent, 50 per cent and 100 per cent non-fusion ovules were found to be present. From this it was evident that the physiology of the gametes was not 100 per cent similar when it comes to gametic fusion. Hence, he indicated the gametic fusion depended on the haploid constitution of the individual gamete.

Knight and Rogers (1955) suggested that a single S locus which was multi allelic for five different alleles, where in a dominance relationship of S1> S2 = S3 > S4 >S5 may exist. Further, more alleles were discovered and reported by Cope (1962) and Glendinning (1967). In Amelonado cocoa a self-compatible (*sf*) allele was reported by Cope (1962). The dominance relationship between the so far discovered alleles in cocoa were determined as S0 = S1 > S2 = S3 > *Sf* > S 4 > S5.

Since selfing of SI cocoa trees led to the formation of 25 per cent, 50per cent and 100 per cent non gamete fused ovules in the ovary, in addition to the previously reported S locus another two loci was hypothesized to be present based on cytological studies conducted by Cope (1958).

Anatomical research by Bouharmont (1960) had revealed that in case of both self-compatible and self-incompatible reactions, the pollen tube grew through the stylar canal without any hindrance at the same rate and both reached the ovary. Till this point no kind of difference was found between the SC and SI interaction.

Cocoa is reported to be the first model of angiosperms to have an incompatibility mechanism, which is based on the success or failure of syngamy (Cope, 1962).

After 24 hours of pollination, in case of the compatible reactions the pollen penetrates into the ovary through the micropile, the fusion of the plasma membrane of the sperm nuclei and the egg nucellus takes place (Cope, 1939 and 1940). This is followed by the union of the male and female nuclei, followed by normal double fertilization and the flower is retained. In case of the self-incompatible reaction, the plasma membranes fail to fuse and syngamy doesn't occur and finally abscission of the flower was observed (Knight and Rogers, 1955).

The SI mechanism in cocoa works in such a manner that non fusion of male and female gametes which possess the same allele in a few ovules, results in loss of the entire ovary even though proper compatible fusion have occurred in other ovules (Cope, 1962).

In case of cocoa SI, a lot many investigations have clearly indicated that the proportion of non-fused ovules (incompatible) and fused ovules (compatible), determine if the floral abscission takes place and whether the fruit sets or not (Glendinning, 1960 and Cope, 1962).

Few studies have shown that the self pollen recognition may occur much earlier than it was reported by Cope (1962). In a study conducted by Aneja (1994) it was observed that self-incompatibility in cocoa occurred at two stages; one, during germination of pollen on the stigmatic surface and two, during the phase of gametophytic fusion (previously reported by Cope, 1962).

Lanaud *et al.* (2017) conducted fine mapping, GWAS (Genome Wide Association Studies) and cytological studies confirming the existing hypotheses that both sporophytic and gametophytic mechanisms take part in regulating the late acting self-incompatibility in cocoa. The existence of two loci independent of each other, one in chromosome 1 (Ch1) and one in chromosome 4 (Ch4) was confirmed. The CH1 locus, which was reported for the first time was observed to act after the pollen tube germination but before the gametic fusion took place. They hypothesized that this level of selection occurred as the result of

the embryo sac preventing the entry of sperm nuclei. They reported Ch4 locus to be involved in the fruit drop.

In cocoa self-fertilization can be achieved by using the mentor pollen affect. Glendinning (1960) successfully obtained selfed seeds from Upper Amazon cocoa (strictly incompatible germplasm) by pollinating a mixture of huge amount of compatible pollen and some amount of self pollen. The selfed seeds were easily identified at the seedling stage with the help of "axil spot gene" a phenotypic marker present in the compatible pollen. Gamma irradiated mentor pollen or heat-treated pollen can also be used to overcome the incompatibility barrier (Ampomah *et al*., 1990).

Higher concentration of CO_2 has been reported to suppress incompatibility in a number of cases, for example: SSI in *B. campestris* L. (Dhaliwal *et al.,* 1981) *B. oleracea* L. *Capitata* (Nakanishi and Hinata, 1973; Taylor, 1982; Palloix *et al.,* 1985) and GSI in *Trifolium repens* L. (Douglas and Connolly, 1989). In cocoa, the fruit set was found to be doubled when selfed flowers were enclosed; the result was hypothesized to be the action of increase in the CO_2 levels (by 8% in 6 hours) from the floral respiration (Aneja *et al*., 1992).

An experiment was conducted by Baker *et al.* (1997) on floral abscission, they reported significant correlation between mode of pollination and alteration in the concentration of phytohormones. They reported increase in the levels of abscisic acid and ethylene in case of incompatible pollen interaction, while remained consistent in compatible reactions. Indole acetic acid levels where found to increase in case of compatible interactions, while it remained consistent in case of incompatible reactions.

Isozyme analysis conducted by Warren *et al.* (1995) revealed one of the loci controlling self-incompatibility to be linked with the enzymes acid phosphatase and isocitrate dehydrogenase loci. Hence these isozyme markers could be used to assess the compatibility type in cocoa seedlings.

2.6 Self-compatibility and yield

Self-compatibility status of the crop is considered as a significant factor contributing to yield in cocoa. Cope (1939) reported that on an average, self-compatibile trees set 66 per cent higher fruits than self-incompatible trees. Yield in self-incompatible clones are significantly dependent on the pollinator insects.

Even though majority of the cocoa populations are self-incompatible in nature, certain studies have shown the occurrence of self-pollination in certain wild types and cultivated clones. Genetic diversity and gene flow analysis conducted by Schawe (2013) provided evidence of a much higher rate of self-compatibility in cultivated cocoa.

Lachenaud *et al.* (2007) observed a relatively higher proportion of self-compatible trees within populations of elite high yielding cultivars

N'Zi et *al*., (2017) conducted compatibility assessment in some elite cocoa cultivars and reported that among them some of the highest yielding clones were self-compatible. These clones could significantly increase the yield potential of polyclonal gardens which are strictly dependent on cross pollination.

2.7 Breakdown of self-incompatibility

In flowering plants, the shift from strictly cross-fertilizing nature (self-incompatible) to partial or predominant self-fertilizing nature (self-compatible) is often observed (Ligic *et al.*, 2005). This transformation does have significant consequence on the genetic make-up of a population and is also considered to be one of the most prevalent transformations found to occur in the angiosperms. The breakdown can be the result of numerous mechanisms.

Doubling of chromosomes or polyploidy has been reported to cause collapse of self-incompatibility in certain species (Miller, 2000). In *Lycopersicon peruvianum,* a diploid self-incompatible species turned to be self-compatible when tetraploids were developed through tissue culture (Chawla *et al.,* 1997).

Chromosomal doubling has also been reported in the breakdown of sporophytic self-incompatibility in Brassicacea. Pandey (1968) demonstrated that cholchicine induced chromosomal doubling can result in conversion of self-incompatible *Nicotiana alata* into self-compatible.

In *Prunus avium*, Sonneveld *et al.* (2005) examined two pollen part mutant haplotypes, one having a major deletion in the *SFB* gene and the other carrying a frame shift mutation which led to the expected loss of self-incompatibility.

In *Prunus cerasus* (Tsukamoto *et al.,* 2006), two stylar-part mutants S (6m2) and S (13m) were identified to be self-compatible. When they were compared with the wild type, the RNase transcript from S (6m2) possessed a base pair deletion which led to frame shift mutation and RNase transcript from S (13m) possessed one base pair substitution that led truncation.

Boyes *et al.* (1991) reported that mutation that occurs on certain genes outside the S locus can also lead to partial or complete self-compatibility. These are called modifier genes or loci. These aren't required for specificity, but play vital roles in rejection of the pollen through the downstream signaling pathways. Mutations in these genes can also lead to the breakdown in self-incompatibility.

Through differential hybridization experiment between SC *N. alata* and SI *N. plumbaginifolia*, McClure *et al.* (1999) identified 101 residue polypeptide possessing an aspargine rich residue at its C-terminal in *Nicotiana spp*. Anti-sense experiments proved the role of this protein in pollen rejection.

Tsuchimatsu *et al.* (2010) characterized the self-compatibility in *Arabidopsis thaliana* as the result of an inversion mutation of 213 base pairs in the male determinant gene, *S locus cysteine rich* (*SCR*). The female specificity gene, *S receptor kinase* (*SRK*), was reported to be still functional in *Arabidopsis thaliana*. They proved this by trans genetically inverting the inversion and reconstituted self-incompatible plants.

2.8 Markers association with self-incompatibility in cocoa

Self-compatibility assessment in cocoa is a highly tedious task. After the plant is taken to field, one has to wait for 3-5 years for the plant to start flowering, hence the elimination of unnecessary genotypes from the breeding program is delayed. Moreover, these genotypes have to be planted in the field to be assessed, which is a drastic wastage of area, resource and time from the breeding program. This assessment process is considered to be error prone due to environmental factors such as rainfall, insect attack, temperature variations *etc* (Da Silva *et al*., 2016).

Mapping techniques such as association mapping, fine mapping, QTL mapping, are the initial steps that are involved in the detection of individual candidate genes responsible for a particular trait.

Crouzillat *et al.* (1996) established a linkage map of cocoa using a back cross population of Catongo variety with the help of 138 molecular markers.

Genes that are directly involved in the mechanism of self-incompatibility in cocoa is yet to be determined. However, marker trait association analyses are of great help in indicating the location of these genes on the chromosomes and also selection of self-incompatible genotypes with high yield. Roayaert *et al.* (2011) have identified genomic regions and molecular markers that are strongly linked with the self-compatible and incompatible trait. They conducted marker-trait association analysis and reported microsatellite markers, mTcCIR222, which displayed strong association with self-compatibility. Three other markers were also reported to be present lying close to the loci, mTcCIR168, mTcCIR115 and mTcCIR158. All of these were reported to be present very close to the proximal end of chromosome 4 (6.7 cM from the proximal end).

Yamada *et al.* (2010) conducted a study to analyze the relationship between self-incompatibility and molecular marker. 68 plants from F_2 population of cross between clones of SCA 6 and ICS 1 and 342 molecular markers were used for the study, out of

which 19 markers were found to be significantly linked with compatibility. The polymorphic markers were mostly found to be present in chromosome four and seven. These results were in favour of the hypothesis put forward by Cope (1962), the role of more than one single locus in controlling self-incompatibility.

Da Silva *et al.* (2016) conducted a genome wide association mapping of the genes responsible for self-incompatibility in cocoa. A characterized population of 295 individuals was genotyped by sequencing. This spawned 5301 single nucleotide polymorphism markers. At the proximal end of chromosome 4, a region of 196 kilobase pairs, a significant number of markers displaying association to incompatibility were found to exist. This is an indication that important candidate genes could be present in that location. They also stated that candidate genes appeared to be present not only in this region but fan out to other regions of the genome as well provided the SNP variations where to be treated as random.

Lanaud *et al.* (2017) conducted fine mapping and GWAS on limited regions of the genomes and developed efficient molecular diagnostic markers on CH4 for the prediction of self-compatibility and fruit set.

2.9 Candidate gene analysis

Using two self-compatible haplotypes (S^4 and S^f) derived from X-ray irradiation Ushijima *et al*. (2004) characterized SFB gene in *Prunus avium* and *Prunus mume*. With the help of DNA sequence analysis, the haplotype ORFs were found to be defective. In S^4, an insertion of four bases on one of its exon leading to frame shift was present and in the other 6.8 kb insertion led to defective transcripts.

Qiao *et al.* (2011) through fine mapping and candidate gene analysis via a gain of function experiment, reported that the *dense and erect panicle 3*(*DEP3*) gene, located on the chromosome 7, which codes platin like phosphor lipase A2 (PLA2) protein in rice controlled the length and branching in panicle internode.

In Chinese soybean, Zhang *et al.* (2013) conducted a bulked segregation analysis (BSA) using SSR markers and found 26 polymorphic markers to be linked with *Phytophthora* resistance, to be present on 17th chromosome. Two co segregating SSR markers were found in this region. This gene/allele was designated as *Rps10*. The gene was then found out to code for two candidate genes encoding serine/ threonine protein kinase. These sequence variations were analyzed and a gene specific marker was designed for making breeding program faster.

By conducting a whole genome re-sequencing study, Lim *et al.* (2014) carried out a QTL mapping and candidate gene analysis on traits related to plant architecture. They reported 54 genes to be present in QTL regions which were linked to six agronomic traits. They narrowed it down to 15 genes based on expression analysis and sequence analyses. Finally, 11 of those genes, SNP polymorphisms were detected that caused frame shift mutations which lead to truncation of the proteins.

Mir *et al.* (2014) conducted candidate gene analysis, QTL mapping and analysis combined with expression profiling and comparative genomics on floral determinacy in pigeon pea. They selected 7 genes on account of previous research conducted in other pulses for this trait. From their studies they concluded that *CcTFL1*is the most likely candidate for determinacy. From their results they developed gene based markers that will speed up the determinacy breeding programs not only in pigeon pea but also in related leguminous plants.

In L*ong Panicle (LP1)* 1 gene, in the third and fifth exon, Liu *et al.* (2016) reported two single nucleotide polymorphisms that lead to changes in the amino acids. The gene codes for a Remorin C-containing protein whose function is unknown. The sequencing analysis of the gene conducted by Liu *et al.* (2016) in two parents and 103 progenies showed that SNP on the third exon was linked with the panicle length.

In cabbage, fine mapping followed by candidate gene analysis of a male sterile mutant (51s) by Han *et al.* (2018) led to the identification of *BoTPD1*as a candidate gene responsible for male sterility. The gene *BoTPD,* which codes for 176 amino acid protein *Tapetum Determinant* was previously reported to play a key role in the anther development and anther cell fate determination. Sequence analysis of *BoTPD1*revealed a 182 base pair insertion in the third intron, which resulted in the disruption of the conserved motifs, responsible for splicing at the 5^{l} end. This would lead to the translation of truncated proteins.

Two mutants *Anthocyanin fruit (Aft)* and *atroviolaceae (atv)* displayed higher levels of pigmentation in the tomato fruits. Colanero *et al.* (2018) confirmed the role of *atroviolacea* gene *Solyc07g052490* which encodes a R3 MYB protein. Whereas *SlMYB-ATV* (Cao *et al.*, 2017), was directly involved in the synthesis of anthocyanin in tomato. MYBs are a class of transcription factors very well known to regulate anthocyanin pigmentation. They sequenced the genome of *atv/atv* followed by candidate gene analysis and detected mutation in R3MYB protein. Candidate gene analysis revealed in a 4 bp insert in the second exon, one SNP in the third exon and the 3^{l} UTR and seven in the first interon. The protein plays a role in feedback mechanism which diminishes the anthocyanin production upon activation by exogenous or endogenous stimuli. The version of R3-MYB TF in *atv* mutation translates a truncated form of the protein, which loses or partially loses its potential in feedback inhibition of anthocyanin pigment synthesis. Three gene specific Indel markers were developed which could further fasten molecular breeding of anthocyanin pigmentation in tomato.

In *Capsicum annum,* a comparative sequence analysis between complete sequences of whole mitochondrial genomes of cytoplasm male sterile line 138A and maintainer line 138B was carried out between the two mitochondrial genomes 14 InDels and 112 SNP where detected. Based on InDels and SNPS, they identified a number ORFs (orf115b, orf262a, orf292a, orf314a, orf157a, orf300a) which were potential candidates for the trait.

Out of these they reported ORF 314a and ORF 300a as strong candidates for the control of male sterility (Wang *et al*., 2019).

Functional analysis of sequence variations by Das *et al*., 2019 identified single nucleotide polymorphism and indel variations on the 5^{l}UTR (untranslated region) and the promoter region *of α tocopherol methyl transferase* (ZmVTE4) gene in maize. This gene directly affects the concentration of α tocopherol in maize kernals. Their analysis revealed 14 SNPs and eight Indels in these regions, which are responsible for the low and high accumulation rates of α tocopherol in the maize kernel. Further, development of co-dominant, allele specific markers are possible from these SNP and InDels, and can be used in molecular breeding for biofortification of maize with α tocopherol.

2.10 Candidate genes for self-incompatibility

A number of fine mappings, linkage and association studies conducted in cocoa previously have pointed out the various loci thought to be involved in the late acting self-incompatibility system governing cocoa. Yamada *et al.* (2010) with the help of polymorphic markers reported regions of chromosome four and chromosome seven to be involved in the mechanism. Roayaert *et al.* (2011) conducted a SSR marker analysis on self-incompatibility in cocoa and reported that markers were to be present in and around a region 6.7 cM from the proximal end of chromosome four to be to be associated with self-incompatibility. Genome wide association studies by Da Silva (2016), carried out using the genotyping by sequencing once again proved that chromosome four was significantly involved in self-incompatibility. A significant number of markers displayed strong association with a region spanning about 196 kilo bases, near the proximal end of chromosome four.

From the results from fine mapping and GWAS by (Lanaud *et al.*, 2017) analyzing targeted regions of the genome, they identified several candidate genes on CH1 and CH4 with the help of Criolo cocoa genome sequence Version VI (Argout *et al.,* 2010). In CH1

they considered three genes to be potential candidates. *Tc01_g007220* which is orthologous to *BAM1*from *Arabidopsis thaliana*, *Tc01_g007270* is an ortholog of *WDR5a* of *A. thaliana* and *Tc01_g007290* which is a putative transmembrane transporter protein. In CH4, a region of 257.27 kb was identified and the following regions were identified as potential candidates: *Tc04_g000160* Voltage-dependent L-type calcium channel of Brassica spp, *Tc04_g000190*, *Tc04_g000230,* Tc04_g000240 and Tc04_g000260 homolog of GEX1 in A. thaliana, *Tc04_g000320* was orthologuous to *PMZ* gene from *A. thaliana* and Tc04_g000330 orthologous to *ARC1* from *Brassica napa.*

2.10.1 *Serine Receptor Kinase (SRK)*

A receptor like protein kinase with an extracellular domain with high similarity with the *S locus glycoprotein (SLG)* is coded by the *Serine Receptor Kinase (SRK)* gene. In the sporophytic self incmpatibility system present in Brassicacea, SRK acts as the female determinant. SRK interacts with the male determinant-SCR/SP11 and also SLG. The interaction sets off a cascade of signaling interactions which ultimately lead to the rejection of pollen (Kakita *et al.,* 2007).

2.10.2 *S locus glycoprotein (SLG)*

The S- haplotype or the S – locus comprising of three tightly linked genes *SLG SRK* and *SCR* is known to be present all the species that have been studied for self-incompatibility in Brassicaceae (*Kusaba et al*., 2001).

The *S* locus glycoprotein (SLG), expressed in the stylar canal, acts as an enhancer of self-incompatibility. It is hypothesized that the self-recognizing interaction leads to the phosphorylation reactions with the arm repeats containing protein (*ARC1*) leading to a signaling cascade (Stone *et al.*, 1999).

2.10.3 *Barely Any Meristem 1* (*BAM1*) and *Barely Any Meristem 2* (*BAM2*)

Tc01_g007220 is orthologous to *A. thaliana Barely Any Meristem 1* (*BAM1*) and *Barely Any Meristem 2* (*BAM2*). BAM1 and BAM2 codes for CLAVATA1 – related leucine rich repeat receptor-like kinases, which are reported to have important role in cell to cell communication during the early stages of anther development, processes such as cell division and cell differentiation (Hord *et al.,* 2006).

Lanaud *et al.* (2017) reported that *Tc01_g007220* was found to be over expressed in self compatibile reaction.

In interior cellular layers of the growing anther, *BAM1*and *BAM2* are produced at a relatively higher rate than the other tissues. They have also been reported to be a part of male pollen and ovule development (Young *et al.,* 2006).

2.10.4 *COMPASS-like H3K4 histone methylase component WDR5a*

Tc01_g007270 is reported to be homologous to the COMPASS-like H3K4 histone methylase component WDR5a from *A. thaliana*. It is transducin WD-40 repeats harbouring protein, playing key roles in numerous biological processes through protein-protein interactions (Stirnimann *et al.*, 2010).

WD40 repeats are well known to be involved in final step of ubiquitination cascade which results in protein degradation, which is also reported to be an important part of self-incompatibility in plants (Wang *et al.*, 2004).

In many cases, it has been reported that WD40 domain containing proteins act as the ubiquitin binding site for the S-locus F box protein responsible for self-incompatibility in *Rosaceae* (Chen *et al.*, 2012).

2.10.5 *Voltage-dependent L-type calcium channel subunit alpha-1F*

Tc04_g000160 is found to be a homolog of voltage -dependent L-type calcium channel subunit alpha-1F gene from *Gossypium arboretum.* This gene has been reported to play a key role in cellular Ca^{2+} influx in animals and plant (Kotturi *et al.*, 2003).

Ca^{2+} ions act as secondary messengers involved throughout the double fertilization process. Denninger *et al.* (2014) showed how Ca^{2+} was involved in all sorts of cellular events during the double fertilization event such as, cross talk between synergids and the pollen tube apex, pollen tube rupture and delivery of sperm into the ovule and finally gamete activation and fusion. It is also reported that they might be involved in the blockage of polyspermy in the egg cell.

Iwono *et al.* (2015) reported SCR/SP11 and the SRK interactions in *Brassicacea* result in a spike in Ca^{2+} concentration in the stigmatic papillae, which is the result of Ca^{2+} influx into the cells via a transmembrane channel or transporter protein.

In *Papaver rhoeas,* Wheeler *et al.* (2010) has reported that the interaction between the incompatible pollen and the pistil determinant proteins results in inhibition of pollen tube growth through Ca^{2+} dependent signaling network, which ultimately results in programmed cell death of the incompatible pollen.

2.10.6 *Gamete Expressed Protein (GEX1)*

Alandete-Saez *et al*., (2011) reported *GEX1* to be expressed in multiple tissues, during various stages of embryo formation in *A. thaliana.* It was observed to be expressed in the embryo sac prior to cellularization, within the egg cell after cellularization, in the zygote soon after fertilization. It was also reported to be expressed in both the vegetative and the sperm cells of the male gametophyte.

2.10.7 Zinc finger AN1 domain-containing stress-associated protein 12 (PMZ)

Lanaud (2017) reported Tc04_g000320 to be orthologous to a Zinc finger AN1 domain-containing stress-associated protein 12 (PMZ) gene from *A. thaliana*. In SC reactions, it was expressed at a slightly higher at 12-24 hours after pollination. PMZ could interact with several other proteins which are involved in Ubiquitin mediated protein degradation.

Zinc Finger proteins (ZnF) are a protein super family that regulates growth and development processes in plants and also in conferring a role in biotic and abiotic stress responses (Feurtado *et al.* 2011; Giri *et al.* 2011). Gupta *et al.* (2012) reported that ZnF domain plays a key part in pathogen-host interactions and this is very much similar to the pistil-pollen interactions observed (Hodgkin *et al.,* 1988).

In almond, Gómez *et al.* (2019) identified candidate genes for gametophytic self-incompatibility components through a transcriptomal study on pollen pistil interactions. They reported a Zinc finger (ZnF) protein constant-like 12-like isoform protein to be one of the important candidates for SI in almond.

2.10.8 *ARM repeat-containing protein* (*ARC1*)

The *ARC1* from *Brassica rapa* is orthologous to *Tc04_g000330* protein. *ARC1* gene in *Brassica rapa* is known to be required for the rejection of self-incompatible pollen by the pistil. It takes part in the downstream reactions of the SRK. Stone *et al.* (2003) showed that the *ARC1* possess the activity of the E3 ubiquitin ligase in the presence of an active SRK. They found that in case of incompatibility, levels of ubiquinated proteins were found to be high. They proposed that *ARC1* is responsible for ubiquitin mediated proteasomal degradation of various compatibility factors which leads to pollen rejection.

Through transgenic RNA interference knock down experiment, the need of ARC1 in the rejection of incompatible pollen was confirmed in *A. lyrata* (Indriolo *et al.,* 2012).

A gain of function experiment conducted in *A. thaliana ARC1* mutants, it was proved that *ARC1* played a critical role in reconstitution of self-incompatibility (Indriolo *et al.*, 2014).

2.11 Hybrid production in cocoa

Physiological and genetic investigations had unveiled that the yield potential of cocoa is not yet fully exploited (Bertus, 2004). Demand for chocolate is increasing at a rate of 15-20 percent every year. And to meet this demand more area has to be brought under cocoa cultivation using improved genetic stock. Development of superior hybrids had significantly contributed to improve cocoa productivity in many countries (Kennedy *et al.*, 1987; Dias *et al.*, 2003). When cocoa hybrids were evaluated for their performance it was observed that they showed wide adaptability, low environmental interaction, improved yield when compared to traditional local cultivars (Dias *et al.*, 2003).

However artificial hybridization in cocoa is very tedious due to the small and fragile flower structure. Success rate is only one percent even with minimum disturbance (Mallika *et al.*, 2000).

Self-incompatibility is a pollination control mechanism which prevents self-fertilization. Hence this can be exploited in hybrid production by avoiding emasculation which is a cumbersome process (Minimol and Amma, 2013). Moreover, emasculation will damage the flower leading to reduction in success rate.

Conventional method to assess self-incompatibility is by selfing 100 flowers per tree. If no fruit set is observed then the plant is classified as self-incompatible (Mallika *et al.*, 2006). This is a time-consuming tedious process which will slow the phase of breeding programme.

Many candidate genes have been reported for self-incompatibility in various other crops (McCormick, 1998). However, the actual sequence variations in these genes are not yet studied in cocoa. Identification of appropriate gene resulting in self-

incompatibility will help to identify the mechanism in early stage and definitely help to fasten the breeding programme.

Self -incompatibility mechanism in cocoa is utilized in hybrid seed production. For this, flower bud due to open on the next day will be protected from insects with pollination hood. The following day morning (8.00-9.30 a.m.), flowers are selfed by keeping the anther of the same flower on the stigmatic lobes. The flowers are again covered with pollination hood and on the next day examined for the fruit set (Minimol and Amma, 2013).

Usually, to examine the self-incompatibility, 100 flowers will be pollinated per plant. The cocoa accession setting fruits on selfing are classified as self-compatible and which do not set fruit as self-incompatible (Mallika *et al.*, 2006).

Materials and methods

3. Materials and Methods

The study entitled 'Candidate gene analysis on self-incompatibility in cocoa' was carried out at Center for Plant Biotechnology and Molecular Biology, College of Horticulture and at Cocoa Research Centre (CRC), Vellanikkara, Thrissur, during 2018-2020.

3.1 Plant material

Two Forastero cocoa genotypes maintained at the Cocoa Research Center, which were contrasting in their self-compatibility status were used to study the variations in the sequence of selected putative genes that could contribute to the self-incompatibility character in cocoa. The genotypes were selected based on a CCRP report (2012), where the compatibility status of the flowers was determined based on the percentage of pod set when the flowers were selfed.

Table 1. List of genotypes selected for the study and their compatibility status

Compatibility status	**Genotype**	**No. of flowers selfed**	**No. of pods set**
Self-compatible	GVI 167 X GIV 18.5	17	4
Self-incompatible	IMC 20	150	0

Budding was done for the selected genotypes during January 2019 at the cocoa Research Center nursery, Vellanikkara, in order to get tender leaves for DNA isolation throughout the period of study. The seedlings of the budded plants are shown in Plate 1.

3.2 Selection of candidate genes and retrieval of their sequence

Nine putative genes were selected for the candidate gene analysis.

Table 2. List of candidate genes selected

Sl. No.	**Putative gene**	**Species/ plant**	**Reference**
1	*Serine Receptor Kinase (SRK)*	*Brassica* sp.	McCormick, 1998
2	*S- Locus Glycoprotein (SLG)*	*Brassica* sp.	McCormick, 1998
3	*Barely Any Meristem 1 (BAM 1)*	*Arabidopsis thaliana*	Lanaud *et al*., 2017
4	*Barely Any Meristem 2 (BAM 2)*	*Arabidopsis thaliana*	Lanaud *et al*., 2017
5	*COMPASS-like H3K4 histone methylase component (WDR5a)*	*Arabidopsis thaliana*	Lanaud *et al*., 2017
6	*Voltage dependent L-type calcium channel subunit (alpha –IF)*	*Gossypium* sp.	Lanaud *et al*., 2017
7	*Gamete Expressed protein (GEX 1)*	*Arabidopsis thaliana*	Lanaud *et al*., 2017
8	*Zinc Finger AN1 domain containing stress associated protein (PMZ)*	*Arabidopsis thaliana*	Lanaud *et al*., 2017
9	*ARM repeat containing protein (ARC1)*	*Brassica* sp.	Lanaud *et al*., 2017
10	*Hapless 2 (Hap 2)*	*Arabidopsis thaliana*	Lanaud *et al*., 2017

3.3 Sequence retrieval of the candidate genes

3.3.1 Retrieval of nucleotide sequences of the candidate genes from respective species

The genomic nucleotide sequences of each of the candidate genes from their respective species were retrieved from the open access database GenBank. A search was

done in the database with the gene name and the respective species name. The gene sequence from the respective plant species were downloaded in FASTA format.

3.3.2 Retrieval of homologous gene sequences in cocoa

Using the nucleotide sequence information of the genes, which were retrieved from the GenBank, gene homologues were identified in cocoa and their sequence for all the nine candidate genes were retrieved. The homologues gene sequences were searched from the cocoa whole genome using the BLASTn search tool available on the Cocoa Genome Hub (Criollo genome version 2.0) (http://cocoa-genome-hub.southgreen.fr/).

3.4 Primer Designing

3.4.1 Selection of regions to be amplified

Regions within the gene carrying maximum exonic sequence were obtained using ORF finder. Using Bioedit, sequences of amplifiable length (around 1 kb) were trimmed and retrieved.

3.4.2 Primer designing

Primers for the trimmed sequences were designed using the online platform Primer3 (Untergasser *et al.*, 2012). Primer sets of optimum size, melting temperatures, GC content and least dimerization capacity, were selected for all the ten candidate genes, targeting the regions carrying maximum exonic information.

3.5 Isolation of DNA

Lemon green coloured leaves, which are in a transitional stage from young to mature leaves, were reported to yield higher quality and quantity of DNA with lesser amount of polyphenols and polysaccharides (Ramírez *et al.*, 2018). Leaves at this stage were used to isolate the genomic DNA.

From the respective genotypes, lemon green coloured leaves were collected and quickly covered using aluminum foil and brought to the laboratory in an ice box. The leaves

were wiped clean using autoclaved distilled water and surface sterilized using 70 per cent ethanol. The leaves were then stored in -80 ^{o}C until used. Since the midrib and veins have been reported to contain high amounts of extra-cellular polysaccharides in their cavities (Clifford *et al.,* 2002), leaf tissue excluding the midrib and veins were used to extract the DNA, in order to minimize the interference of polysaccharides during DNA extraction.

Modified version of the CTAB method (Doyle and Doyle, 1987) was used to extract genomic DNA from cocoa leaves.

Reagents used

A. CTAB extraction buffer (5X)
    - 5 per cent CTAB (w/v)
    - 100 mM Tris Base (pH 8.0)
    - 20 mM EDTA (pH 8.0)
    - 1.4 M NaCl
    - 1 per cent polyvinyl pyrrolidone (PVP)
    - 0.2 per cent β-mercaptoethanol
B. Liquid nitrogen
C. Chloroform:Isoamyl alcohol (24:1 ratio v/v)
D. 70 per cent ethanol
E. 100 per cent ethanol
F. Ice cold Iso-propanol

Procedure

- CTAB extraction buffer (5x) was preheated to 65 ^{o}C in a 50 ml Oakridge centrifuge tube in a water bath
- 0.75g of fresh leaf tissue was ground into fine powder using pre-cooled pestle and mortar, using liquid nitrogen and transferred immediately into the Oakridge tube containing pre-heated CTAB extraction buffer.

- The tubes were incubated at 65 °C for 45 minutes with occasional gentle shaking.
- After incubation, tubes were set on ice for 10-15 minutes and then equal volume of chloroform-isoamyl alcohol was added into the tube and mixed by gentle inversion.
- The mixture was centrifuged at 12,000 rpm at room temperature for 20 minutes (Kubota 6500).
- Centrifugation has separated the contents in the tube into three distinct layers. Uppermost aqueous layer was pipetted out into a new sterile Oakridge tube very carefully by omitting the as much of the viscous polysaccharides, which is also present within the aqueous phase itself.
- The washing was repeated by adding equal amount of chloroform-isoamyl alcohol (24:1) to the Oakridge tube and centrifuging again at 12,000 rpm at room temperature for 15 minutes. The supernatant was carefully pipetted out into a new sterile Oakridge tube.
- Two fifth volume of ice cold isopropanol was added to the tube, mixed gently and incubated at -20 °C for one hour for complete precipitation of the DNA.
- Following incubation, the tubes were centrifuged at 4 ^{o}C, 12,000 rpm for 15 minutes and the supernatant was carefully discarded without disturbing the pellet.
- The pellet obtained was washed with 10 ml of 70 per cent ethanol by centrifuging at 10,000 rpm for 5 minutes and the alcohol was decanted.
- The pellet was air dried, dissolved completely in 50 µL of autoclaved distilled water and finally stored in -20 ^{o}C

3.6 Quality assessment by agarose gel electrophoresis

The quality of the isolated DNA was assessed by gel electrophoresis on a 0.8 per cent agarose gel (Sambrook *et al.,* 1989).

Equipment

- Electrophoresis unit- Bio-Rad power pack, gel casting tray, comb
- Gel documentation system- BioRad Gel DOC imaging system

Reagents

A. Agarose (Promega) 0.8 per cent (w/v)
B. Tris Acetic acid EDTA (TAE) buffer - 50X

   - Tris base - 242 g

   - Glacial acetic acid - 57.1 mL

   - 0.5 mM EDTA - 100 mL

C. Loading dye (Bangalore Genei)- 6X
D. Ethidium bromide - stock concentration 10 $mgmL^{-1}$; working concentration 0.5 μgmL^{-1}

Procedure

- 0.8 per cent agarose gel was prepared by dissolving 0.48 g of agarose in 60 mL of 1X TAE and heated using a microwave oven
- After cooling to about 42-45 ^{o}C at room temperature 5 µL of ethidium bromide from a stock solution of 10 $mgmlL^{-1}$ was added to the solution and mixed thoroughly.
- The 60 ml casting tray was fixed in the gel casting apparatus and the comb was placed at around one inch from the top of the tray.
- The warm solution was carefully poured inside the casting tray to a depth of about 5-6 mm depth and allowed to cool and solidify for about 30-40 minutes.
- Once solidified, the casting apparatus was dismantled and the gel tray was placed inside the horizontal electrophoresis unit and the unit was filled with 1X TAE buffer solution to a point where all the wells were submerged.
- 4 µL suitable molecular marker was loaded into the first well.

- Samples to be loaded for electrophoresis was prepared by mixing 1 μL of 6X gel loading dye and 5 μL of DNA sample and the entire 6 μL of the entire mixture was loaded into the wells.
- Tank was closed and the anode and cathode were connected to the power pack and electrophoresis was carried out at a constant voltage of 70 V until the tracking dye reached two thirds of length of the gel.
- Once the tracking dye reached two thirds of the length of the gel, the gel was taken out of the unit and viewed under a UV trans-illuminator for the presence of DNA.

3.7 Gel documentation

The DNA quality and purity were further analyzed with the help of NanoDrop spectrophotometer. Ethidium bromide binds to the hydrogen bonds between the double helix and this dye gets excited at a radiation of 260 nm and it re-emits yellowish orange coloured light in the range of 590 nm wavelength.

The gel containing the electrophoresed DNA was observed under a UV trans-illuminator. The gel was observed for intactness of the bands. The image was documented and saved in the gel documentation system (BioRad Gel DOC-It™ imaging system).

3.8 Assessment of DNA by spectrophotometer

The quantity and quality of the isolated DNA was examined by spectrophotometer **(**NanoDrop® ND-1000). Nucleic acid displays maximum absorbance at the wavelength of 260 nm, whereas proteins show highest absorbance at 280 nm. The absorbance were recorded at both the wavelengths and the purity was determined from the OD_{260}/OD_{280}. A ratio value in between 1.8 and 2.0 indicates that DNA is pure and relatively free from proteins and RNA. A ratio less than 1.8 is indicative of protein contamination and a value of more than 2.0 is indicative of RNA contamination.

Using the following relation, the quantity of DNA present in the sample was calculated:

1 OD at 260 nm = 50 μg DNA mL^{-1}

Hence the OD_{260} obtained from the sample X 50 gives the total quantity of DNA in μgmL^{-1}

Procedure

- NanoDrop spectrophotometer was connected to the system and the operating software ND-1000 was run.
- The nucleic acid option was selected. The sampling arm was lifted and the upper and lower measurement pedestal was wiped carefully and 1 μl of distilled water was pipetted out and placed onto the lower measurement pedestal. Sampling arm was closed and spectral measurement was initiated in the operating software.
- The reading was set to zero with blank sample.
- The sampling arm was once again opened. Both the pedestals were wiped clean. 1 μl of the DNA sample was pipetted and placed onto the measurement pedestal and the option measure was selected. Upon completion of measurement, the sampling arm and the upper and lower pedestals were thoroughly wiped using a soft laboratory wipe and the next sample was measured.

3.9 Standardization of amplifying conditions for the designed primers

Effective amplification of any primer is dependent on the on the proportion of the various components in the reaction mixture of a polymerase chain reaction. The reaction mixture consists of the DNA, which acts as the template, an assay buffer, required to maintain the optimum pH, $MgCl_2$, which acts as an essential co-factor for the enzymatic activity of the polymerase, dNTPs, which are the building blocks of the PCR product, the primer, short stretch of oligonucleotides which determines the target region and finally the polymerase which extends the primers by synthesizing long chains of nucleic acid.

Another important factor that affects the amplification is the temperatures of various stages in the thermal cycles. PCR consists of three stages (i) denaturation, where the samples are heated up to 92-98 ^{o}C in order to separate the strands (ii) annealing, where the samples are cooled down to about 48-78 ^{o}C usually annealing temperature will not go to 78. 68 may be fine. For the primers to bind to the complementary regions on the template DNA and (iii) extension, where the temperature is brought up to 72 ^{o}C and the polymerase extends the primers to form the PCR product. The process is repeated multiple times.

3.9.1 Standardization of annealing temperatures of individual primers using gradient PCR

The nature of the sequence of the primer and its length determines the annealing temperature for the thermal cycling reaction. Only an approximate value for this temperature can be calculated from the Tm and hence the annealing temperature for each primer sets have to be standardized. Gradient PCR, a variant of the normal PCR was used to optimize the annealing temperature, where the intended product was amplified to the maximum extend, nonspecific products were nil and primer dimerization at its minimum.

The PCR amplification was carried out in a 20 μL reaction mixture and the mixture composition are listed out below,

- Genomic DNA (40 ngμL^{-1}) - 2.0 μL
- 10X *Taq* assay buffer A (with $MgCl_2$) - 2.0 μL
- dNTP mix (10mM each) - 1.8 μL
- Forward primer (100 pM) - 0.8 μL
- Reverse primer (100 pM) - 0.8 μL
- *Taq* DNA polymerase (3U) - 0.4 μL
- Autoclaved distilled water -12.2 μL

Total volume - 20 μL

The following program was used to perform the PCR,

95^{o}C for 5 minutes	- Initial denaturation
95^{o}C for 30 seconds	- Denaturation
Gradient range for 30 seconds	- Primer annealing
72 ^{o}C for 1 minute	- Primer extension
72 ^{o}C for 8 minutes	- Final extension

4 ^{o}C for infinity to keep the samples on hold

A gradient range between the +5 ^{o}C and -5 ^{o}C of the average of calculated annealing temperature of the forward and reverse primers for a primer set was used.

Upon completion of the reaction, the entire 20 μL reaction mixture was mixed with 3 μL of 6X gel loading dye and was loaded into a 1.4 % agarose gel and electrophoresed. The optimum temperature for the reaction was determined based on the quality of the desired band and lack of undesired bands.

3.10 Sample preparation for sequencing

While preparing the PCR products for sequencing, a reaction volume of 30 μL was prepared and the composition of the reaction mixture are listed below,

• Genomic DNA (40 ngμ L^{-1})	- 3.0 μL
• 10X *Taq* assay buffer B (without $MgCl_2$)	- 3.0 μL
• $MgCl_2$	- 3.0 μL
• dNTP mix (10mM each)	- 2.7 μL
• Forward primer (100 pM)	- 1.2 μL
• Reverse primer (100 pM)	- 1.2 μL
• *Taq* DNA polymerase (3U)	- 0.6 μL
• Autoclaved distilled water	- 15.3 μL
Total volume	- 30.0 μL

The following program was used to perform the PCR,

94 ^{o}C for 5 minutes	- Initial denaturation
94 ^{o}C for 30 seconds	- Denaturation
Optimum temperature for 30 seconds	- Primer annealing
72 ^{o}C for 1 minute	- Primer extension
72 ^{o}C for 8 minutes	- Final extension

4 ^{o}C for infinity to keep the samples on hold

Upon termination of the PCR cycle, the amplification was confirmed by loading 10 μL of the PCR product in a 1.4 per cent agarose gel, followed by electrophoresis at 80V for one hour. Following the completion of time, the presence of the band was checked under the UV trans-illuminator. If the bands were present, the samples were stored in -20 $^{\circ}$C and sequenced.

3.11 Sequencing, data analysis and SNP search

3.11.1 Sequencing of PCR product

PCR products of all the primer sets from the self-compatible and self-incompatible genotypes were sequenced by outsourcing (AgriGenome Private Limited, Cochin). Sequencing was carried out using the Sanger sequencing method (Sanger *et al*, 1977).

3.11.2 Data analysis

3.11.2.1 Assembly of contiguous sequence (contig)

CAP3 sequence assembly program, maintained by PRABI-Doua, was used to build contiguous sequences from a set of overlapping sequences (http://doua.prabi.fr/software/cap3) (Huang and Madan, 1999). Using CAP3, the forward and reverse sequence of a primer set

was assembled into one single contig. The same was done for the forward and reverse sequence of all the nine sequenced primer sets.

3.11.2.2 Computational confirmation of the contig obtained

To confirm that the sequenced genes were the target itself, assembled contigs were subjected to BLASTn against cocoa whole genome database at cocoa genome hub (Criollo genome version 2.0) (http://cocoa-genome-hub.southgreen.fr/) (Argout *et al.*, 2010).

3.11.2.3 Sequence variation analysis

Once the contigs were prepared for the self-compatible sequence and self-incompatible sequences for a particular primer, the sequences were aligned along with the reference sequence (self–compatible). Multiple sequence alignment was done using the online tool Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/), which is maintained by European Bioinformatics Institute **(**Goujon *et al.*, 2010**).** The software aligns the sequences according to the identity shared between the sequences**.** The identical bases at any position are marked with an asterisk (*) symbol, non-identical bases at any point had no symbols and the InDels with gaps ( ). The non-identical bases and InDels were considered as SNPs. All the SNPs were identified and noted down.

3.11.2.4 Characterization of SNPs

To make sure that it is not a sequencing error, the SNPs were confirmed by referring to the chromatogram in AB1 file and observing the peak at which the SNP was found. The exact location of the SNP was searched by using the flanking region (8-10 nucleotide). The Base calling at the SNP locus was observed by carefully examining the peaks. The base calling at every SNP locus was thoroughly analyzed in both the self-compatible and self-incompatible contigs.

a. G VI 167 X G IV 18.5

b. IMC20

Plate 1. Budded seedlings of genotypes

Plate 2. Leaf stage used to isolate DNA (Lemon coloured)

Results and Discussion

4. RESULTS AND DISCUSSION

The study entitled “Candidate gene analysis on self-incompatibility in cocoa” was carried out at Center for Plant Biotechnology and Molecular Biology, College of Horticulture and Cocoa Research Center, Vellanikkara, Thrissur, during 2018-2020. The results of the study are presented and discussed in this chapter.

4.1 Planting material

The study was carried out using two Forestero genotypes whose compatibility status was predetermined, a self-compatible hybrid G VI 167 X G IV 18.5 and a self-incompatible genotype IMC 20. G VI 167 X G IV 18.5 was found to set four pods when seventeen flowers were selfed and hence considered self-compatible. IMC 20 set no pods when 150 flowers were selfed and hence considered self-incompatible.

4.2 Sequence retrieval and primer designing

4.2.1 Retrieval of nucleotide sequences of the candidate genes from respective species

Genomic and mRNA nucleotide sequences of the putative genes reported previously for their involvement in self-incompatibility mechanism were retrieved from the NCBI GenBank.

All the sequences, except *SLG* (*S- Locus Glycoprotein*) from *Brassica* sp., were retrieved from their reported host plants itself. The homologue for the gene *S- Locus Glycoprotein* was retrieved from *Arabidopsis thaliana*. mRNA sequence 1461 base pairs long coding for *SLG* was retrieved.

Serine Receptor Kinase (*SRK*), sequence was retrieved from the whole genome sequence of *Brassica oleracea* var. *oleracea*. The sequence of *SRK* gene was present in chromosome 6 in between the region of 2114759-32116266.

Barely Any Meristem 1 (*BAM* 1) sequence was retrieved from the whole genome assembly of *Arabidopsis thaliana.* The gene was present on chromosome 5 in between the regions 26636063-26639834.

Barely Any Meristem 2 (*BAM 2*) sequence was retrieved from the whole genome assembly of *Arabidopsis thaliana.* The gene was present on chromosome 3 in between the regions 18372845-18376540.

COMPASS-like H3K4 histone methylase component (*WDR5a*) sequence was retrieved from the whole genome assembly of *Arabidopsis thaliana.* The gene was present on chromosome 3 in between the regions 18368841 **-** 18370990.

Voltage dependant L-type calcium channel subunit (Alpha –IF) was retrieved from the genome assembly data of Gossypium turneri isolate. The gene was present in Chromosome 10 in between the region 45111222-45114646.

Gamete Expressed protein (*GEX 1*) was retrieved from the whole genome assembly of *Arabidopsis thaliana*. The gene was present on chromosome 5 in between the regions 22832541 - 22835766.

Zinc Finger AN1 domain containing stress associated protein (*PMZ*) sequence was retrieved from the whole genome assembly of *Arabidopsis thaliana.* The gene was present on chromosome 3 in between the regions 18372845-18376540.

ARM repeat containing protein (*ARC1*) sequence was retrieved from *Brassica rapa*. The mRNA sequence of 3311 base pair was retrieved.

Hapless 2 (*Hap 2*) sequence was retrieved from *A. thaliana* whole genome assembly. The gene was present on chromosome 4 in between the regions 7063255-7066878.

The NCBI reference sequence ID numbers and the annotated region (range) hosting the gene within the entire sequence of all the genes used in this study are presented in Table 2.

4.2.2 Retrieval of homologous gene sequences in cocoa

The homologous genes present in cocoa were retrieved using the host plant gene from the cocoa whole genome through BLASTn at Cocoa Genome Hub. Retrieved homologous gene sequences were analyzed.

Description of all the homologous genes retrieved from cocoa including the size of the gene, the chromosome which it lies on, its position within the chromosome and the number of exons present within the gene are presented in Table 3. The homologous genes identified from BLAST search results are presented in Figures (1-10).

4.2.2.1 *Serine Receptor Kinase (SRK)*

A receptor like protein kinase with an extracellular domain with high similarity with the *S locus glycoprotein (SLG)* is coded by the *Serine Receptor Kinase (SRK)* gene. In the sporophytic self-incmpatibility system present in Brassicacea, SRK acts as the female determinant. SRK interacts with the male determinant-SCR/SP11 and also SLG. The interaction sets off a cascade of signaling interactions which ultimately lead to the rejection of pollen (Kakita *et al.,* 2007).

From the BLAST result, the gene homologous to *Serine Receptor Kinase (SRK)* reported in *Brassica* sp. was found to be present in Chromosome 7 (Figure 1). It was present between the regions 24617907 and 24620867. The gene was 2,961 base pairs long and contained seven coding exonic regions.

4.2.2.2 *S locus glycoprotein (SLG)*

The S-haplotype or the S-locus comprising of three tightly linked genes *SLG*, *SRK* and *SCR* was present in all the species studied for self-incompatibility in Brassicaceae (Kasuba *et al*., 2001).

The *S locus glycoprotein* (*SLG*), expressed in the stylar canal, acts as an enhancer of self-incompatibility. It is hypothesized that the self-recognizing interaction leads to the phosphorylation reactions with the arm repeats containing protein (*ARC1*) leading to a signaling cascade (Stone *et al.*, 1999).

From the BLAST result, the gene homologous to *S locus glycoprotein (SLG)* reported in *Brassica* sp. was found to be present in chromosome 7 (Figure 2). It was present between the regions 4000353 and 4003605. The gene was 3,253 base pairs long and had nine exons, the maximum number among the candidate genes used in the study.

4.2.2.3 *Barely Any Meristem 1* (*BAM1*) and *Barely Any Meristem 2* (*BAM2*)

BAM1 and BAM2 codes for CLAVATA1 related leucine rich repeat receptor-like kinase proteins, which are reported to have important roles in cell to cell communication during the early stages of anther development, processes such as cell division and cell differentiation in various crops (Hord *et al.*, 2006).

From the BLAST search it was evident that the genes *BAM1* and *BAM2* which were reported as two separate genes in *Arabidopsis thaliana* were found to show similarity with only one common homologous gene in cocoa (Figure 3 and Figure 4). The homologous gene found in cocoa was present in chromosome 1 in between the regions 4013885 and 4021362. The gene was found to be 7,478 base pairs long containing four coding exonic regions.

DeYoung *et al.* (2006) reported that interior cellular layers of the growing anther secreted *BAM1*and *BAM2* at a relatively higher rate than the other tissues. They have also

been reported to be a part of male pollen and ovule development. Lanaud *et al.* (2017) had reported the homologue to be over expressed in self-compatible interaction in cocoa.

4.2.2.4 *COMPASS-like H3K4 histone methylase component* (*WDR5a*)

Tc01_g007270 is reported to be homologous to the *COMPASS-like H3K4 histone methylase component WDR5a* from *A. thaliana*. It is a transducin WD-40 repeats harbouring protein, playing key roles in numerous biological processes through protein-protein interactions (Stirnimann *et al.*, 2010). Wang *et al.* (2004) reported WD40 repeats to be involved in final step of ubiquitination cascade which results in protein degradation, which is also reported to be an important part of self-incompatibility in plants.

From the BLAST result, the gene homologous to *COMPASS-like H3K4 histone methylase component WDR5a* was found to be present in Chromosome 1 (Figure 5). It was present between the regions 4055118 and 4057603. The gene was 2,486 base pairs long and contained only two exons.

WD40 domain containing proteins that act as the ubiquitin binding site for the S-locus F box protein are responsible for self-incompatibility in *Rosaceae* (Chen *et al.*, 2012).

4.2.2.5 *Voltage-dependent L-type calcium channel subunit* (*Alpha-1F*)

Voltage-dependent L-type calcium channel subunit alpha-1F gene from *Gossypium arboretum* has been reported to play a key role in cellular Ca^{2+} influx in animals and plants (Kotturi *et al.*, 2003).

Denninger et al. (2014) showed how Ca^{2+} was involved in all sorts of cellular events during the double fertilization. In case of self-incompatibility, Iwano *et al.* (2015) reported SCR/SP11 and the SRK interactions in Brassicacea resulted in a spike in Ca^{2+} concentration in the stigmatic papillae. This was due to the result of Ca^{2+} influx into the cells through a trans membrane channel.

Through BLAST search, it was detected in cocoa that the gene homologous to the *Voltage-dependent L-type calcium channel subunit alpha-1F* gene from *Gossypium arboretum*, is present on chromosome 4 (Figure 6) at 130379…133389. The gene was 3,011 base pairs long with three exonic regions.

In *Papaver rhoeas*, Wheeler *et al*, (2010) reported that in case of incompatible interaction, Ca^{2+} dependent signaling network results in inhibition of pollen tube growth through programmed cell death of the incompatible pollen.

4.2.2.6 *Gamete Expressed Protein (GEX1)*

Alandete-Saez *et al.* (2011) reported *GEX1* to be expressed in multiple tissues in *A. thaliana*, during various stages of embryo formation. It was observed to be expressed in the embryo sac prior to cellularization and within the egg cell after cellularization and in the zygote soon after fertilization. It was also reported to be expressed in both the vegetative and the sperm cells of the male gametophyte.

From the BLAST search conducted, it was found that the gene homologous to *Gamete Expressed Protein (GEX1)* reported in *A. thaliana* was present on the chromosome 4 of cocoa, between 206256 and 208665 (Figure 7). The gene was 2,410 base pairs long, containing six exonic regions.

4.2.2.7 *Zinc finger AN1 domain-containing stress-associated protein 12 (PMZ)*

Gupta *et al.* (2012) reported that ZnF domain play a key part in pathogen-host interactions and this is very much similar to the pistil-pollen interactions observed (Hodgkin *et al.*, 1988).

In self-compatibility reactions, it got expressed at a slightly higher level at 12-24 hours after pollination. PMZ can interact with several other proteins which are involved in Ubiquitin mediated protein degradation (Lanaud *et al.*, 2017).

In almond, Gómez *et al.* (2019) identified candidate genes for gametophytic self-incompatibility components through a transcriptome study on pollen pistil interactions.

They reported a Zinc finger (ZnF) protein constant-like 12-like isoform protein to be one of the important candidates for self-incompatibility in almond.

In cocoa, the gene homologous to the *Zinc finger AN1 domain-containing stress-associated protein 12* (*PMZ*) from *Arabidopsis thaliana* was present on chromosome 4 (Figure 8), between 244000 and 245003. Among the candidate genes, this was the smallest with only 1,004 base pairs and two coding regions.

4.2.2.8 *ARM repeat-containing protein* (*ARC1*)

ARC1 gene in *Brassica rapa* is required for the rejection of self-incompatible pollen by the pistil. It takes part in the downstream reactions of the *SRK*. Stone *et al.* (2003) proposed that *ARC1* is responsible for ubiquitin mediated proteasomal degradation of various compatibility factors which leads to pollen rejection.

Through transgenic RNA interference knock down experiment, the need of *ARC1* in the rejection of incompatible pollen was confirmed in *A. lyrata* (Indriolo *et al.*, 2012). In a functional experiment conducted in *A. thaliana ARC1* mutants, Indriolo *et al*, (2014) proved that *ARC1* played a critical role in reconstitution of self-incompatibility.

In cocoa, gene homologous to the *ARM repeat-containing protein* (*ARC1*) gene from *Arabidopsis thaliana* was present on chromosome 6 (Figure 9), between 22707730 and 22711009. The gene was 3,280 base pairs with three exons.

4.2.2.9 *Hapless 2* (*Hap2*)

Hap2 has key roles in proper differentiation of sperm nuclei, fertilization, and signaling events in membrane fusion during fertilization (Wong and Johnson, 2010; Takahashi *et al*., 2018). HAP2 is a conserved protein identified in genomes of all major eukaryotic taxa, except fungi (Liu *et al*., 2008; Steele and Dana, 2009).

BLAST search had shown that the gene homologous to the *Hapless 2* (*Hap2*) from *Arabidopsis thaliana* was present on chromosome 3 of cocoa, between 33793949 and 33801670 (Figure10). This was longest among the candidate genes with 7,722 base pairs and 18 small coding regions.

Table 3. Description of all the sequences of genes retrieved from reported plant species

Sl. No.	Putative gene	Species/ plant	Sequence ID	Chromosome and position
1	*SLG*	*Brassica* sp.	XM_013773032.1	Chr 7 : 1-1461
2	*SRK*	*A. Thaliana*	NC_027753.1	Chr 6 : 2114759-32116266
3	*BAM 1*	*A. thaliana*	LR782546.1	Chr 1 : 26636063 - 26639834
4	*BAM 2*	*A. thaliana*	LR782544.1	Chr 1 : 18372845 - 18376540
5	*WDR5a*	*A. thaliana*	LR782544.1	Chr 1 : 18368841 - 18370990
6	*Alpha –IF*	*Gossypium* sp.	CP032577.1	Chr 4 : 45111222 - 45114646
7	*GEX 1*	*A. thaliana*	LR782546.1	Chr 4 : 22832541 - 22835766
8	*PMZ*	*A. thaliana*	LR782544	Chr 4 : 10518349 - 10519702
9	*ARC1*	*Brassica* sp.	XM_009143118.3	Chr 6 : 776 - 3194
10	*Hapless 2*	*A. thaliana*	LR782544.1	Chr 3 : 33793949 - 33801670

Table 4: Description of the homologous gene sequences retrieved from cocoa genome through BLAST analysis

Sl. No.	Putative Gene	Chromosome	Chromosomal Location	Number of exons	Gene length (bp)
1	*SRK*	6	24617907..24620867	7	2,961
2	*SLG*	7	4000353..4003605	9	3,253
3	*BAM 1*	1	4013885..4021362	4	7,478
4	*BAM 2*	1	4013885..4021362	4	7,478
5	*WDR5a*	1	4055118..4057603	2	2,486
6	*Alpha –IF*	4	130379..133389	3	3,011
7	*GEX 1*	4	206256..208665	6	2,410
8	*PMZ*	4	244000..245003	2	1,004
9	*ARC1*	6	22707730..22711009	3	3,280
10	*Hapless 2*	3	33793949..33801670	18	7,722

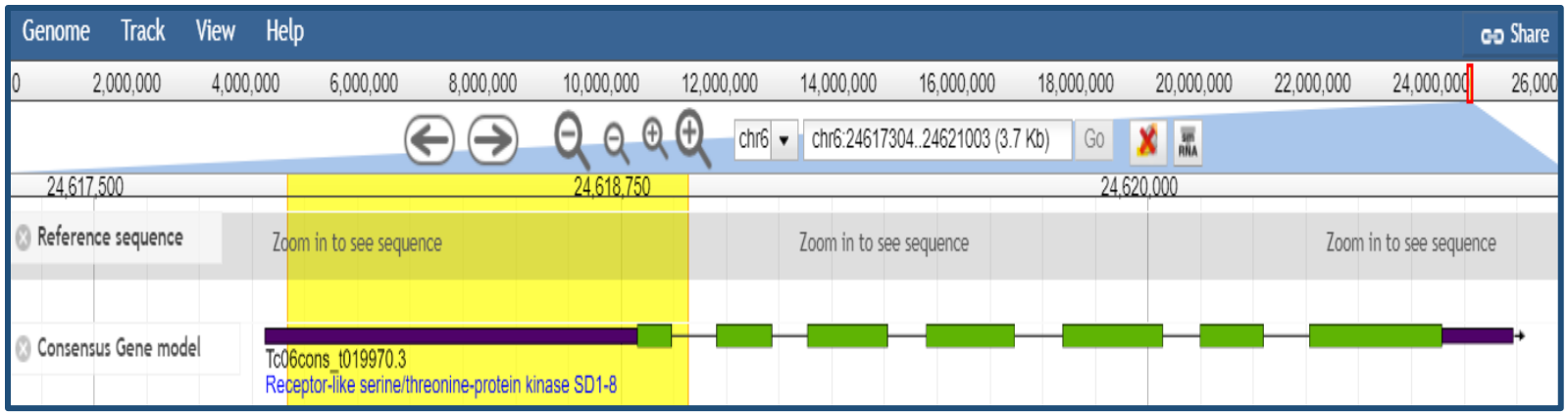


Figure 1. BLAST result for *Serine Receptor Kinase (SRK)*

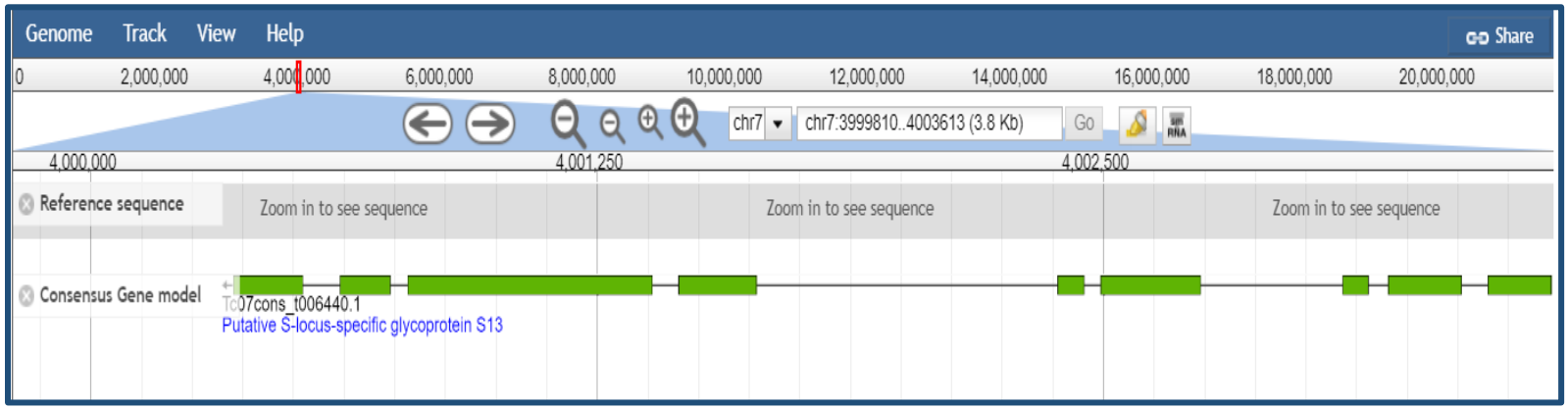


Figure 2. BLAST result for *S locus specific glycoprotein (SLG)*

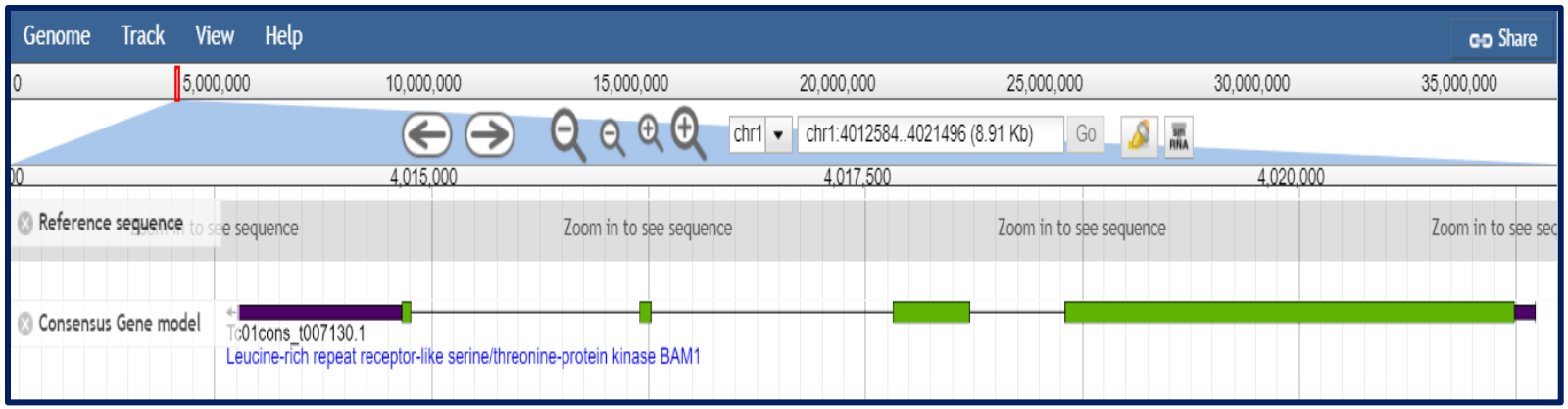


Figure 3. BLAST result for *Barely Any Meristem 1* (*BAM1*)

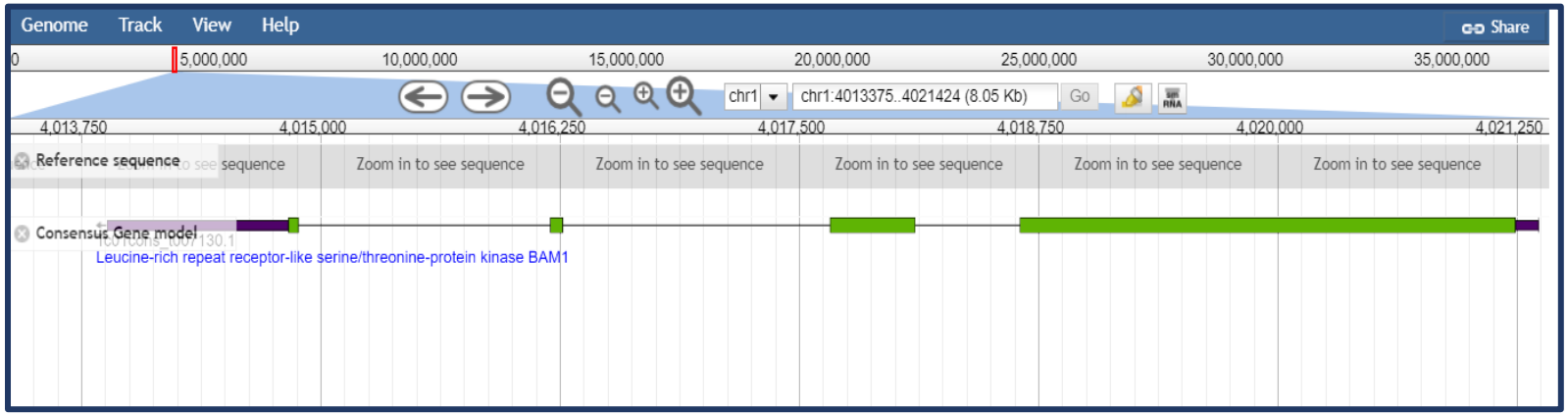


Figure 4. BLAST result for *Barely Any Meristem 2* (*BAM2*)

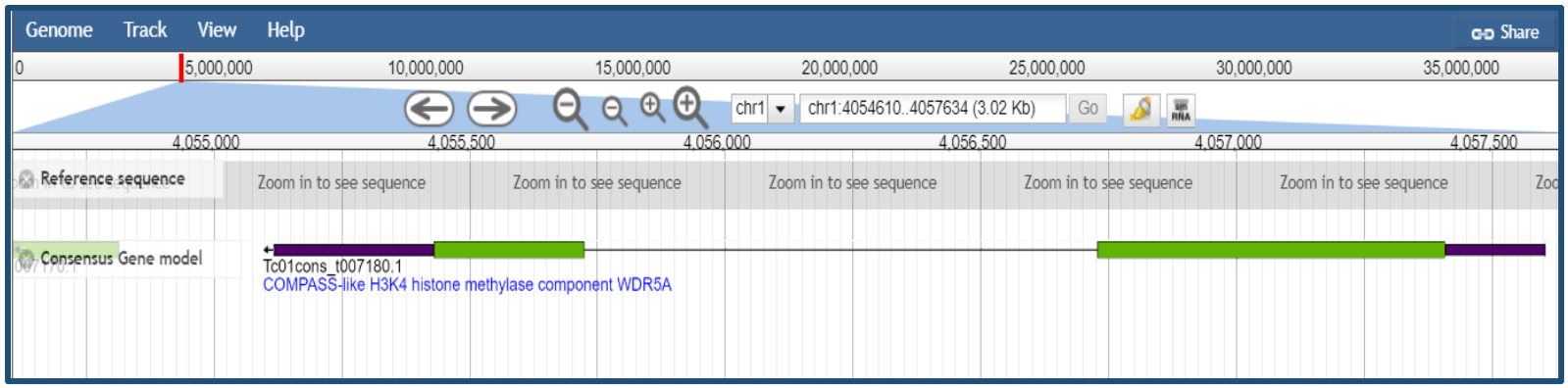


Figure 5. BLAST result for *COMPASS-like H3K4 histone methylase component (WDR5a)*

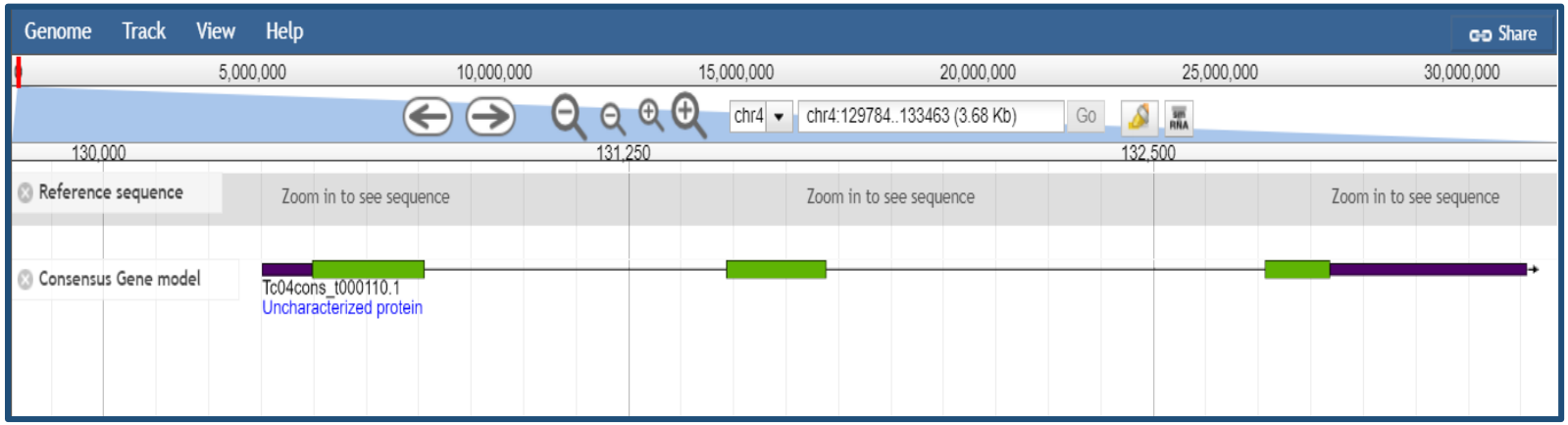


Figure 6. BLAST result for *Voltage-dependent L-type calcium channel subunit alpha-1F*

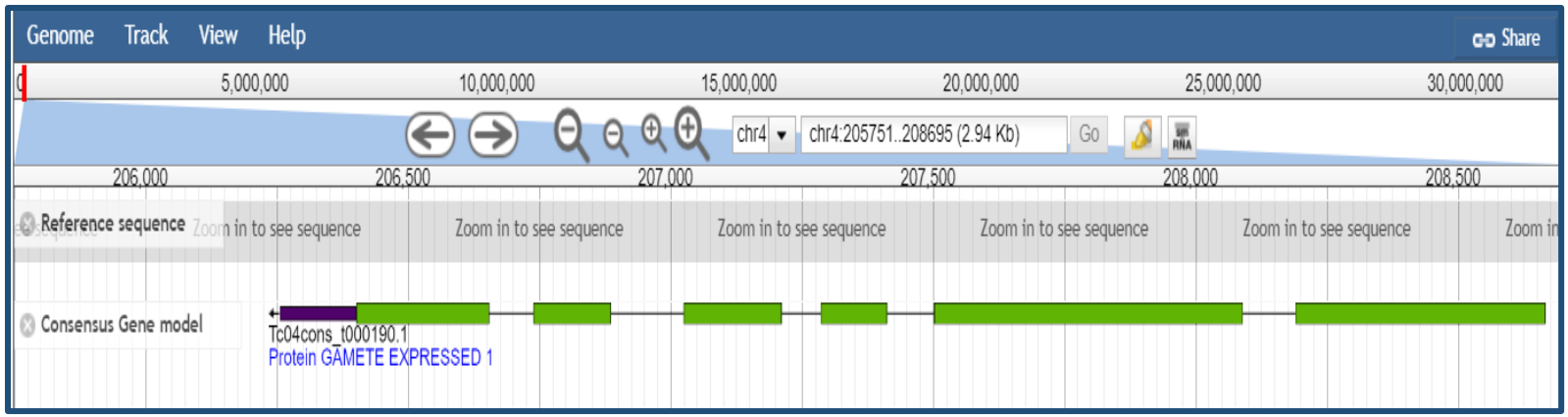


Figure 7. BLAST result for *Gamete Expressed Protein (GEX1)*

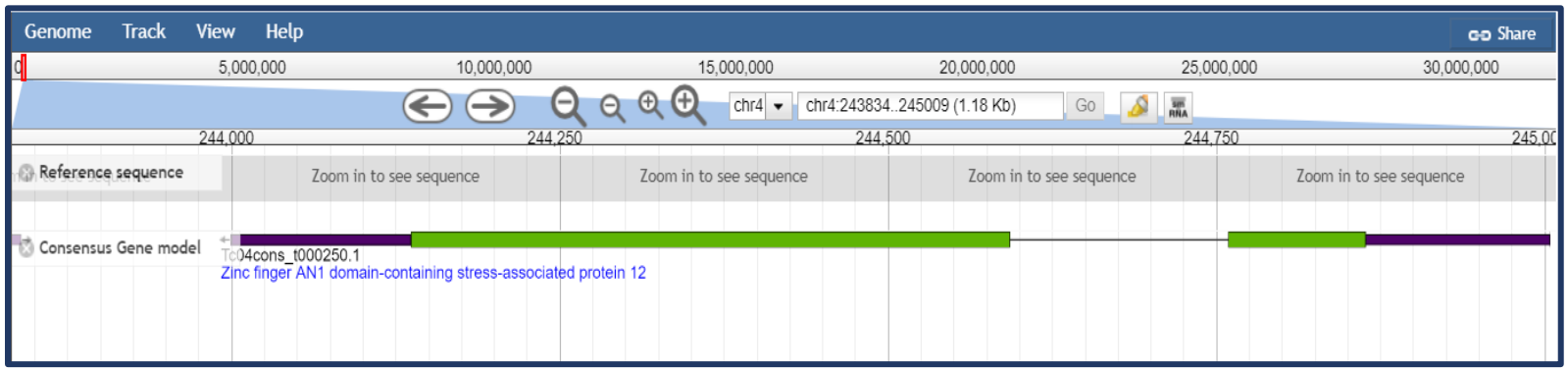


Figure 8. BLAST result for *Zinc finger AN1 domain-containing stress-associated protein 12 (PMZ)*

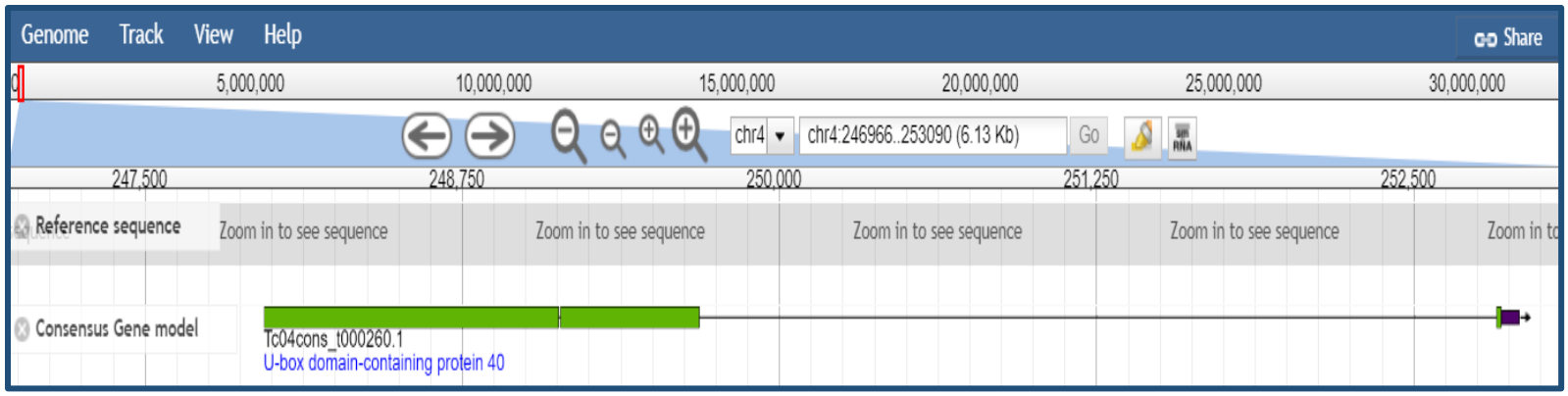


Figure 9. BLAST result for *ARM repeat-containing protein* (*ARC1*)

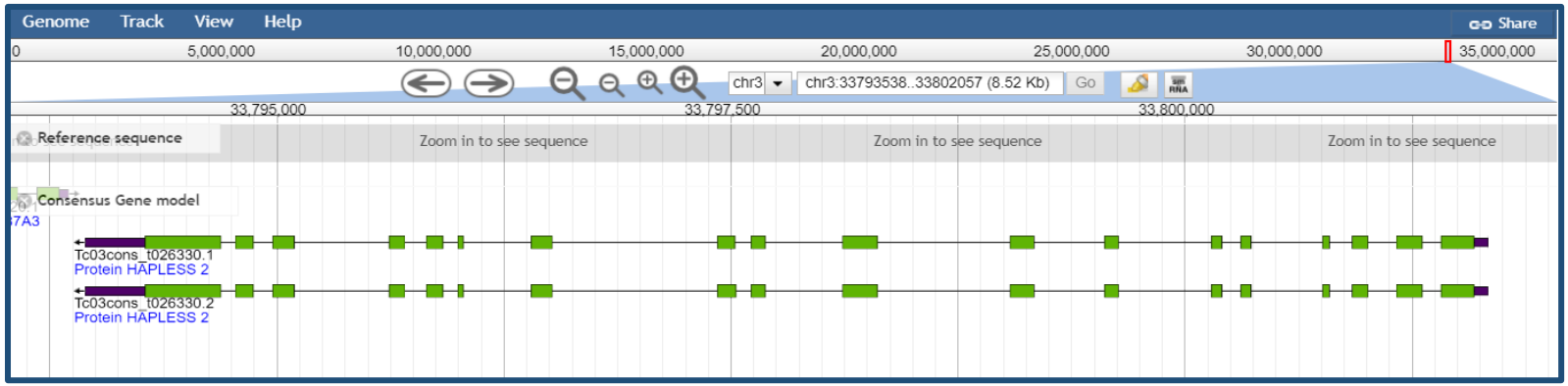


Figure 10. BLAST result for *Hapless 2* (*Hap2*)

4.3 Primer designing

4.3.1 Selection of region and trimming of sequence

The sizes of all the genes except *PMZ*, were larger ranging from 2486 bp to 7,478 bp, containing long intronic regions. The processivity of *Taq* polymerase is limited to about 1500 bp and hence, the primers were designed to amplify around 1000 bp. Smaller sized genes were split into two or three sequences, each in the range of 1000 bp. For larger genes, coding regions were selected using ORF finder. The "SMART BLAST" option in ORFfinder was used to confirm if the ORF selected was coding for intended protein or not. Using BioEdit the selected regions were trimmed.

4.3.2 Primer designing

Using Primer3 software, 17 primer sets of optimal lengths, GC content, melting temperature and least primer dimerization were designed (Table 5).

Table 5. Primer designing

Sl No.	Primer name	Forward primer	Reverse primer	Product size	T_m(°C)
1	***SLG*_1**	5'GCAAGCCCTTTCTTTCCCAT 3'	5' AGTAGCCCAAAAGCCTGACA 3'	967	58.73-F 59.23-R
2	***SLG*_2**	5'AGTCTGTCCTCGTGTTCACA 3'	5'ATCAGAGCTCACCATGGCAT 3'	995	58.60-F 59.16-R
3	***SLG*_3**	5'AGGTTTGACATCATGGAAAAGC 3'	5'CATCCATGTCCTTTGCCACT 3'	664	57.47-R 58.16-R
4	***SRK_1***	5'ACTGTGTGCTGCATGTTACC 3'	5' GCCTACCCTCTGTCACTTCC 3'	953	58.76-F 59.46-R
5	***SRK* _2**	5'TTTCAGGAGATGGTGGGAAC 3'	5'GCTTGCTTTGAGATCTCTATGGA3'	956	57.12-F 58.3-R
6	***SRK* _3**	5'TGCTTGATGGAGAAATGAACCC 3'	5'ACCAATACCCAAAACTGCCAC3'	883	58.91-F 59.03-R
7	***Bam1_1***	5'CATATCCAGCGGCAAAGTCC 3'	5'TTCTCCTCCCCATTTCCCAC 3'	1097	59.05-F 59.00-R
8	***Bam1_2***	5'GCAAGGCCAAAATCAGCAAC 3'	5'GGTCGAATCCCAGCTGAGAT 3'	1049	58.85-F 59.24-R
9	***Wdr5a*_1**	5'TTAAACCCTCCCCTAAGCCC 3'	5'GCTCCTCCATTTCCAGCCTA 3'	975	58.70-F 59.16-R
10	***Wdr5a*_2**	5'AGCATGTGTTTGAGAGTCGC 3'	5'GAGAAGAGGGTCGAGAGTGT 3'	1062	58.85-F 58.17-R

Table 5. Primer designing (continued)

Sl. No	Primer name	Forward primer	Reverse Primer	Product size	T_m (°C)
11	***Alpha 1F***	5'CGTCATCAGCATCAACTCCG3'	5'AAAGAGCCAAGAAGCCTCCT3'	1003	59.07-F 58.93-R
12	***Alpha 1F***	5'GGCCCAGAAATGGTGAACTG3'	5'CTCCAAACAGTCAGGACCCT3'	1123	59.11-F 58.94-R
13	***GEX1***	5'TCTGTGGCTGAATTCTCGGT3'	5'TCGTAGACCCTCAAGTGCTG3'	985	59.02-F 59.11-R
14	***GEX1***	5'CAATGTTGGCAGCTCAGGTT3'	5'TGGGTTTGTGCTTGTGTGTT3'	953	59.04-F 58.74-R
15	***PMZ***	5'CGATTCCAGGCAAGACTTGG3'	5'TTTCCTTCGTGCACATTCCC3'	956	58.91-F 58.76-R
16	***ARC1***	5'ACCCTGATTTCTCCACCGTT3'	5'GCCGAAGCCAAGTTACACAA3'	1012	58.94-F 59.05-R
17	***Hapless 2***	5'TCATCACTGGAGAAACTTGCA3'	5'ATTGGAGTTCTCAGGTGGCA3'	841	57.51-L 58.93-R

4.4 Molecular characterization

4.4.1 Isolation of DNA

Isolation of DNA from cocoa has always been a cumbersome task compared to other crops (Jose and Usha, 2000). Cocoa belongs to Malvaceae family that contain very high concentration of mucilage in the leaf tissue. Presence of mucilage in the sample interferes with DNA recovery during isolation. Further, mucilage in the isolated DNA interferes with primer annealing during thermal cycling. Apart from mucilage, there exists polyphenols and tannins, which were also reported to affect the outcome of plant molecular analyses (Bayer *et al*., 1999).

During the study, DNA was isolated from three different stages – reddish brown tender leaves, lemon yellow coloured leaves and mature leaves. Two different protocols were experimented, a modified version of Dellaporta *et al.* (1983) protocol and a modified version of Doyle and Doyle (1987). Lemon yellow leaves with its mid rib and thick veins removed was found to yield the highest quantity and the best quality of DNA. In leaves at this stage, the mucilage content was found to be very less. Doyle and Doyle (1987) protocol proved to be more efficient than the Dellaporta protocol. Care was taken during the pipetting out of the aqueous layer to avoid the intake of mucilage.

Cetyl Trimethyl Ammonium Bromide (CTAB) was used as the cell lysis buffer. It has been established to be the best detergent used for isolation of plant genomic DNA. Phenolic compounds when oxidized into quinones, binds with the DNA through hydrogen bonding and makes the compound heavier and unable to be extracted during isolation. In order to prevent the oxidation of polyphenols into polyquinones, the samples were kept in ice and transferred into -20 ^{o}C as soon as possible and DNA was isolated as early as possible. Grinding with the help of liquid nitrogen also contributed to the prevention of oxidation of polyphenols. In case of coca, a slightly higher concentration of PVP helped in reduction of mucilage and phenolics to a greater extend (Gallego and Martinez, 1996). Too much concentration of the same was also found to reduce the quantity of DNA isolated drastically. The β–mercaptoethanol acts as a reducing agent which helped in reducing proteins and phenolic compounds.

Two chloroform isoamyl alcohol (24:1) washes were found sufficient to get rid of mucilage, protein, pigments and other impurities in the sample. Ice cold propanol was used to precipitate the nucleic acids in the sample. Ethanol wash was given to get rid of the excess salts and other contaminants in the sample. The protocol yielded average sized yellowish-brown stained pellets. The colour was predominantly due to the minor concentrations of polyphenolic compounds present in the pellet. The pellet dissolved readily in distilled water, which indicated good quality DNA.

4.4.2 Quantity and quality assessment of DNA

Quality and quantity of DNA used during molecular analysis plays a vital role in its outcome. For the assessment of quality and quantity, agarose gel electrophoresis and UV spectrophotometry methods, respectively were used.

The isolated DNA were run on 0.8 per cent agarose gel with Ethidium Bromide (EtBr). EtBr acts as an intercalater and binds in between base pairs. When observed under UV radiation, DNA intercalated with EtBr can be visualized as a yellowish orange band. Wettasinghe and Peffley (1998) reported that if sample had an intact band of high molecular weight, with very minimum smearing and minimum contamination by RNA, it is of good quality. The samples isolated in the study showed intact high molecular weight bands with very less smearing. Minor RNA contamination was cleared by RNase treatment. The gel image of the isolated DNA from the self-compatible and self-incompatible genotypes are shown in Plate 3.

Using UV spectrophotometer (NanoDrop ND-1000®) the quantity of the isolated DNA was measured. For quantification, optical density values at 260 nm and 280 nm were taken. For samples of good quality DNA, the A_{260}/A_{280} ratio will be within the range of 1.8-2.0 (Meena, 2014; Thakur e*t al*., 2014). The absorbance ratios of the present samples fell within this range. (Table 6).

Lanaud *et al.*, (1995) had reported that 40 ngμL^{-1} concentration of cocoa DNA is optimum for PCR reactions. Hence, the samples were diluted to this concentration. After dilution, a RAPD marker was run and good amplification was obtained. Hence, it was concluded that the samples are fit for further molecular analyses.

Table 6. Assessment of DNA quality and quantity using UV spectrophotometer

Genotype	Absorbance at 260nm(A_{260})	Absorbance at 280nm(A_{280})	Optical Density ($A_{260/280}$)	Quantity of DNA (ng/µl)
G VI 167 X G IV 18.5	11.69	5.90	1.98	748.30
IMC 20	9.60	4.82	1.99	725.10

4.4.3 Standardization of amplification conditions for the designed primers

Since the primers were self-designed, the temperatures for the thermal cycling had to be optimized and standardized. Apart from the temperature, the optimal concentration of the primers used in the reaction had to be standardized. When primers recede the optimum concentration, the amplification is minimized, whereas when the concentration exceeds the optimum, results in primer dimer formation. Primer dimers are a potential threat to the quality of sequence obtained via Sanger sequencing (Brownie *et al.,* 1997; Poritz and Ririe, 2014). Hence, the least concentration at which the amplification is maximum and there is no primer dimerization was selected as the optimum primer concentration.

The annealing temperature is a pivotal factor in a PCR reaction. The annealing temperature is theoretically 5^{o}C lesser than the melting temperature of the primer. However, the optimum experimental value varies quite often by a small margin from the theoretical value and hence it was necessary to standardize annealing temperature (Rychlik *et al.,* 1990). Gradient PCR, a variant of normal PCR, was used to standardize the annealing temperature of each primer combination. For each primer, gradient PCR with a temperature range of ±5^{o}C of the average of theoretical annealing temperature of the forward and reverse primers was run. Upon completion, the products were electrophoresed in a 1.4 per cent agarose gel (Liu *et al.,* 2013). The temperature at which

maximum amplification and least dimerization was selected as the optimum temperature.

4.4.4 Amplification, sequencing and sequence analysis of the genes

4.4.4.1 Amplification and sequencing

Eight genes were amplified using their respective primer combinations at their optimum annealing temperature and using optimum primer combination. Even after a wide range of temperature and various concentration of primers were tried to amplify, no amplification was obtained in *SLG* and *Hapless.* The optimized PCR conditions for various primer combinations are shown in Table 7.

Table 7. Optimized PCR conditions for various primer combinations.

Sl. No.	**Gene**	**Annealing temperature** (^{o}C)	**Optimum primer concentration** (μM)	**Product size** (bp)
1	*SRK*	61.4	0.5	953
2	*SLG*	No amplification	-	967
3	*BAM 1*	61.9	0.5	1097
4	*BAM 2*	59.8	0.4	1049
5	*WDR5a*	60.2	0.5	975
6	*Alpha –IF*	62.2	0.4	1003
7	*GEX 1*	57.0	0.4	985
8	*PMZ*	59.5	0.4	956
9	*ARC1*	57.1	0.4	1012
10	*HAPLESS*	No amplification	-	851

4.4.4.2 Analysis of sequencing results

4.4.4.2.1 Analysis of the sequences

The quality and length of the sequence obtained after sequencing reaction was thoroughly analyzed. In a Sanger sequencing reaction, small number of base pairs (around 50-60 base pairs) may be compromised at the 5’ and 3’ ends. This is due to the

noise that arise as a result of signal pick up issues in the beginning and the end of the sequencing reaction. The trimmed length of the forward sequence and reverse sequence were observed. The Quality Value (QV), an established measure for determining quality of sequencing data, was analyzed for each sequence, for reliability. QV score greater than 20 indicated that the probability of a base being miscalled is not greater than one per cent. This is the acceptable standard for a good sequencing reaction. Except one read, the reverse run of *PMZ,* all the other reads had a QV score greater than 20. In case of *PMZ,* since the sequence generated in the forward run was of good quality, the forward sequence alone was used for analysis.

4.4.4.2.2 Contig assembly

Once the sequences were analyzed, the forward and reverse sequences were assembled into a single contiguous sequence (contig) using CAP3. Two genes *SLG* and *Hapless 2* sequencing could not be carried out, since no amplification was obtained for both the genes.

4.4.4.2.3 Computational confirmation of the contigs

This was done by BLASTn against the cocoa whole genome. This confirmed that the PCR products sequenced were actual target genes.

4.4.4.2.4 Sequence variation analysis

From the analysis of multiple sequence alignment, a total of 35 variants were discovered, out of which 33 were SNPs and two were sequencing errors.

4.4.4.3 Identification of SNPs and InDels

The SNPs and InDels were identified by sequence alignment. To confirm the variant as an SNP and to rule out any sequencing error, the chromatogram in AB1 file was crosschecked. When the AB1 file was examined at the SNP loci, most of SNP loci were found hetrozygous. The reference genome used in the study is a self-compatible criollo type - Belizean (B97-61/B2) (Argout *et al.*, 2010; Lachenaud and Motamayor, 2017) which is homozygous in nature. The amplification profile obtained, length and quality of the sequence, contig assembly, variation analysis and SNP characterization for each gene are discussed in the following subsection.

Table 8. Sequence analysis of candidate genes

Gene	Genotype	Sequence	Length	Q.V Score	Contig
SRK	Self- compatible	Forward	797	54	850
		Reverse	850	51	
	Self-incompatible	Forward	883	53	883
		Reverse	837	53	
Bam1	Self- compatible	Forward	864	54	929
		Reverse	766	51	
	Self-incompatible	Forward	720	53	720
		Reverse	871	53	
Bam2	Self- compatible	Forward	874	55	1000
		Reverse	871	53	
	Self-incompatible	Forward	864	54	996
		Reverse	847	52	
WDR5a	Self- compatible	Forward	942	52	948
		Reverse	864	51	
	Self-incompatible	Forward	871	54	871
		Reverse	732	52	
Alpha-1F	Self- compatible	Forward	872	53	954
		Reverse	885	53	
	Self-incompatible	Forward	871	54	962
		Reverse	837	51	
PMZ	Self- compatible	Forward	720	53	0
		Reverse	0	0	
	Self-incompatible	Forward	874	55	0
		Reverse	32	53	
Gex1	Self- compatible	Forward	864	54	963
		Reverse	847	52	
	Self-incompatible	Forward	942	52	982
		Reverse	797	54	
ARC1	Self- compatible	Forward	953	54	964
		Reverse	883	51	
	Self-incompatible	Forward	864	53	912
		Reverse	871	53	

4.4.4.3.1 Serine Receptor Kinase *(SRK)*

4.4.4.3.1.1 Amplification of the primer

The primer set used to amplify targeted the first 953 bp of the gene. It covered the 5' UTR region and the first exon of the gene. PCR was run with the annealing temperature of 61.4^{o}C, which was determined by the gradient PCR experiment. The optimum concentration of forward and reverse primers of reaction mixture was found to be 0.5 μM. Upon agarose gel electrophoresis, successful amplification was observed in both the genotypes and a single band in the size of the expected product size of 953 base pairs was obtained, with minimum primer dimers (Table 7). The gel image is portrayed in Plate 4.

4.4.4.3.1.2 Analysis of the sequences

In self-compatible genotype, trimmed sequence of the forward run was of 797 base pairs and had a QV score of 54. The length of the reverse sequence was of 850 base pairs and the QV score was 51. Both the sequences were of good quality (Table 8). In self-incompatible genotype, trimmed sequence of the forward run was of 883 base pairs and had a QV score of 53. The length of the reverse sequence was of 837 base pairs and the QV score was 53 (Table 8).

4.4.4.3.1.3 Contig assembly

The contig obtained after assembling the forward and reverse sequence of the *SRK* primer combination was 850 bp long for the self-compatible genotype and 883 bp long for the self-incompatible genotype (Table 8).

4.4.4.3.1.4 Computational confirmation

BLASTn analysis of the contigs against the cocoa genome database has confirmed that the sequence obtained was exactly of the targeted region (Figure 11).

4.4.4.3.1.5 Multiple sequence alignment

Sequences of self-compatible and self-incompatible genotypes were aligned with the reference sequence (compatible). Four variants were found when the sequences were analyzed. Alignment of remaining three sequences is presented in Figure 12.

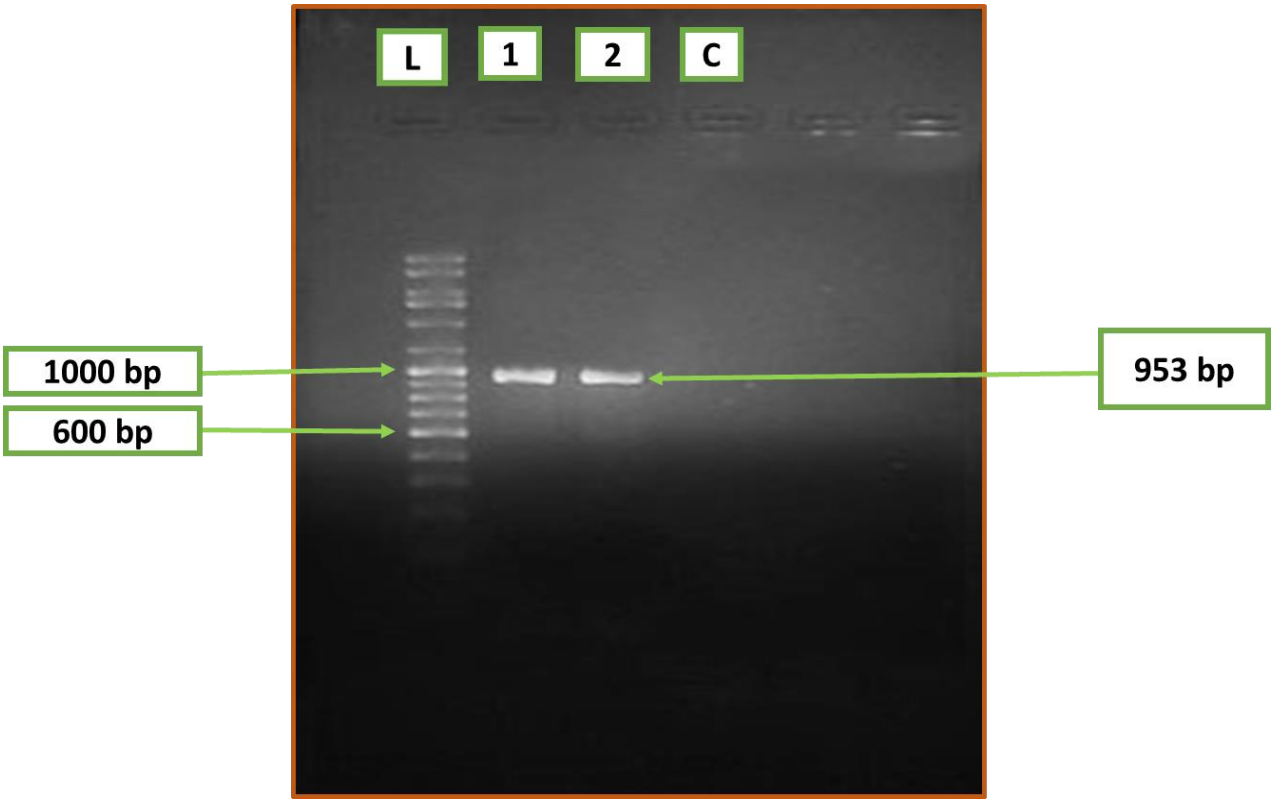


Plate 4. Amplification of *SRK* gene in self-compatible and self-incompatible genotypes

1 Amplification in compatible genotype

2 Amplification in incompatible genotype

C Control/ Blank

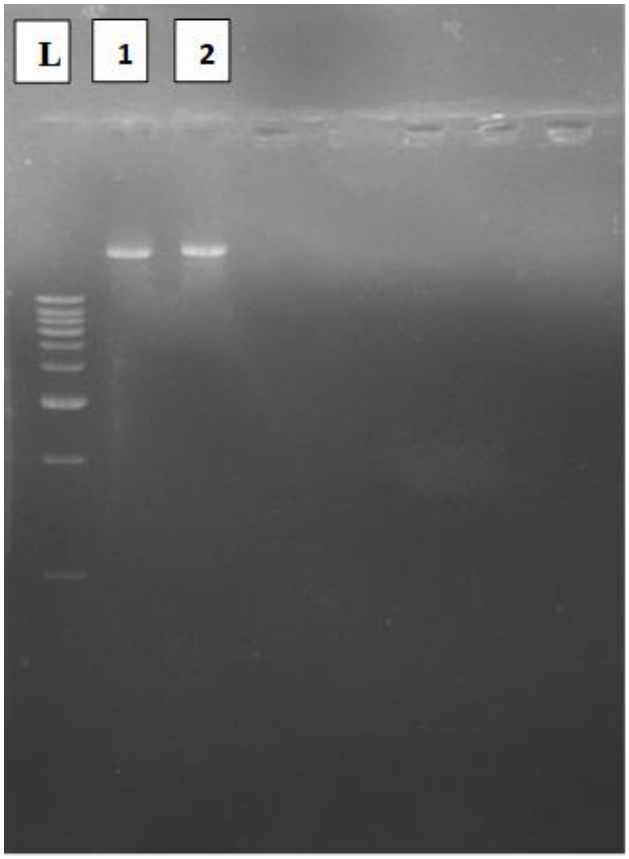


Plate 3. Gel profile of the DNA isolated from genotypes

1- Self-compatible genotype - G VI 167 X G IV 18.5

2- Self-incompatible genotype - IMC 20

L- 3 kb Ladder

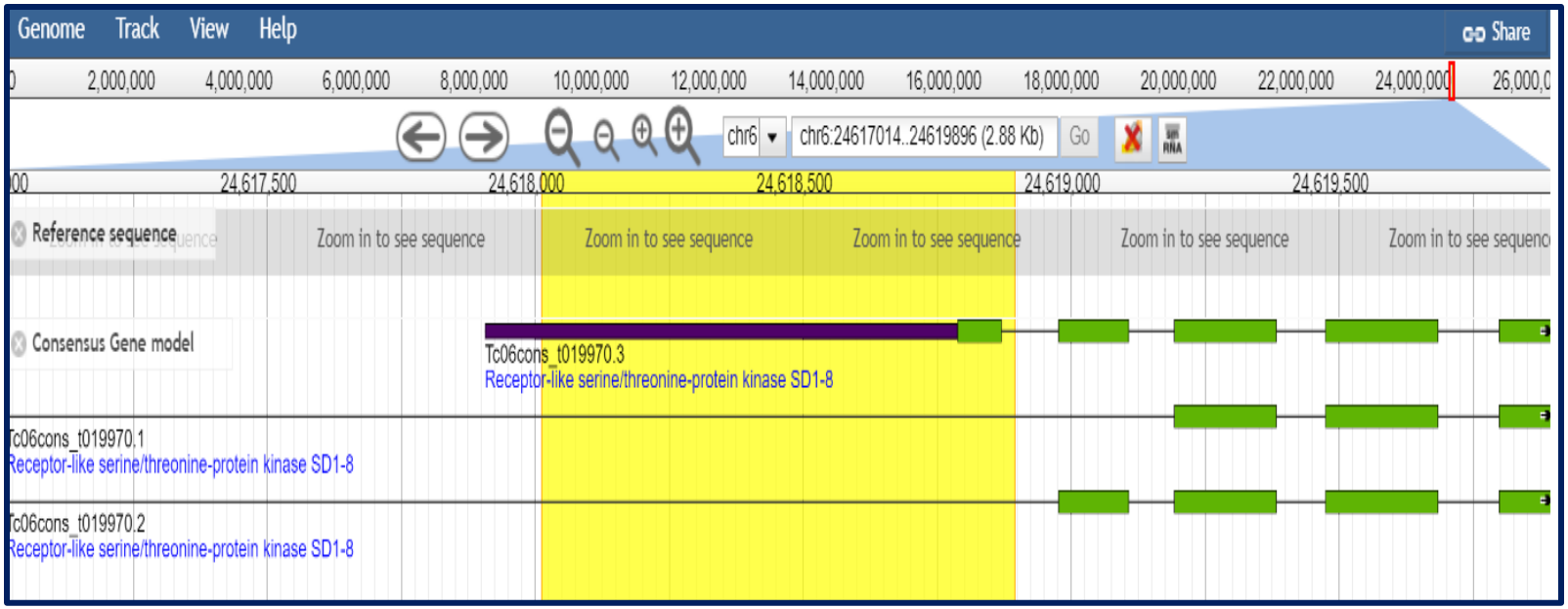


Figure 11. Computational confirmation of *SRK* region sequenced by BLAST search.

Figure 12 (a). Multiple sequence alignment of *SRK* gene

Figure 12 (b). Multiple sequence alignment of *SRK* gene

4.4.4.3.1.6 Characterization of SNPs

SNP 1

Self-compatible SNP1 locus was heterozygous. Though thymine was the base called at the locus, there was another significant cytosine peak too (T/C). In self-incompatible genotype, there was a single cytosine peak. The chromatographic peaks present in both the sequences at the self-compatible and self-incompatible loci are shown in Figure 14 (a) and (b), respectively. Since the reference genome also possessed cytosine at the locus, which is identical to the base call at the self-incompatible genotype, the SNP is insignificant and cannot be linked to self-compatibility trait.

SNP 2

The chromatogram of the SNP 2 locus in self-compatible genotype had a homozygous guanine (G) and self-incompatible genotype was heterozygous. The base that was called at the SNP locus by the system was adenine, guanine peak present in self-compatible genotype was also present in self–incompatible genotype. The chromatographic peaks present in self-compatible and self-incompatible genotypes are shown in Figure 15 (a) and 15 (b), respectively. The reference sequence possessed a guanine (G) in the same locus.

Guanine was found to be common in both the compatible sequences (genotype used in this study and the reference genotype) as well as one of the two alleles of self-incompatible sequence. However, the allele carrying the adenine base was present only in self-incompatible genotype. Hence this SNP may be linked to self-incompatibility.

SNP 3

The SNP 3 was identical to the SNP 2. At this locus, the self-compatible genotype possessed a homozygous guanine (G) whereas self-incompatible genotype was heterozygous. The base that was called at the SNP locus by the system was adenine, however chromatogram had shown that the guanine peak present in self-compatible is also present in self-incompatible. Hence the self-incompatible genotype was heterozygous for adenine and guanine (A/G). The reference sequence possessed a guanine (G) in the same locus. The chromatographic peak present in self-compatible and self-incompatible loci are

shown in Figure 16 (a) and 16 (b), respectively. Guanine was common in both the compatible (genotype used in this study and the reference genotype) as well as one of the two alleles of self-incompatible sequence. Allele carrying the adenine base was present only in the self-incompatible genotypes. Hence, similar to SNP 2, the adenine present in the incompatible genotype sequence could be linked to self-incompatibility trait.

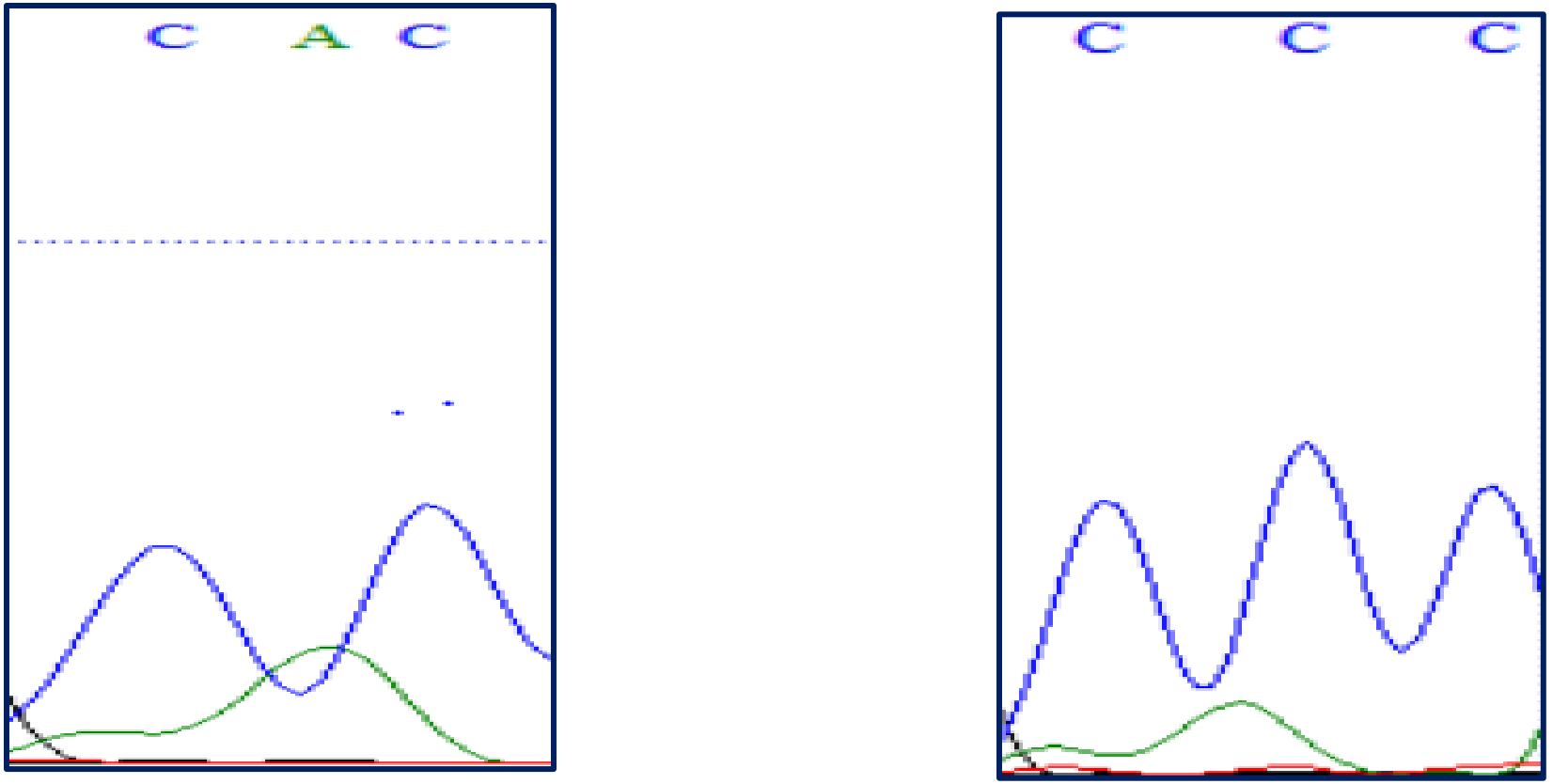
Kakita *et al.* (2007) reported that in sporophytic self-incompatibility system present in Brassicacea, *SRK* acts as the female determinant. *SRK* interacts with the male determinant-*SCR/SP11* and also *SLG*. The interaction sets off a cascade of signaling interactions which ultimately lead to the rejection of pollen. In this study, two SNPs (SNP 2 and SNP 3) were discovered in *SRK* that could be possibly linked to self-incompatibility. The details of various mismatches observed are summarized in Table 9.

Table 9. List of SNPs and base called at their respective loci (*SRK*)

SI. No.	**Self-compatible**	**Self-incompatible**	**Reference**
Error	-	-	-
SNP – 1	C/T	C	C
SNP – 2	G	A/G	G
SNP – 3	G	A/G	G

4.4.4.3.2 *S- locus Glycoprotein*

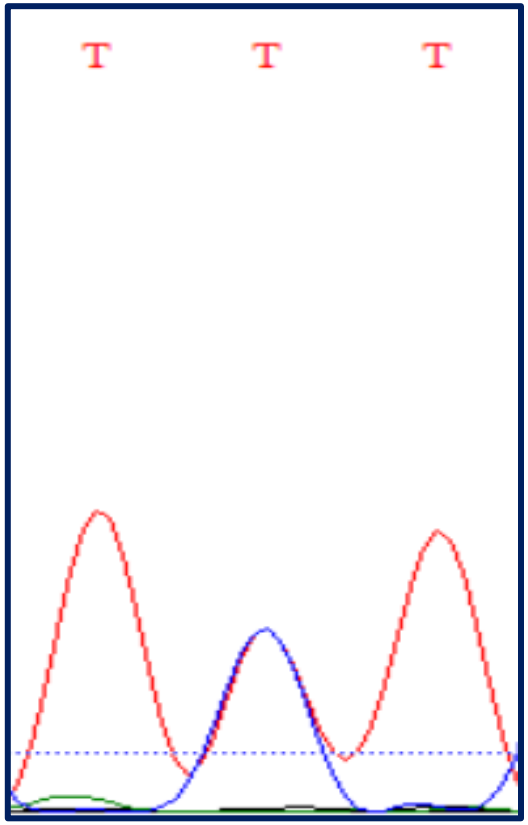
No amplification was obtained using the *SLG* primer. Wide range of temperatures and different concentration of the PCR components were tried. However, no successful amplification was obtained.



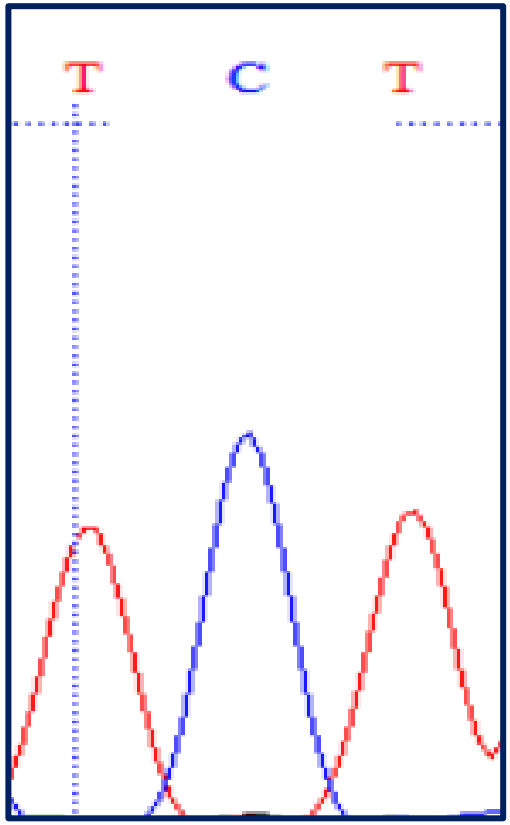
a. Self-compatible

b. Self-incompatible

Figure 13. Base call at Error 1 (*SRK*)

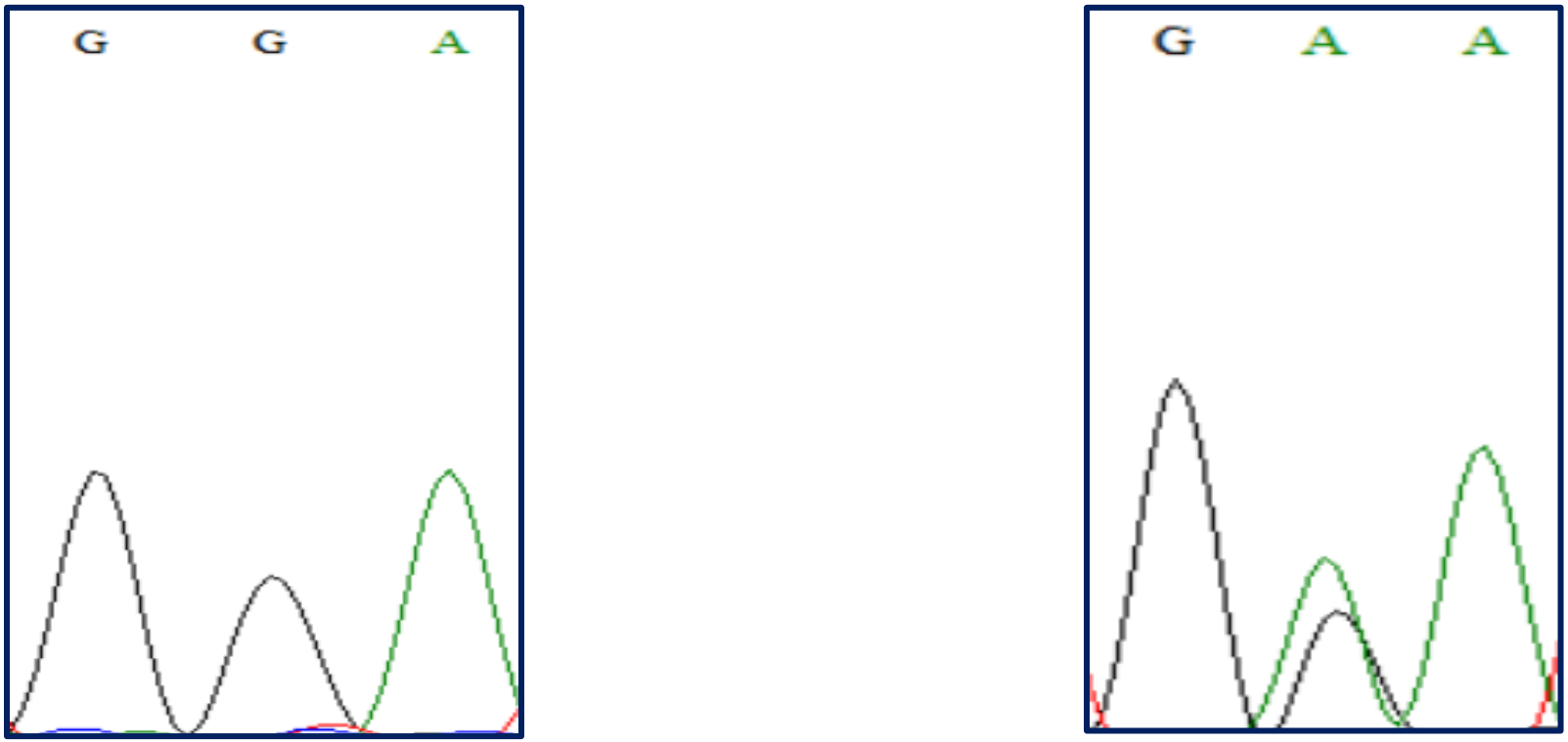


a. self-compatible



b. Self-incompatible

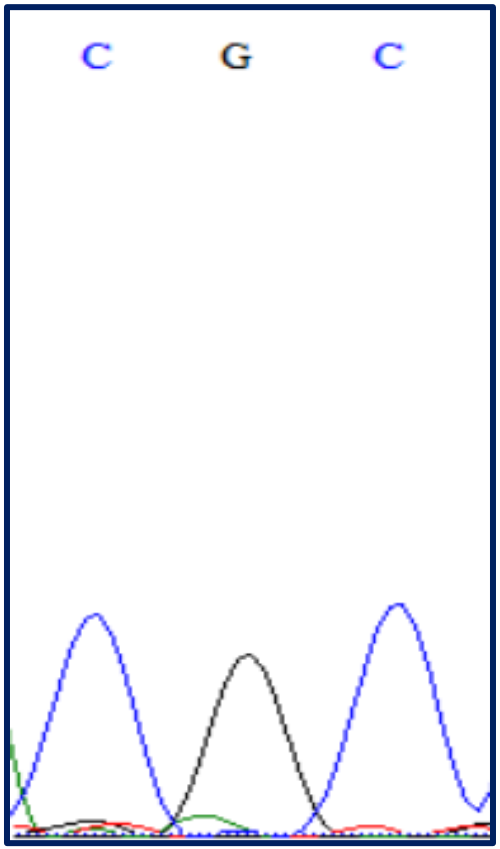
Figure 14. Base call at SNP 1 (*SRK*)



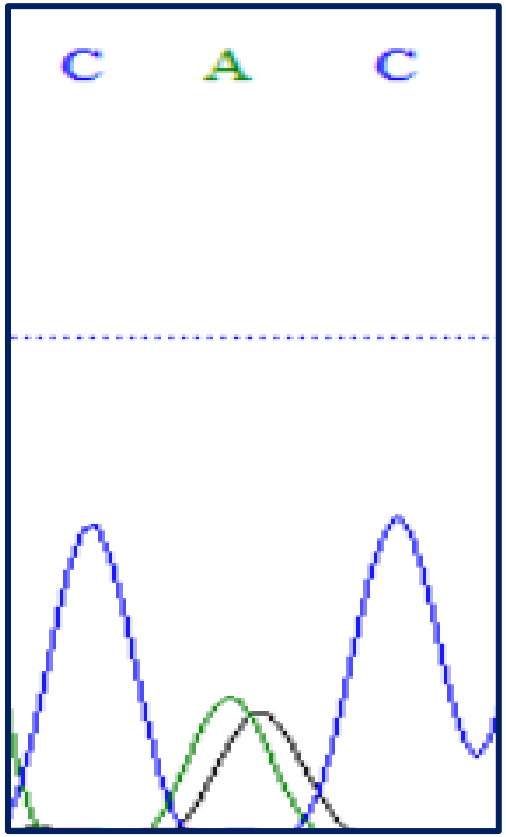
a. Self-compatible

b. Self-incompatible

Figure 15. Base call at SNP 2 (*SRK*)



a. Self-compatible



b. Self-incompatible

Figure 16. Base call at SNP 3 (*SRK*)

4.4.4.3.3 *Barely any meristem 1 (BAM 1)*

4.4.4.3.3.1 Amplification of the primer

The primer set (*BAM1*) used to amplify the gene has targeted 1097 bp on exon 1, the largest exon in the gene. PCR was run with an annealing temperature of 61.9^{o}C, which was determined by the gradient PCR experiment. The optimum concentration of forward and reverse primers of reaction mixture was found to be 0.5 μM. Upon agarose gel electrophoresis, single band at 1097 bp was obtained (Table 7). The gel image is portrayed in Plate 5. The PCR product was sequenced.

4.4.4.3.3.2 Analysis of the sequences

In self-compatible genotype, trimmed sequence of the forward run was of 864 bp and had a QV score of 54. The length of the reverse sequence was of 766 bp and the QV score was 51 (Table 8). In self-incompatible genotype, trimmed sequence of the forward run was of 720 bp and had a QV score of 53. The length of the reverse sequence was of 871 base pairs and the QV score was 53 (Table 8).

4.4.4.3.3.3 Contig assembly

The contig obtained after assembling the forward and reverse sequence of the BAM1 primer was 929 bp long for the self-compatible genotype and 720 bp long for the self-incompatible genotype (Table 8).

4.4.4.3.3.4 Computational confirmation

BLASTn search using the contig against the cocoa genome database has confirmed that the sequence is the targeted region, the first part of exon 1 (Figure 17).

4.4.4.3.3.5 Multiple sequence alignment

Sequences of self-compatible and self-incompatible genotypes were aligned with the reference sequence (compatible). There was no variation between the sequences of self-compatible and the self-incompatible genotypes (Figure. 18).

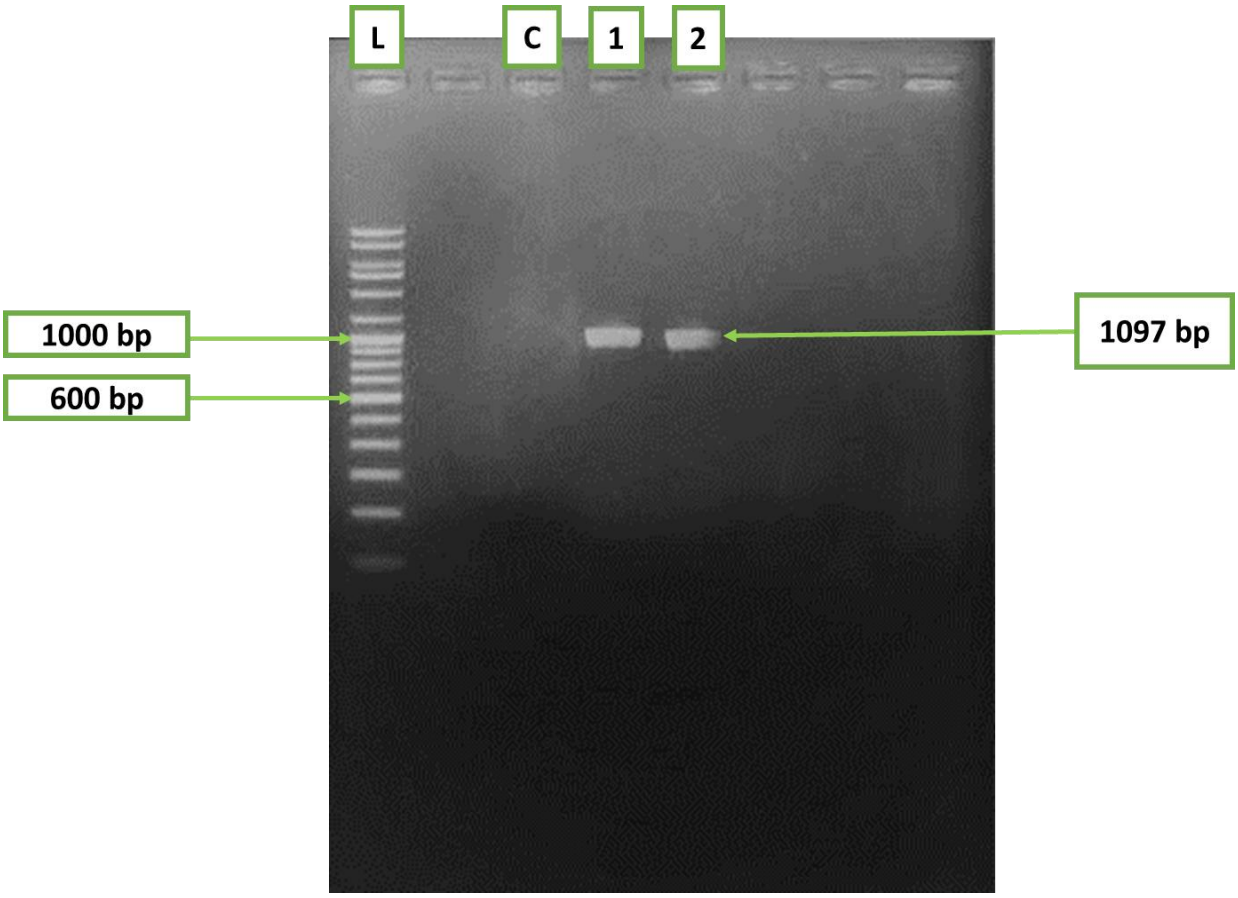


Plate 5. Amplification of *BAM1* gene

1 - Amplification in self-compatible genotype

2 - Amplification in self-incompatible genotype

C - Control/ Blank

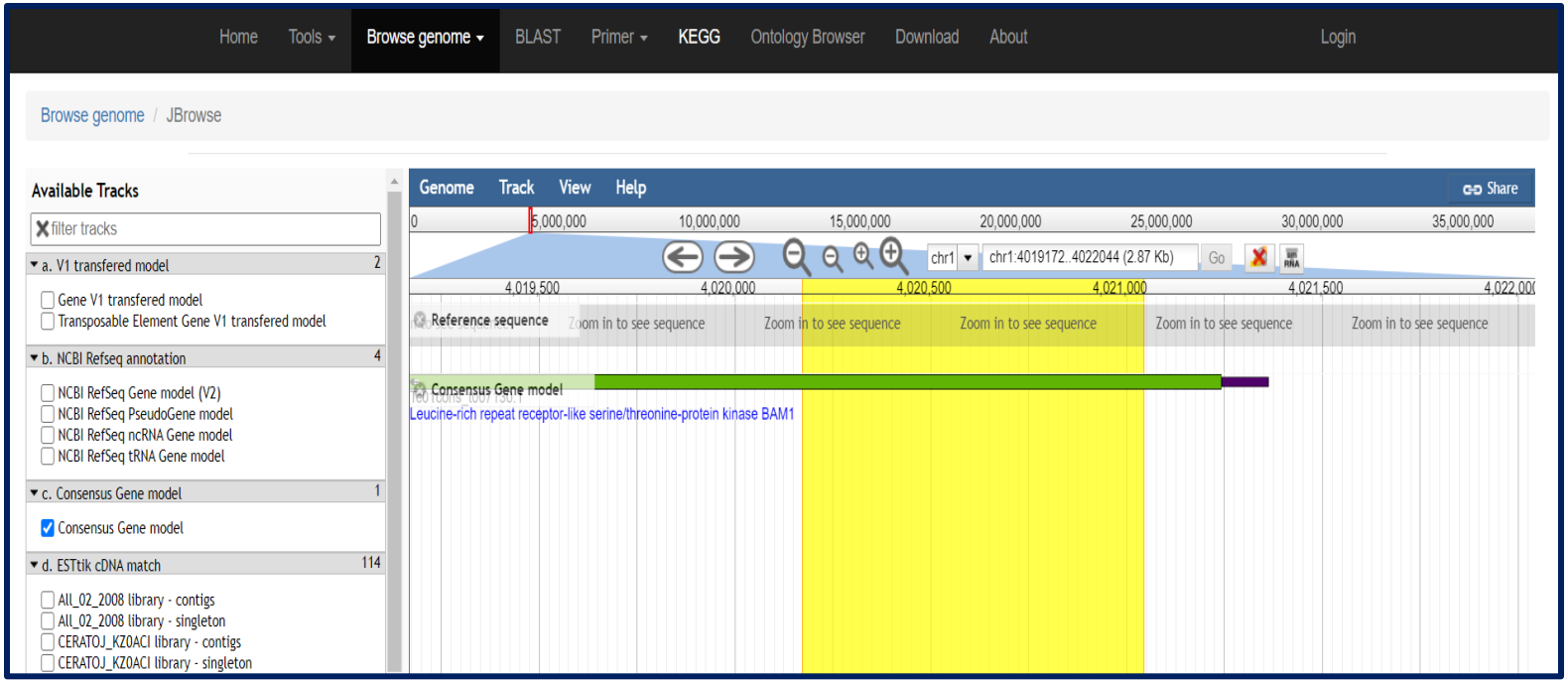
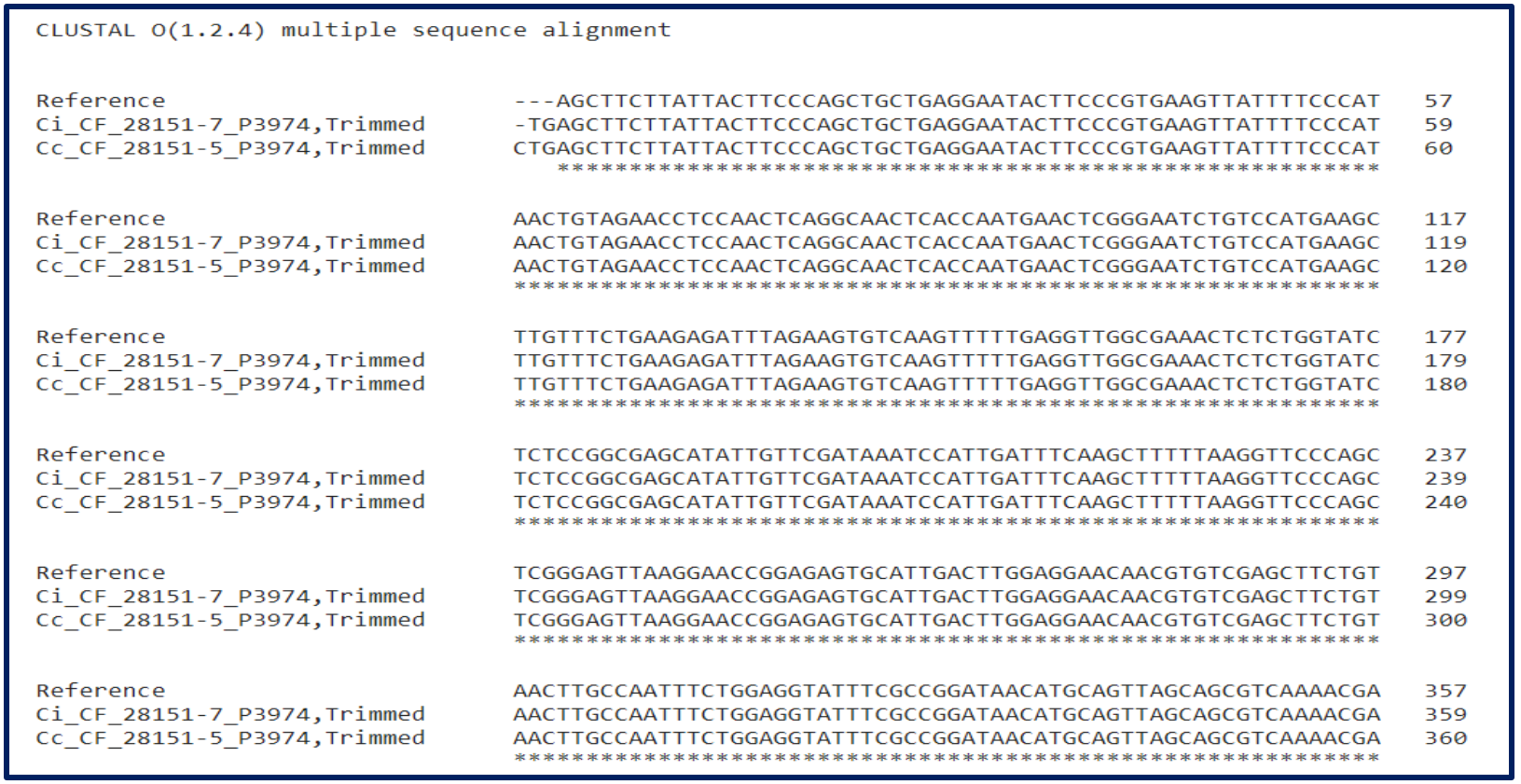


Figure 17. Computation confirmation of the *BAM1* region sequenced by BLAST search.



```
CLUSTAL O(1.2.4) multiple sequence alignment


Reference                      ---AGCTTCTTATTACTTCCCAGCTGCTGAGGAATACTTCCCGTGAAGTTATTTTCCCAT   57
Ci_CF_28151-7_P3974,Trimmed    -TGAGCTTCTTATTACTTCCCAGCTGCTGAGGAATACTTCCCGTGAAGTTATTTTCCCAT   59
Cc_CF_28151-5_P3974,Trimmed    CTGAGCTTCTTATTACTTCCCAGCTGCTGAGGAATACTTCCCGTGAAGTTATTTTCCCAT   60
                                  ***********************************************************

Reference                      AACTGTAGAACCTCCAACTCAGGCAACTCACCAATGAACTCGGGAATCTGTCCATGAAGC   117
Ci_CF_28151-7_P3974,Trimmed    AACTGTAGAACCTCCAACTCAGGCAACTCACCAATGAACTCGGGAATCTGTCCATGAAGC   119
Cc_CF_28151-5_P3974,Trimmed    AACTGTAGAACCTCCAACTCAGGCAACTCACCAATGAACTCGGGAATCTGTCCATGAAGC   120
                               ***********************************************************

Reference                      TTGTTTCTGAAGAGATTTAGAAGTGTCAAGTTTTTGAGGTTGGCGAAACTCTCTGGTATC   177
Ci_CF_28151-7_P3974,Trimmed    TTGTTTCTGAAGAGATTTAGAAGTGTCAAGTTTTTGAGGTTGGCGAAACTCTCTGGTATC   179
Cc_CF_28151-5_P3974,Trimmed    TTGTTTCTGAAGAGATTTAGAAGTGTCAAGTTTTTGAGGTTGGCGAAACTCTCTGGTATC   180
                               ***********************************************************

Reference                      TCTCCGGCGAGCATATTGTTCGATAAATCCATTGATTTCAAGCTTTTTAAGGTTCCCAGC   237
Ci_CF_28151-7_P3974,Trimmed    TCTCCGGCGAGCATATTGTTCGATAAATCCATTGATTTCAAGCTTTTTAAGGTTCCCAGC   239
Cc_CF_28151-5_P3974,Trimmed    TCTCCGGCGAGCATATTGTTCGATAAATCCATTGATTTCAAGCTTTTTAAGGTTCCCAGC   240
                               ***********************************************************

Reference                      TCGGGAGTTAAGGAACCGGAGAGTGCATTGACTTGGAGGAACAACGTGTCGAGCTTCTGT   297
Ci_CF_28151-7_P3974,Trimmed    TCGGGAGTTAAGGAACCGGAGAGTGCATTGACTTGGAGGAACAACGTGTCGAGCTTCTGT   299
Cc_CF_28151-5_P3974,Trimmed    TCGGGAGTTAAGGAACCGGAGAGTGCATTGACTTGGAGGAACAACGTGTCGAGCTTCTGT   300
                               ***********************************************************

Reference                      AACTTGCCAATTTCTGGAGGTATTTCGCCGGATAACATGCAGTTAGCAGCGTCAAAACGA   357
Ci_CF_28151-7_P3974,Trimmed    AACTTGCCAATTTCTGGAGGTATTTCGCCGGATAACATGCAGTTAGCAGCGTCAAAACGA   359
Cc_CF_28151-5_P3974,Trimmed    AACTTGCCAATTTCTGGAGGTATTTCGCCGGATAACATGCAGTTAGCAGCGTCAAAACGA   360
                               ***********************************************************
```

Figure 18 (a). Multiple sequence alignment of *BAM1* gene

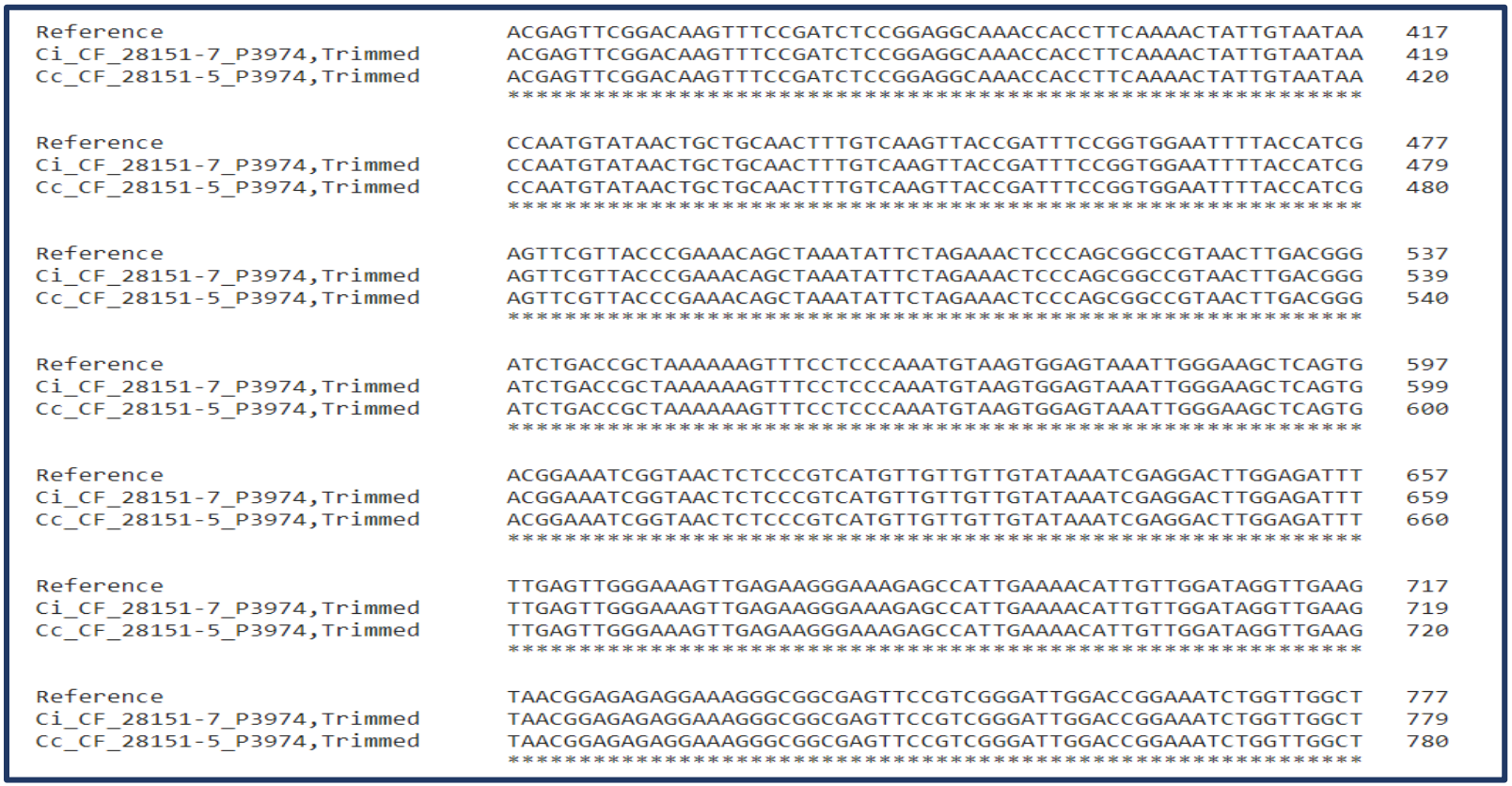


Figure 18 (b). Multiple sequence alignment of *BAM1*gene

4.4.4.3.4 *Barely any meristem 2 (BAM 2)*

4.4.4.3.4.1 Amplification of the gene

The primer set (*BAM2*) was designed to amplify 1049 bp on exon 1, the largest exon in the gene. PCR was run with an annealing temperature of 59.8 ^{o}C, as determined by the gradient PCR experiment. The optimum concentration of forward and reverse primers of reaction mixture was 0.5 μM (Table 7). PCR resulted in a single band of the expected size, with minimum primer dimers. The gel image is portrayed in Plate 6.

4.4.4.3.4.2 Analysis of the sequences

In self-compatible genotype, trimmed sequence of the forward run was of 874 bp and had a QV score of 55. The length of the reverse sequence was of 871 bp and the QV score was 53 (Table 8). In self-incompatible genotype, trimmed sequence of the forward run was of 864 bp and had a QV score of 54. The length of the reverse sequence was of 847 bp and the QV score was 52 (Table 8).

4.4.4.3.4.3 Contig assembly

The contig obtained after assembling the forward and reverse sequence of the *BAM2* primer was 1000 bp long for the self-compatible genotype and 996 bp long for the self-incompatible genotype (Table 8).

4.4.4.3.4.4 Computational confirmation

BLASTn search using the contig against the cocoa genome database has confirmed that the sequence is of the targeted region, second part of exon 1 (Figure 19).

4.4.4.3.4.5 Multiple sequence alignment

Using Clustal omega, multiple sequence alignment was done for the sequences of self-compatible genotype, self-incompatible genotype and the reference sequence (compatible). After alignment a total of four variants were found to be present between the self-compatible and self-incompatible genotypes (Figure 20).

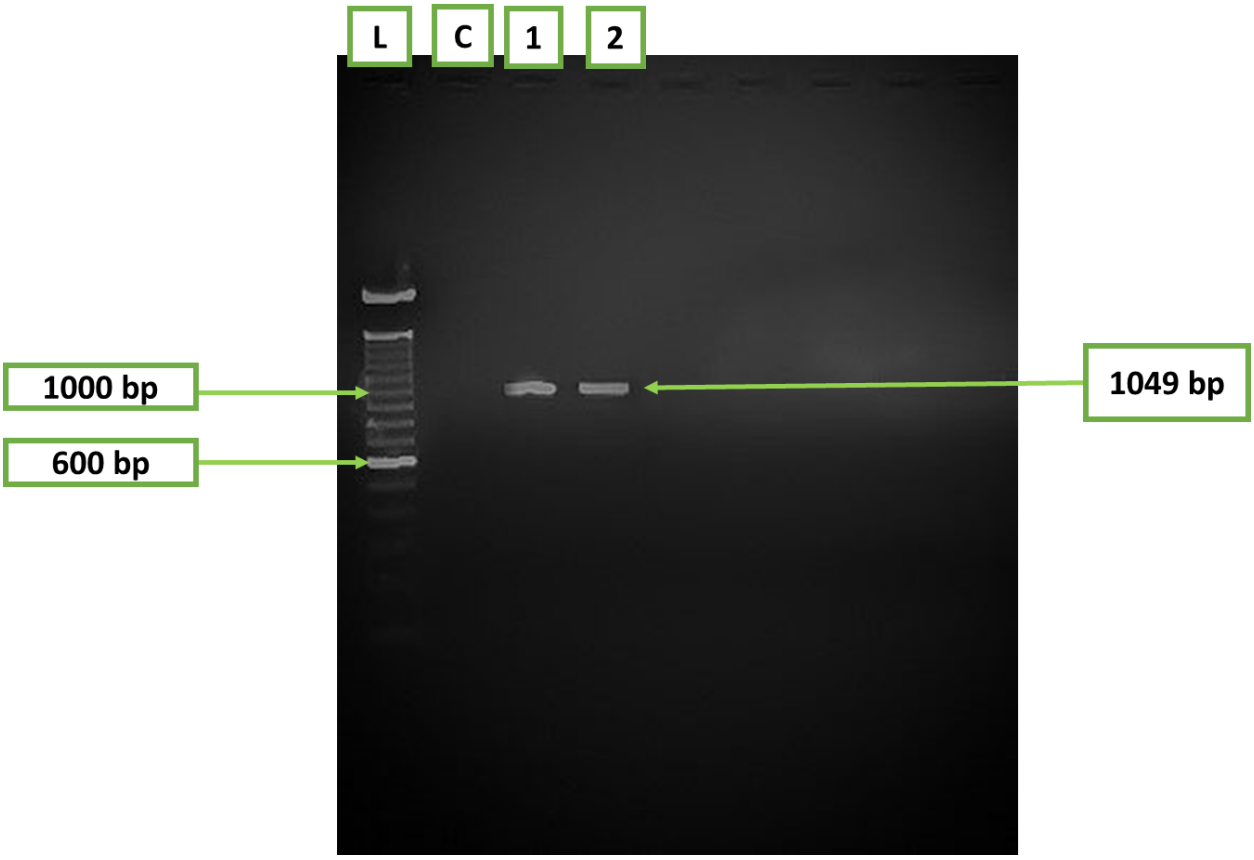


Plate 6. Amplification of *BAM2* gene

1- Amplification in self-compatible genotype

2- Amplification in self-incompatible genotype

C- Control/ Blank

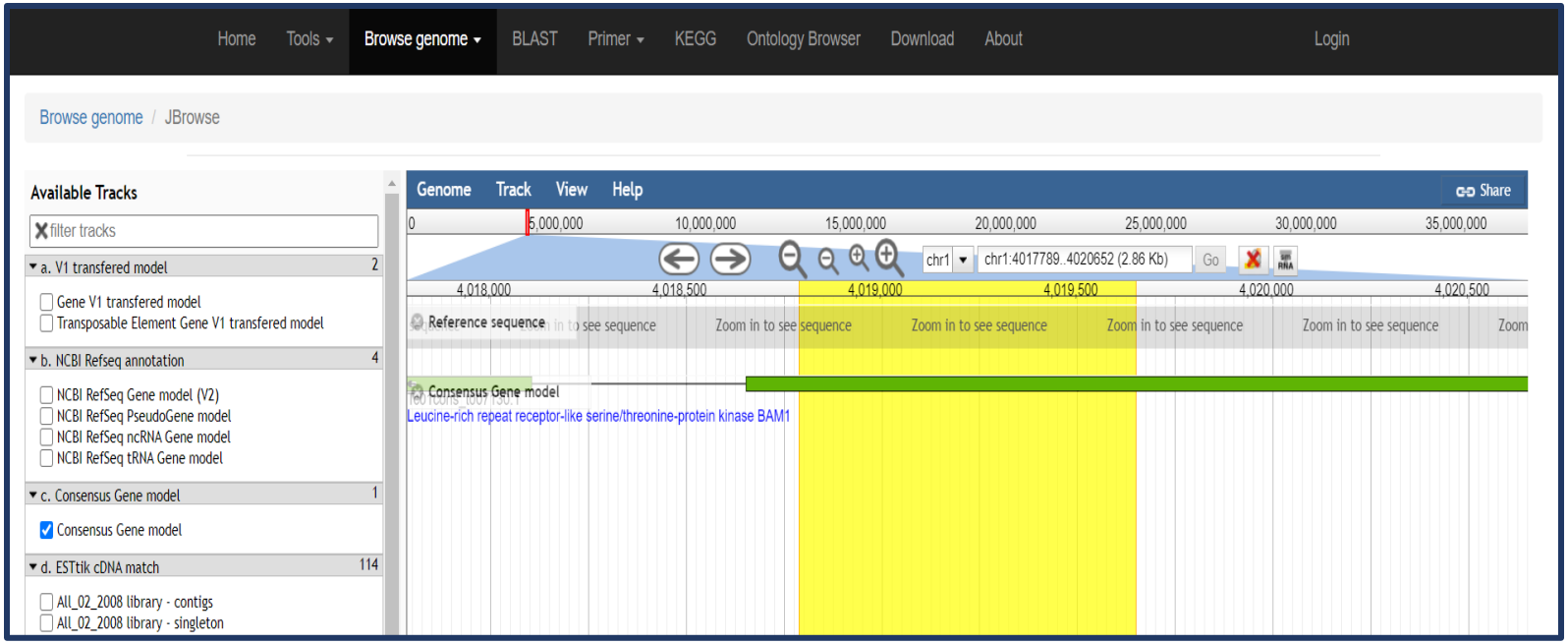


Figure 19. Computational confirmation of *BAM2* region sequenced by BLAST search.

Figure 20 (a). Multiple sequence alignment of *BAM2*

Figure 20 (b). Multiple sequence alignment of *BAM2*

Figure 20 (c). Multiple sequence alignment of *BAM2*

4.4.4.3.4.6 Characterization of SNPs

After alignment a total of four variants were found to be present between the self-compatible and self-incompatible genotypes. Each of the mismatches have been looked into deeply and characterized.

SNP 1

SNP1 locus was heterozygous in both self-compatible genotype and self-incompatible genotype. The loci were heterozygous for adenine and guanine (A/G). The chromatographic peak present in both the sequence at the self-compatible and self-incompatible locus is shown in Figure 21 (a) and 21 (b) respectively. Hence SNP1 loci cannot be considered to be linked to self-incompatibility trait because both the alleles are found to be present in both the genotypes, hence considered insignificant.

SNP 2

SNP 2 locus was heterozygous in both self-compatible genotype and self-incompatible genotype. The loci were heterozygous for adenine and thymine (A/T). The chromatographic peak present in both the sequence at the self-compatible and self-incompatible locus is shown in Figure 22 (a) and 22 (b), respectively. Hence SNP 2 loci cannot be considered to be linked to the self-incompatibility trait because both the alleles are found to be present in both the genotypes. Hence SNP 2 is considered to be insignificant.

SNP 3

SNP3 locus was heterozygous in both self-compatible genotype and self-incompatible genotype. The loci were heterozygous for cytosine and guanine (C/G). The chromatographic peak present in both the sequence at the self-compatible and self-incompatible locus is shown in Figure 23 (a) and 23 (b), respectively. Hence SNP 3 loci

cannot be considered to be linked with self-incompatibility trait because both the alleles are found to be present in both the genotypes. Hence the SNP is considered insignificant.

SNP 4

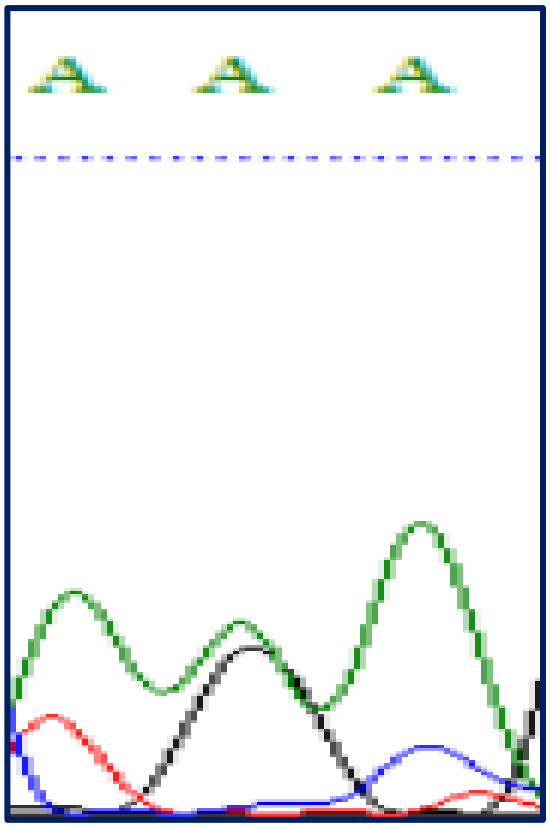
SNP 4 locus was heterozygous in both self-compatible genotype and self-incompatible genotype. The locus was heterozygous for adenine and guanine (A/G). The chromatographic peak present in both the sequence at the self-compatible and self-incompatible locus is shown in Figures. 24 (a) and 24 (b) respectively. Hence SNP 4 loci cannot be considered to be linked with self-incompatibility trait because both the alleles were found to be present in both the genotypes. Hence considered insignificant.

All the SNP loci identified within this region were found to be heterozygous in both, self-compatible genotype and the self-incompatible genotype. As a result this region is considered not to be associated with the self-incompatibility in cocoa. Hence the four SNPs discovered in this region can be considered insignificant.

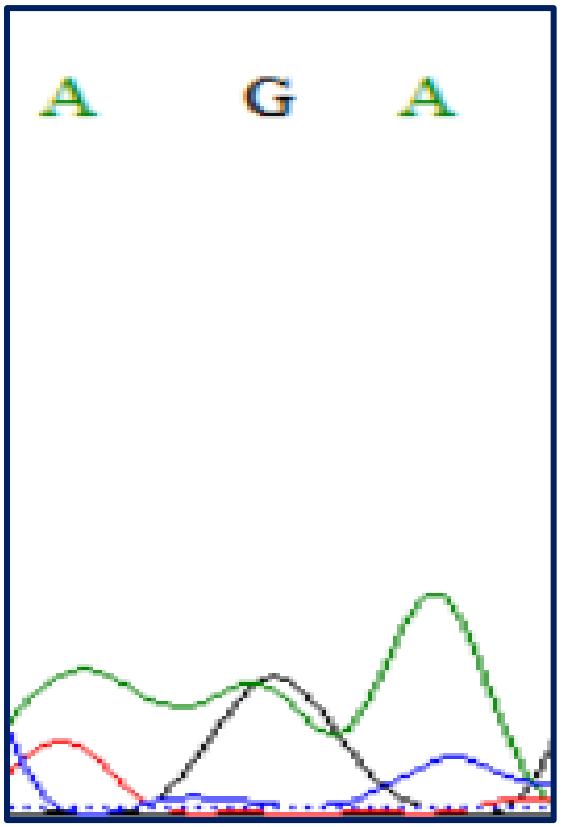
BAM1 and *BAM2* codes for *CLAVATA1* – related leucine rich repeat receptor-like kinases, which were reported to have important role in cell to cell communication during the early stages of anther development, processes such as cell division and cell differentiation (Hord *et al.*, 2006). Lanaud *et al.* (2017) reported that *Tc01_g007220* (*BAM 1*) locus was found to be over expressed in self-compatible reaction. However, in the present study, there were no variation in the sequences that was discovered which could significantly contribute to or be linked to self-incompatibility. This is due to the fact that only a small portion of the gene sequence was used for the study. Further analysis of whole gene sequence may yield potential regions of SNP variations.

Table 10. List of SNPs and base called at their respective loci (*BAM2*)

SI. No.	**Self-compatible**	**Self-incompatible**	**Reference**
SNP 1	A/G	G/A	A
SNP 2	T/A	A/T	A
SNP 3	C/G	G/C	C
SNP 4	G/A	A/G	A

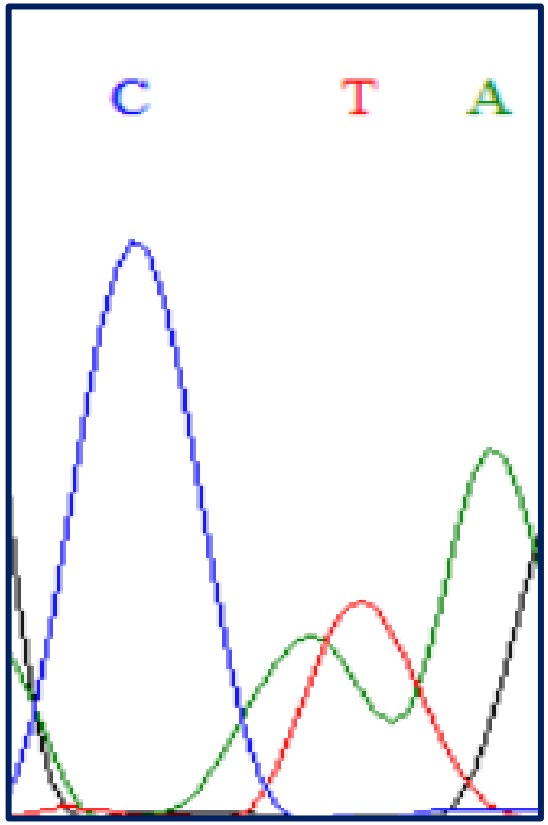


a.Self-compatible

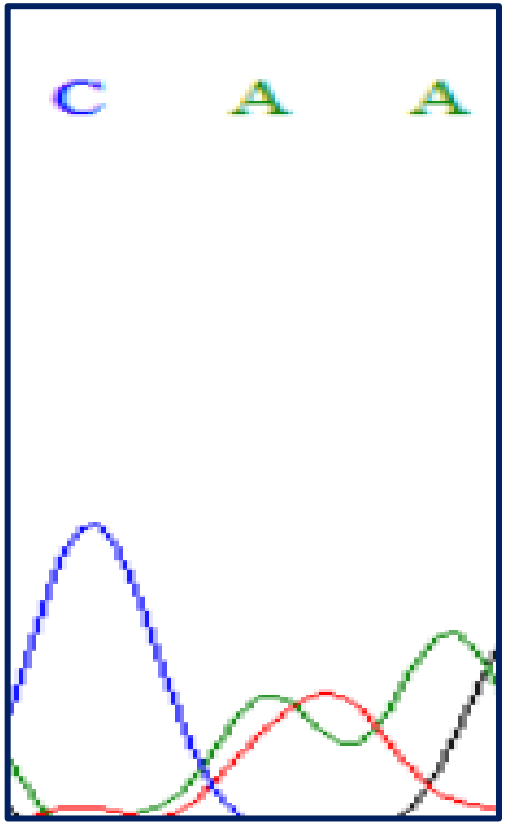


b. Self-incompatible

Figure 21. Base call at SNP 1 (*BAM2*)

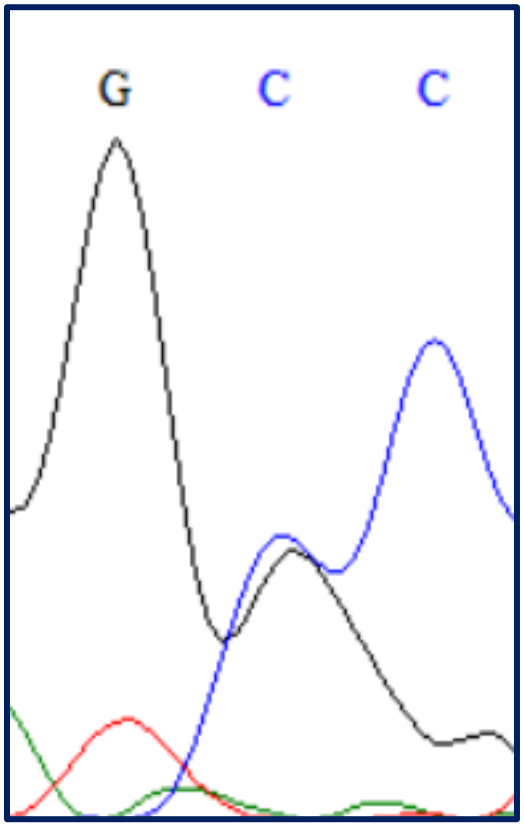


a. Self-compatible

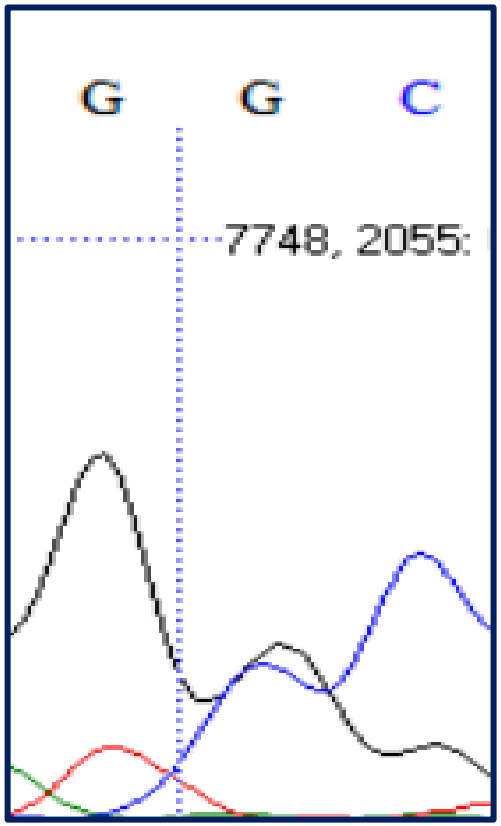


b. Self-incompatible

Figure 22. Base call at SNP 2 (*BAM2*)

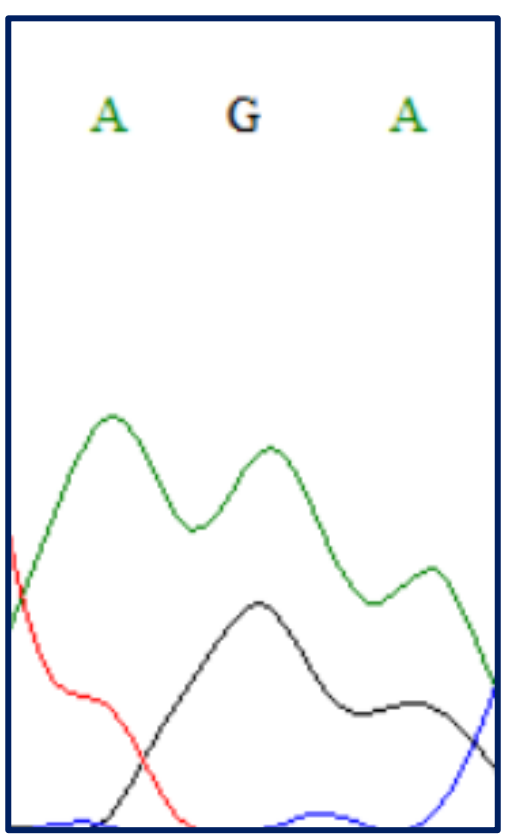


a. Self-compatible

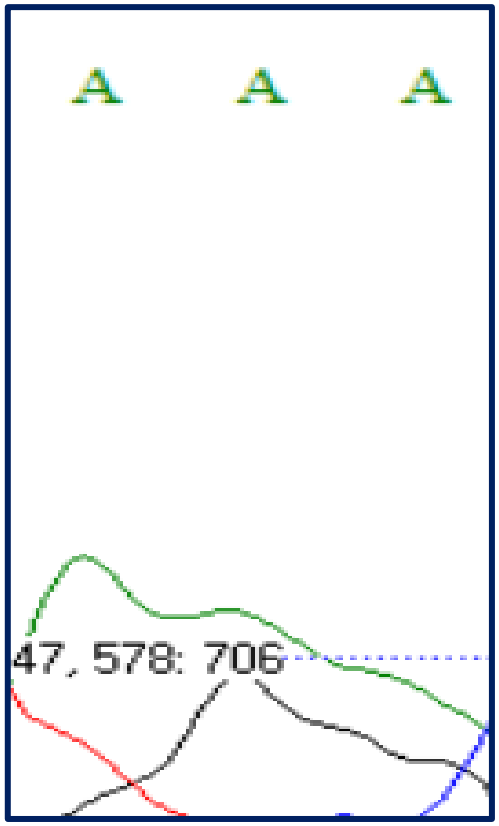


b. Self-incompatible

Figure 23. Base call at SNP 3 (*BAM2*)



a. Self-compatible



b. Self-incompatible

Figure 24. Base call at SNP 4 (*BAM2*).

4.4.4.3.5 *COMPASS-like H3K4 histone methylase component (WDR5a)*

4.4.4.3.5.1 Amplification of the primer

The primer set used to amplify targeted the first 975 bp of the gene. It covered entire first exon out of the two exons of the gene. PCR was run with the annealing temperature of 60.2^{o}C, which was determined by the gradient PCR experiment. The optimum concentration of forward and reverse primers of reaction mixture was found to be 0.5 μM. Upon agarose gel electrophoresis, successful amplification was observed in both the genotypes and a single band in the size of the expected product size of 975 base pairs was obtained, with minimum primer dimers (Table 7). The gel image is portrayed in Plate 7.

4.4.4.3.5.2 Analysis of the sequences

In self-compatible genotype, trimmed sequence of the forward run was of 942 base pairs and had a QV score of 52. The length of the reverse sequence was of 864 base pairs and the QV score was 51. (Table 8). In self-incompatible genotype, trimmed sequence of the forward run was of 871 base pairs and had a QV score of 54. The length of the reverse sequence was of 732 base pairs and the QV score was 52. (Table 8).

4.4.4.3.5.3 Contig assembly

The contig obtained after assembling the forward and reverse sequence of the *WDR5a* primer combination was 948 bp long for the self-compatible genotype and 871 bp long for the self-incompatible genotype (Table 8).

4.4.4.3.5.4 Computational confirmation

BLASTn analysis of the contigs against the cocoa genome database has confirmed that the sequence obtained was exactly of the targeted region (Figure 25).

4.4.4.3.5.5 Multiple sequence alignment

Sequences of self-compatible and self-incompatible genotypes were aligned with the reference sequence (compatible). Four variants were found when the sequences were analyzed. The alignment of the two sequences are presented in Figure 26.

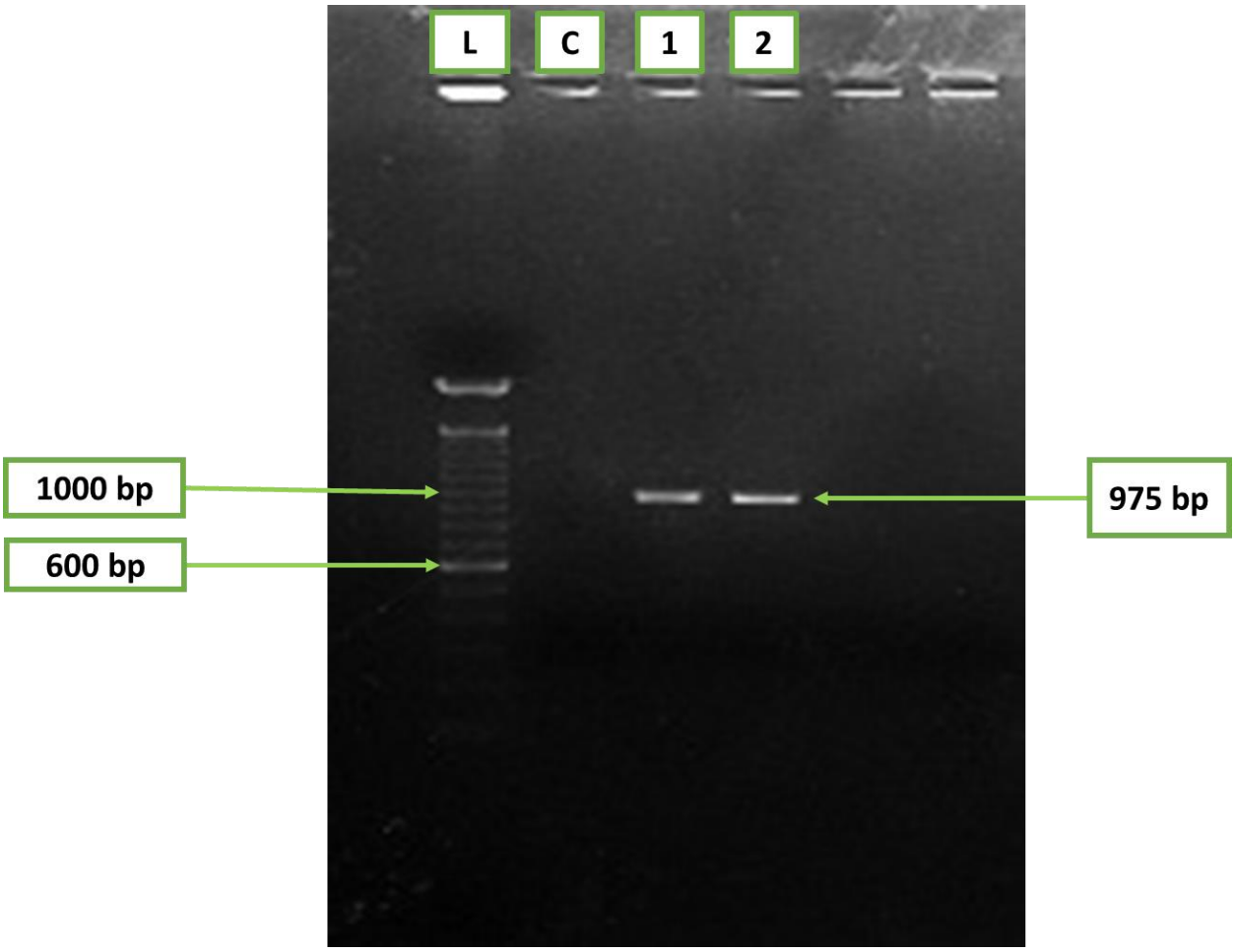


Plate 7. Amplification of *WDR5a*

1- Amplification in compatible genotype
2- Amplification in incompatible genotype
C. Control/ Blank

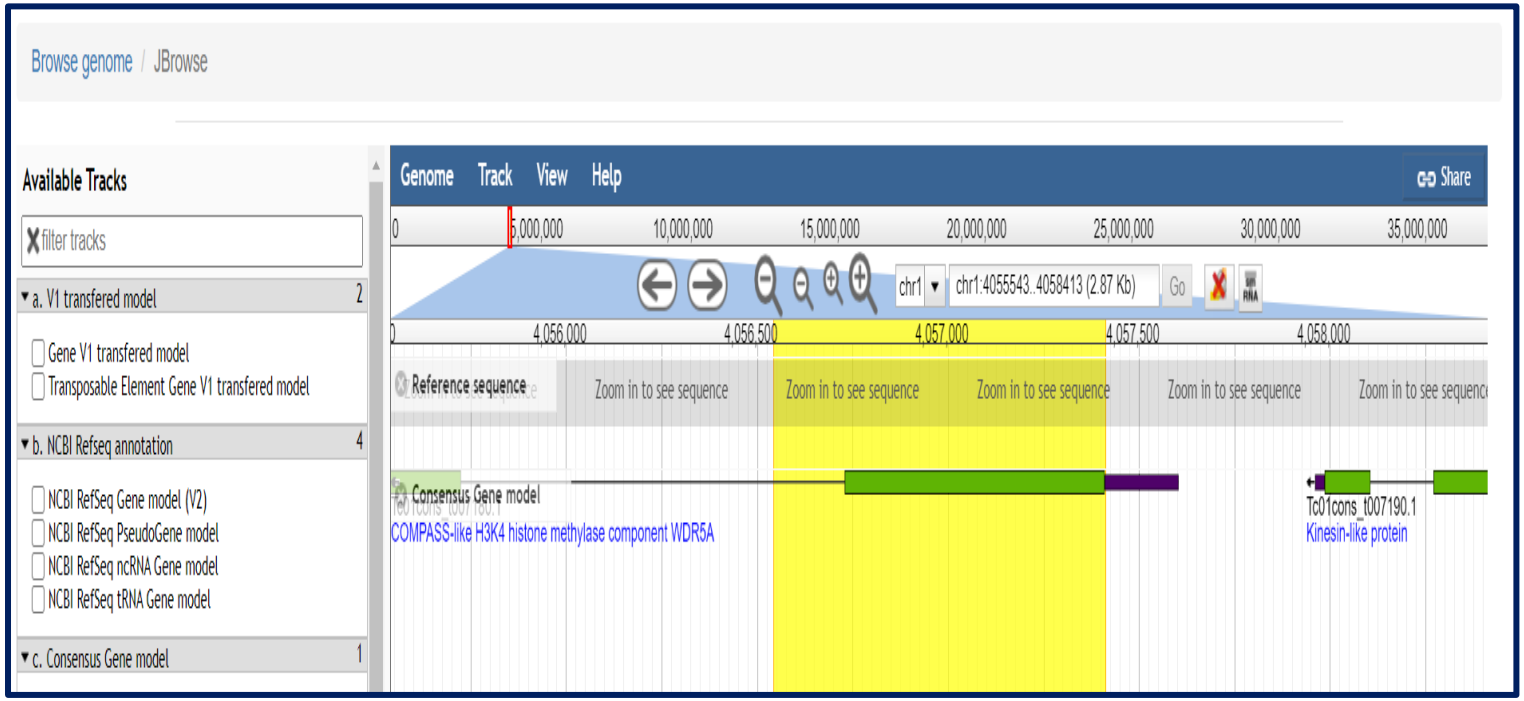


Figure 25. Computational confirmation of the *WDR5a* region sequenced by BLAST search.

Figure 26 (a). Multiple sequence alignment of *WDR5a*

```
EIncompatible_WDR.F_25720-11_P3586,Trimmed      CTATGTCTTCTGCGCCAACTTTAACCCGCAATCTAACATGATCGTTTCGGGTTCTTTCGA    420
Reference                                       CTATGTCTTCTGCGCCAACTTTAACCCGCAATCTAACATGATCGTTTCGGGTTCTTTCGA    420
ECompatible_WDR.F_25720-9_P3590,Trimmed         CTATGTCTTCTGCGCCAACTTTAACCCGCAATCTAACATGATCGTTTCGGGTTCTTTCGA    418
                                                ************************************************************

EIncompatible_WDR.F_25720-11_P3586,Trimmed      CGAAACGGTGCGTGTTTGGGACGTGAAAACGGGCAAATGTTTGAAAGTTTTACCGGCTCA    480
Reference                                       CGAAACGGTGCGTGTTTGGGACGTGAAAACGGGCAAATGTTTGAAAGTTTTACCGGCTCA    480
ECompatible_WDR.F_25720-9_P3590,Trimmed         CGAAACGGTGCGTGTTTGGGACGTGAAAACGGGCAAATGTTTGAAAGTTTTACCGGCTCA    478
                                                ************************************************************

EIncompatible_WDR.F_25720-11_P3586,Trimmed      TTCGGATCCAGTGACAGCCGTTGATTTTAATCGAGATGGATCCCTCATTGTTTCGAGTAG    540
Reference                                       TTCGGATCCAGTGACAGCCGTTGATTTTAATCGAGATGGATCCCTCATTGTTTCGAGTAG    540
ECompatible_WDR.F_25720-9_P3590,Trimmed         TTCGGATCCAGTGACAGCCGTTGATTTTAATCGAGATGGATCCCTCATTGTTTCGAGTAG    538
                                                ************************************************************

EIncompatible_WDR.F_25720-11_P3586,Trimmed      TTACGATGGGCTTTGTCGGATTTGGGATGCTGGAACTGGACATTGTATGAAGACTTTGAT    600
Reference                                       TTACGATGGGCTTTGTCGGATTTGGGATGCTGGAACTGGACATTGTATGAAGACTTTGAT    600
ECompatible_WDR.F_25720-9_P3590,Trimmed         TTACGATGGGCTTTGTCGGATTTGGGATGCTGGAACTGGACATTGTATGAAGACTTTGAT    598
                                                ************************************************************

EIncompatible_WDR.F_25720-11_P3586,Trimmed      TGATGATGAGAACCCACCTGTTTCATTTGTTAAATTTTCGCCTAATGGGAAATTTATTCT    660
Reference                                       TGATGATGAGAACCCACCTGTTTCATTTGTTAAATTTTCGCCTAATGGGAAATTTATTCT    660
ECompatible_WDR.F_25720-9_P3590,Trimmed         TGATGATGAGAACCCACCTGTTTCATTTGTTAAATTTTCGCCTAATGGGAAATTTATTCT    658
                                                ************************************************************
```

Figure 26 (b). Multiple sequence alignment of *WDR5a.*

Figure 26 (c). Multiple sequence alignment of *WDR5a*.

4.4.4.3.5.6 Characterization of SNP

SNP 1

In self-incompatible SNP1 locus was heterozygous. Thymine guanine were the bases called at the locus (T/G). Whereas in the self-compatible genotype, there was only one guanine peak. The chromatographic peak present in both the sequence at the self-compatible and self-incompatible locus is shown in Figure 27 (a) and 27 (b) respectively. The reference genome possessed guanine at the same locus

Thymine was present only in one of the alleles of self-incompatible genotypes, whereas guanine was present in both the compatible genotypes (reference genotype and the genotype used in this study). The allele possessing adenine was only specific to self-incompatible and hence this locus is a potential candidate that could be linked to self-incompatibility.

SNP 2

The self-incompatible genotype possessed a homozygous adenine (A), whereas the self-compatible genotype was found to be heterozygous for thymine and adenine (A/T). The chromatographic peak present in both the sequence at the self-compatible and self-incompatible locus is shown in Figure 28 (a) and 28 (b) respectively. The reference sequence possessed a thymine (T) in the same locus.

Here the allele found to be present in the self-incompatible genotype at homozygous condition was also present in the self-compatible genotype. Hence it seems unlikely that this locus could be linked with the trait.

SNP 3

In self-incompatible genotype possessed a homozygous cytosine (C). Self-compatible genotype was found to be heterozygous, a guanine and cytosine peak was observed at the SPN locus (C/G). The chromatographic peak present in both the sequence at the self--compatible and self-incompatible locus is shown in Figure 29 (a) and 29 (b) respectively. The reference sequence possessed a cytosine (C) in the same locus.

Cytosine the base found to be homozygous at self--incompatible locus was common in one of the alleles in self-compatible genotype and also the reference sequence. Hence SNP3 in considered insignificant and not linked to the trait.

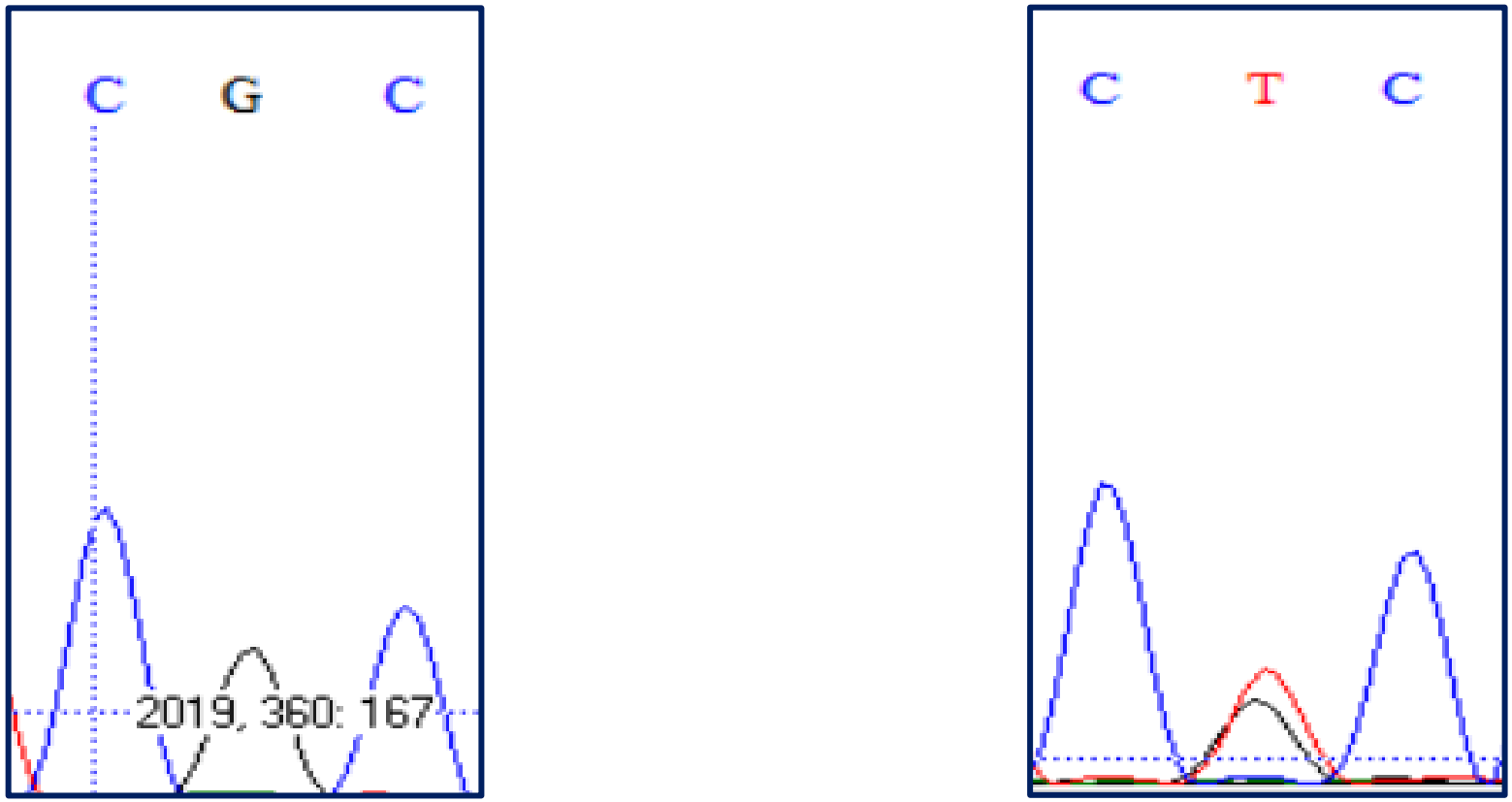
SNP 4

SNP 4 loci was heterozygous in both self-compatible genotype and self-incompatible genotype. The loci were heterozygous for both adenine and cytosine (A/C). The chromatographic peak present in both the sequence at the self-compatible and self-incompatible locus is shown in Figure 30 (a) and 30 (b) respectively. Hence SNP 4 loci cannot be considered to be linked to self-incompatibility trait because both the alleles are found to be present in both the genotypes. Therefore SNP 4 locus is considered to be insignificant.

Wang *et al.,* (2004) reported WD40 repeats to be involved in final steps of ubiquitination cascade which results in protein degradation, which is also reported to be an important part of self-incompatibility in plants. Chen *et al*. (2012) reported that WD40 domain containing proteins act as the ubiquitin binding site for the S-locus F box protein responsible for self-incompatibility in *Rosaceae.* In present study, in WDR5a, SNP 1 has been discovered that could be potentially linked to self-incompatibility trait. The self-compatible genotype is homozygous for guanine, whereas self-incompatible genotype is heterozygous for guanine and thymine. The compatible reference genome possessed a guanine. The allele possessing adenine was only specific to self-incompatible genotype and hence this locus is a potential candidate that could be linked to self-incompatibility.

Table 11. List of SNPs and base called at their respective loci (*WDR5a*)

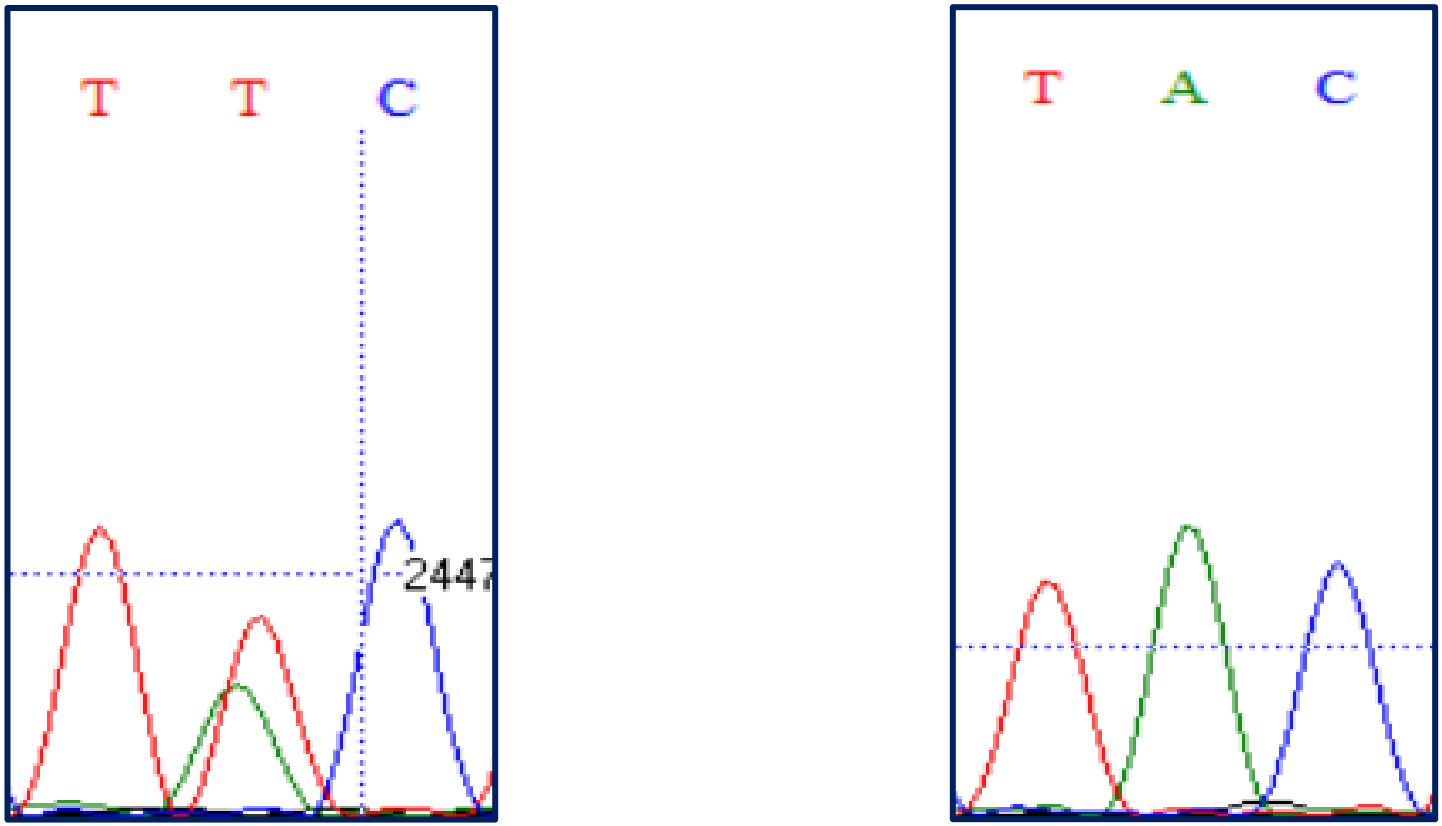
SI. No.	Self-compatible	Self-incompatible	Reference
SNP - 1	G	T/G	G
SNP - 2	T/A	A	T
SNP - 3	G/C	C	C
SNP - 4	C/A	C/A	A



a. Self-compatible

b. Self-incompatible

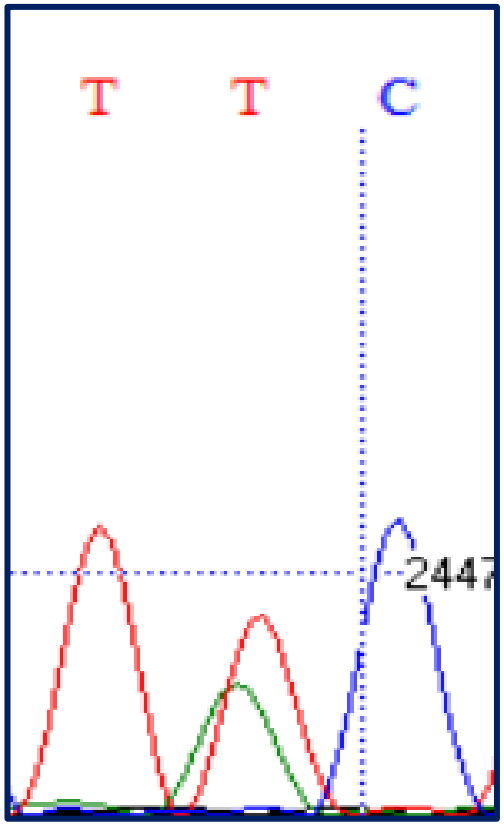
Figure 27. Base call at SNP 1(*WDR5a*)



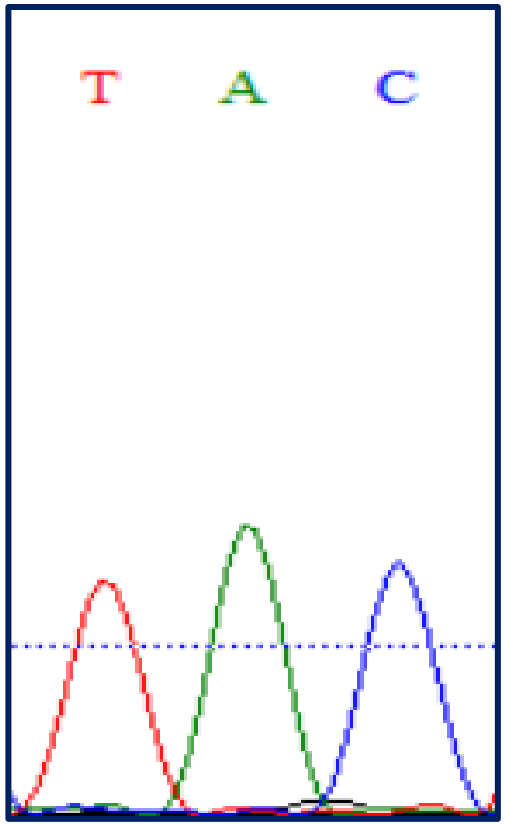
a. Self-compatible

b. Self-incompatible

Figure 28. Base call at SNP 2 (*WDR5a*)

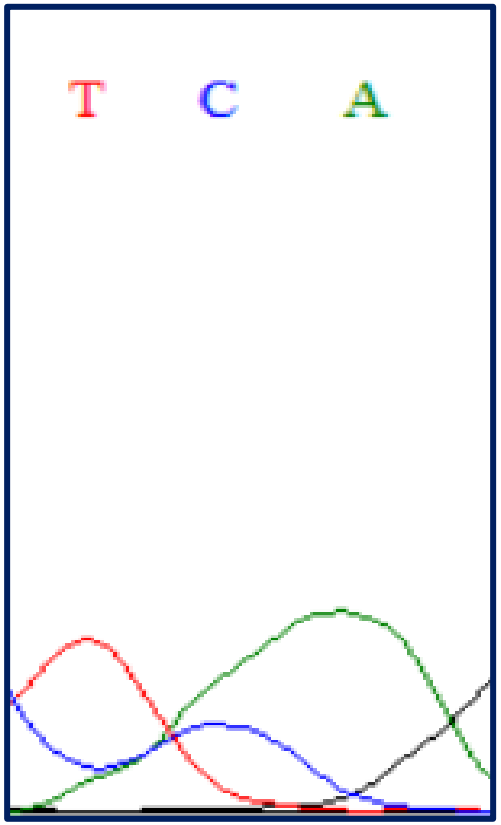


a. Self-compatible

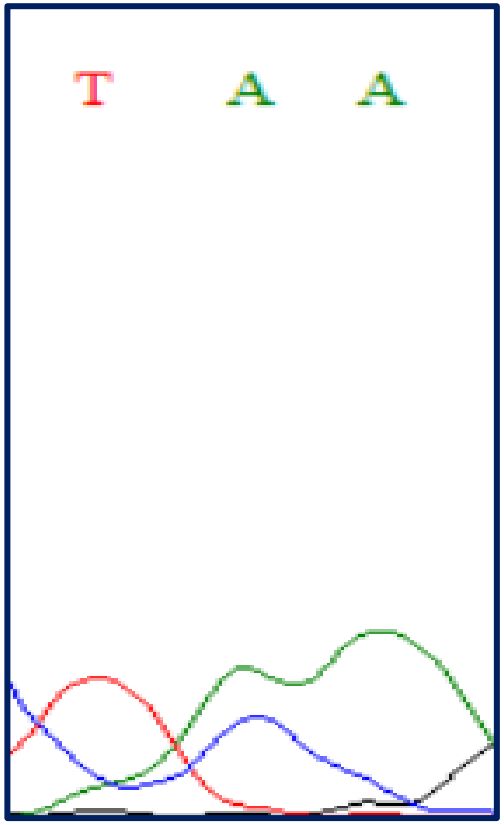


b. Self-incompatible

Figure 29. Base call at SNP 3 (*WDR5a*)



a. Self-compatible



b. Self-incompatible

Figure 30. Base call at SNP 4 (*WDR5a*)

4.4.4.3.6 *Voltage-dependent L-type calcium channel subunit* (*Alpha-1F*)

4.4.4.3.6.1 Amplification of the primer

The primer set used to amplify targeted the first 1003 bp of the gene. It covered a portion on first exon, out of the three exons 5'. PCR was run with the annealing temperature of 62.2^{o}C, which was determined by the gradient PCR experiment. The optimum concentration of forward and reverse primers of reaction mixture was found to be 0.4 μM. Upon agarose gel electrophoresis, successful amplification was observed in both the genotypes and a single band in the size of the expected product size obtained, with minimum primer dimers (Table 7). The gel image is portrayed in Plate 8. The PCR products were sequenced.

4.4.4.3.6.2 Analysis of the sequences

In self-compatible genotype, trimmed sequence of the forward run was of 872 base pairs and had a QV score of 53. The length of the reverse sequence was of 885 base pairs and the QV score was 53 (Table 8).

In self-incompatible genotype, trimmed sequence of the forward run was of 871 base pairs and had a QV score of 54. The length of the reverse sequence was of 837 base pairs and the QV score was 51 (Table 8).

4.4.4.3.6.3 Contig assembly

The contig obtained after assembling the forward and reverse sequence of the *Alpha-1F* primer combination was 954 bp long for the self-compatible genotype and 962 bp long for the self-incompatible genotype (Table 8).

4.4.4.3.6.4 Computational confirmation

BLASTn analysis of the contigs against the cocoa genome database has confirmed that the sequence obtained was exactly of the targeted region (Figure 31).

4.4.4.3.6.5 Multiple sequence alignment

Sequences of self-compatible and self-incompatible genotypes were aligned with the reference sequence (compatible). Four variants were found when the sequences were analyzed. The alignment of the three sequences is presented in Figure 32.

4.4.4.3.6.6 Characterization of SNP

SNP 1

Self-incompatible SNP 1 locus was heterozygous cytosine and adenine (C/A). In self-compatible genotype, one homozygous adenine peak was present (A). The chromatographic peak present in both the sequence at the self-compatible and self-incompatible locus is shown in Figure 33 (a) and (b) respectively. The reference genome possessed adenine at the locus (A).

Cytosine is present only in one of the alleles of self-incompatible genotype, whereas adenine is present in both compatible genotypes (reference genotype and the genotype used in this study). Hence this SNP 1 can be considered to be linked with self-incompatibility.

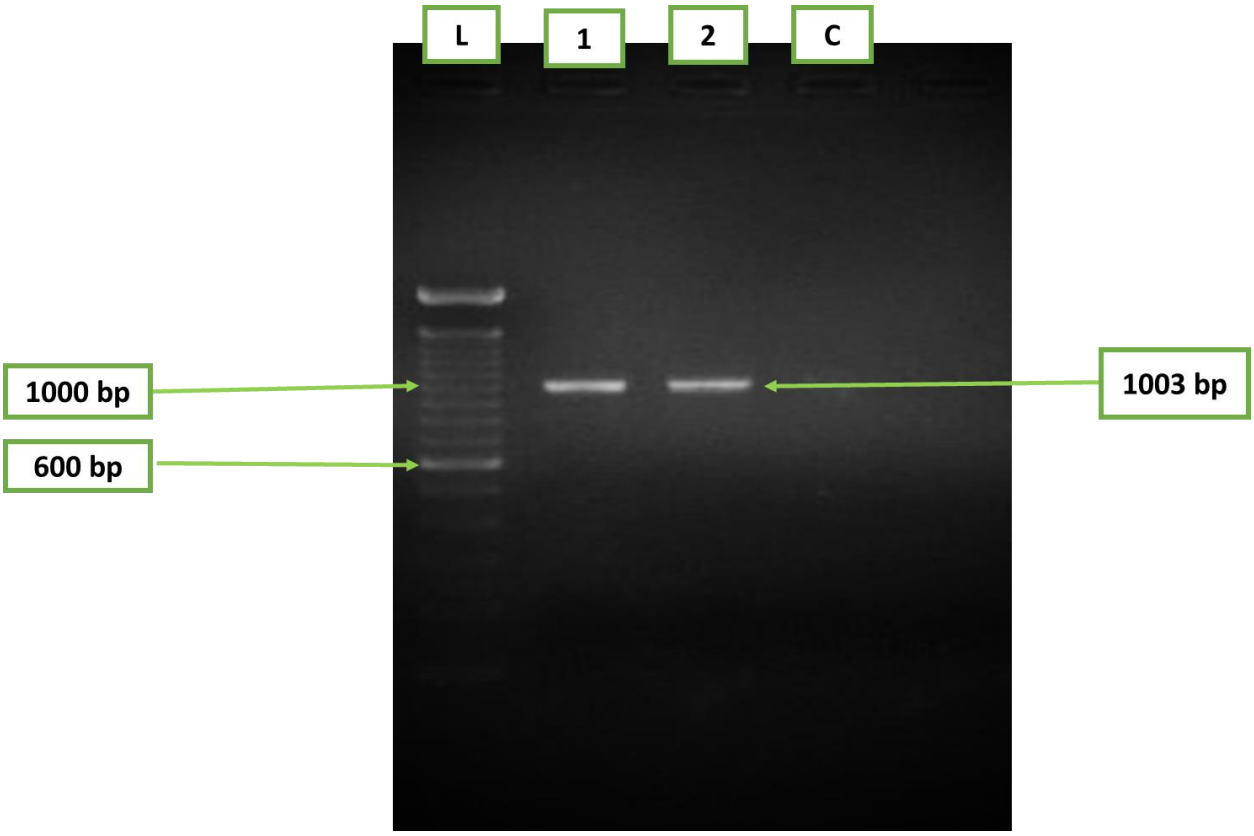


Plate 8. Amplification of *Alpha-1F*

1- Amplification in compatible genotype
2- Amplification in incompatible genotype
C. Control/ Blank

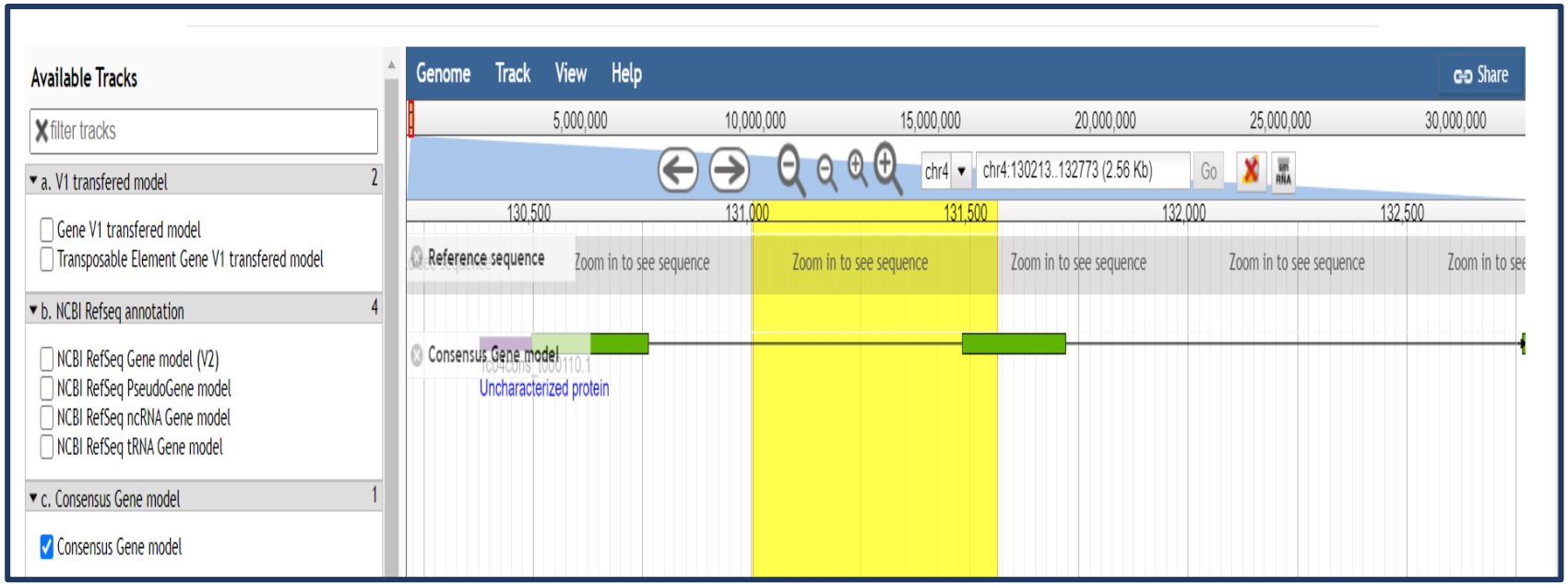


Figure 31. Computational confirmation of *Alpha-1F* region sequenced by BLAST search.

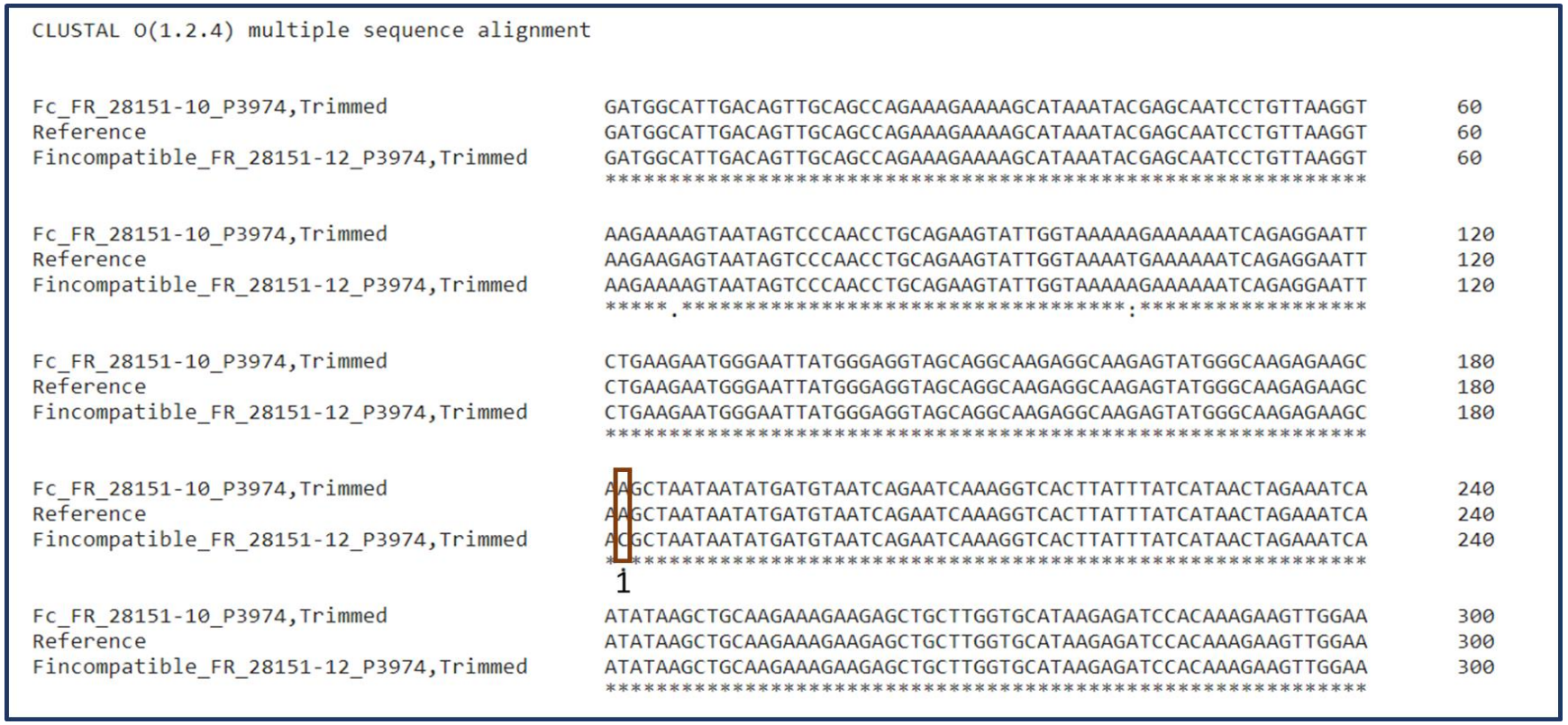


Figure 32 (a). Multiple sequence alignment of *Alpha-1F*

Figure 32 (b). Multiple sequence alignment of *Alpha-1F*

SNP 2

Self-incompatible SNP 2 locus was heterozygous. Thymine (T) was the base called at the locus, there was a significant guanine peak too (T/G). In self-compatible genotype, there was only single guanine peak (G). The chromatographic peak present in both the sequence at the self-compatible and self-incompatible locus is shown in Figure 34 (a) and 34 (b) respectively. The reference genome possessed guanine at the locus.

Thymine is present only in one of the alleles of self-incompatible genotypes, whereas guanine is present in both the compatible genotypes (reference genotype and the genotype used in this study). Hence this SNP may be linked and be considered for further analysis.

SNP 3

At the locus thymine was found in one of the alleles of self-incompatible genotypes, whereas only adenine is present in both compatible genotypes (reference genotype and the genotype used in this study). Hence this SNP 3 can be considered to be linked with self-incompatibility and be used for further analysis.

SNP 4

From chromatogram, it was evident that self-incompatible SNP 4 locus was heterozygous. Cytosine (C) was the base called at the locus, another significant guanine peak was noticed (C/G). In self-compatible genotype, there was only one clear single guanine peak (G). The chromatographic peak present in both the sequence at the self-compatible and self-incompatible locus is shown in Figure 36 (a) and 36 (b) respectively. The reference genome possessed guanine at the locus.

Cytosine was present only in of self-incompatible genotype, while guanine is present in both compatible genotypes at a homozygous state (reference genotype and the

genotype used in this study). Hence SNP 4 can be considered to be linked with self-incompatibility and used for further study.

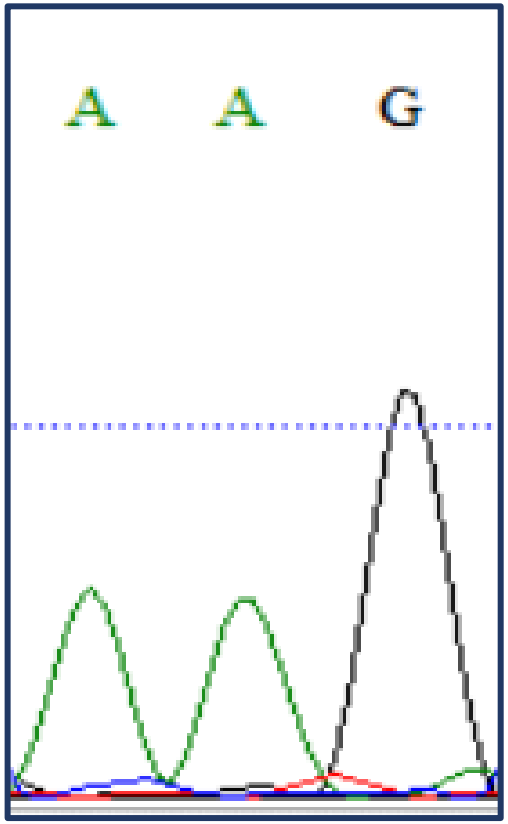
Alpha – 1F gene has been reported to play a key role in cellular Ca2+ influx in animals and plants (Kotturi *et.al.*, 2003). Ca^{2+} ions act as secondary messengers involved throughout the double fertilization process. Denninger et al. (2014) showed how Ca^{2+} was involved in all sorts of cellular events during the double fertilization event such as, cross talk between synergids and the pollen tube apex, pollen tube rupture and delivery of sperm into the ovule and finally gamete activation and fusion. It was also reported that they might be involved in the blockage of polyspermy in the egg cell.

In *Papaver rhoeas*, Wheeler *et al.* (2010) had reported that the interaction between the incompatible pollen and the pistil determinant proteins results in inhibition of pollen tube growth through Ca^{2+} dependant signaling network, which ultimately results in programmed cell death of the incompatible pollen.

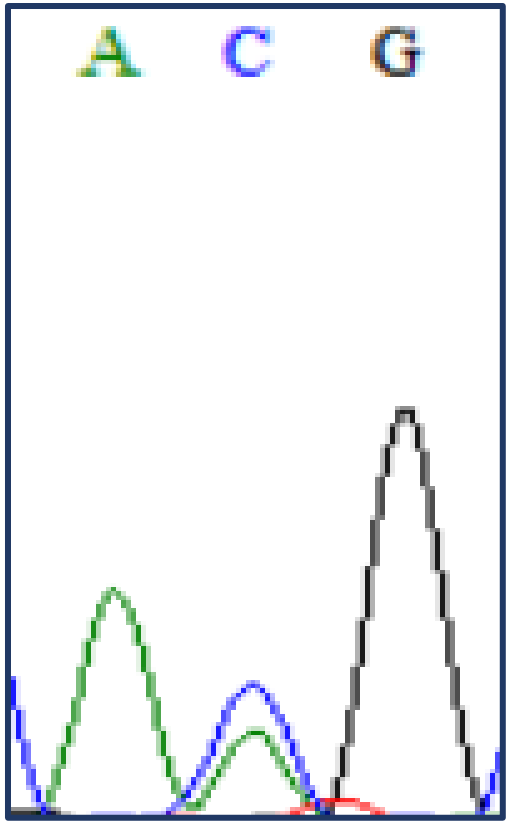
All the four discovered SNPs within *Alpha-1F* are potential candidates that could be linked to self-incompatibility trait.

Table 12. List of SNPs and base called at their respective loci (*Alpha-1F*)

SI. No	**Self-Compatible**	**Self-incompatible**	**Reference**
SNP – 1	A	C/A	A
SNP – 2	G	T/G	G
SNP – 3	T	A/T	T
SNP- 4	G	C/G	G

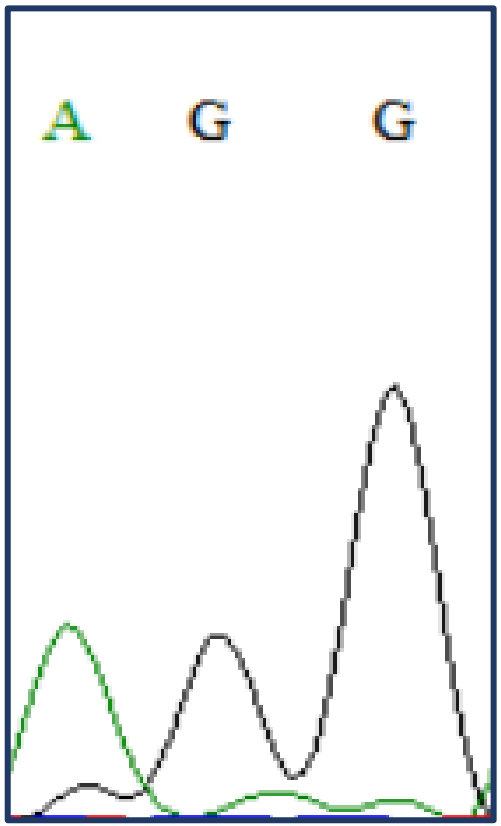


a. Self-compatible

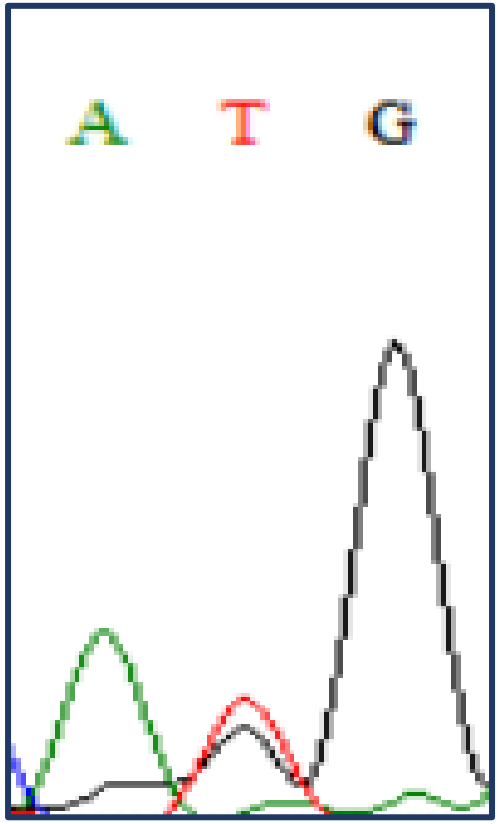


b. Self-incompatible

Figure 33. Base call at SNP 1(*Alpha-1F*)

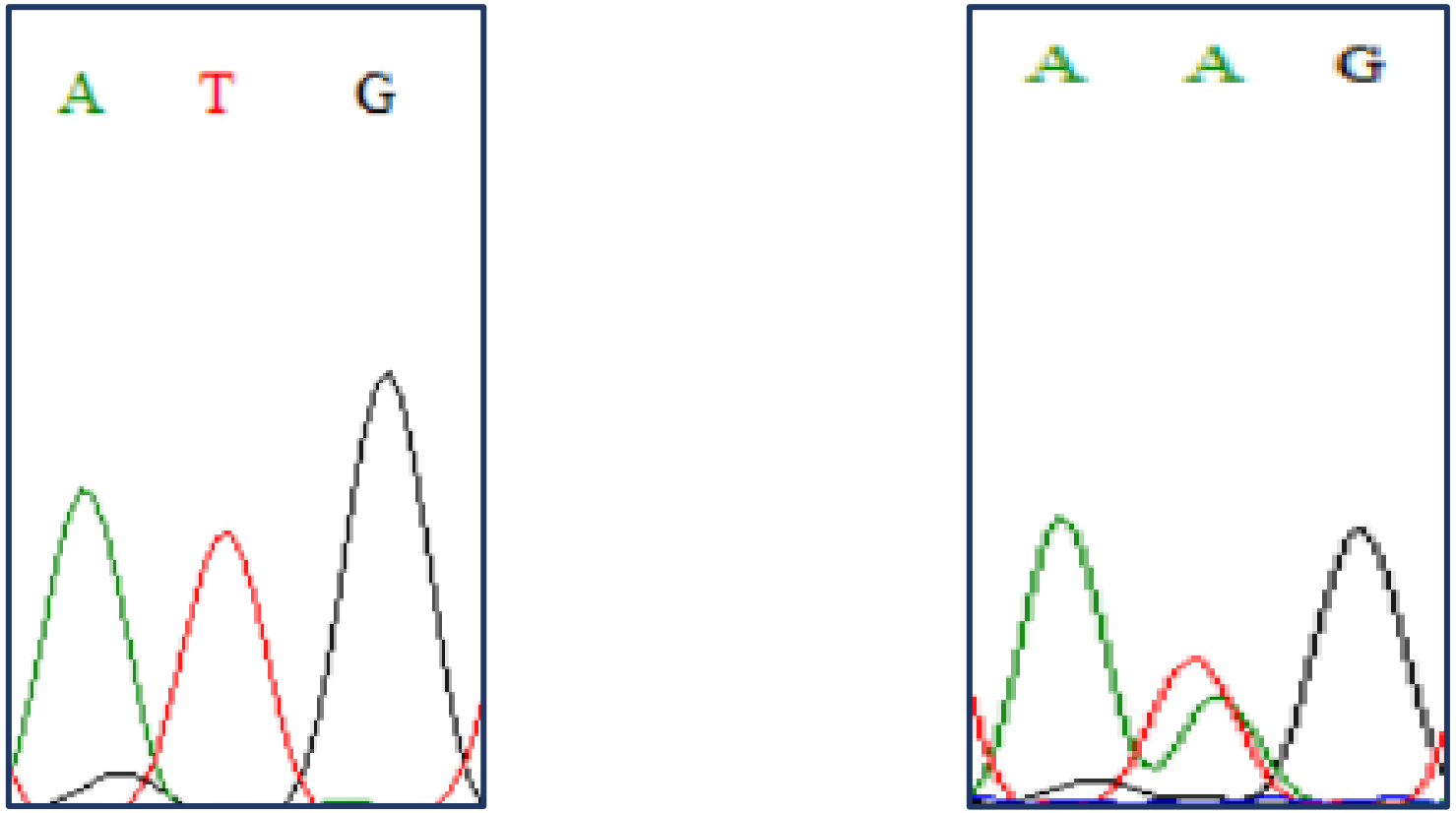


a. Self-compatible



b. Self-incompatible

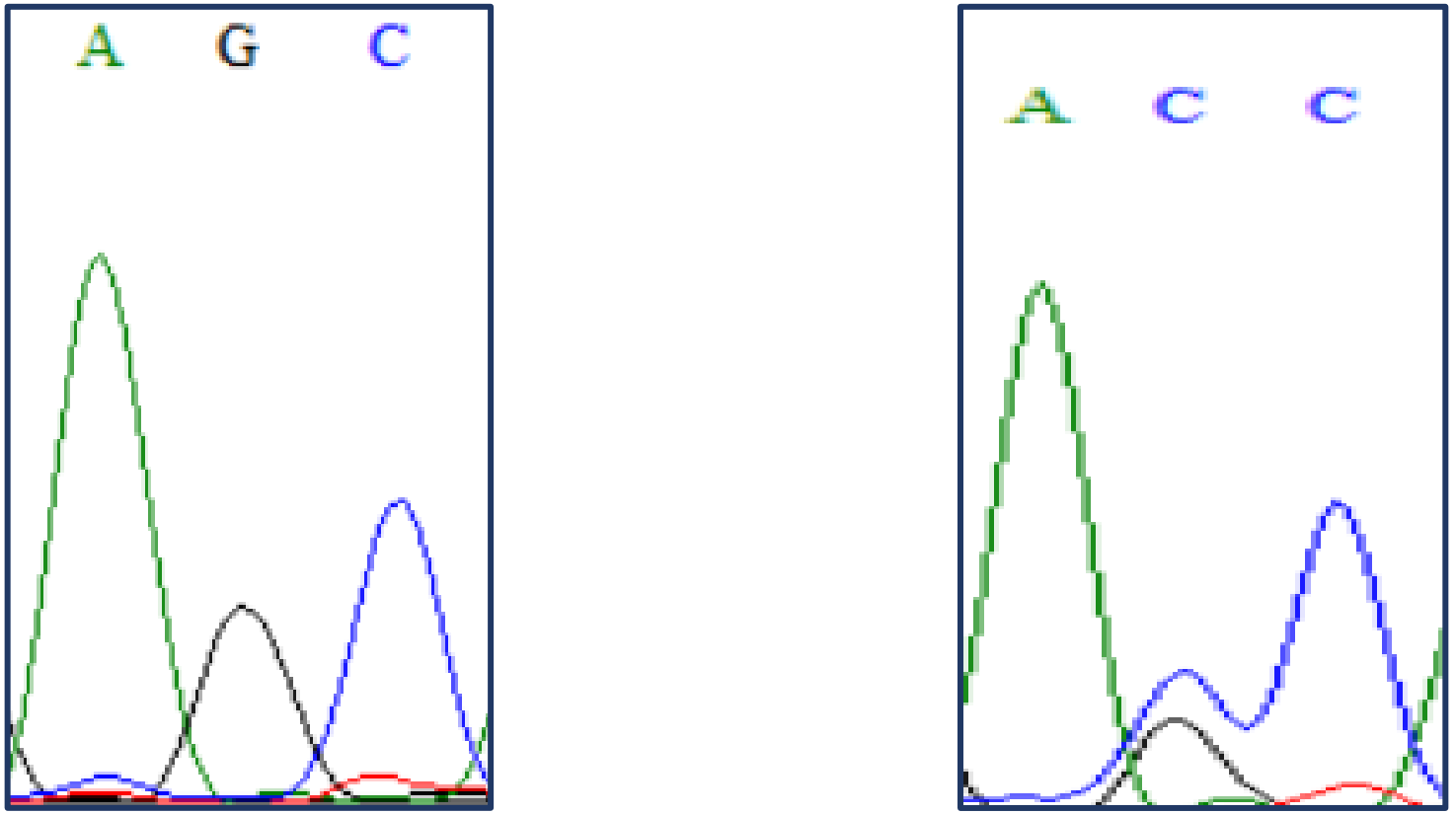
Figure 34. Base call at SNP 2 (*Alpha-1F*)



a. Self-compatible

b. Self-incompatible

Figure 35. Base call at SNP 3 (*Alpha-1F*)



a. Self-compatible

b. Self-incompatible

Figure 36. Base call at SNP 4 (*Alpha-1F*)

4.4.4.3.7 Gamete expression protein *(GEX 1)*

4.4.4.3.7.1 Amplification of the primer

The primer combination used to amplify targeted the first 985 bp on exon 5 and exon 6. PCR was run with the annealing temperature of 57.0 ^{o}C, which was determined by the gradient PCR experiment. The optimum concentration of forward and reverse primers of reaction mixture was found to be 0.4 μM. Upon agarose gel electrophoresis, successful amplification was observed in both the genotypes and a single band in the size of the expected product size of 985 base pairs was obtained, with minimum primer dimers (Table 7). The gel image is portrayed in Plate 9. The PCR products were sequenced.

4.4.4.3.7.2 Analysis of the sequences

In self-compatible genotype, trimmed sequence of the forward run was of 864 base pairs and had a QV score of 54. The length of the reverse sequence was of 847 base pairs and the QV score was 52 (Table 8).

In self-incompatible genotype, trimmed sequence of the forward run was of 942 base pairs and had a QV score of 52. The length of the reverse sequence was of 797 base pairs and the QV score was 54 (Table 8).

4.4.4.3.7.3 Contig assembly

The contig obtained after assembling the forward and reverse sequence of the *ARC1* primer was 963 base pairs long for the self-compatible genotype and 982 base pairs long for the self-incompatible genotype (Table 8).

4.4.4.3.7.4 Computational confirmation

BLASTn analysis of the contigs against the cocoa genome database has confirmed that the sequence obtained was exactly of the targeted region (Figure 37).

4.4.4.3.7.5 Multiple sequence alignment

Sequences of self-compatible and self-incompatible genotypes were aligned with the reference sequence (compatible). 14 variants were found to be present between the self-compatible and self-incompatible genotypes. This was the region with the highest mismatches in this study. The alignment of the two sequences is presented in Figure 38.

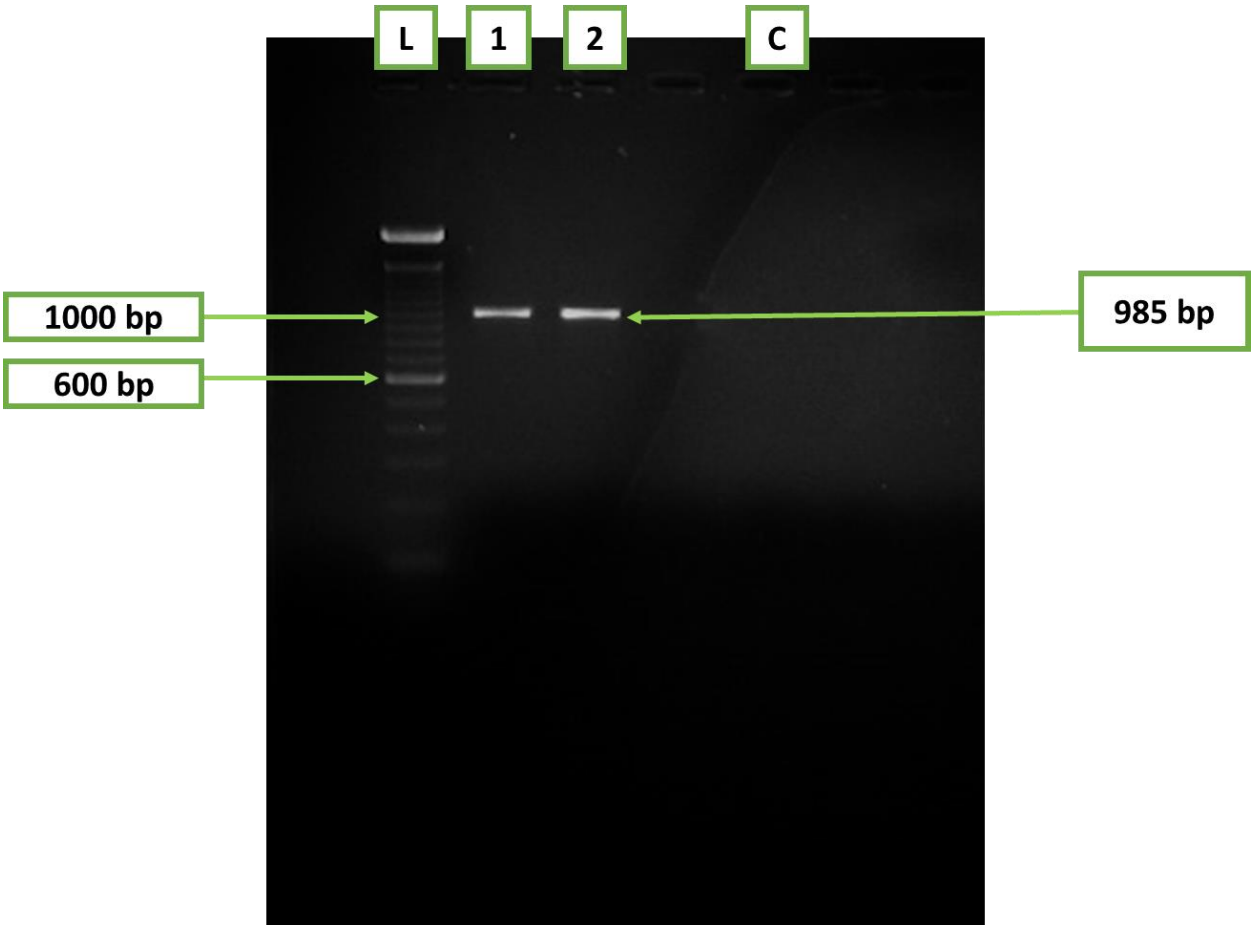


Plate 9. Amplification of *GEX1*

1- Amplification in compatible genotype
2- Amplification in incompatible genotype
C. Control/ Blank

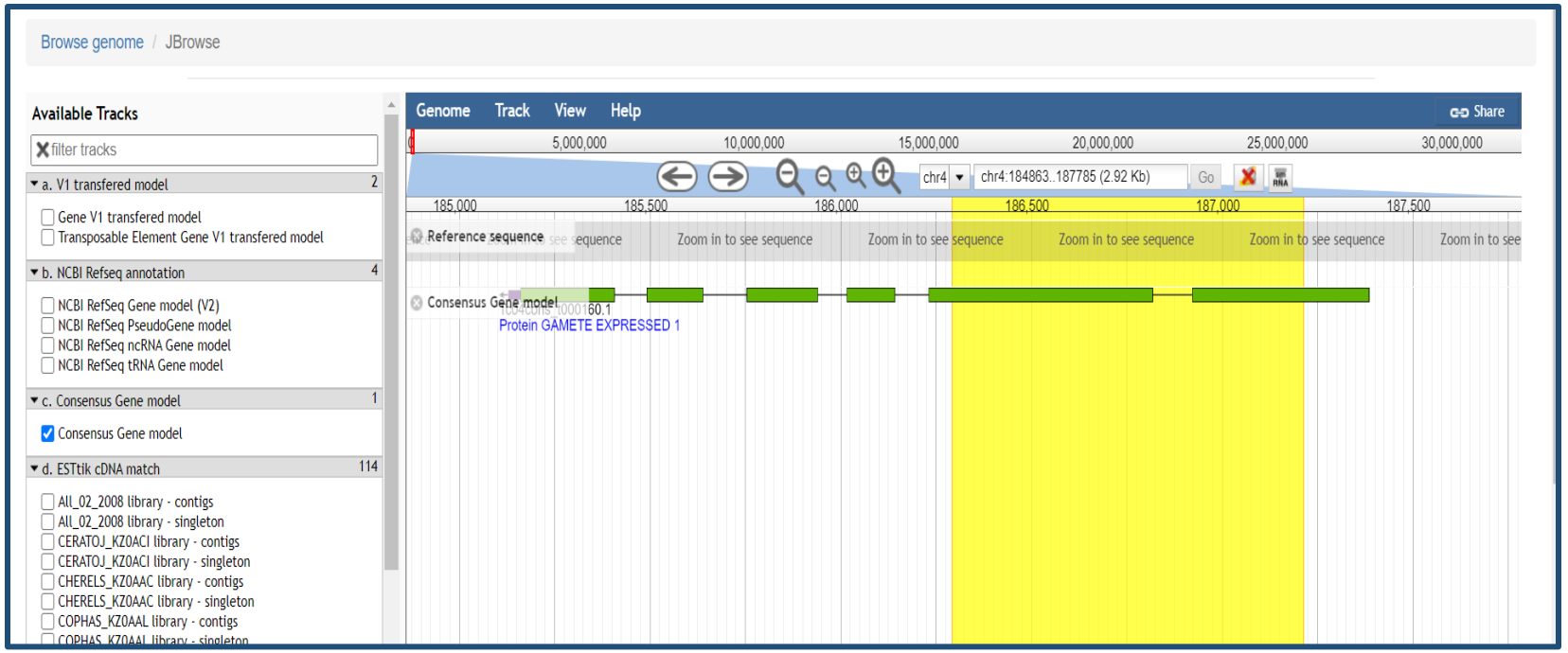


Figure 37. Computational confirmation of *GEX1* region sequenced by BLAST search

Figure 38 (a). Multiple sequence alignment of *GEX1*

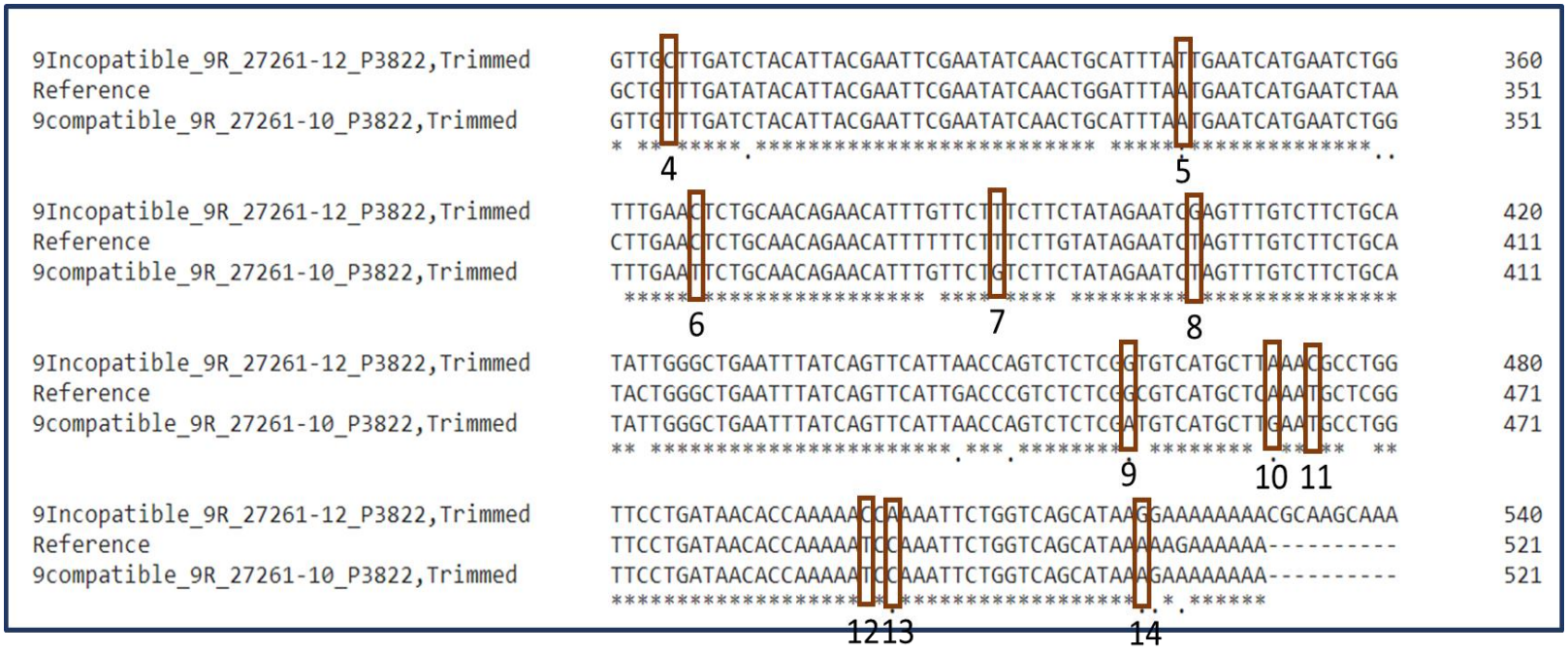


Figure 38 (b). Multiple sequence alignment of *GEX1*

4.4.4.3.7.6 Characterization of SNP

Out of the 14 variants, 13 variants were found to be SNPs and one variant was the result of sequencing error.

SNP 1

SNP 1 locus was heterozygous in self-compatible, cytosine and thymine were the bases present (C/T). Whereas in self-incompatible it was homozygous for thymine (T). The chromatographic peaks present in both the sequence at the self-compatible and self-incompatible SNP loci are shown in Figure 39 (a) and (b) respectively.

Cytosine was present in the sequence of reference genome. SNP1 loci is not considered for further analysis because the self-incompatible locus is homozygous in nature.

SNP 2

SNP 2 locus was heterozygous in self-compatible, cytosine and thymine were the peaks observed (C/T). Whereas in self-incompatible it was homozygous for thymine (T). The chromatographic peaks present in both the sequence at the self-compatible and self-incompatible SNP loci are shown in Figure 40 (a) and (b) respectively.

Cytosine was present in the sequence of reference genome. SNP 2 loci cannot be considered for further analysis because the self-incompatible locus is homozygous (C) in nature.

SNP 3

In self-compatible SNP 3 locus was heterozygous for adenine and thymine (A/T). Whereas in self-incompatible it was homozygous for adenine (A). The chromatographic peaks present in both the sequence at the self-compatible and self-incompatible SNP loci are shown in Figure 41 (a) and (b) respectively.

Thymine was present in the sequence of reference genome. SNP 3 loci cannot be considered to be linked with self-incompatibility because the homozygous allele A present in self-incompatible was also present in self-compatible genotypes. Hence SNP 3 was considered to be insignificant.

SNP 4

SNP 4 locus was heterozygous in self-compatible. The locus was heterozygous for both cytosine and thymine (C/T). Whereas in self-incompatible it was homozygous for cytosine (C). The chromatographic peaks present in both the sequence at the self-compatible and self-incompatible SNP loci are shown in Figure 42 (a) and (b) respectively.

Cytosine was present in the sequence of reference genome, since it is homozygous for cytosine (C) at the self-incompatible locus.

SNP 5

In self-compatible genotype SNP 5 locus was heterozygous for adenine and thymine (A/T). Whereas in self-incompatible it was homozygous for thymine (T). The chromatographic peaks present in both the sequence at the self-compatible and self-incompatible SNP loci are shown in Figure 43 (a) and (b) respectively.

Adenine was present in the sequence of reference genome. SNP 5 loci cannot be considered, since the self-incompatible locus is homozygous for thymine (T).

SNP 6

SNP 6 locus was heterozygous in self-compatible. The locus was heterozygous for both cytosine and thymine (C/T). Whereas self-incompatible was homozygous for cytosine (C). The chromatographic peaks present in both the sequence at the self-compatible and self-incompatible SNP loci are shown in Figure 44 (a) and (b) respectively.

Cytosine was present in the sequence of reference genome. SNP 6 loci cannot be considered for further analysis because the self-incompatible locus is homozygous for cytosine (C).

SNP 7

SNP 7 locus was heterozygous in self-compatible. The locus was heterozygous for both guanine and thymine (G/T). Whereas self-incompatible was homozygous for thymine (T). The chromatographic peaks present in both the sequence at the self-compatible and self-incompatible SNP loci are shown in Figure 45 (a) and (b) respectively.

Thymine was present in the sequence of reference genome. SNP 7 loci cannot be considered, since the self-incompatible locus is homozygous for thymine (T).

SNP 8

In self-compatible SNP 8 locus was for guanine and thymine (G/T). Whereas in self-incompatible it was homozygous for thymine (T). The chromatographic peaks present in both the sequence at the self-compatible and self-incompatible SNP loci are shown in Figure 46 (a) and (b) respectively.

Thymine was present in the sequence of reference genome. Since self-incompatible locus is homozygous for thymine (T), SNP 8 loci cannot be considered for further analysis.

SNP 9

SNP 9 locus was heterozygous in self-compatible having both adenine and guanine (G/A). The self-incompatible was homozygous for guanine (G). The chromatographic peaks present in both the sequence at the self-compatible and self-incompatible SNP loci are shown in Figure 47 (a) and (b) respectively.

Guanine was present in the sequence of reference genome. SNP 9 loci cannot be considered, since the self-incompatible locus is homozygous for guanine (G).

SNP 10

SNP 10 locus was heterozygous in self-compatible. The locus was heterozygous for both adenine and guanine (G/A). Whereas in self-incompatible it was homozygous for Adenine (A). The chromatographic peaks present in both the sequence at the self-compatible and self-incompatible SNP loci are shown in Figure 48 (a) and (b) respectively.

Adenine was present in the sequence of reference genome. SNP 10 loci cannot be considered for further analysis because the self-incompatible locus is homozygous for adenine (A).

SNP 11

Self-compatible SNP 11 locus was heterozygous for cytosine and thymine (C/T). Whereas in self-incompatible it was homozygous for cytosine (C). The chromatographic peaks present in both the sequence at the self-compatible and self-incompatible SNP loci are shown in Figure 49 (a) and (b) respectively.

Thymine was present in the sequence of reference genome. SNP 11 loci cannot be considered, since the self-incompatible locus is homozygous for cytosine (C).

SNP 12

SNP 12 locus was heterozygous for both cytosine and thymine (C/T). Whereas in self-incompatible it was homozygous for cytosine (C). The chromatographic peaks present in both the sequence at the self-compatible and self-incompatible SNP loci are shown in Figure 50 (a) and (b) respectively.

Thymine was present in the sequence of reference genome. SNP 12 loci cannot be considered for further analysis because the self-incompatible locus is homozygous for cytosine (C).

Error 1

In case of the self-compatible locus there were three peaks present as a result of sequencing error. Hence that locus could not be used further for analysis (Figure 51 (a)). Therefore, it is considered as an error.

SNP 13

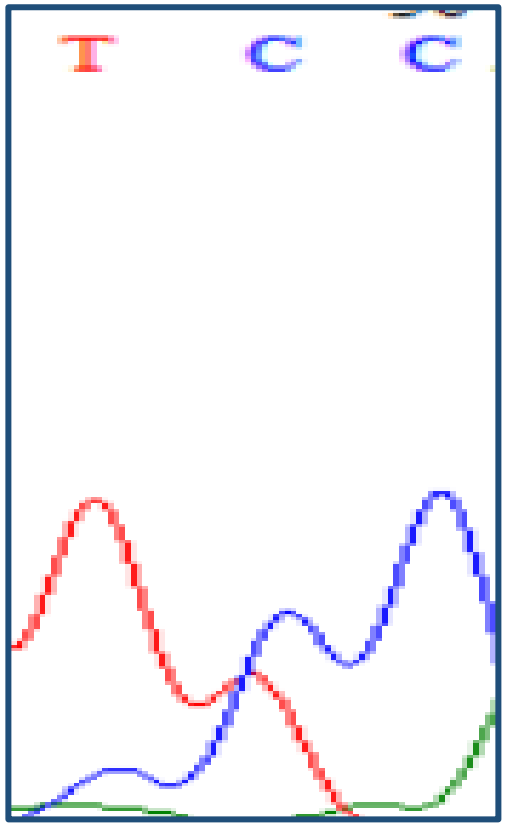
SNP 13 locus was heterozygous in self-compatible for adenine and guanine (G/A). Whereas in self-incompatible it was homozygous for Adenine (A). The chromatographic peaks present in both the sequence at the self-compatible and self-incompatible SNP loci are shown in Figure 52 (a) and (b) respectively.

Adenine was present in the sequence of reference genome. SNP 13 loci cannot be considered, since the self--incompatible locus is homozygous for adenine (A).

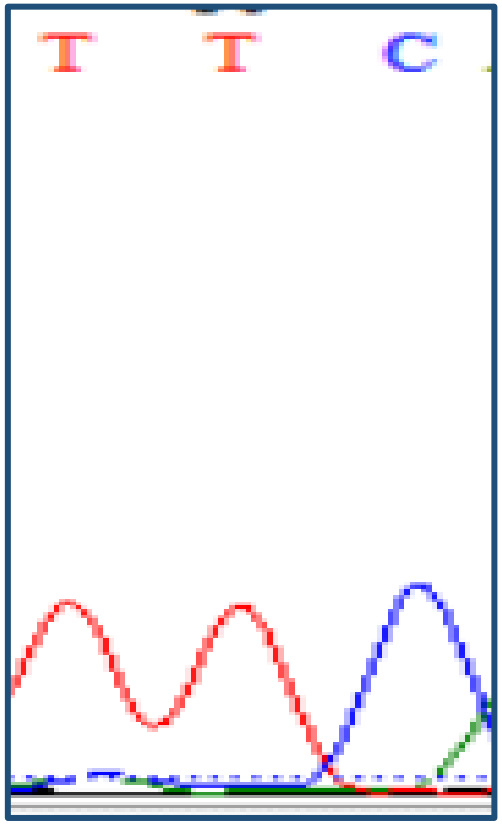
Alandete-Saez *et al.* (2011) reported *GEX1* to be expressed in multiple tissues, during various stages of embryo formation in *A. thaliana.* It was observed to be expressed in the embryo sac prior to cellularization, within the egg cell after cellularization, in the zygote soon after fertilization. It was also reported to be expressed in both the vegetative and the sperm cells of the male gametophyte. During this study, *GEX1* was the gene in which maximum number of SNPs were discovered. A total of 13 SNPs were discovered in *GEX1*.All the SNPs except SNP 13 was found to be heterozygous in the self--incompatible genotype G VI 167 X G IV 18.5. However, all such SNPs are of least significance since all the SNPs discovered were found to be heterozygous in self-compatible genotype and homozygous in self-incompatible genotype.

Table 13. List of SNPs and base called at their respective loci (*GEX1*)

SI. no	**Self-Compatible**	**Self-incompatible**	**Reference**
SNP 1	C/T	T	C
SNP 2	C/T	T	C
SNP 3	T/A	A	T
SNP 4	T/C	C	C
SNP 5	A/T	T	A
SNP 6	T/C	C	C
SNP 7	G/T	T	T
SNP 8	T/G	G	T
SNP 9	A/G	G	G
SNP 10	G/A	A	A
SNP 11	T/C	C	T
SNP 12	T/C	C	T
Error	-	-	-
SNP 13	G/A	A	A

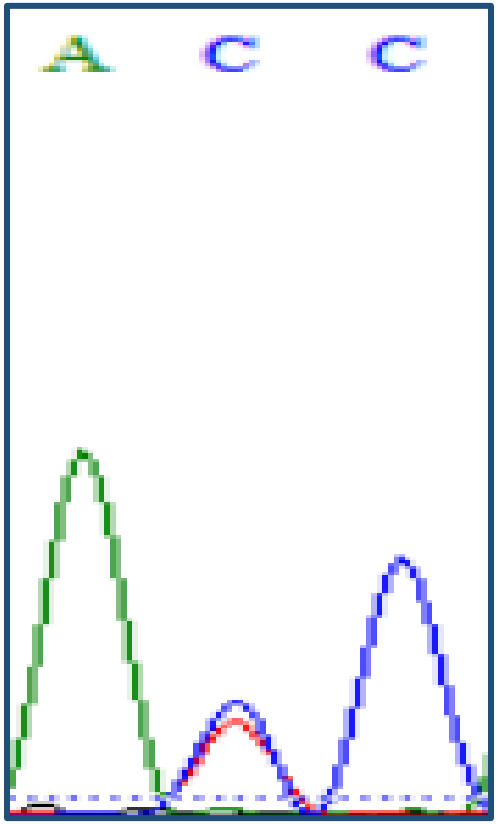


a. Self-compatible

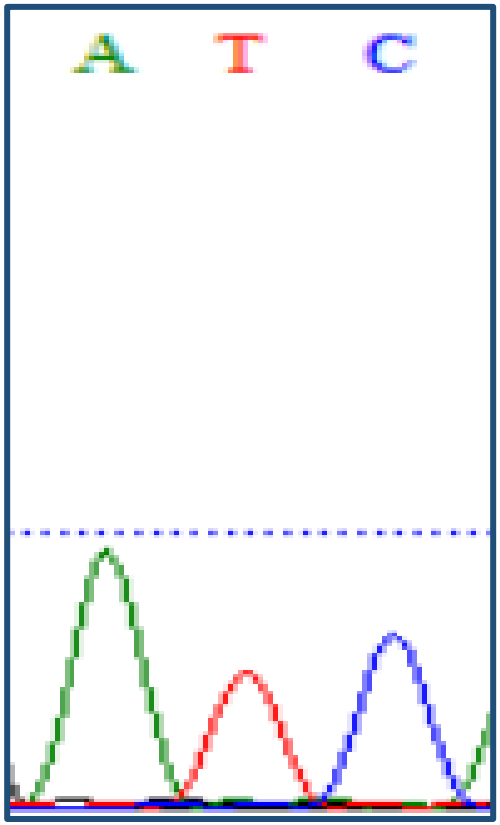


b. Self-incompatible

Figure 39. Base call at SNP 1 (*GEX1*)

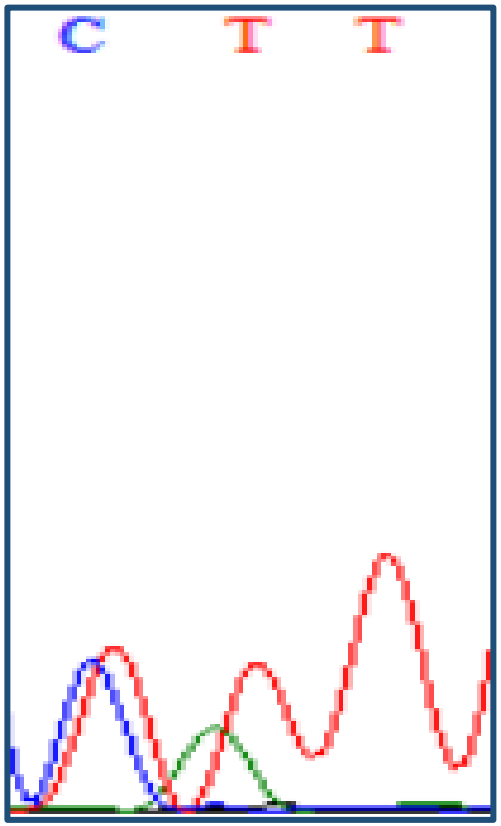


a. Self-compatible

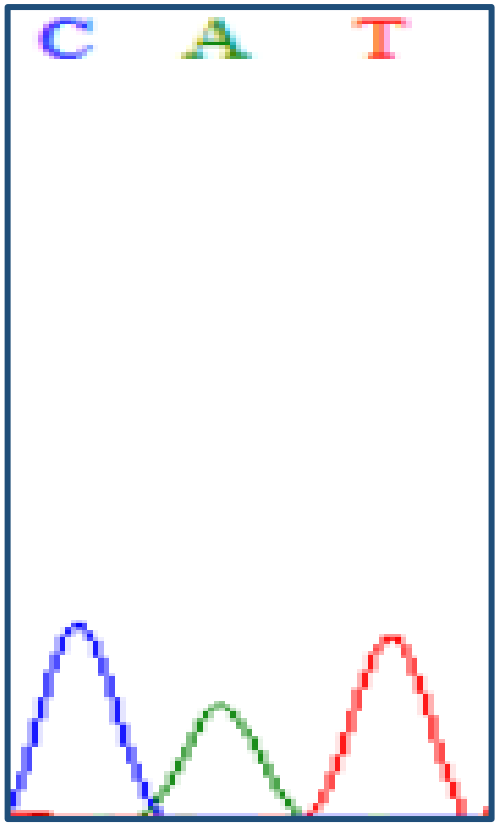


b. Self-incompatible

Figure 40. Base call at SNP 2 (*GEX1*)

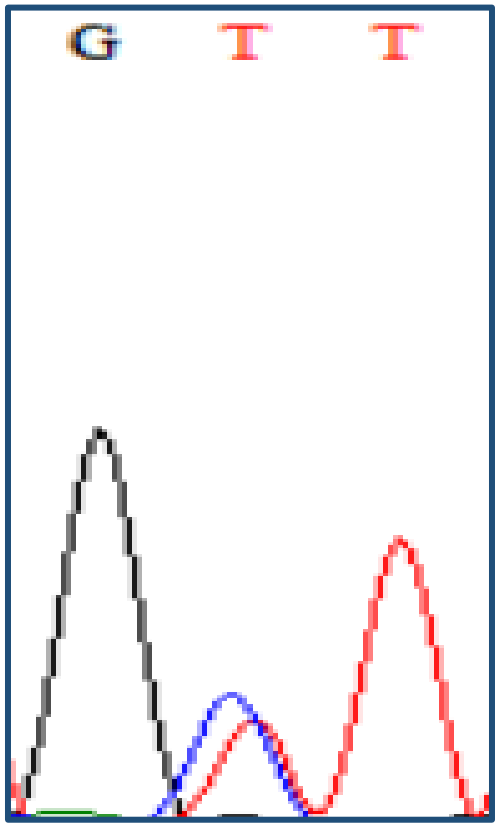


a. Self-compatible

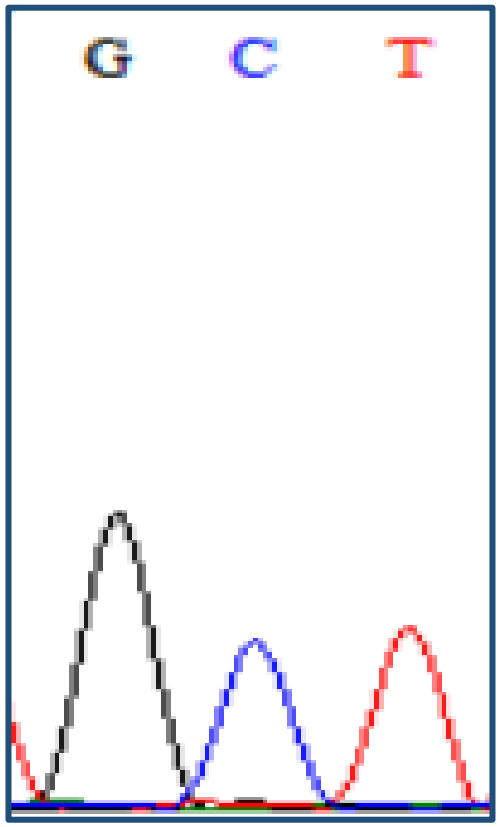


b. Self-incompatible

Figure 41. Base call at SNP 3 (*GEX1*)

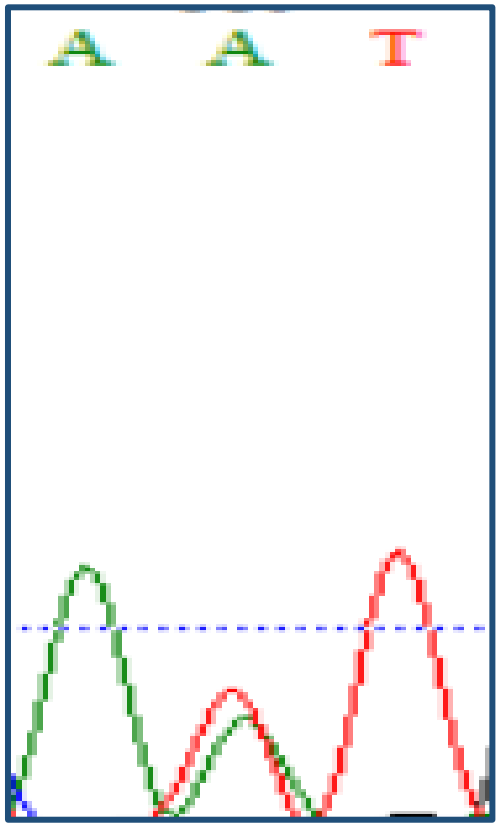


a. Self-compatible

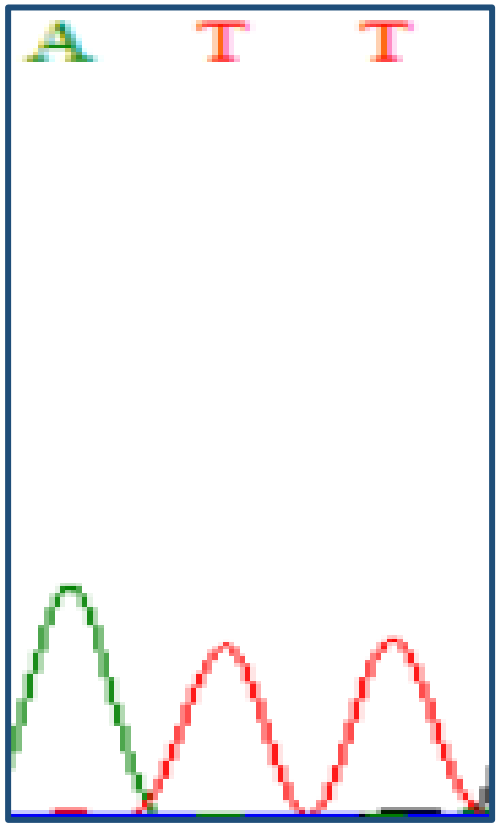


b. Self-incompatible

Figure 42. Base call at SNP 4 (*GEX1*)

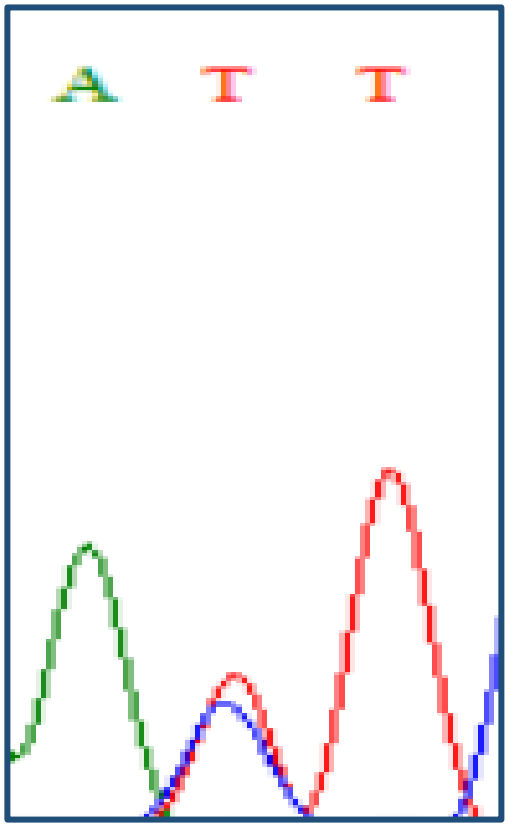


a. Self-compatible

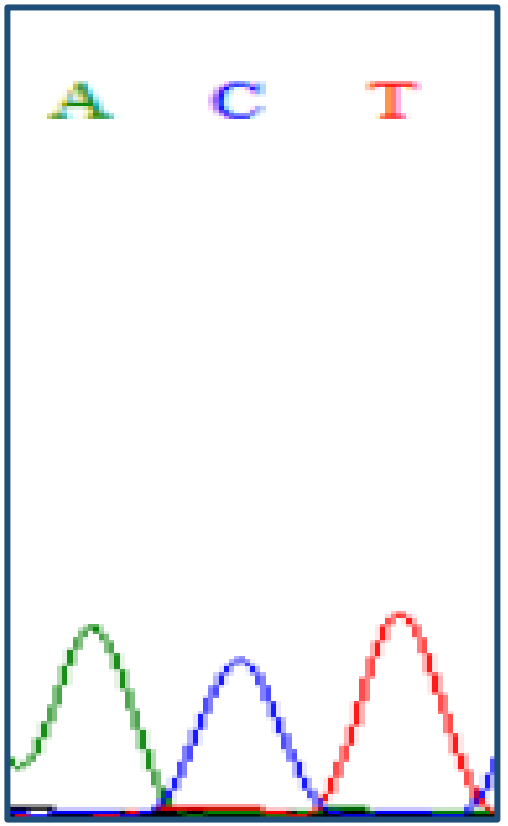


b. Self-incompatible

Figure 43. Base call at SNP 5 (*GEX1*)

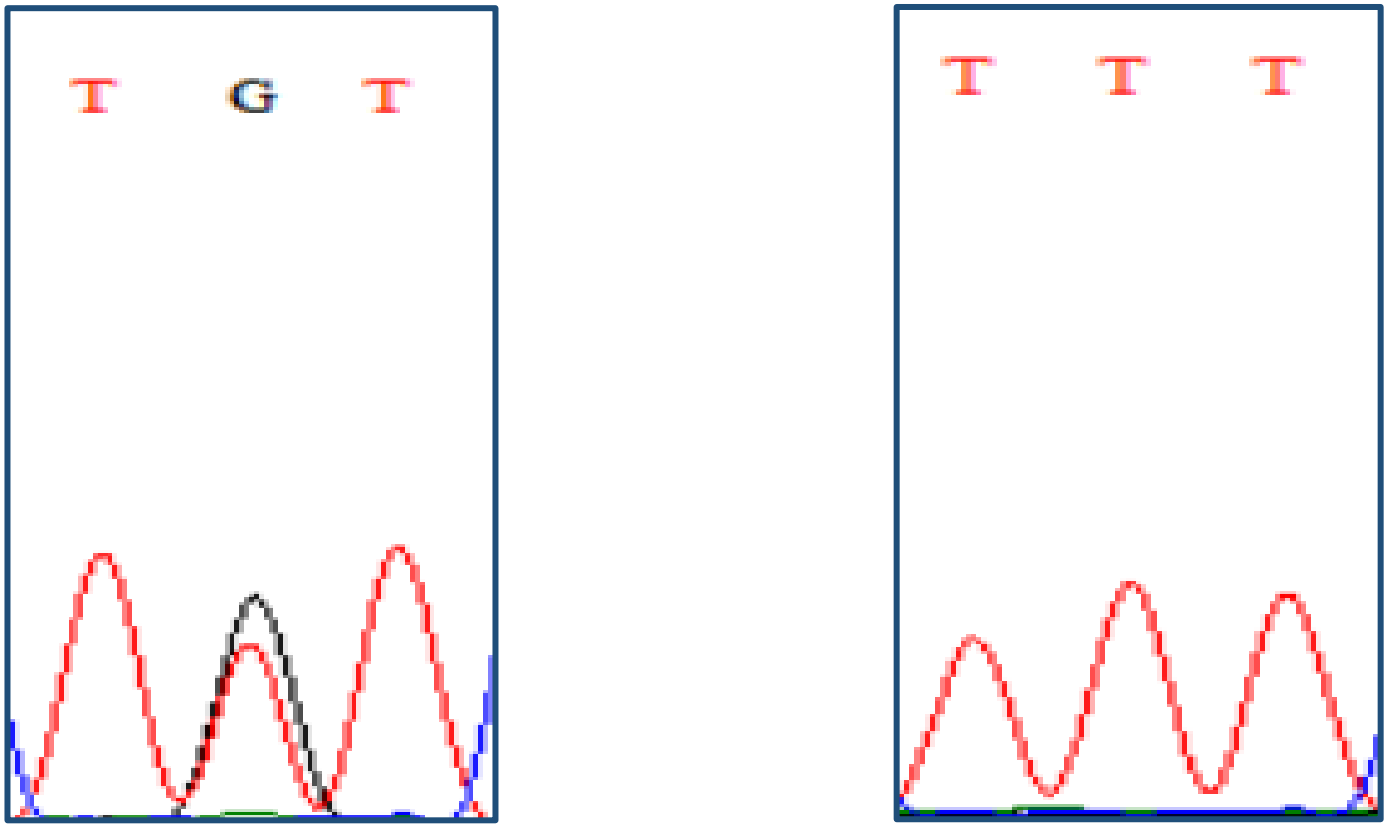


a. Self-compatible



b. Self-incompatible

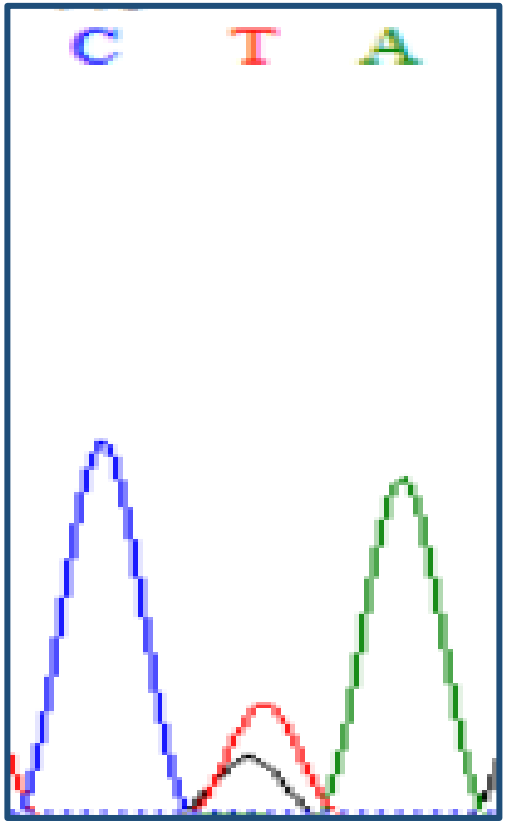
Figure 44. Base call at SNP 6 (*GEX1*)



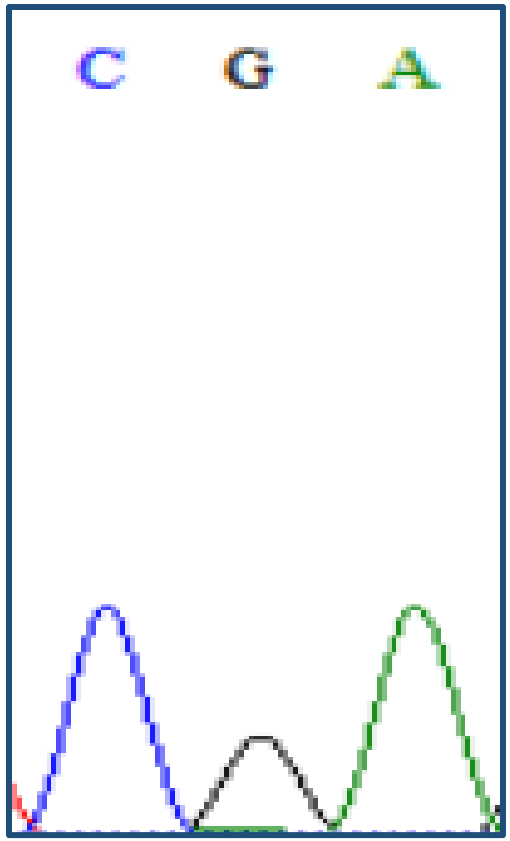
a. Self-compatible

b. Self-incompatible

Figure 45. Base call at SNP 7 (*GEX1*)

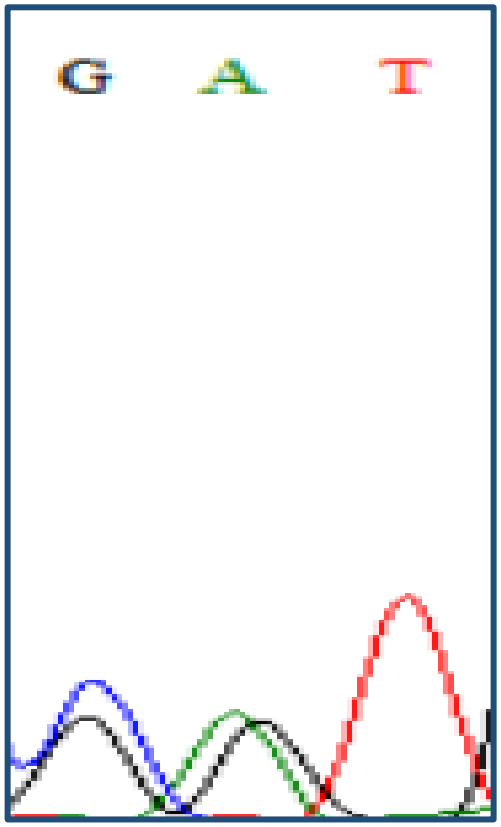


a. Self-compatible

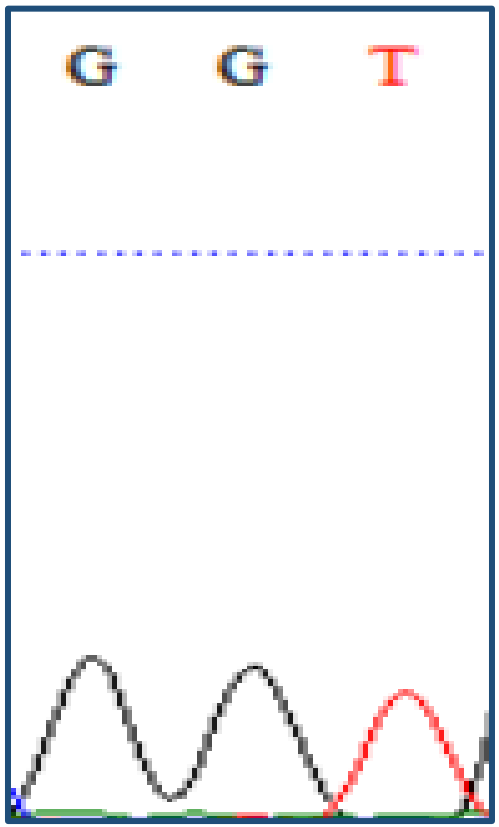


b. Self-incompatible

Figure 46. Base call at SNP 8 (*GEX1*)

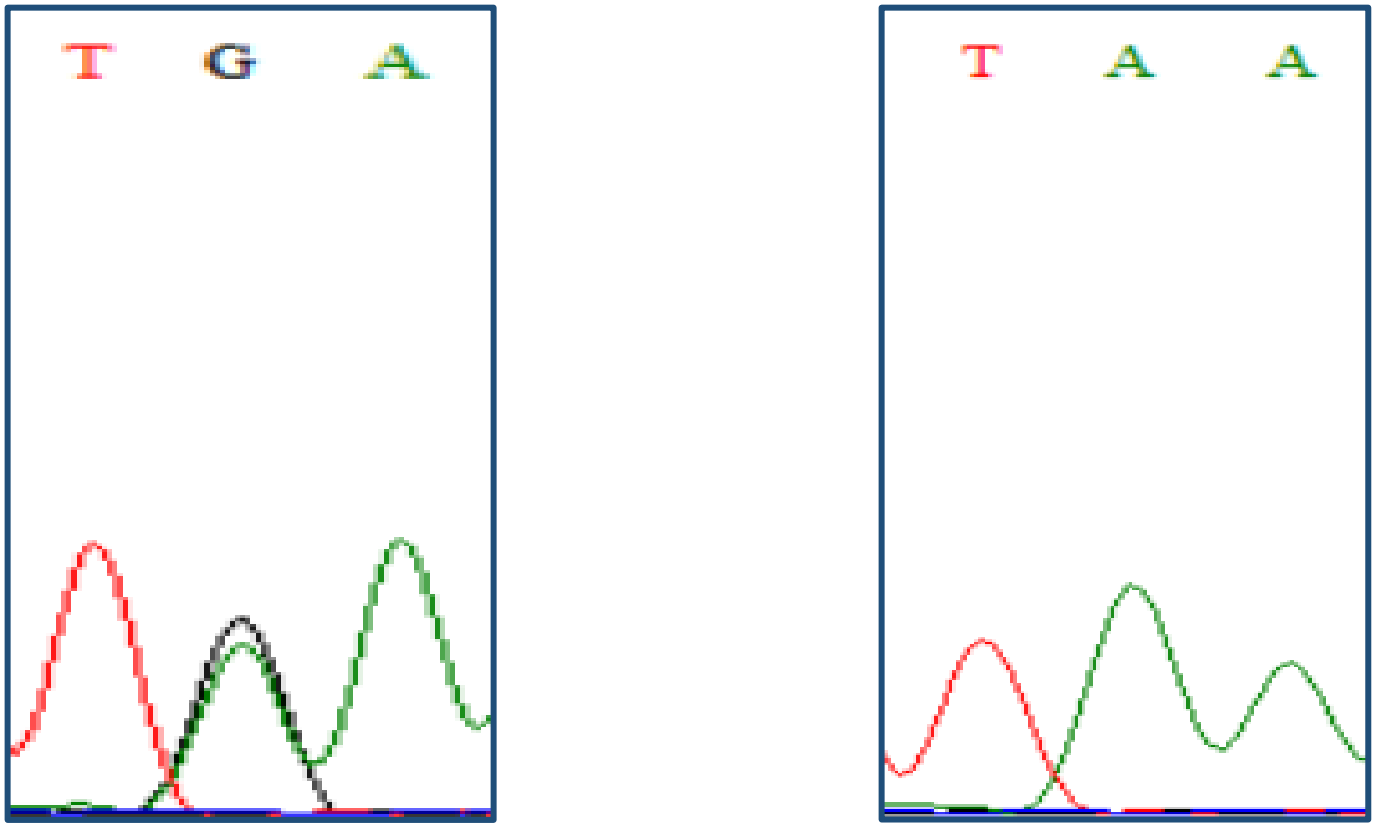


a. Self-compatible



b. Self-incompatible

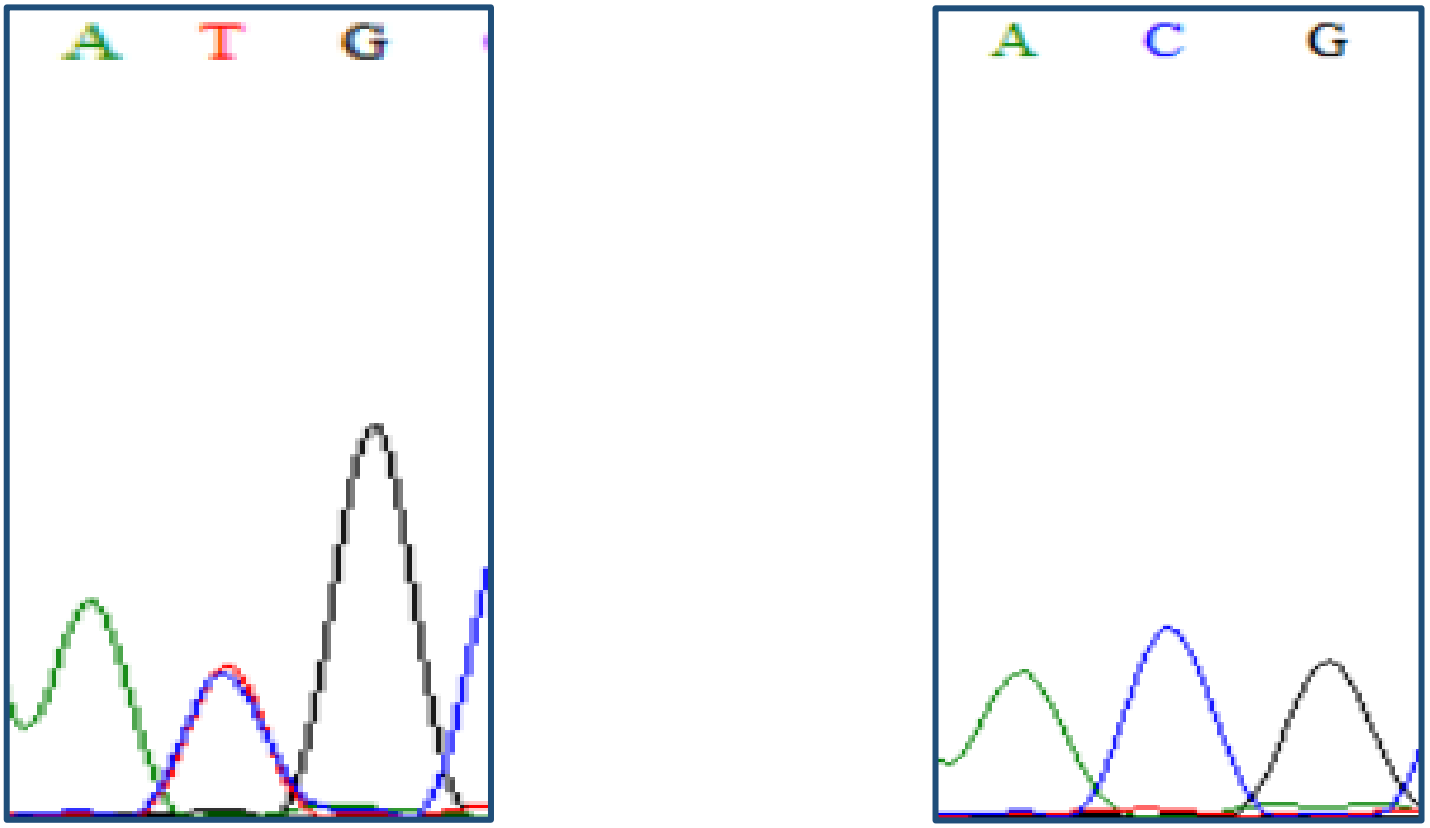
Figure 47. Base call at SNP 9 (*GEX1*)



a. Self-compatible

b. Self-incompatible

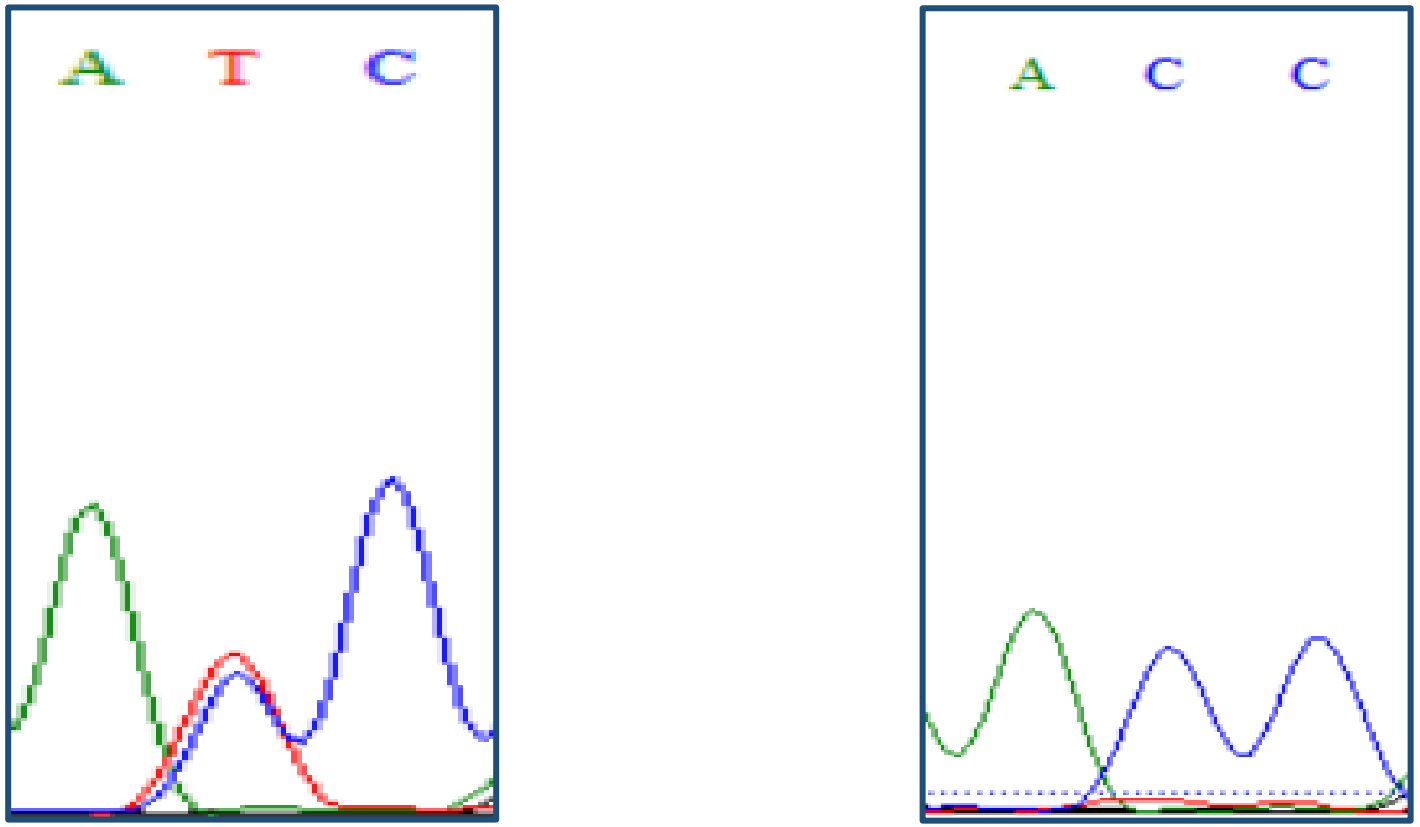
Figure 48. Base call at SNP 10 (*GEX1*)



a. Self-compatible

b. Self-incompatible

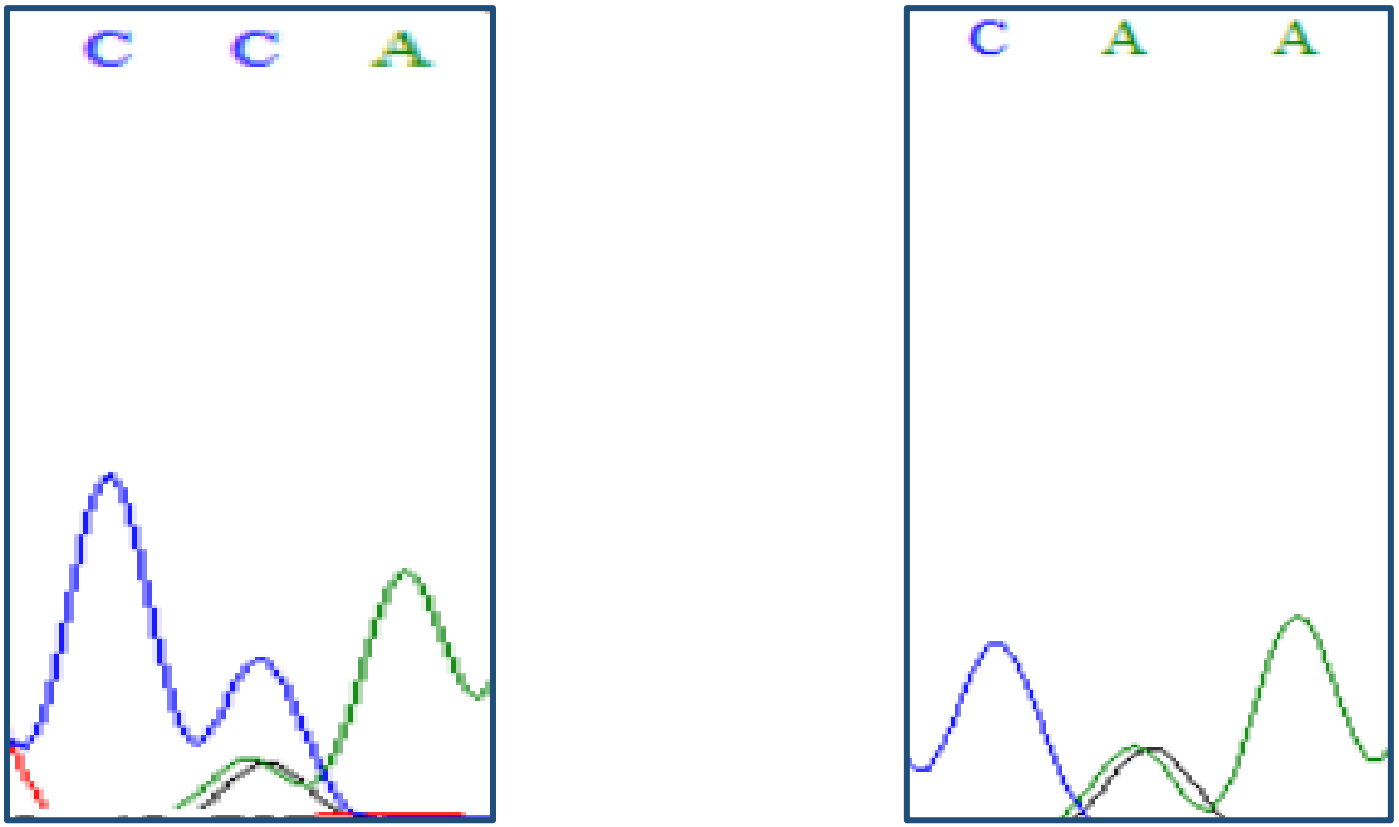
Figure 49. Base call at SNP 11 (*GEX1*)



a. Self-compatible

b. Self-incompatible

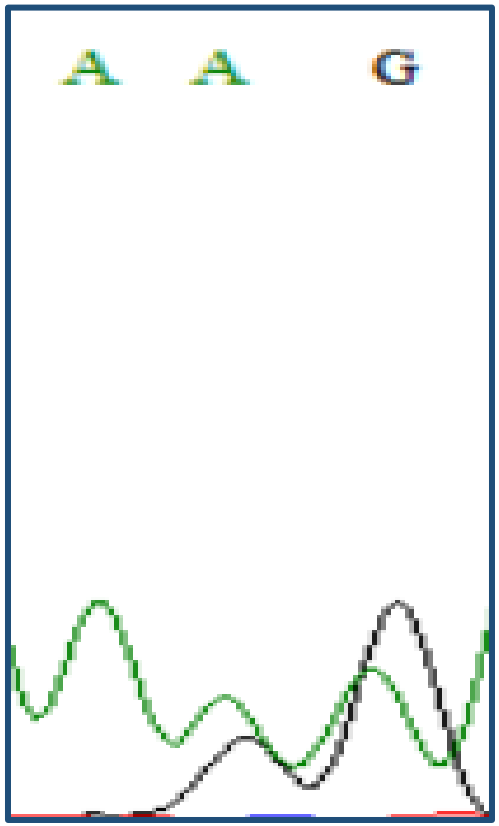
Figure 50. Base call at SNP 12 (*GEX1*)



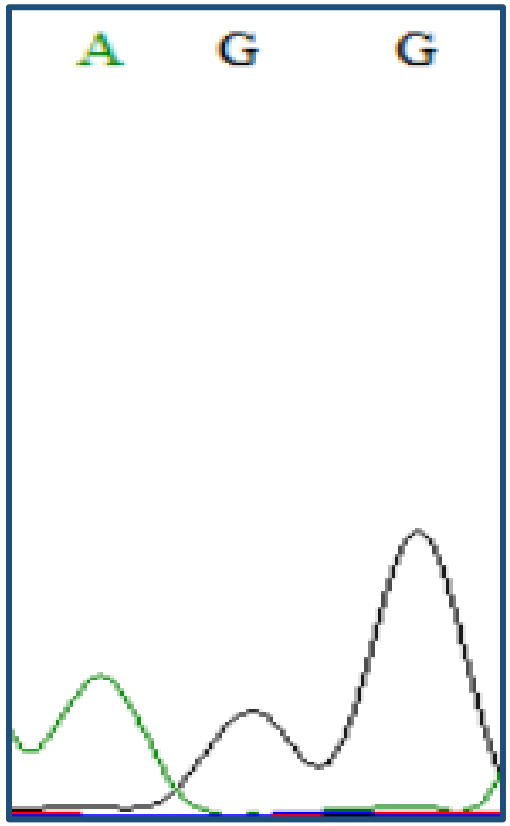
a. Self-compatible

b. Self-incompatible

Figure 51. Base call at Error 1 (*GEX1*)



a. Self-compatible



b. Self-incompatible

Figure 52. Base call at SNP 13 (*GEX1*)

4.4.4.3.8 *Zinc finger AN1 domain-containing stress-associated protein 12 (PMZ)*

4.4.4.3.8.1 Amplification of the primer

The primer combination used to amplify targeted the first 956 bp of the gene covering the entire gene. PCR was run with the annealing temperature of 59.5^{o}C, which was determined by the gradient PCR experiment. The optimum concentration of forward and reverse primers of reaction mixture was found to be 0.4 μM. Upon agarose gel electrophoresis, successful amplification was observed in both the genotypes and a single band in the size of the expected product was obtained, with minimum primer dimers (Table 7). The gel image is presented in Plate 10. The products were sequenced.

4.4.4.3.8.2 Analysis of the sequences

In self-compatible genotype, trimmed sequence of the forward run was of 720 base pairs and had a QV score of 53. The length of the reverse sequence was of 0 base pairs and the QV score was 0. (Table 8).

In self-incompatible genotype, trimmed sequence of the forward run was of 874 base pairs and had a QV score of 55. The length of the reverse sequence was of 32 base pairs and the QV score was 53 (Table 8).

The reverse sequences of both the genes were of poor quality.

4.4.4.3.8.3 Contig assembly

Since the reverse sequence quality was poor for both the genotypes contig was not assembled.

4.4.4.3.8.4 Computational confirmation

BLASTn analysis of the contigs against the cocoa genome database has confirmed that the sequence obtained was exactly of the targeted region (Figure 53).

4.4.4.3.8.5 Multiple sequence alignment

Sequences of self-compatible and self-incompatible genotypes were aligned with the reference sequence (compatible). Three variants were found when the sequences were analyzed. The alignment of the two sequences is presented in Figure 54.

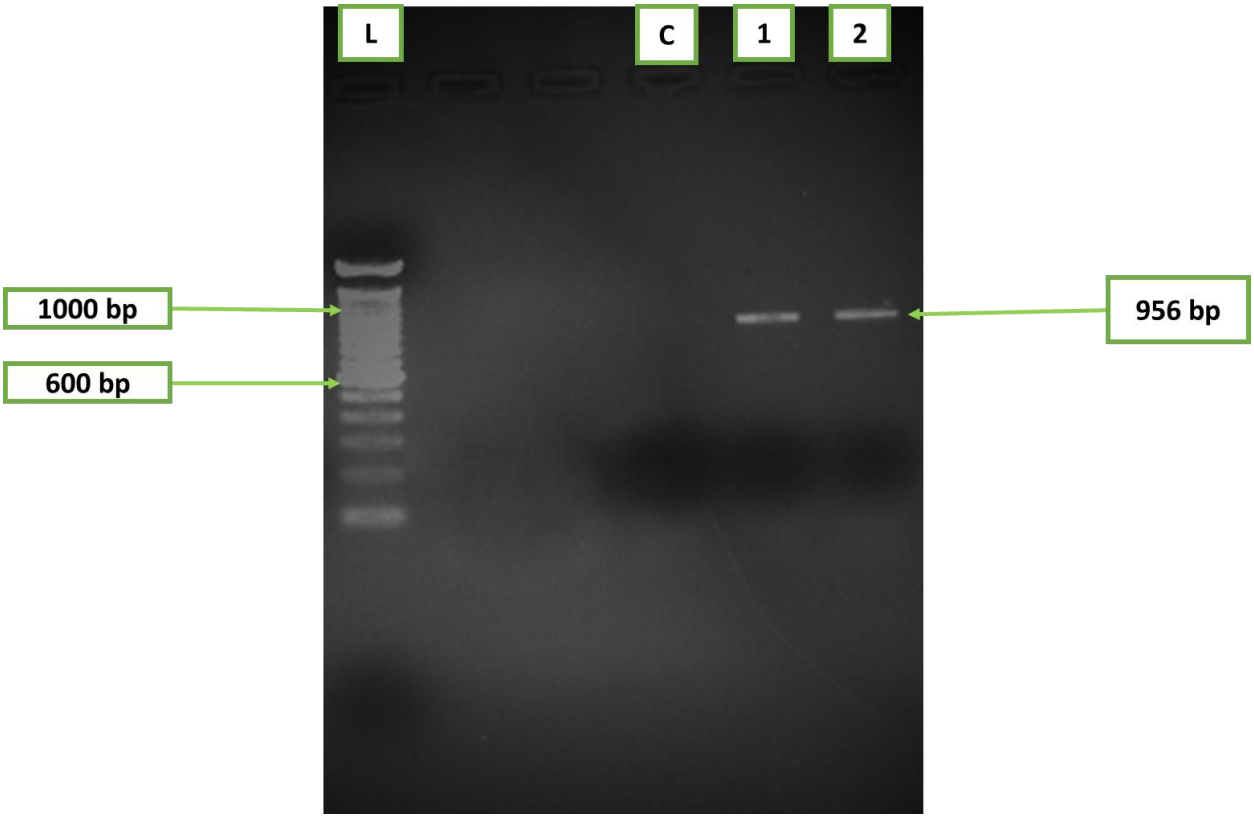


Plate 10. Amplification of *PMZ*

1- Amplification in compatible genotype
2- Amplification in incompatible genotype
C. Control/ Blank

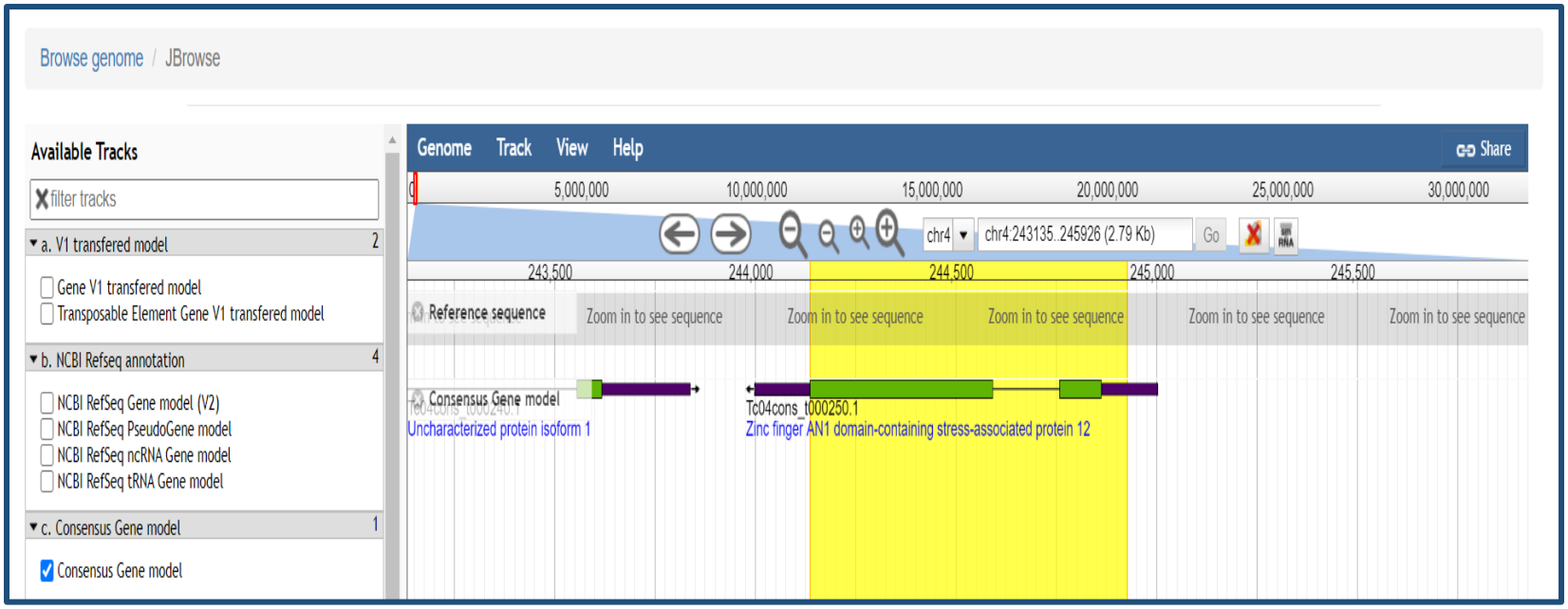


Figure 53. Computational confirmation of *PMZ* region sequenced by BLAST search.

```
CLUSTAL O(1.2.4) multiple sequence alignment


Referencechr4                          -------TCTCTGGCCTCCGCCCAAAACATCCTAAATTCCGATAAATCTTTGTTGACTTT        53
7Compatible_7F_27261-1_P3822,Trimmed   CCTCCCCTCTCTGGCCTCCGCCCAAAACATCCTAAATTCCGATAAATCTTTGTTGACTTT        60
7Incompatible_7F_27261-3_P3822,Trimmed -------TCTCTGGCCTCCGCCCAAAACATCCTAAATTCCGATAAATCTTTGTTGACTTT        53
                                              *****************************************************


Referencechr4                          AAGCATGGGCGGAGGAACAGAAGCTTTTCCTGATTTGGGAAGACATTGCCAACATTCAGA        113
7Compatible_7F_27261-1_P3822,Trimmed   AAGCATGGCTGGAGGAACAGAAGCTTTTCCTGATTTGGGAAGACATTGCCAACATTCAGA        120
7Incompatible_7F_27261-3_P3822,Trimmed AAGCATGGCTGGAGGAACAGAAGCTTTTCCTGATTTGGGAAGACATTGCCAACATTCAGA        113
                                       ********  *****************************************************


Referencechr4                          TTGTCATCAATTAGATTTTCTTCCTTTCAAATGCGACGGTTGTCACAAGGTACATTGCTT        173
7Compatible_7F_27261-1_P3822,Trimmed   TTGTCATCAATTAGATTTTCTTCCTTTCAAATGCGACGGTTGTCACAAGGTACATTGCTT        180
7Incompatible_7F_27261-3_P3822,Trimmed TTGTCATCAATTAGATTTTCTTCCTTTCAAATGCGACGGTTGTCACAAGGTACATTGCTT        173
                                       ************************************************************


Referencechr4                          ATTACTATTGTTTTCTTTCGTAATTCTATTCACATGGATCAAATAAAACGCCTCTGATAG        233
7Compatible_7F_27261-1_P3822,Trimmed   ATTACTATTGTTTTCTTTCGTAATTCTATTCACATGGATCAAATAAAACGCCTCTGATAG        240
7Incompatible_7F_27261-3_P3822,Trimmed ATTACTATTGTTTTCTTTCGTAATTCTATTCACATGGATCAAATAAAACGCCTCTGATAG        233
                                       ************************************************************


Referencechr4                          AAGAAAATCCGGATCACATTAGGAAAAAGAAGGTTACGTGACAATATTTTTCTACTCTTG        293
7Compatible_7F_27261-1_P3822,Trimmed   AAGAAAATCCGGATCACATTAGGAAAAAGAAGGTTACGTGACAATATTTTTCTACTCTTG        300
7Incompatible_7F_27261-3_P3822,Trimmed AAGAAAATCCGGATCACATTAGGAAAAAGAAGGTTACGTGACAATATTTTTCTACTCTTG        293
                                       ************************************************************
```

Figure 54 (a). Multiple sequence alignment of *PMZ*.

Figure 54 (b). Multiple sequence alignment of *PMZ*

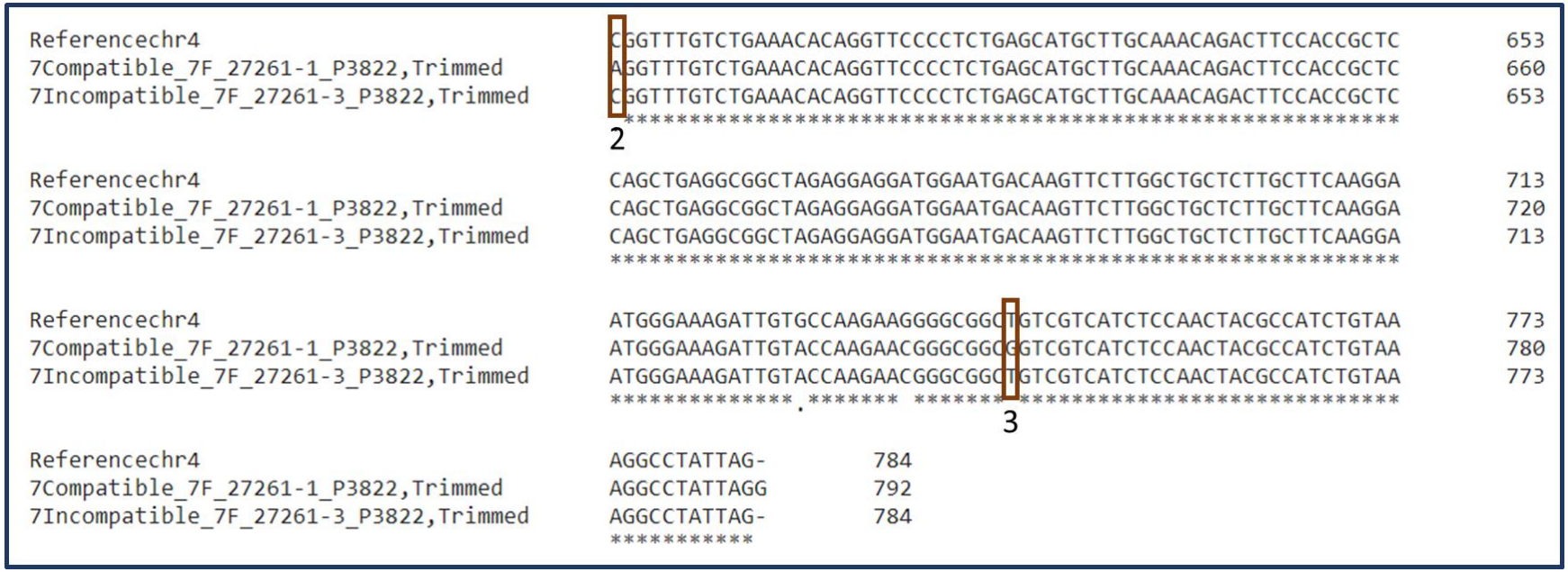


Figure 54 (c). Multiple sequence alignment of *PMZ*

4.4.4.3.8.6 Characterization of SNP

SNP 1

SNP1 locus was heterozygous in both self-compatible genotype and self-incompatible genotype. The loci were heterozygous for cytosine and guanine (C/G). The chromatographic peak present in both the sequence at the self-compatible and self-incompatible locus is shown in Figure 55 (a) and (b) respectively.

SNP1 loci cannot be considered to be linked the self-incompatibility trait because both the alleles were found to be present in both the genotypes, hence considered insignificant.

SNP 2

SNP 2 locus heterozygous in self-compatible genotype and self-incompatible genotype. The loci were heterozygous for both adenine and cytosine (A/C). The chromatographic peak present in both the sequence at the self-compatible and self-incompatible locus is shown in Figure 56 (a) and (b) respectively.

SNP 2 locus cannot be considered to be linked to the self-incompatibility trait because both the alleles were found to be present in both the genotypes. Hence SNP 2 is considered to be insignificant.

SNP 3

When the chromatogram of the SNP 3 loci was examined, it was confirmed that SNP3 locus was heterozygous in both self-compatiblegenotype and self-incompatiblegenotype. The loci were heterozygous for thymine and guanine (T/G). The chromatographic peak present in both the sequence at the self-compatibleand self-incompatiblelocus is shown in Figure 57 (a) and (b) respectively.

SNP 3 loci cannot be considered to be linked the self-incompatibility trait because both the alleles were found to be present in both the genotypes, hence considered insignificant.

In cocoa, Lanaud (2017) reported Tc04_g000320 to be orthologous to a Zinc finger AN1 domain-containing stress-associated protein 12 (PMZ) gene from *A. thaliana*. In SC reactions, it was expressed at a slightly higher at 12-24 hours after pollination. *PMZ* could interact with several other proteins which are involved in Ubiquitin mediated protein degradation.

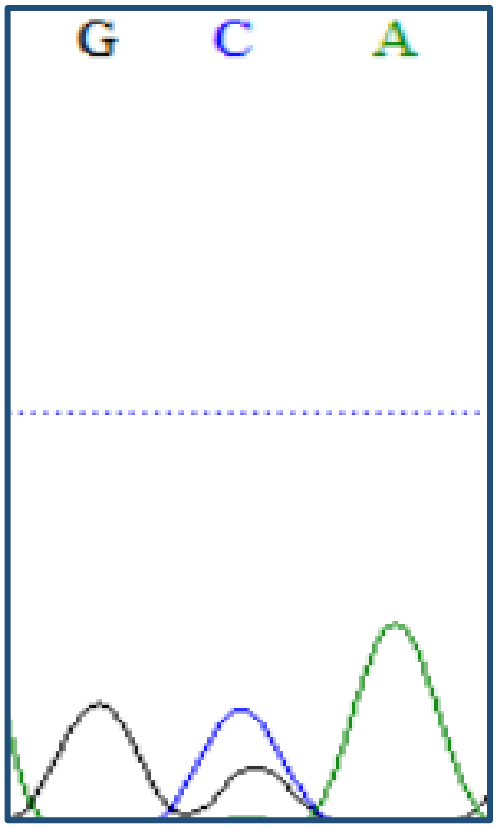
Zinc Finger proteins (ZnF) are a protein super family that regulates growth and development processes in plants and also in conferring a role in biotic and abiotic stress responses (Giri *et al.* 2011). Gupta *et al.* (2012) reported that ZnF domain play a key part in pathogen-host interactions and this is very much similar to the pistil-pollen interactions observed (Hodgkin *et al.,* 1988).

In almond, Gómez *et al.* (2019) identified candidate genes for gametophytic self-incompatibility components through a transcriptomal study on pollen pistil interactions. They reported a Zinc finger (ZnF) protein constant-like 12-like isoform protein to be one of the important candidates for SI in almond.

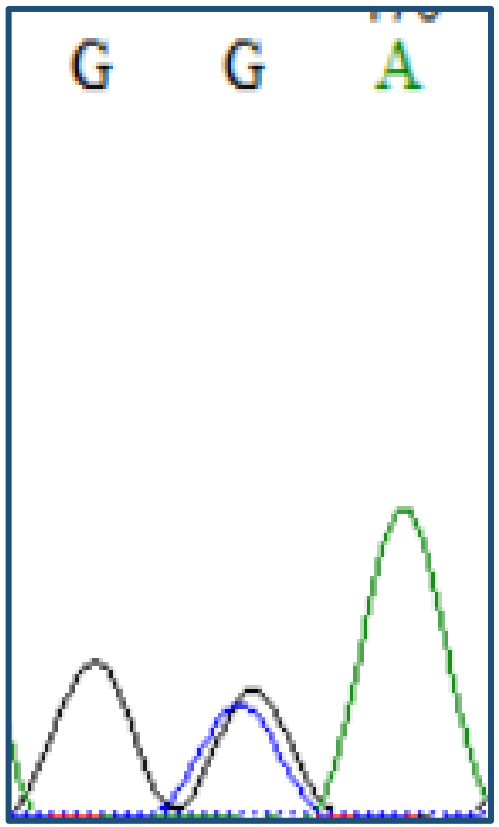
However, in the present study, all the SNP loci identified within this region were found to be heterozygous and common in both, self-compatible genotype and the self-incompatible genotype. As a result, this region is considered not to be associated with the self-incompatibility in cocoa. Hence the three SNPs discovered in this region can be considered insignificant.

Table 14. List of SNPs and base called at their respective loci (*PMZ*)

SI. No	**Self-Compatible**	**Self-incompatible**	**Reference**
SNP - 1	C/G	G/C	G
SNP - 2	A/C	C/A	C
SNP - 3	G/T	T/G	T

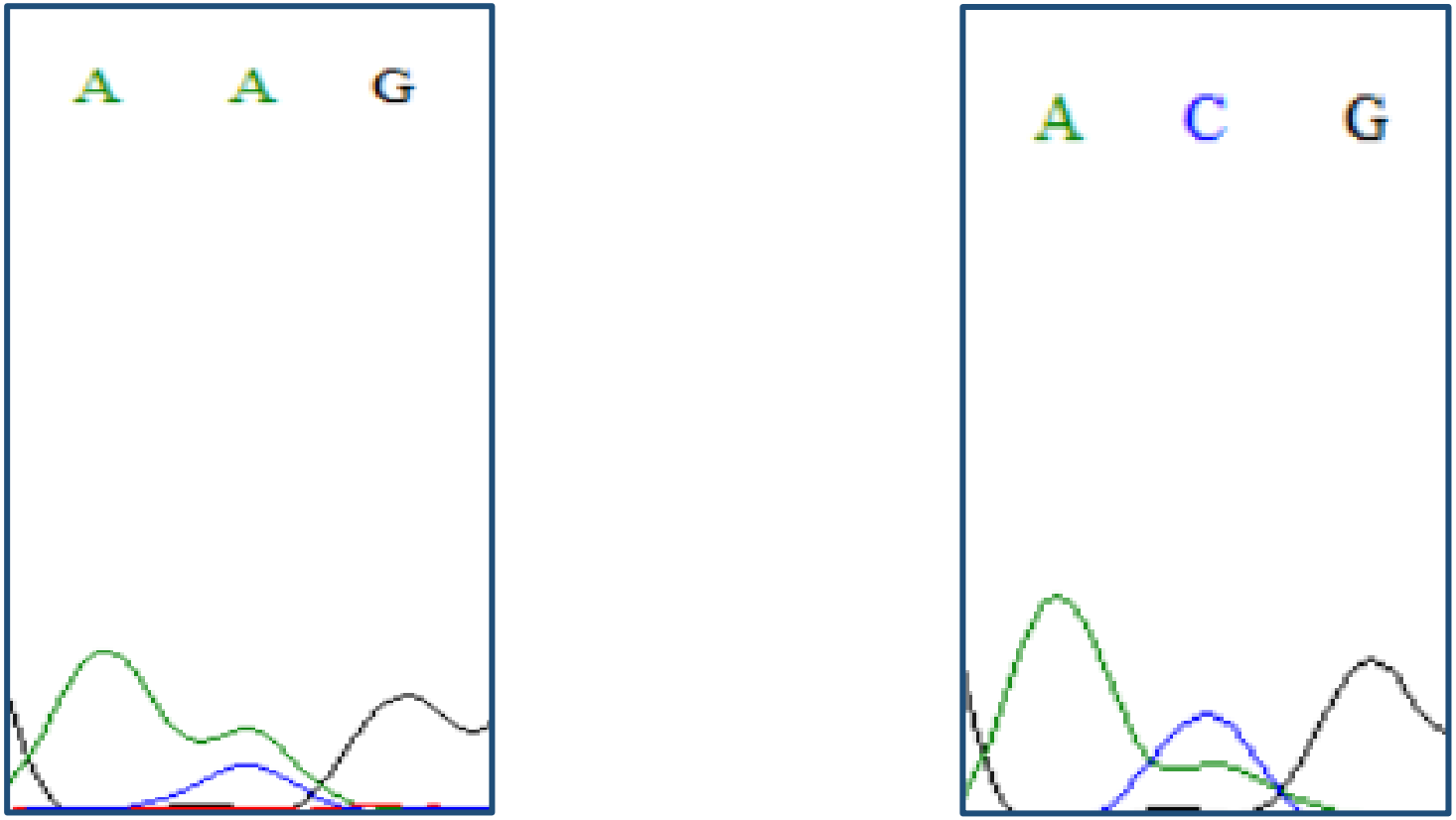


a. Self-compatible



b. Self-incompatible

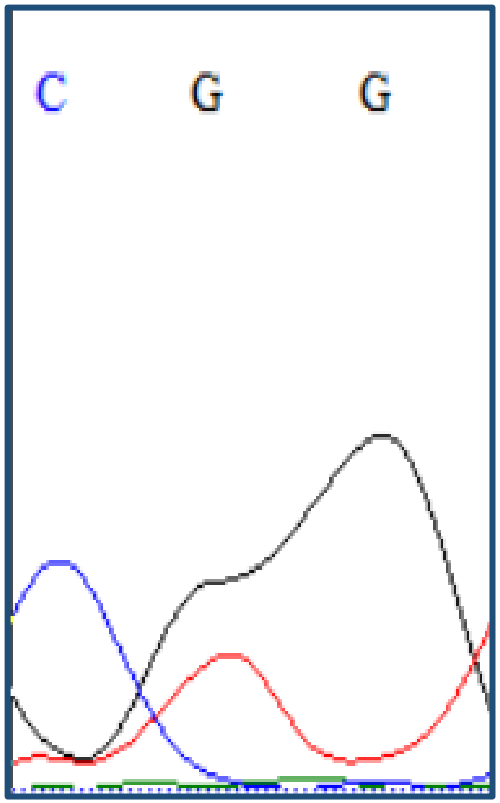
Figure 55. Base call at SNP 1(*PMZ*)



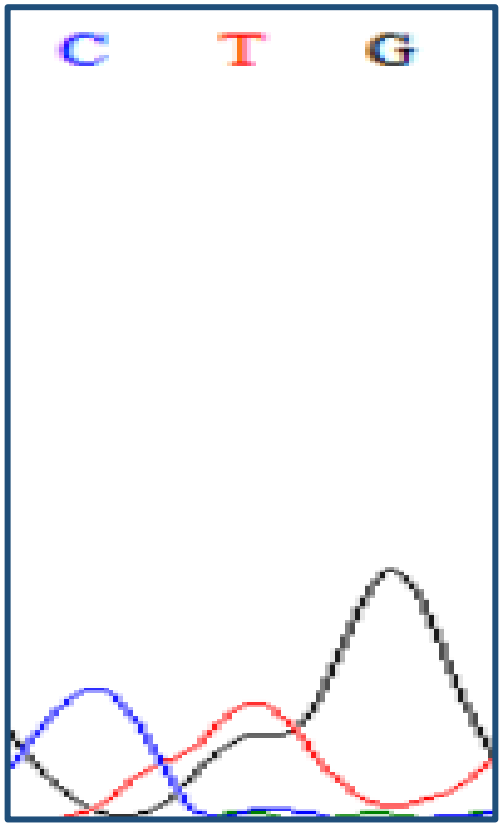
a. Self-compatible

b. Self-incompatible

Figure 56. Base call at SNP 2 (*PMZ*)



a. Self-compatible



b. Self-incompatible

Figure 57. Base call at SNP 3 (*PMZ*)

4.4.4.3.9 *ARM repeat-containing protein (ARC1)*

4.4.4.3.9.1 Amplification of the primer

The primer combination used to amplify targeted the first 1012 bp of the gene, targeting exon 1, the largest exon of the gene. PCR was run with the annealing temperature of 57.1^{o}C, which was determined by the gradient PCR experiment. The optimum concentration of forward and reverse primers of reaction mixture was found to be 0.4 μM. Upon agarose gel electrophoresis, successful amplification was observed in both the genotypes and a single band in the size of the expected product size was obtained, with minimum primer dimers (Table 7). The gel image is portrayed in Plate 11. The PCR products were sequenced.

4.4.4.3.9.2 Analysis of the sequences

In self-compatible genotype, trimmed sequence of the forward run was of 953 base pairs and had a QV score of 54. The length of the reverse sequence was of 883 base pairs and the QV score was 51 (Table 8).

In self-incompatible genotype, trimmed sequence of the forward run was of 864 base pairs and had a QV score of 53. The length of the reverse sequence was of 871 base pairs and the QV score was 53 (Table 8).

4.4.4.3.9.3 Contig assembly

The contig obtained after assembling the forward and reverse sequence of the *ARC1* primer was 964 base pairs long for the self-compatible genotype and 912 base pairs long for the self-incompatible genotype (Table 8).

4.4.4.3.9.4 Computational confirmation

BLASTn analysis of the contigs against the cocoa genome database has confirmed that the sequence obtained was exactly of the targeted region (Figure 58).

4.4.4.3.9.5 Multiple sequence alignment

Sequences of self-compatible and self-incompatible genotypes were aligned with the reference sequence (compatible). No variation between the sequences of self-compatible and the self-incompatible genotypes were found. The alignment of the three sequences is presented in Figure 59.

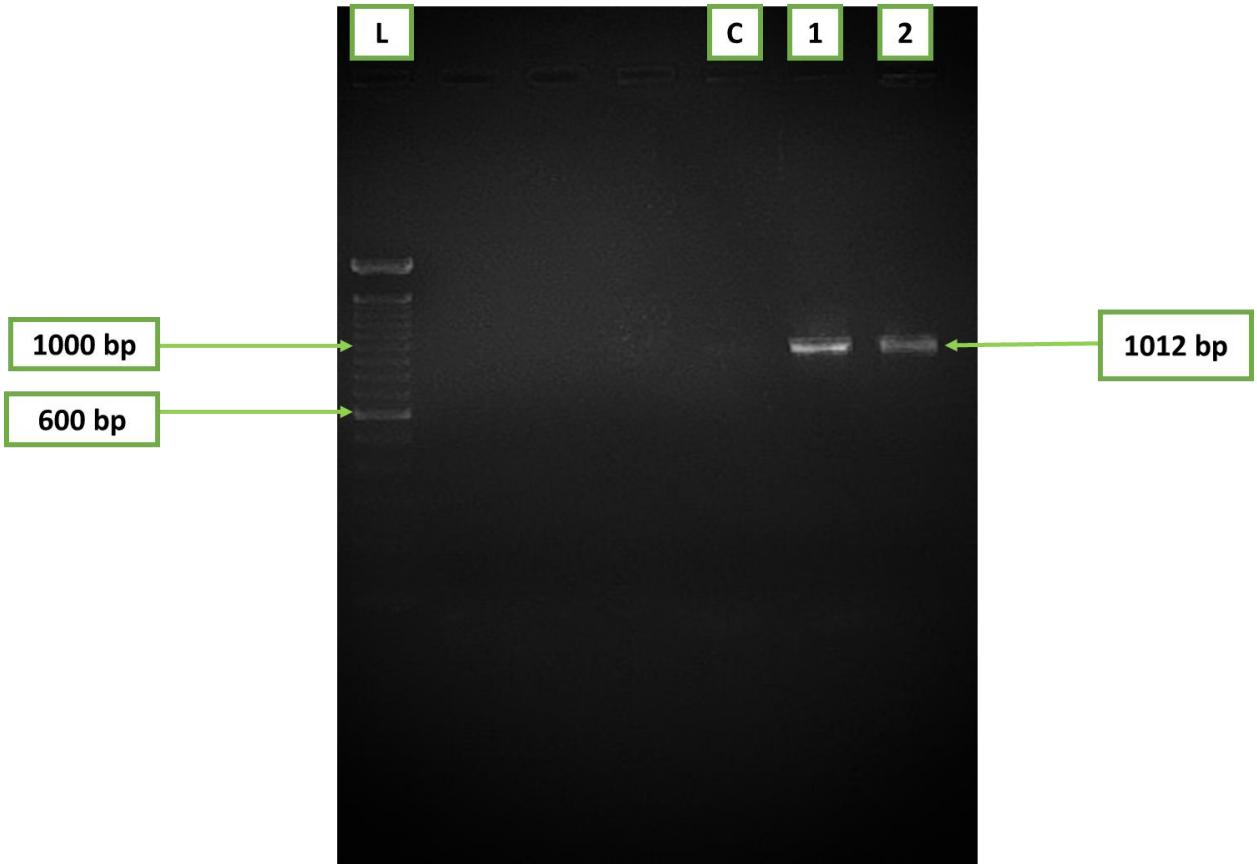


Plate 11. Amplification of *ARC1*

1- Amplification in compatible genotype
2- Amplification in incompatible genotype
C. Control/ Blank

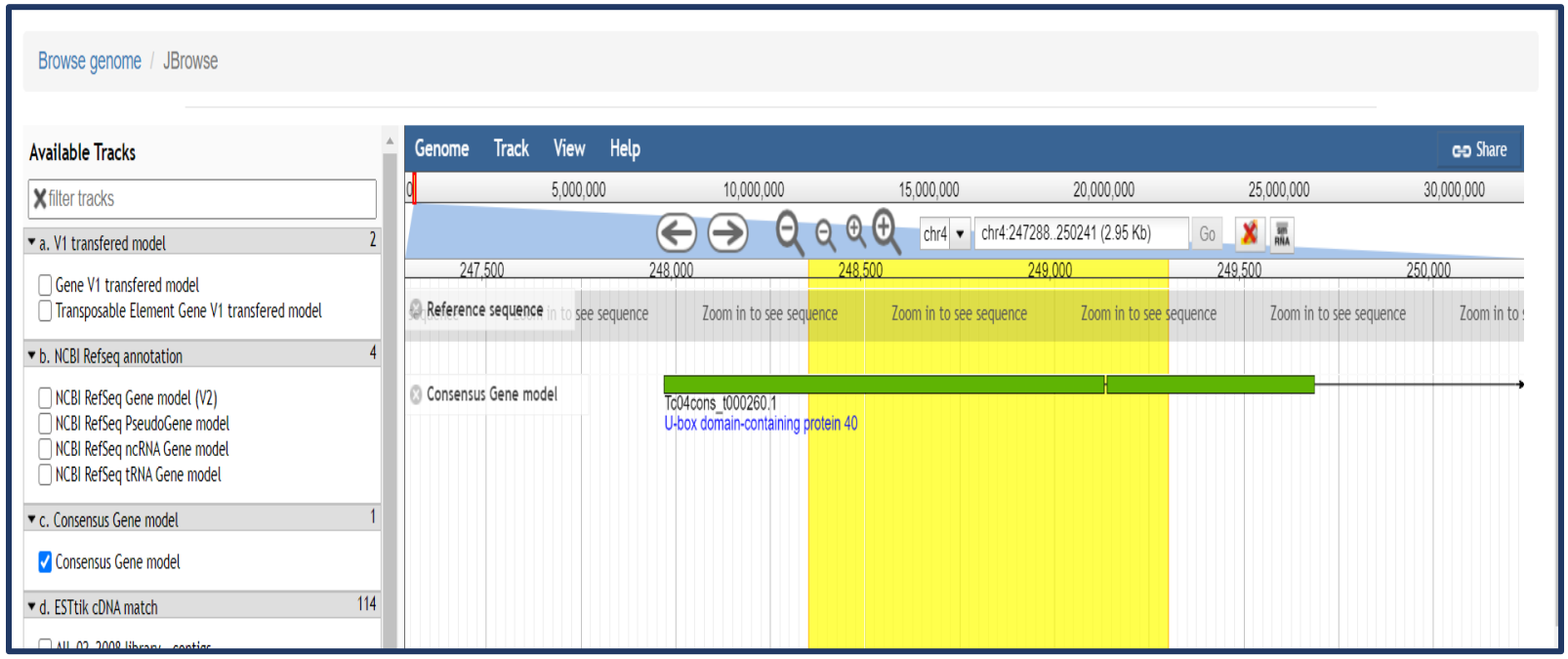


Figure 58. Computational confirmation of *ARC1* region sequenced by BLAST search.

```
8b_8F_27261-7_P3822,Trimmed      --TT-CTCACTGGTGCCAAAAACATTCCTTAAATCCTCCAAAACCCCTGGATTTCTCCAC   57
Reference                        CATTCTCAACTGGTGCCAAAAACATTCCTTAAATCCTCCAAAACCCCTGGATTTCTCCAC   60
8a_8F_27261-5_P3822,Trimmed      -----CTCACTGGTGCCAAAAACATTCCTTAAATCCTCCAAAACCCCTGGATTTCTCCAC   55
                                       .*************************************************

8b_8F_27261-7_P3822,Trimmed      AGCAGAGGAACTTGTCCGTACATTAATAGCCAAAAATCCAAATACCCTGATCAAAAAAGA   117
Reference                        AGCAGAGGAACTTGTCCGTACATTAATAGCCAAAAATCCAAATACCCTGATCAAAAAAGA   120
8a_8F_27261-5_P3822,Trimmed      AGCAGAGGAACTTGTCCGTACATTAATAGCCAAAAATCCAAATACCCTGATCAAAAAAGA   115
                                 ************************************************************

8b_8F_27261-7_P3822,Trimmed      ACAGGAAAAGGTGTCTAATTCAGAGAAAGAGTTGATTCAAGGAGTTGAAGATATTCCTTC   177
Reference                        ACAGGAAAAGGTGTCTAATTCAGAGAAAGAGTTGATTCAAGGAGTTGAAGATATTCCTTC   180
8a_8F_27261-5_P3822,Trimmed      ACAGGAAAAGGTGTCTAATTCAGAGAAAGAGTTGATTCAAGGAGTTGAAGATATTCCTTC   175
                                 ************************************************************

8b_8F_27261-7_P3822,Trimmed      AGTGAAATTCGACCATGCAGCCACTCAGTTAAGTCGGAGGCCTCCTCAATTCCACTCGAG   237
Reference                        AGTGAAATTCGACCATGCAGCCACTCAGTTAAGTCGGAGGCCTCCTCAATTCCACTCGAG   240
8a_8F_27261-5_P3822,Trimmed      AGTGAAATTCGACCATGCAGCCACTCAGTTAAGTCGGAGGCCTCCTCAATTCCACTCGAG   235
                                 ************************************************************

8b_8F_27261-7_P3822,Trimmed      CTCGGAGGAATCGATTGCTGCTGCTGACGCCGCTGCCGCCGCCACCAACGTTCTAACACC   297
Reference                        CTCGGAGGAATCGATTGCTGCTGCTGACGCCGCTGCCGCCGCCACCAACGTTCTAACACC   300
8a_8F_27261-5_P3822,Trimmed      CTCGGAGGAATCGATTGCTGCTGCTGACGCCGCTGCCGCCGCCACCAACGTTCTAACACC   295
                                 ************************************************************
```

Figure 59 (a). Multiple sequence alignment of *ARC1*.

Figure 59 (b). Multiple sequence alignment of *ARC1*

ARC1 gene in *Brassica napa* is known to be required for the rejection of self-incompatiblepollen by the pistil. It takes part in the downstream reactions of the SRK. Stone *et al.* (2003) showed that the *ARC1* possess the activity of the E3 ubiquitin ligase in the presence of an active SRK. They found that in case of incompatibility, levels of ubiquinated proteins were found to be high. They proposed that *ARC1* is responsible for ubiquitin mediated proteasomal degradation of various compatibility factors which leads to pollen rejection.

Through transgenic RNA interference knock down experiment, the need of ARC1 in the rejection of incompatible pollen was confirmed in *A. lyrata* (Indriolo *et al.,* 2012). A gain of function experiment conducted in *A. thaliana ARC1* mutants, it was proved that *ARC1* played a critical role in reconstitution of self-incompatibility (Indriolo *et al.*, 2014). However, in this study, no variations were discovered when the self-compatible genotype sequence and the self-incompatible genotype sequence were aligned. The variation may be underlying in some other part of the gene, which has to be sequenced and examined in detail.

4.4.4.3.10 *Hapless 2*

Similar to *SLG* no amplification was obtained using the *Hapless 2* primer. Wide range of temperatures and different concentration of the PCR components were tried. However, no successful amplification was obtained.

4.4.4.4 Potential candidate SNPs that could be linked to self-incompatibility

SNPs that are heterozygotic in the self-incompatible genotype, homozygotic in the self-compatible genotype, with the reference genome (compatible) identical to the base present in the homozygotic self-compatible genotype loci are potential candidates that can be linked to self-incompatibility trait. Seven such SNPs that have the potential to be linked to the trait have been discovered in this study. Two SNPs (SNP 2 and SNP 3) were discovered in *SRK*, one (SNP 1) in *WDR5a* and four SNPs (SNP 1, SNP 2, SNP 3, SNP 4) were discovered in *Alpha-1F*. SNPs potentially linked to the trait self incompatibly are presented in Table 15.

Table 15. Potential SNPs that could be linked to self-incompatibility

Sl. No.	Gene	SNP#	Self-compatible	Reference	Self-incompatible
1	***SRK***	SNP2	G	G	G **A**
2	***SRK***	SNP3	G	G	G **A**
3	***WDR***	SNP1	G	G	G **T**
4	***Alpha-1f***	SNP1	A	A	A **C**
5	***Alpha-1f***	SNP2	G	G	G **T**
6	***Alpha-1f***	SNP3	T	T	T **A**
7	***Alpha-1f***	SNP4	G	G	G **C**

Cai *et al.* (2015) developed an SNP based marker in rice that can be used to screen genotypes for their grain amylose content in rice. A single G to T transversion in the first intron of *waxy* gene was found to decrease the amylose content in rice. CTPP based markers were developed which can be used in various molecular breeding programs.

In sugarbeet, Bakooie *et al.* (2018) developed a SNP based marker based on a locus on the gene *R6m-1* which can be used to differentiate root knot nematode resistant genotypes from susceptible genotypes. This marker was found to be very useful for marker assisted breeding programs for resistance to root knot nematode.

In rye grass, a gametophyic self-incompatible species, Shinozuka (2010), conducted physical mapping and discovered nine genes that are present in the Z locus. The conducted nucleotide diversity assessment between self-compatible and self-incompatible genotypes and identified two potential genes, *LpTC116908* and *LpDUF247*, as plausible candidates for the male and female determinants of the S-Z SI system.

From the present study, SNP based markers, such as Confronted Two Primer PCR markers can be designed such that the 3' end of a primer falls complementary to the base on the allele that is present only in the allele in self-incompatible genotype, at the same time absent in the self-compatible genotypes allele at this locus (Table 15) (Waters *et al.,* 2008). In such a condition, amplification will be observed only in the self-incompatible genotype. The marker can be validated by screening a larger population of self-compatible and self-incompatible genotypes.

4.4.5 Designing of CTTP primers for the potential markers

Polymerase chain reaction with confronting two-pair primers (PCR-CTPP) is an inexpensive, time-saving and reliable method for biallelic genotyping (Hamajima *et al*. 2000, 2002; Hamajima, 2001). It is applicable to single nucleotide polymorphisms (SNPs) where an appropriate restriction enzyme is not available (where CAPS can't be developed), as well as to insertion or deletion polymorphisms.

Prim-SNPing is a web based open source software by which Confronted Two Primer PCR with optimum conditions can be developed quickly and easily (Chang *et al.,* 2009). The advantage of CTPP primer designing is that there requires no additional step of restriction digestion, like in the case of CAPS. Using Prim-Snping, with the help of nucleotides (200-300bp) flanking on both sides of the SNP loci efficient primers can be designed.

The same can be explained with the example of SNP 3 in *Alpha-1F*. In the gene *Alpha-1F,* SNP 3 was homozygous for thymine (T) in self-compatible genotype and the reference (self-compatible) sequence (Table 15). Self-incompatible locus was heterozygotic for thymine and adenine (T/A). Primer can be designed such that the 3'end of the primer binds and amplifies the allele in self-compatible that has adenine (A) at the locus. Whereas the allele with thymine (T) does not get amplified. Hence amplification will be only observed in self-incompatible genotype. Using Prim-Snping software, SNP based primers can be designed. 300 nucleotides flanking both sides of the SNP loci must be given as input and the SNP will be given in square brackets in the middle (NNNNNNNNN[T/A]NNNNNNNN). The software directly

provides the primer sequences. The sequence of the forward and reverse primers of SNP 3 of *Alpha-1F* are given below.

Forward Primer: 5' **GACAAGAGATTGGCATGAAAAATAA** 3'

Reverse Primer: 5' **CTTTCATGCTGCAAAATTTTCTGT** 3'

Such primers can be designed for all the seven potential SNPs identified across the gene in this study. These primers can be validated in a large population of self-incompatible and self-compatible genotypes.

Summary

5. Summary

The study entitled 'Candidate gene analysis on self-incompatibility in cocoa (*Theobroma cacao* L.)' was conducted in the Department of Plant Biotechnology, College of Horticulture, Vellanikkara, during, 2018-2020. The objective of the study was to analyze the sequence variations in putative genes that contribute to self-incompatibility in Forestero cocoa maintained at Cocoa Research Center.

The results of the study are as follows:

- Doyle and Doyle method with small modification was found to be optimum for DNA isolation from lemon coloured young cocoa leaves.

- Spectrophotometer NanoDrop® ND-1000 was used for the quantification of isolated DNA. The UV absorbance ratio ranged from 1.80 – 2.0, which indicated the presence of good quality DNA suitable for molecular characterization.

- Ten candidate genes were selected for this study *viz*., *Serine Receptor Kinase (SRK), S locus glycoprotein (SLG), Barely Any Meristem 1* (*BAM1*), *Barely Any Meristem 2* (*BAM2*), *COMPASS-like H3K4 histone methylase component (WDR5a), Voltage-dependent L-type calcium channel subunit* (*Alpha-1F*), *Gamete Expressed Protein (GEX1), Zinc finger AN1 domain-containing stress-associated protein 12 (PMZ), ARM repeat-containing protein* (*ARC1*) and *Hapless 2.*

- Genomic nucleotide sequences of the candidate genes from reported host plant species were retrieved from the NCBI GenBank database. Using this sequence information, homologous gene sequences of the candidate genes present in cocoa were retrieved from the cocoa whole genome database (cocoa genome hub).

- Major exonic regions within each of the candidate genes were identified using ORF finder and primers were designed for each one of the candidate genes with the help of Primer 3.

- Genomic DNA was isolated from a self-compatible genotype and a self-incompatible genotype, the candidate genes were PCR amplified using the designed primers.

- Amplified products were sequenced and the variations in the sequence between the self-incompatible and self-compatible genotypes were analyzed in comparison with a reference genome that is self-compatible.

- Total of 31 SNPs were discovered between the self-incompatible and self-compatible genotypes.

- All the SNPs were found to be heterozygous at the SNP locus either in self-compatible or self-incompatible genotype or both.

- The highest number of SNPs were found to be present in *Gamete Expressed Protein (GEX1)* gene, 13 SNPs were found in the targeted region.

- Four SNPs were found in *Barely Any Meristem 2* (*BAM2*), *COMPASS-like H3K4 histone methylase component (WDR5a)* and *Voltage-dependent L-type calcium channel subunit* (*Alpha-1F*); three were found in *Zinc finger AN1 domain-containing stress-associated protein 12 (PMZ*) and *Serine Receptor Kinase (SRK).*

- No variation was found in *Barely Any Meristem 1* (*BAM1*), and *ARM repeat-containing protein* (*ARC1*).

- No amplification was obtained in *S locus glycoprotein* and *Hapless 2*.

- Seven potential candidate SNPs that could be linked to self-incompatibility have been discovered in this study. Two SNPs (SNP 2 and SNP 3) were discovered in *SRK*, one (SNP 1) in *WDR5a* and four SNPs (SNP 1, SNP 2, SNP 3, SNP 4) were discovered in *Alpha-1F*.

- From the potential SNPs identified to be linked to self-incompatibility in this study, markers may be developed, validated and used to genotype self-

compatible and self-incompatible genotypes at a very early stage. Based on these SNPs, allele specific PCR genotyping techniques such as CTPP or SSCP can be used. Designing primer combinations such that its 3’ end forms mismatch with the compatible DNA template will act refractory to the primer, hence amplification will be obtained only in self-incompatible genotypes. Once the markers are developed, they need to be validated in a larger population of self-incompatible and self-compatible genotypes.

References

REFERENCES

Alandete Saez, M., Ron, M., Leiboff, S., and McCormick, S. 2011. *Arabidopsis thaliana* GEX1 has dual functions in gametophyte development and early embryogenesis. *Plant J.* 68(4): 620-632.

Allen, A. M. and Hiscock, S. J. 2008. Evolution and phylogeny of self-incompatibility systems in angiosperms. In: *Self-incompatibility in flowering plants*. Springer, Berlin, Heidelberg. pp.73-101.

Ampomah, A.Y., Novak, F.J., Klu, G.Y. P. and Lamptey, T.V.O. 1990. Use of irradiated pollen as mentor pollen to induce self-fertilization of two self-incompatible Upper Amazon cacao clones. *Euphytica, 51*(3): 219-225.

Aneja, M., Gianfagna, T., Ng, E. and Badilla, I. 1994. Carbon dioxide treatment partially overcomes self-incompatibility in a cacao genotype. *Hort. Sci.* 29(1):15-17.

Aneja, M., Gianfagna, T., Ng, E. and Badilla, I., 1992. Carbon dioxide and temperature influence pollen germination and fruit set in cocoa. *Hort. Sci*. 27(9):1038-1040.

Argout, X., Salse, J., Aury, J. M., Guiltinan, M. J., Droc, G., Gouzy, J., Allegre, M., Chaparro, C., Legavre, T., Maximova, S. N., and Abrouk, M. 2010. The genome of *Theobroma cacao*. *Nat. Genet.* 43(2): 101-108.

Baker, R. P., Hasenstein, K. H. and Zavada, M. S. 1997. Hormonal changes after compatible and incompatible pollination in *Theobroma cacao* L. *Hort. Sci.* 32(7): 1231-1234.

Bakooie, M., Pourjam, E., Mahmoudi, S.B., Safaie, N., and Naderpour, M. 2018. Development of an SNP marker for sugar beet resistance/susceptible genotyping to root-knot nematode. *Journal of Agricultural Sience and Technology*, 17: 443-454.

Barrett, S.C.H. 1988. Self-incompatibility systems. *Plant reproductive ecology patterns and strategies. Oxford Univ. Press, Oxford*, pp.98-124.

Barrett, S.C.H. 1992. Heterostylous genetic polymorphisms: model systems for evolutionary analysis. In *Evolution and function of heterostyly*, Springer, Berlin, Heidelberg, pp. 1-29.

Barrett, S.C.H. 2002. Sexual interference of the floral kind. *Heredity* 88(2): 154-159.

Bayer, C., Fay, M. F., De Bruijn, A. Y., Savolainen, V., Morton, C. M., Kubitzki, K., and Chase, M. W. 1999. Support for an expanded family concept of Malvaceae within a recircumscribed order Malvales: A combined analysis of plastid atp B and rbc L DNA sequences. *Bot. J. Linn. Soc.* 129(4): 267-303.

Bertus, E. 2004. How to select superior mother trees in heterogenous progenies for more productive clone or hybrid varieties. *Ingenic Newsletter* 9: 2-3.

Bouharmont, J. 1960. Cytological investigations on fruiting and incompatibility in cacao. *Ser. Sci. Inst. Nat. Etude Agron. Congo*. 89: 117

Boyes, D. C., Chen, C. H., Tantikanjana, T., Esch, J. J. and Nasrallah, J. B. 1991. Isolation of a second S-locus-related cDNA from *Brassica oleracea*: genetic relationships between the S locus and two related loci. *Genet., 127*(1): 221-228.

Brewbaker, J.L. 1957. Pollen cytology and self-incompatibility systems in plants. *J. Hered.,* 48: 217-277.

Brownie, J., Shawcross, S., Theaker, J., Whitcombe, D., Ferrie, R., Newton, C., and Little, S. 1997. The elimination of primer-dimer accumulation in PCR. *Nucleic Acids Res.* 25(16): 3235-3241.

Cai, H., Xu, D., Zhou, L., Cheng, J., Zhang, Z., Wu, J., and You, A. 2015. Development of PCR-based CNP marker of rice Waxy gene with confronting two-pair primers. *Russ. J. Genet.* 51 (7): 673-676.

Candozo, T. and Pango, M. 2004. The SCF ubiquitin ligase: insights into a molecular machine. *Nature Rev. Mol. Cell Biol.* 5(9): 739-751.

Cao, X., Qiu, Z., Wang, X., Van Giang, T., Liu, X., Wang, J., Wang, X., Gao, J., Guo, Y., Du, Y., and Wang, G. 2017. A putative R3 MYB repressor is the candidate gene underlying atroviolacium, a locus for anthocyanin pigmentation in tomato fruit. *J. Exp. Bot.* 68(21-22): 5745-5758.

CCRP report. 2012. Cadbury-KAU co-operative Cocoa Research Project, Kerala Agricultural University, Thrissur: 103.

Chang, H.W., Chuang, L.Y., Cheng, Y.H., Hung, Y.C., Wen, C.H., Gu, D.L. and Yang, C.H., 2009. Prim-SNPing: a primer designer for cost-effective SNP genotyping. *Biotechniques*, *46*(6), pp.421-431.

Charlesworth, D. 1985. Multi-allelic self-incompatibility polymorphisms in plants. *Bioessays* 17(1): 31-38.

Chawla, B., Bernatzky, R., Liang, W. and Marcotrigiano, M. 1997. Breakdown of self-incompatibility in tetraploid *Lycopersicon peruvianum*: inheritance and expression of S-related proteins. *Theor. Appl. Genet. 95*(5-6): 992-996.

Chen, G., Zhang, B., Liu, L., Li, Q., Xie, Q., and Xue, Y. 2012. Identification of a ubiquitin-binding structure in the S-locus F-box protein controlling S-RNase-based self-incompatibility. *J. Genet. Genomics* 39(2): 93-102.

Clifford, S. C., Arndt, S. K., Popp, M., and Jones, H.G. 2002. Mucilages and polysaccharides in *Ziziphus* species (Rhamnaceae): localization, composition and physiological roles during drought-stress. *J. Exp. Bot.* 53(366): 131-138.

Colanero, S., Perata, P., and Gonzali, S. 2018. The atroviolacea gene encodes an R3-MYB protein repressing anthocyanin synthesis in tomato plants. *Front. Plant Sci.* 9: 830.

Cope, F. W. 1939. Studies in the mechanism of self-incompatibility in cocoa I. Eighth Annual Report of Cocoa Research, Trinidad, pp 20-21.

Cope, F. W. 1940. Studies in the mechanism of self-incompatibility in cocoa II. Ninth Annual Report of Cocoa Research, Trinidad, pp 19-23.

Cope, F. W. 1958. Incompatibility in Theobroma cacao. *Nature*, 181(4604): 279-279.

Cope, F. W. 1962. The effects of incompatibility and compatibility on genotype proportions in populations of Theobroma cacao L. *Heredity*, 17(2): 183-195.

Cope, F. W. 1962. The mechanism of pollen incompatibility in *Theobroma cacao* L. *Heredity*, 17(2):157-182.

Crouzillat, D., Lerceteau, E., Petiard, V., Morera, J., Rodríguez, H., Walker, D., Phillips, W., Ronning, C., Schnell, R., Osei, J., and Fritz, P. 1996. *Theobroma cacao* L.: a genetic linkage map and quantitative trait loci analysis. *Theor. Appl. Genet.* 93(1-2):205-214.

Da Silva, M. R., Clément, D., Gramacho, K. P., Monteiro, W. R., Argout, X., Lanaud, C., and Lopes, U. 2016. Genome-wide association mapping of sexual incompatibility genes in cacao (*Theobroma cacao* L.). *Tree Genet. Genomes* 12(3): 1-62.

Das, A. K., Chhabra, R., Muthusamy, V., Chauhan, H. S., Zunjare, R. U., and Hossain, F. 2019. Identification of SNP and InDel variations in the promoter and 5′ untranslated regions of γ-tocopherol methyl transferase (ZmVTE4) affecting higher accumulation of *α*-tocopherol in maize kernel. *Crop J.* 7(4): 469-479.

DCCD [Directorate of Cashewnut and Cocoa Development]. 2018. Area Production and Productivity of Cocoa in India [on-line]. Available: http://dccd.gov.in/Content.aspx?mid=1072&tid=1[01 July 2020].

de Nettancourt, D. 1997. Incompatibility in angiosperms. *Sex. Plant Reprod*. 10(4): 185-199.

de Nettancourt, D. 2001. *Incompatibility and incongruity in wild and cultivated plants*. Springer Science & Business Media.

Dellaporta, S. L., Wood, J., and Hicks, J. B. 1983. A rapid method for DNA extraction from plant tissue. *Plant Mol. Biol. Rep*: 19-21.

Denninger, P., Bleckmann, A., Lausser, A., Vogler, F., Ott, T., Ehrhardt, D. W., Frommer, W. B., Sprunck, S., Dresselhaus, T., and Grossmann, G. 2014. Male–female communication triggers calcium signatures during fertilization in Arabidopsis. *Nat. Commun.* 5(1): 1-12.

DeYoung, B.J., Bickle, K.L., Schrage, K.J., Muskett, P., Patel, K. and Clark, S.E., 2006. The CLAVATA1-related BAM1, BAM2 and BAM3 receptor kinase-like proteins are required for meristem function in Arabidopsis. *The Plant Journal*, *45*(1), pp.1-16.

Dhaliwal, A. S., Malik, C. P. and Singh, M. B. 1981. Overcoming incompatibility in *Brassica campestris* L. by carbon dioxide, and dark fixation of the gas by self- and cross-pollinated pistils. *Ann. Bot*. 48(2): 227-234.

Dias, L. A. S., Marita, J., Cruz, C. D., Barros, E.G., and Salomao, T. M. F. 2003. Genetic distances and its association with heterosis in cocoa. *Braz. Arch. Bio. Tech.* 46: 339-347.

Douglas, G. C. and Connolly, V. 1989. Self-fertilization and seed set in *Trifolium repens* L. by in situ and in vitro pollination. *Theor. Appl. Genet.* 77(1): 71-75.

Doyle, J. J. and Doyle, J. L. 1987. Isolation of plant DNA from fresh tissue. *Phytochem. Bull*. 12: 13-15.

East, E.M. and Mangelsdorf, A.J. 1925. A new interpretation of the hereditary behavior of self-sterile plants. *Proceedings of the National Academy of Sciences of the United States of America*, *11*(2), p.166.

East, E.M. and Mangelsdorf, A.J. 1925. A new interpretation of the hereditary behavior of self-sterile plants. *Proceedings of the National Academy of Sciences of the United States of America*, 11(2), p.166.

Emerson, S. 1938. The genetics of self-incompatibility in *Oenothera organensis*. *Genetics* 23(2): 190-202.

Feurtado, J. A., Huang, D., Wicki-Stordeur, L., Hemstock, L. E., Potentier, M. S., Tsang, E. W.,–and Cutler, A.J. 2011. The Arabidopsis C2H2 zinc finger indeterminate domain1/enhydrous promotes the transition to germination by regulating light and hormonal signaling during seed maturation. *The Plant Cell* 23(5): 1772-1794.

Franklin, F. C. H., Tong, F. V. E. and Wheeler, M. J. 2001. The molecular and genetic basis of pollen–pistil interactions. *New Phytologist*, 151(3): 565-584.

Gallego, F. J. and Martinez, I. 1996. Molecular typing of rose cultivars using RAPDs. *J. Hort. Sci.* 71(6): 901-908.

Gestrel, D.U. 1950. Self-incompatibility studies in guayule II. Inheritance. *Genetics* 35(4): 482.

Gibbs, P.E. 1986. Do homomorphic and heteromorphic self-incompatibility systems have the same sporophytic mechanism?. *Plant Syst. Evol.* 154(3-4): 285-323.

Gibbs, P.E. and Bianchi, M.B. 1999. Does late-acting self-incompatibility (LSI) show family clustering? Two more species of Bignoniaceae with LSI: *Dolichandra cynanchoides* and *Tabebuia nodosa*. *Ann. Bot.* 84(4): 449-457.

Giri, J., Vij, S., Dansana, P. K., and Tyagi, A.K. 2011. Rice A20/AN1 zinc-finger containing stress associated proteins (SAP1/11) and a receptor like cytoplasmic kinase (OsRLCK253) interact via A20 zinc finger and confer abiotic stress tolerance in transgenic Arabidopsis plants. *New Phytol.* 191(3): 721-732.

Glendinning, D. R., 1960. Selfing of self-incompatible cocoa. *Nature*, 187(4732): 170-170.

Glendinning, D. R., 1967. Incompatibility alleles of cocoa. *Nature*, 213(5073): 306-306.

GOI [Government of India]. 2016. Agricultural Statistics at a Glance 2015 [online]. Available: https://eands.dacnet.nic.in/PDF/Agricultural_Statistics_At_Glance-2015.pdf [1 June 2020].

GOI [Government of India]. 2019. Agricultural Statistics at a Glance 2016 [online]. Available: http://agricoop.gov.in/sites/default/files/agristatglance2018.pdf [01 June 2020].

Gómez, E. M., Buti, M., Sargent, D. J., Dicenta, F., and Ortega, E. 2019. Transcriptomic analysis of pollen-pistil interactions in almond (*Prunus dulcis*) identifies candidate genes for components of gametophytic self-incompatibility. *Tree Genet. Genomes* 15(4): 1-53.

Goujon, M., Mc William, H., Li, W., Valentin, F., Squizzato, S., Paern, J., and Lopez, R. 2010. A new bioinformatics analysis tools framework at EMBL–EBI. *Nucleic Acids Res. 38*(2): 695-W699.

Gupta, S .K., Rai, A. K., Kanwar, S. S., and Sharma, T. R. 2012. Comparative analysis of zinc finger proteins involved in plant disease resistance. *PLoS One*, 7(8), p.e42578.

Hamajima, N. 2001. PCR-CTPP: a new genotyping technique in the era of genetic epidemiology. *Expert Rev. Mol. Diagn.* 1(1): 119-123.

Hamajima, N., Saito, T., Matsuo, K., and Tajima, K. 2002. Competitive amplification and unspecific amplification in polymerase chain reaction with confronting two-pair primers. *J. Mol. Diagn.* 4(2): 103-107.

Hamajima, N., Saito, T., Matsuo, K., Kozaki, K.I., Takahashi, T., and Tajima, K. 2000. Polymerase chain reaction with confronting two-pair primers for polymorphism genotyping. *Jpn. J. Cancer Res.* 91(9): 865-868.

Han, F., Yuan, K., Kong, C., Zhang, X., Yang, L., Zhuang, M., Zhang, Y., Li, Z., Wang, Y., Fang, Z., and Lv, H. 2018. Fine mapping and candidate gene identification of the genic male-sterile gene ms3 in cabbage 51S. *Theor. Appl. Genet.* 131(12): 2651-2661.

Hatakeyama, K., Takasaki, T., Suzuki, G., Nishio, T., Watanabe, M., Isogai, A. and Hinata, K., 2001. The S receptor kinase gene determines dominance relationships in stigma expression of self-incompatibility in Brassica. *The Plant J.*, 26(1): 69-76.

Heslop-Harrison, Y. and Shivanna, K.R. 1977. The receptive surface of the angiosperm stigma. *Ann. Bot.* 41(6): 1233-1258.

Hiscock, S.J. and McInnis, S.M. 2003. The diversity of self-incompatibility systems in flowering plants. *Plant Biol.* 5(01): 23-32.

Hiscock, S.J., Kües, U., and Stahl, U. 1995. Recombination: Sexual Reproduction in Plants: Self-Incompatibility as a Mechanism Promoting Outbreeding and Gene Flow. In *Progress in Botany* . Springer, Berlin, Heidelberg, pp. 275-300.

Hodgkin, T., Lyon, G. D., and Dickinson, H. G. 1988. Recognition in flowering plants: A comparison of the Brassica self incompatibility system and plant pathogen interactions. *New Phytol.* 110(4): 557-569.

Hord, C. L., Chen, C., DeYoung, B. J., Clark, S. E. and Ma, H., 2006. The BAM1/BAM2 receptor-like kinases are important regulators of Arabidopsis early anther development. *The Plant Cell* 18(7): 1667-1680.

Hua, Z., Meng, X. and Kao, T.H., 2007. Comparison of *Petunia inflata* S-locus F-box protein (Pi SLF) with Pi SLF–like proteins reveals its unique function in S-RNase–based self-incompatibility. *The Plant Cell*, *19*(11): pp.3593-3609.

Huang, X. and Madan, A., 1999. CAP3: A DNA sequence assembly program. *Genome Res.* 9(9): 868-877.

Hughes, M.B. and Babcock, E.B., 1950. Self-incompatibility in *Crepis foetida* (L.) subsp. rhoeadifolia (Bieb.) Schinz et Keller. *Genetics* 35(5): 570.

Indriolo, E., Safavian, D., and Goring, D. R. 2014. The ARC1 E3 ligase promotes two different self-pollen avoidance traits in Arabidopsis. *The Plant Cell* 26(4): 1525-1543.

Indriolo, E., Tharmapalan, P., Wright, S. I., and Goring, D. R. 2012. The ARC1 E3 ligase gene is frequently deleted in self-compatible Brassicaceae species and has a conserved role in *Arabidopsis lyrata* self-pollen rejection. *The Plant Cell* 24(11): 4607-4620.

Iwano, M., Ito, K., Fujii, S., Kakita, M., Asano-Shimosato, H., Igarashi, M., Kaothien-Nakayama, P., Entani, T., Kanatani, A., Takehisa, M., and Tanaka, M. 2015. Calcium signalling mediates self-incompatibility response in the Brassicaceae. *Nat. Plants* 1(9): 1-9.

Jordan, N. D., Franklin, F. C. H. and Tong, F. V. E. 2000. Evidence for DNA fragmentation triggered in the self-incompatibility response in pollen of *Papaver rhoeas*. *The Plant J.*, 23(4): 471-479.

Jose, J. and Usha, R. 2000. Extraction of geminiviral DNA from a highly mucilaginous plant (*Abelmoschus esculentus*). *Plant Mol. Biol. Rep.* 18(4): 349-355.

Kakizaki, Y. 1930. Studies on the genetics and physiology of self-and cross-incompatibility in the common cabbage (*Brassica oleracea* L. var. capitata L.). *Jap. J. Bot.* 5.

Kakita, M., Murase, K., Iwano, M., Matsumoto, T., Watanabe, M., Shiba, H., Isogai, A. and Takayama, S., 2007. Two distinct forms of M-locus protein kinase localize to the plasma membrane and interact directly with S-locus receptor kinase to transduce self-incompatibility signaling in *Brassica rapa. The Plant Cell*, *19*(12), pp.3961-3973.

Kao, T.H. and McCubbin, A.G. 1996. How flowering plants discriminate between self and non-self pollen to prevent inbreeding. *Proceedings of the National Academy of Sciences*, 93(22), pp.12059-12065.

Kennedy, A. J., Lockwood, G., Mossu, G., Simmonds, N. W., and Tan, G.Y. 1987. Cocoa Breeding: Past, Present and Future. *Cocoa Growers Bulletin* 38: 5-22.

Kenrick, J., Kaul, V., and Williams, E.G. 1986. Self-incompatibility in *Acacia retinodes*: site of pollen-tube arrest is the nucellus. *Planta* 169(2): 245-250.

Knight, R. and Rogers, H. 1955. Incompatibility in *Theobroma cacao*. *Heredity*, *9*(1): 69-77.

Knight, R. and Rogers, H. H., 1953. Sterility in *Theobroma cacao* L. *Nature*, 172(4369): 164-164.

Kotturi, M. F., Carlow, D. A., Lee, J. C., Ziltener, H. J., and Jefferies, W. A. 2003. Identification and functional characterization of voltage-dependent calcium channels in T lymphocytes. *J. Biol. Chem.* 278(47): 46949-46960.

Kubo, K. I., Entani, T., Takara, A., Wang, N., Fields, A. M., Hua, Z., Toyoda, M., Kawashima, S. I., Ando, T., Isogai, A. and Kao, T. H.. 2010. Collaborative non-self-recognition system in S-RNase–based self-incompatibility. *Sci.*, *330* (6005): 796-799.

Kusaba, M., Dwyer, K., Hendershot, J., Vrebalov, J., Nasrallah, J. B. and Nasrallah, M. E., 2001. Self-incompatibility in the genus Arabidopsis: characterization of the S locus in the outcrossing *A. lyrata* and its autogamous relative *A. thaliana*. *The Plant Cell*, 13(3): 627-643.

Kusaba, M., Tung, C.W., Nasrallah, M. E. and Nasrallah, J. B. 2002. Monoallelic expression and dominance interactions in anthers of self-incompatible *Arabidopsis lyrata*. *Plant Physiol.*, 128(1): 7-20.

Lachenaud, P. and Motamayor, J.C. 2017. The Criollo cacao tree (*Theobroma cacao* L.): a review. *Genet. Resour. Crop Evol.* 64(8): 1807-1820.

Lachenaud, P., Paulin, D., Ducamp, M. and Thevenin, J. M., 2007. Twenty years of agronomic evaluation of wild cocoa trees (*Theobroma cacao* L.) from French Guiana. *Sci. Hortic.*, 113(4): 313-321.

Lanaud, C., Fouet, O., Legavre, T., Lopes, U., Sounigo, O., Eyango, M. C., Mermaz, B., Silva, D. M. R., Solorzano, R. G., Argout, X. and Gyapay, G., 2017. Deciphering the *Theobroma cacao* self-incompatibility system: from genomics to diagnostic markers for self-compatibility. *J. Exp. Bot.*, 68(17): 4775-4790.

Lanaud, C., Risterucci, A. M., N'goran, J. A. K., Clement, D., Flament, M. H., Laurent, V. and Falque, M. 1995. A genetic linkage map of *Theobroma cacao* L. *Theor. Appl. Genet*. 91: 987-99.

Lawrence, M. J., Afzal, M. and Kenrick, J. 1978. The genetical control of self-incompatibility in *Papaver rhoeas*. *Heredity*, 40(2): 239-253.

Li, X., Paech, N., Nield, J., Hayman, D., and Langridge, P. 1997. Self-incompatibility in the grasses: evolutionary relationship of the S gene from Phalaris coerulescens to homologous sequences in other grasses. *Plant Mol. Biol.* 34(2): 223-232.

Ligic, B., Bohs, L. and Kohn, J. R., 2005. Ancient polymorphism reveals unidirectional breeding system shifts. In: *Proceedings of the National Academy of Sciences*, 103(5): 1359-1363.

Lim, J. H., Yang, H. J., Jung, K. H., Yoo, S. C., and Paek, N. C. 2014. Quantitative trait locus mapping and candidate gene analysis for plant architecture traits using whole genome re-sequencing in rice. *Mol. Cells*, 37(2):1-149.

Liu, E., Liu, Y., Wu, G., Zeng, S., Tran Thi, T. G., Liang, L., Liang, Y., Dong, Z., She, D., Wang, H., and Zaid, I. U. 2016. Identification of a candidate gene for panicle length in rice (*Oryza sativa* L.) via association and linkage analysis. *Front. Plant Sci.* 7: 596.

Liu, Y., Wu, T., Song, J., Chen, X., Zhang, Y., and Wan, Y. 2013. A mutant screening method by critical annealing temperature-PCR for site-directed mutagenesis. *BMC Biotech.* 13(1): 1-8.

Liu, Z. Q., Xu, G. H., and Zhang, S. L. 2008. Pyrus pyrifolia stylar S-RNase induces alterations in the actin cytoskeleton in self-pollen and tubes *in vitro*. *Protoplasma* 232(1): 1-61.

Lundqvist, A. 1990. The complex S-gene system for control of self-incompatibility in the buttercup genus Ranunculus. *Hereditas* 113(1): 29-46.

Lundqvist, A. 1991. Four-locus S-gene control of self-incompatibility made probable in *Lilium martagon* (Liliaceae). *Hereditas* 114(1): 57-63.

Malhotra, S. K. and Hubali, V. 2016. *Cashew and Cocoa- Production to Marketing*. Directorate of Cashew and Cocoa Development, Kochi, 140p.

Mallika, V. K., Amma, P. S., Namboothiri, R., and Nair, R. V. 2006. Morphology and genetic analysis of inbreds and inbred crosses in cocoa (*Theobroma cacao* L.). *J. Plant. Crops* 34(3): 229-235.

Mallika, V. K., Prasannakumari Amma, S., and Koshy, A. 2000. Comparison of yield in hybrids and clones of cocoa (*Theobroma cacao* L). *J. Plant. Crops* 28(3): 222-225.

Mather, K. 1944. Genetical control of incompatibility in angiosperms and fungi. *Nature*, 153: 392-394.

Matton, D. P., Luu, D. T., Xike, Q., Laublin, G., O'Brien, M., Maes, O., Morse, D. and Cappadocia, M. 1999. Production of an S RNase with dual specificity suggests a novel hypothesis for the generation of new S alleles. *The Plant Cell*, 11(11): 2087-2097.

McClure, B., Mou, B., Canevascini, S., and Bernatzky, R. 1999. A small asparagine-rich protein required for S-allele-specific pollen rejection in Nicotiana. *Proceedings of the National Academy of Sciences*, 96(23):13548-13553.

McCormick, S. 1998. Self-incompatibility and other pollen-pistil interactions. *Curr. Opin. Plant Biol.* 1: 18-25.

McCubbin, A.G. and Kao, T.H. 2000. Molecular recognition and response in pollen and pistil interactions. *Annu. Rev. Cell Dev. Biol.* 16(1): 333-364.

Meena, P. 2014. Molecular characterization and DNA fingerprinting of selected cashew (Anacardium occidentale L.) varieties of KAU. M.Sc. thesis, Kerala Agricultural University, Thrissur, 120 p.

Miller, J. S. and Venable, D. L., 2000. Polyploidy and the evolution of gender dimorphism in plants. *Sci.*, 289(5488): 2335-2338.

Minimol J. S. and Amma, P. S. 2013. Self-incompatibility assessment in vascular streak dieback (VSD) disease resistant hybrids of cocoa (*Theobroma cacao* L.). *Asian J. Horti.* 8(1):114-117.

Mir, R. R., Kudapa, H., Srikanth, S., Saxena, R. K., Sharma, A., Azam, S., Saxena, K., Penmetsa, R. V., and Varshney, R. K. 2014. Candidate gene analysis for determinacy in pigeonpea (*Cajanus* spp.). *Theor. Appl. Genet.* 127(12): 2663-2678.

N'Zi, J. C., Kahia, J., Diby, L. and Kouamé, C. 2017. Compatibility of ten elite cocoa (*Theobroma cacao* L.) clones. *Hortic.*, 3(3): 45.

Nakanishi, T. and Hinata, K., 1973. An effective time for CO_2 gas treatment in overcoming self-incompatibility in Brassica. *Plant Cell Physiol*, 14(5): 873-879.

Palloix, A., Herve, Y., Knox, R. B. and Dumas, C. 1985. Effect of carbon dioxide and relative humidity on self-incompatibility in cauliflower, *Brassica oleracea*. *Theor. Appl. Genet.*, 70(6): 628-633.

Pandey, K. K. 1968. Colchicine-induced changes in the self-incompatibility behaviour of Nicotiana. *Genet.*, 39(1): 257-271.

Poritz, M. A. and Ririe, K. M. 2014. Getting things backwards to prevent primer dimers. *J. Mol. Diagn.* 16(2): 159-162.

Poulter, N. S., Wheeler, M. J., Bosch, M. and Tong, F. V. E. 2010. Self-incompatibility in *Papaver*: identification of the pollen S-determinant PrpS. *Biochem. Soc. Trans.* 38(2): 588-592.

Pound F. J. 1932. Studies of fruitfulness in cocoa II, evidence for partial self-sterility. 1st Annual Report on Cocoa Research, Trinidad, pp. 24-28.

Qiao, Y., Piao, R., Shi, J., Lee, S. I., Jiang, W., Kim, B. K., Lee, J., Han, L., Ma, W., and Koh, H. J. 2011. Fine mapping and candidate gene analysis of dense and erect panicle 3, DEP3, which confers high grain yield in rice (*Oryza sativa* L.). *Theor. Appl. Genet.* 122(7): 1439-1449.

Henao Ramírez, A.M., Salazar Duque, H.J. and Urrea Trujillo, A.I., 2018. Quality of cocoa (*Theobroma cacao* L.) DNA from foliar tissue at different stages of development. *Acta Agronómica*, *67*(2), pp.311-318.

Royaert, S., Phillips-Mora, W., Leal, A.M.A., Cariaga, K., Brown, J.S., Kuhn, D.N., Schnell, R.J. and Motamayor, J.C., 2011. Identification of marker-trait associations for self-compatibility in a segregating mapping population of *Theobroma cacao* L. Tree genetics & genomes, 7(6), pp.1159-1168.

Rychlik, W. J. S. W., Spencer, W. J., and Rhoads, R. E. 1990. Optimization of the annealing temperature for DNA amplification *in vitro*. *Nucleic Acids Res.* 18(21): 6409-6412.

Sage, T. L., Bertin, R. I. and Williams, E. G. 1994. Ovarian and other late-acting self-incompatibility systems. In: *Genetic control of self-incompatibility and reproductive development in flowering plants*. Springer, Dordrecht, pp. 116-140.

Sage, T.L. and Sampson, F.B. 2003. Evidence for ovarian self-incompatibility as a cause of self-sterility in the relictual woody angiosperm, *Pseudowintera axillaris* (Winteraceae). *Ann. Bot.* 91(7): 807-816.

Sambrook, J., Fritsch, E. F., and Maniatis, T. 1989. *Molecular cloning: a laboratory manual* (Ed. 2). Cold spring harbor laboratory press.

Sanger, F., Nicklen, S., and Coulson, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the national academy of sciences*, 74(12), pp.5463-5467.

Sassa, H., Hirano, H. and Ikehashi, H. 1993. Identification and characterization of stylar glycoproteins associated with self-incompatibility genes of Japanese pear, *Pyrus serotina* Rehd. *Mol. Gen. Genet.* 241(1-2): 17-25.

Schawe, C. C., Durka, W., Tscharntke, T., Hensen, I. and Kessler, M. 2013. Gene flow and genetic diversity in cultivated and wild cacao (*Theobroma cacao*) in Bolivia. *Am. J. Bot.* 100(11): 2271-2279.

Schopfer, C. R., Nasrallah, M. E. and Nasrallah, J. B. 1999. The male determinant of self-incompatibility in Brassica. *Sci.*, 286(5445): 1697-1700.

Sears, E. R. 1937. Cytological phenomena connected with self-sterility in the flowering plants. *Genet.* 22(1): 130.

Seavey, S.R. and Bawa, K.S. 1986. Late-acting self-incompatibility in angiosperms. *Bot. Rev.* 52(2): 195-219.

Shiba, H., Iwano, M., Entani, T., Ishimoto, K., Shimosato, H., Che, F. S., Satta, Y., Ito, A., Takada, Y., Watanabe, M. and Isogai, A. 2002. The dominance of alleles controlling self-incompatibility in Brassica pollen is regulated at the RNA level. *The Plant Cell*, 14(2): 491-504.

Shinozuka, H., Cogan, N.O., Smith, K.F., Spangenberg, G.C. and Forster, J.W., 2010. Fine-scale comparative genetic and physical mapping supports map-based cloning strategies for the self-incompatibility loci of perennial ryegrass (*Lolium perenne* L.). *Plant molecular biology*, *72*(3), p.343.

Silva, N.F. and Goring, D.R. 2001. Mechanisms of self-incompatibility in flowering plants. *Cell. Mol. Life Sci.* 58(14): 1988-2007.

Sonneveld, T., Tobutt, K. R., Vaughan, S. P. and Robbins, T. P. 2005. Loss of pollen-S function in two self-compatible selections of *Prunus avium* is associated with deletion/mutation of an S haplotype–specific F-Box gene. *The Plant Cell*, 17(1): 37-51.

Steele, R. E. and Dana, C. E. 2009. Evolutionary history of the HAP2/GCS1 gene and sexual reproduction in metazoans. *PLoS ONE* 4: 7680.

Stirnimann, C. U., Petsalaki, E., Russell, R. B., and Müller, C. W. 2010. WD40 proteins propel cellular networks. *Trends Biochem. Sci.* 35(10): 565-574.

Stone, S. L., Anderson, E. M., Mullen, R. T., and Goring, D. R. 2003. ARC1 is an E3 ubiquitin ligase and promotes the ubiquitination of proteins during the rejection of self-incompatible Brassica pollen. *The Plant Cell* 15(4): 885-898.

Stone, S. L., Arnoldo, M. and Goring, D. R. 1999. A breakdown of Brassica self-incompatibility in ARC1 antisense transgenic plants. *Sci.*, 286(5445): 1729-1731.

Takahashi, T., Mori, T., Ueda, K., Yamada, L., Nagahara, S., Higashiyama, T., Sawada, H., and Igawa, T. 2018. The male gamete membrane protein DMP9/DAU2 is required for double fertilization in flowering plants. *Development* 145(23).

Takayama, S. and Isogai, A., 2005. Self-incompatibility in plants. *Annu. Rev. Plant Biol.*, *56*: 467-489.

Taylor, J. P. 1982. Carbon dioxide treatment as an effective aid to the production of selfed seed in kale and Brussels sprouts. *Euphytica*, 31(3): 957-964.

Thakur, P. P., Mathew, D., Nazeem, P. A., Abida, P. S., Indira, P., Girija, D., Shylaja, M. R., and Valsala, P. A. 2014. Identification of allele specific AFLP markers linked with bacterial wilt [Ralstonia solanacearum (Smith) Yabuuchi, et al.] resistance in hot peppers (Capsicum annuum L.). Physiol. Mol. Plant Pathol. 87: 19-24.

Tsuchimatsu, T., Suwabe, K., Shimizu-Inatsugi, R., Isokawa, S., Pavlidis, P., Städler, T., Suzuki, G., Takayama, S., Watanabe, M., and Shimizu, K. K., 2010. Evolution of self-compatibility in Arabidopsis by a mutation in the male specificity gene. *Nature* 464(7293): 1342-1346.

Tsukamoto, T., Hauck, N. R., Tao, R., Jiang, N. and Iezzoni, A. F. 2006. Molecular characterization of three non-functional S-haplotypes in sour cherry (*Prunus cerasus*). *Plant Mol. Biol.*, 62(3): 371.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., and Rozen, S.G. 2012. Primer3—new capabilities and interfaces. *Nucleic Acids Res.* 40(15): e115-e115.

Ushijima, K., Yamane, H., Watari, A., Kakehi, E., Ikeda, K., Hauck, N.R., Iezzoni, A. F., and Tao, R. 2004. The S haplotype specific F box protein gene, SFB, is defective in self-compatiblehaplotypes of Prunus avium and P. mume. *Plant J.* 39(4): 573-586.

Wang, L., Dong, L., Zhang, Y. E., Zhang, Y., Wu, W., Deng, X., and Xue, Y. 2004. Genome-wide analysis of S-Locus F-box-like genes in *Arabidopsis thaliana*. *Plant Mol. Biol.* 56(6): 929-945.

Wang, P., Lu, Q., Ai, Y., Wang, Y., Li, T., Wu, L., Liu, J., Cheng, Q., Sun, L., and Shen, H. 2019. Candidate gene selection for cytoplasmic male sterility in pepper (*Capsicum annuum* L.) through whole mitochondrial genome sequencing. *Int. J. Mol. Sci.* 20(3): 578.

Warren, J. and Misir, S. 1995. Isozyme markers for self-compatibility and yield in *Theobroma cacao* (cacao). *Heredity, 74*(4): 354-356.

Waters, D. L., Bundock, P. C., and Henry, R.J. 2008. Genotyping by allele specific PCR. *Plant Genotyping II: SNP Technology*. CABI, Wallingford, UK, pp.88-97.

Wettasinghe, R. and Peffley, E.B., 1998. A rapid and efficient extraction method for onion DNA. *Plant breeding*, *117*(6), pp.588-589.

Wheeler, M. J., Vatovec, S., and Franklin-Tong, V. E. 2010. The pollen S-determinant in Papaver: comparisons with known plant receptors and protein ligand partners. *J. Exp. Bot.* 61(7): 2015-2025.

Whitehouse, H.L. 1950. Multiple-allelomorph incompatibility of pollen and style in the evolution of the angiosperms. *Ann. Bot.* 14(54): 199-216.

Williams, E. G., Kaul, V. and Rouse, J. L. 1986. Early events in the embryo sac after intraspecific and interspecific pollinations in *Rhododendron kawakamii* and *R. retusum*. *Can. J. Bot.*, 64(2): 282-291.

Wong, J. L. and Johnson, M. A. 2010. Is HAP2-GCS1 an ancestral gamete fusogen? *Trends Cell Biol*., 20: 134-141.

Yamada, M. M., Faleiro, F. G., Clement, D., Lopes, U. V., Pires, J. L., and Melo, G. R. P. 2010. Relationship between molecular markers and incompatibility in *Theobroma cacao* L. *Agrotropica* 22: 71-74.

Young, B. J., Bickle, K. L., Schrage, K. J., Muskett, P., Patel, K., and Clark, S. E. 2006. The CLAVATA1 related BAM1, BAM2 and BAM3 receptor kinase like proteins are required for meristem function in Arabidopsis. *Plant J.* 45(1): 1-16.

Zavada, M.S. 1984. The relation between pollen exine sculpturing and self-incompatibility mechanisms. *Plant Syst. Evol.* 147(1-2): 63-78.

Zhang, J., Xia, C., Duan, C., Sun, S., Wang, X., Wu, X. and Zhu, Z., 2013. Identification and candidate gene analysis of a novel Phytophthora resistance gene Rps10 in a Chinese soybean cultivar. *PloS one*, 8(7), p.e69799.

ANNEXURE I

List of laboratory equipment used for the study

Refrigerated centrifuge	:	Kubota, Japan
Horizontal electrophoresis system	:	BIO-RAD, USA
Thermal cycler	:	Proflex
Gel documentation system	:	Gel Doc^{TM} XR+BIORAD
Micropipettes	:	Eppendorf
Spectrophotometer	:	Nanodrop® ND-1000, USA

ANNEXURE II

Reagents required for DNA isolation

1. 5x CTAB extraction buffer (100ml)

CTAB (Cetyl trimethyl ammonium bromide)	:	5 g
Trisbase	:	1.21 g
EDTA	:	0.745 g
NaCl	:	8.18 g
PVP	:	1.0 g

Adjusted the pH to 8 and made up final volume up to 100 ml.

2. EB1 (Extraction Buffer 2) (100ml)

0.5M EDTA pH 8.0 : 2ml 1M

Tris pH: 8.0 : 5ml

3. 20 % SDS (do not autoclave) (100 ml)

SDS : 20 g

dH_2O : 100ml

4. 5M KOAC (Potassium acetate) (100 ml)

KOAC : 49.1 g

dH_2O : 100ml

5. 3M NaOAC (Sodium acetate) (100ml)

NaOAC : 24.6 g

dH_2O : 100 ml

6. Chloroform- Isoamyl alcohol (24:1v/v)

To chloroform (24 parts), isoamyl alcohol (1 part) was added and mixed thoroughly.

7. Chilled isopropanol

Isopropanol was stored in refrigerator at 0ºC and was used for the study.

8. Ethanol (70 %)

To the 70 parts of absolute ethanol (100 %), 30 parts of sterile distilled water was added to make 70 per cent ethanol.

9. TE buffer (pH 8, 100ml)

Tris HCl (10 mM) : 0.1576 g EDTA (1 mM) : 0.0372g

(The solution was prepared, autoclaved and stored at room temperature)

ANNEXURE III

Composition of buffers and dyes used for gel electrophoresis

a. TAE Buffer50X

Tris base	:	242 g
Glacial acetic acid	:	57.1 ml 0.5M
EDTA (pH 8.0)	:	100 ml

b. Loading dye6X

0.25 per cent bromophenol blue

0.25 per cent xylene cyanol 30 per cent glycerol in water

c. Ethidium bromide

The dye was prepared as a stock solution of 10 mg/ml in water and was stored at room temperature in a dark bottle

CANDIDATE GENE ANALYSIS ON SELF-INCOMPATIBILITY IN COCOA

(*Theobroma cacao* L.)

By

SHARAT PRABHAKARAN

(2018-11-165)

ABSTRACT OF THE THESIS

Submitted in partial fulfilment of the requirement for the degree of

Master of Science in Agriculture

(PLANT BIOTECHNOLOGY)

Faculty of Agriculture

Kerala Agricultural University, Thrissur

CENTRE FOR PLANT BIOTECHNOLOGY AND MOLECULAR BIOLOGY

COLLEGE OF HORTICULTURE

VELLANIKKARA, THRISSUR - 680656

KERALA, INDIA

2020

ABSTRACT

Cocoa is a perennial tree with typical plant habit and specific fruit characteristics. It is highly influenced by climate changes and growing environment, which makes long term and dynamic breeding programme necessary (Malhotra and Hubali, 2016). Physiological and genetic investigations have unveiled that the yield potential of cocoa is not yet fully exploited (Bertus, 2004). Demand for chocolate is increasing at a rate of 15-20 per cent every year. To meet this demand, more area has to be brought under cocoa cultivation using improved genetic stock. Development of superior hybrids have significantly contributed to improve the cocoa productivity in many countries (Kennedy *et al*., 1987; Dias *et al*., 2003). Cocoa hybrids showed wide adaptability, low environmental interaction and improved yield, when compared to traditional cultivars (Dias *et al*., 2003).

Self-incompatibility is a pollination control mechanism which prevents self-fertilization. Hence, this can be exploited in hybrid production by avoiding emasculation, which is a cumbersome process (Minimol and Amma, 2013). Moreover, emasculation will damage the flowers leading to reduced success rate. Conventionally, the self-incompatibility is measured by selfing 100 flowers per tree. If no fruit set is observed, then the plant is classified as self-incompatible (Mallika *et al.,* 2006). This is a tedious process which will reduce the pace of breeding programme.

In various other crops, many candidate genes have been reported for self-incompatibility (McCormick, 1998). However, the actual sequence variations in candidate genes are yet to be studied in cocoa. Identification of appropriate genes involved in self-incompatibility will help to identify its mechanism at an early stage and quicken the breeding programme.

In this study, 10 candidate genes *viz. Serine Receptor Kinase (SRK), S Locus Glycoprotein (SLG), Barely Any Meristem 1* (*BAM1*), *Barely Any Meristem 2* (*BAM2*), *COMPASS-like H3K4 histone methylase component (WDR5a), Voltage-dependent L-type calcium channel subunit (alpha-1F), Gamete Expressed Protein (GEX1), Zinc finger AN1 domain-containing stress-associated protein 12 (PMZ), ARM repeat-containing protein* (*ARC1*) and *Hapless 2* (McCormic, 1998; Lanaud *et al*., 2017) were

characterized. These genes were reported to have involved in self-incompatibility in other crops.

Genomic nucleotide sequences from reported host plant species were retrieved from the NCBI GenBank database. Using this information, homologous gene sequences of the candidate genes in cocoa were retrieved. Primer sets targeting major exonic regions for each of the candidate genes were designed. Genomic DNA was isolated from self-compatible genotype (GVI-167 x GIV-18.5) and self-incompatible genotype (IMC20) and the candidate genes were PCR amplified. Amplified products were sequenced and the variations in the sequences between the self-incompatible and self-compatible genotypes were analyzed, in comparison with a self-compatible reference genome (Argout *et al.*, 2010).

Between the self-incompatible and self-compatible genotypes, a total of 31 different SNPs were discovered among the genes studied. All of them were found to be heterozygous at the locus either in self-compatible or self-incompatible genotype. The maximum number of SNPs, a total of 12, were found in *GEX1* gene. Four SNPs each were found in genes *SRK*, *BAM2*, *WDR5a* and *Alpha1F* whereas three SNPs were found in *PMZ*. No variation was seen in *BAM1* and *ARC1*.

SNP locus homozygous in self-compatible and heterozygous in self-incompatible, with the corresponding locus of self-compatible reference genome can be used as potential candidate for developing markers to distinguish them. Such SNPs are identified and recommended for further validation.