Genetics of self-incompatibility in perennial ryegrass (*Lolium perenne* L.)

by

Chloé Manzanares

A thesis submitted to The University of Birmingham for the degree of DOCTOR OF PHILOSOPHY

School of Biosciences

The University of Birmingham

February 2013

Abstract

Self-incompatibility (SI) prevents pollination by "self" pollen and promotes outbreeding and diversity. SI is a widespread mechanism among angiosperms, well known in many species and largely genetically governed by a single locus, *S*-locus. Grass SI is a unique system as it is controlled by two loci, *S* and *Z*, mapped in linkage group 1 and 2 respectively but the mechanism remains elusive despite many studies.

In this project, SI in perennial ryegrass (*Lolium perenne* L.) is studied at the gene level for the *S*-locus. Using a fine-mapping approach, large perennial ryegrass mapping populations were genotyped using High Resolution Melting (HRM) method in order to isolate a small genomic region of 0.11 cM. The genomic *S*-region was sequenced in order to assess the genes within it and the transcripts from pollen and stigma were sequenced in order to study the gene expression in this region. The project identified four stigma *S*-candidate genes and potentially six pollen *S*-candidate genes. An initial study of the allelic diversity of an F-box Transporter Inhibitor Response gene was conducted but indicated high sequence conservation. Moreover, using closely linked markers, the allelic diversity in a breeding population was assessed for both *S* and *Z*. This assay demonstrated that allelic diversity is maintained in a half-sib family recurrent selection programme and that even if the SI genes are unknown, it is possible to predict their genotypes.

Finally, self-compatibility (SC) has been reported in many grasses. Using self-compatible perennial ryegrass populations, two additional loci, *F* and *T*, have been investigated, by a mapping approach as well as observation of self-pollination under the microscope. For both loci, a distortion segregation was observed on linkage group 3 (*F*) and 5 (*T*) and initial marker recombination maps were created. The study of the self-compatibility phenotype concluded that SC is governed by a single locus, which is not *F*.

Acknowledgements

At first, I would like to thank Teagasc (Ireland) for the Walsh fellowship, for giving me the opportunity to be part of this project, and IBERS (Aberystwyth University) for allowing me to undertake my studies at their institute and make use of their ever expanding facilities.

I would like to thank my supervisors Dr Susanne Barth (Teagasc) and Professor Chris Franklin (Birmingham University) for their help and guidance through my PhD. I would like to give a particular thank you to my third supervisor, Dr Daniel Thorogood (IBERS) who helped and guided me from the beginning, sharing with me his passion for self-incompatibility, plant breeding and pub quizzes.

During my PhD, I have benefitted from the advice and help of other scientists and I would like to thank them: Dr. Torben Asp, Dr. Stephen Byrne, Dr. Michael Abberton, Dr. Ian Armstead and Dr Matthew Hegarty. This project would not be what it is if it was for the help of Prof. Dr. Bruno Studer. Thank you for guiding me through my research, and believing in me, even now that I am part of his team at the ETH in Zurich.

A PhD is a hard and long journey but it was an enjoyable one thanks to Matthew, Charlotte, Jim, Ellen, Alan, Kirsten, Ann, Teri, Andi, Caerwyn, Markku, Karen, Jose, Michal, Florence, Deepmala and the potting shed team. I would also like to thank my friends from home who supported me despite the distance: the blonde one, Pauline, the ginger one, Caroline, Antoine alias Chou, Charlotte and mon Jules. Du fond du cœur, merci les amis !!

A special thank you to my friend, colleague, housemate, officemate Maciej Bisaga. You are and you will stay my BOMF. I hope we will have many more opportunities to gossip around a cup of coffee, a glass of wine or a shot of vodka!

I would like to thanks my lovely partner, Dr (!!!) Steven Yates for his support. We started together, you might have finished before but you supported me until the end, and I love you even more for that.

Finally, I would like to thank all my family and especially my mum for her weekly Wednesday phone calls and her faith in me, and my little sister Eudoxie for her joy.

All of you made me what I am today and I will be forever grateful for it. Milles Merci.

Contents

Table of Contents

List of figures

List of tables

List of abbreviations

$[Ca^{2+}]_i$	intracellular cytosolic free calcium concentration
AFLP	Amplified Fragment Length Polymorphism
ANOVA	Analysis of Variance
ARC1	Armadillo Repeat motif-Containing protein 1
ARF	Auxin Response Factor
AS	Alternative splicing
ASO	Allele-Specific Oligonucleotide hybridization
ATP	Adenosine Tri-Phosphate
BAC	Bacterial Artificial Chromosome
BCV	Biological Coefficient of Variation
BLAST	Basic Local Alignment Search Tool
CCR1	Cinnamoyl-CoA Reductase gene
cDNA	complementary DNA
CDPK	Ca^{2+}-Dependent Protein Kinase
CDS	Coding DNA Sequence
cM	centiMorgan
CMS	Cytoplasmic Male Sterility
CTAB	Cetyl Trimethylammonium Bromide
DNA	Deoxyribonucleic Acid
dNTP	Deoxyribonucleotide Triphosphate
ds	double-stranded
ds-DNA	double-stranded DNA
dsRBD	double-stranded RNA Binding Domain
dsRBP	double-stranded RNA Binding Protein
dsRBM	ds-RNA Binding Motif
DTS	Daniel Thorogood *S*-population
DTT	Dithiothreitol
DUF	Domain of Unknown Function
EC	Electrophoresis by Capillarity
EDTA	Ethylenediaminetetraacetic Acid
EMS	Ethyl Methanesulphonate
EST	Expressed Sequence Tag
GOT/3	Glutamate Oxalo-acetate Transminase 3
GSI	Gametophytic Self-Incompatibility
HRM	High Resolution Melting
HV	Hypervariable region
IAA	Indole-3-acetic Acid
IBERS	Institute of Biological, Environmental and Rural Sciences
ILGI	International *Lolium* Genome Initiative
InDel	Insertion/Deletion

LG	Linkage Group
logCPM	$\log_2$ Count Per Million
logFC	$\log_2$ Fold Change
LRR	Leucine Rich Region
MAS	Marker Assisted Selection
MID	Multiplex Identifier
MLPK	*M*-locus Protein Kinase
mRNA	Ribonucleic Acid messanger
MS	Mass Spectrometry
NB-ARC	Nucleotide-Binding – APAF-1, R protein and CED-4
NCBI	National Center for Biotechnology Information
NGS	Next Generation Sequencing
OP	Open-Pollination
PCD	Programmed Cell Death
PCR	Polymerase Chain Reaction
PE	Paired-End
PGI-2	Phosphoglycoisomerase 2
PiCUL1-G	*Petunia inflata* Cullin 1
PiSBP1	*Petunia inflata* S-RNase Binding Protein 1
PiSCF	*Petunia inflata* S-locus F-box protein
PrpS	*Papaver rhoeas* pollen *S*
PrsS	*Papaver rhoeas* stigma *S*
QTL	Quantitative Trait Locus
Rbx1	Ring-HC finger protein
RFLP	Restriction Fragment Length Polymorphism
RIL	Recombinant Inbred Line
RNA	Ribonucleic acid
RNAi	RNA interference
RNase	Ribonuclease
RPKM	Reads per Kb of exon model
rpm	rotation per minute
SC	Self-compatibility
SCF	Skp1-Cullin-F-box
SCR	*S*-locus cystei-rich protein
SD	Segregation Distortion
SDS	Sodium Dodecyl Sulfate
SI	Self-Incompatibility
Skp1	S-phase kinase-associated protein
SLB	S-locus F-box protein
SLF	S-locus F-box protein
SLG	*S*-locus glycoprotein
SNP	Single Nucleotide Polymorphism

SP	Self-Pollination
SP11	*S*-locus pollen protein
SRK	*S*-locus receptor kinase
ss-DNA	Single strand DNA
SSH	Suppression Subtractive Hybridization
SSI	Sporophytic Self-Incompatibility
SSK1	SLF-interacting Skp1-like 1
ssp	sub-specie
SSR	Simple Sequence Repeats
STS	Sequence Tagged Sites
Ta	Annealing Temperature
TE-Rnase	Tris-EDTA Rnase A buffer
THL	Thioredoxin *h* protein
TILLING	Targeting Induced Local Lesions in Genomes
TIR1	Transport Inhibitor Response 1
Tm	Melting Temperature
TPR	Tetratricopeptide

Chapter 1:

Introduction

1.1. Perennial ryegrass

Perennial ryegrass (*Lolium perenne* L.) is a tufted grass which belongs to the family of Poaceae, broadly named grasses, and is an important plant for agriculture, livestock farming and landscaping. The plant is widely distributed throughout the world, with a preference for temperate regions, and is native to Europe, temperate Asia and North Africa (Terrell E., 1968). In Western Europe, permanent grassland represents 17 % of the total land area and in the U.K., grassland represents around 50% of the total land area (12 Mha), of which 1.7 Mha is sown with *Lolium* species (Humphreys *et al*., 2006). *Lolium* is usually grown in mixture with grasses, clovers or lucerne. Due to its large habitat range, perennial ryegrass is used largely in agriculture, as a forage grass but also a grazing grass for milk and meat production; 80% of the world's cow milk production and 70% of the world's beef and veal meat are produced from grasslands (Wilkins and Humphreys, 2003). It is also an important component of recreational turf, in gardens and on sports pitches as well as a potential as a feedstock for sustainable bioenergy.

Lolium perenne is a diploid (2n = 2x = 14) with an estimated genome size of 2.6 Gbp (Kopecký *et al*., 2010). Its genome has not been published yet but its sequencing is on the way. Therefore, for any genetic work, information from the genomes of closely related fully sequenced plant species are used. *Lolium perenne* is part of the Poaceae family, grouping monocotyledonous flowering plants, which contains numerous agriculturally important plants, among which rice, wheat, barley, oat, maize, millet and sugarcane to name only few. Among the grass plants, the genomes of rice (The Rice Genome Sequencing Project, 2005), *Sorghum* (Paterson *et al*., 2009), *Brachypodium* (The International *Brachypodium* Initiative, 2010) and maize (Maize Genome Sequencing

Project, 2009) are available, and can be used in syntenic studies to infer genetic organisation in *Lolium perenne*.

Breeding programmes for *Lolium perenne* varieties, after focusing on the yield and the dry matter content are now turning their focus towards traits affecting the feeding value (dry matter digestibility, crude protein content, water-soluble carbohydrate, fibre contents) as well as good persistency and tolerance/resistance to biotic and abiotic stresses. Perennial ryegrass is an outbreeding plant, due to its self-incompatibility (SI) system, which has for consequence its high genetic load with many recessive alleles, not expressed in heterozygous forms. Because of its SI mechanism, varieties of perennial ryegrass bred traditionally are heterogeneous and not resulting from a cross between two pure lines as it is the case for self-compatible species such as wheat. An alternative breeding strategy would be to produce partially inbred lines with a restricted range of SI genotypes. These, when crossed with other lines that have different SI genotypes would be more compatible between lines than within thus producing an 'F1' population that results from the majority of seed being derived from an inter-population rather than an intra-population cross. In this way the population will consist of advantageous heterozygous combinations, and therefore maximize the potential hybrid vigour. Knowledge of the SI system may well lead to insights into manipulating the breeding systems of ryegrass but also major grass crops as they are sharing the same SI system.

1.2. Introduction to self-incompatibility

Reproduction is a biological process which is a fundamental feature of all known life. Reproduction can either be sexual or asexual, with the creation of gametes, or by

vegetative reproduction (branching or tillering), respectively. Sexual reproduction requires the involvement of both female and male sexes. The two sexual organs can be borne on two separate plants in dioecious species (e.g. European holly), or on the same plant in monoecious species. In monoecious species, anthers (male organ) and pistil (female organ) can be borne on the same flower (hermaphrodite) or separately (imperfect flower). The reproduction can either involve one individual when there is a self-reproduction (self-pollination) or two separate individuals, cross-pollination between two different plants. Asexual progeny genomes are identical to the parent plant unless somatic mutations occur. Sexual reproduction however, enables recombination of genes during the creation of the gamete and therefore increases the genetic variation in the progeny; even so genetic diversity will be reduced in the case of a self-pollination (inbreeding) compared to a cross-pollination (outbreeding) which is the merger of two haplotypes (Holsinger, 2000).

The evolution of the reproduction towards the sexual reproduction is thought to be due to the environmental adaptation advantages (biotic and abiotic stresses adaptation) of having a wider range of phenotypes in the progeny (Maynard Smith, 1971; Pandey, 1977). But self-pollination, because of its inbreeding nature, can cause inbreeding depression, leading to the fixation of genes and standardisation of populations that can reduce the vigour of plants over successive generations. To prevent inbreeding and maintain diversity, many angiosperms have developed one of several systems of self-incompatibility (SI).

SI was first defined by some authors (Darwin, 1876; Sutton, 1918; East and Mangelssdorf, 1925) as the inability of fertile plants to reproduce after selfing. It was then defined by Dreux de Nettancourt (1977) as “the inability of a fertile hermaphrodite seed plant to produce zygotes after self-pollination”. There are various possibilities for a plant to express SI. Heteromorphic SI is largely used by plants to prevent self-pollination, over 167

genera have been found to use it (Ganders, 1979), and results of the flower shape or gamete maturity. Herkogamy is one type of heteromorphic SI, where a spatial separation between anthers and stigma prevents the pollination, like in the classic pin-thrum dimorphism of *Primula* species (Stirling, 1932). The style length is bigger than the anther in pin-flower, placing the stigma (pollen receptor) above the anthers (pollen producer); but the situation is opposite in thrum-flower, with the anthers above the stigma. Another type of heteromorphic SI is dichogamy which is a time separation between the maturity of the pollen and the pistil (protandry, when the pollen is mature before the receptive stage of the stigma and protogyny which is the opposite), such as the avocado tree (Darwin, 1862). The molecular genetics of heteromorphic SI is not completely understood (McCubbin et al., 2006) even if, in the case of the herkogamy of *Primula* ssp., a single *S*-locus has been identified, with at least seven genes involved in SI (Kurian and Richards, 1997).

In these systems, there is also a physiological barrier to self-fertilisation. When cross-pollination is not encouraged by the morphology of the flower, outcrossing is entirely dependent on physiological recognition of pollen by styles: homomorphic SI. In these systems, the self-pollen will be rejected by the stigma whereas pollen from a different plant (non-self) may germinate and lead to fertilization of the female gamete. Homomorphic SI is a widespread mechanism that prevents inbreeding and occurs in over half of the angiosperm (de Nettancourt, 2001). It is genetically controlled by one or more multi-allelic loci; SI relies on complex interactions between the pistil and the pollen where the SI loci consist of at least one pollen and one stigmatic component that, essentially, are tightly linked in a supergene complex (Silva and Goring, 2001).

1.3. Homomorphic self-incompatibility

1.3.1. Gametophytic and sporophytic SI.

Homomorphic SI is classified into two groups, gametophytic (GSI) and sporophytic (SSI) (deNettancourt, 1977), according to the pollen genotype that is recognised by the stigma. In most cases, SI is genetically controlled by a single multi-allelic locus, the *S*-locus, mainly composed of two genes, one of which will be expressed in pollen and the other one in stigma/stylar tissues, their interaction leading to the compatibility or not of the pollination. The compatibility mechanism between the *S*-pollen and the *S*-stigma/style will be dependent of the part of the pollen that is recognised by the stigma. In the case of GSI, the haploid pollen genome (gamete) is recognised by the stigma but in SSI, the diploid exine of the pollen grain (sporophyte) will determine the phenotype of the SI reaction (deNettancourt, 1977). For a pollination to be compatible, the pollen and stigma *S*-alleles have to been different, but because of the ploïdy of the pollen in SSI, both alleles from the pollen parent have to be different from the female for the pollination to be compatible (see Figure 1.1).

In a compatible pollination, the pollen tube growth will initiate when the pollen is deposited on the stigmatic surface of the pistil. The pollen tube will grow between the cells of the stigma and through the transmitting tract of the style until it reaches the ovary. Then, the sperm cells of the pollen will be injected into the ovule and fertilisation will lead to the formation of the embryo (Palanivelu and Preuss, 2000). In a SI reaction, the "self" pollen is rejected during the pollination process, at varying development stages depending on the species and the SI type, either on the stigma surface, during the growth of the pollen tube in the style, in the ovule or even post-fertilization. In the case of SSI, the pollen is rejected rapidly at the stigma surface, soon after attempts at germination, and pollen tubes only

make cursory attempts to penetrate the style (deNettancourt, 1977). Gametophytic SI systems, on the other hand, have been characterised in Rosaceae and Solanaceae as a stylar reaction: the arrest of incompatible pollen tubes occurs in the style after considerable pollen-tube growth and not on the stigma surface (de Nettancourt, 2001). But this relation between the site of the pollen rejection and the type of SI is not absolute as for example, in *Papaver rhoeas* and in grasses, the response of the GSI occurs rapidly, at the stigma surface (Jordan *et al*., 2000a; Shivanna *et al*., 1982; Heslop-Harrison, 1982).

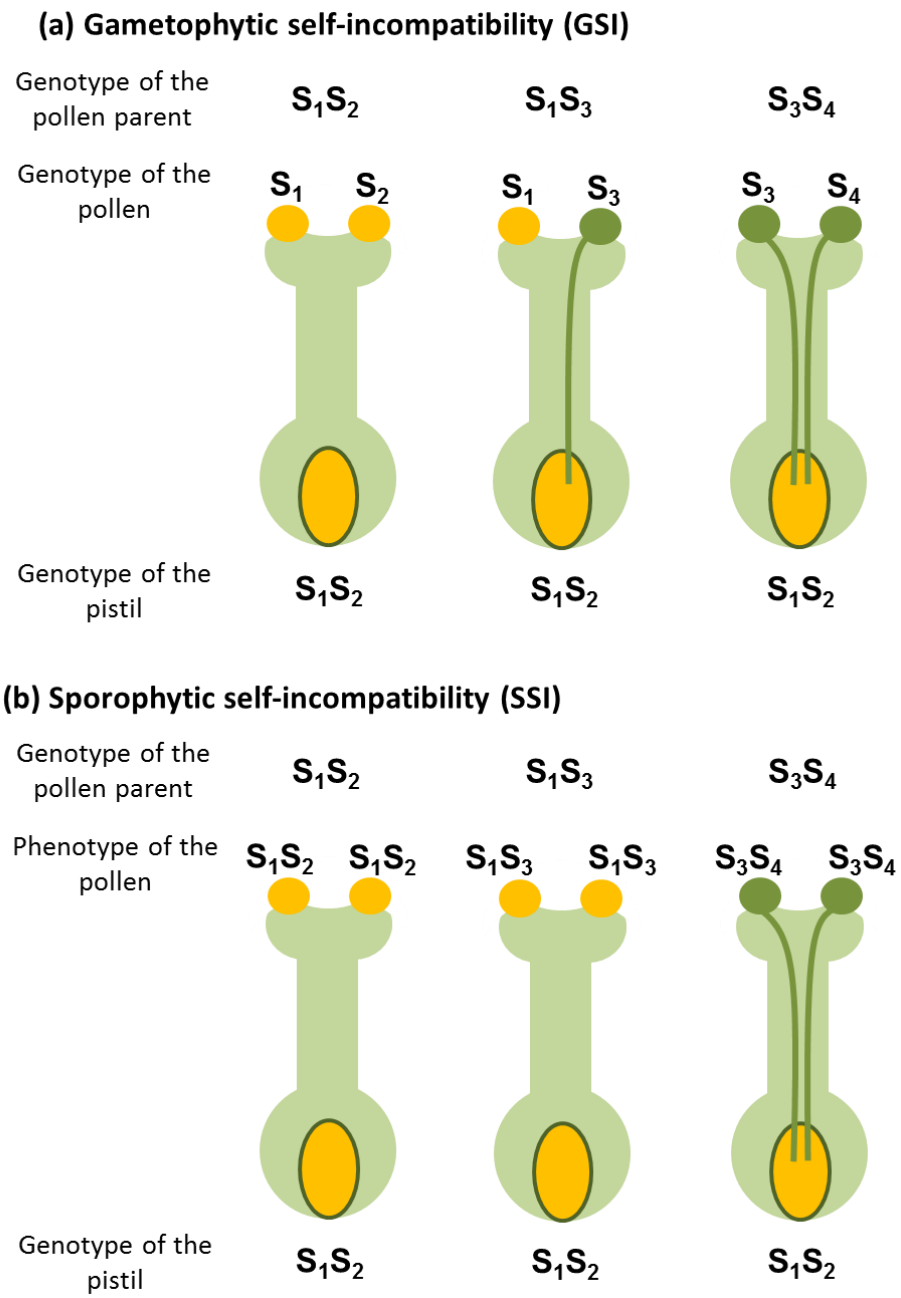


Figure 1.1: Diagram of the single locus gametophytic SI (GSI) and sporophytic SI (SSI) mechanisms. (a) In the case of GSI, the genotype of the pollen that will be recognised is haploid and for a successful pollination, the haplotype of the pollen needs to be different from both of the pistil. In this case, only pollen with *S*-alleles different from S_1 and S_2 will be compatible; a male plant with the alleles S_1S_3 will be partially compatible, through its pollen S_3. (b) In the case of SSI, the haplotype of the pollen responsible for the phenotype of the reaction is diploid, with the same *S*-alleles as the male parent. For a successful pollination, both *S*-alleles of the pollen have to be different from the pistil *S*-alleles. In this case, the male plant S_1S_3 is fully incompatible as the allele S_1 is common between the male and the female plants.

The mechanisms involved in pollen tube arrest are diverse and the site of action of the SI gene products is not always understood. The stigma's morphology and physiology are different between species but these characteristics can be grouped into the two SI groups: SSI and GSI. Dry stigma is found to be typical of SSI (Brewbaker, 1959) unlike wet stigma, characteristics to GSI (Heslop-Harrison and Shivanna, 1977), which is thought to contain promoters of pollen-tube growth (Franklin *et al*., 1995). But whatever the type of SI, the SI reaction is thought to occur on or in the pollen grain or the pollen-tube, as compatible and incompatible pollen can germinate simultaneously in the pistil (Linskens, 1965). However, exceptions are possible as in the case of the Brassica SSI system, the rejection of the pollen is thought to be due to the stigma protein degradation, localized near the self-pollen (Stone *et al*., 2003).

SSI and GSI are different in the mechanism by which the pollen and the pistil genotypes interact but they also differ in characteristics specific to one another. A few SI species are well studied today and can be used as models to study other SI systems. For the SSI system, present in families like Asteraceae, Betulaceae and Brassicaceae, only one mechanism has been extensively described to date, the mechanism in Brassicaceae where both pollen and pistil components has been isolated and identified (Hinata and Nishio, 1978; Stein *et al*., 1991). For GSI, two main systems have been extensively studied: the S-RNase system in Solanaceae (Lee *et al*., 1994; Murfett *et al*., 1994) and in Rosaceae (Cheng *et al*., 2006) and the system in *Papaver rhoeas* (Franklin-Tong & Franklin, 1992).

1.3.2. Sporophytic self-incompatibility

The most described SSI system is the *Brassica* system, studied largely in *Brassica* species. But SSI is not unique to these species even so researches were more focused on them. Currently, the SSI system in *Ipomoea trifida* (Convolvulaceae family) (Kowyama *et al*., 2000) and *Senecio squalidus* (Oxford Ragwort, Asteraceae family) (Hiscock, 2000; Hiscock *et al*., 2003a) are being actively studied and the results so far have shown that, like GSI, SSI can be characterised by different mechanisms, governed by different genes, and therefore novel SSI mechanisms have been described (reviewed by Hiscock and Tabah, 2003b). However, up to date, only the *Brassica* system is largely understood and both pollen and pistil *S* genes have been identified and downstream genes involved in the SI response have been characterized.

In *Brassica*, the pistil *S* genes involved in the SI response have been identified as an *S*-locus glycoprotein (SLG) (Hinata and Nishio, 1978) and an *S*-locus receptor kinase (SRK) (Stein *et al*., 1991). The SLG protein is located in the stigma papillae cell wall whereas the SRK receptor has an extra-cellular domain, a transmembrane-spanning domain and a cytoplasmic serine/threonine kinase domain, but is also located in the papillae cells (Stein *et al*., 1991). The role of SLG in the SI response is not fully known but it seems that the protein enhances the SRK mechanism (Takasabi *et al*., 2000) and is not essential to SI: in the self-incompatible *Arabidopsis lyrata*, SLG is absent (Schierup *et al*., 2001) and in *Brassica oleracea*, SLG genes are not functional (Suzuki *et al*., 2000). The intracellular domain of the SRK receptor was found to react with two thioredoxin *h* protein, THL1 and THL2 (Bower *et al*., 1996). When THL1/2 is attached to the SRK, the kinase activity of the SRK is suppressed, inhibiting the cascade of reactions responsible for pollen rejection; allowing the pollen to germinate. The male determinant is the *S*-locus

cysteine-rich protein (SCR), also called *S*-locus pollen protein 11 (SP11) (Schopfer *et al*., 1999; Takayama *et al*., 2000). This small protein (less than 10 kDa) is anchored on the pollen coat and its *S*-allele will be recognised by the extra-cellular domain of the SRK receptor (Kachroo *et al*., 2001; Takayama *et al*., 2001), but the exact mechanism of recognition between those two proteins is not well known. SCR/SP11 and SRK are closely linked, suppressing recombination which would otherwise lead to the breakdown of SI function (Casselman *et al*., 2000).

In the case of a self-pollination, the recognition of the self-pollen will lead to a cascade reaction (see Figure 1.2). First, the protein THL1/2 will be released, inducing the SRK kinase activity (Cabrillac *et al*., 2001). SRK will phosphorylate another component, ARC1 (Armadillo repeat motif-containing protein) which is a protein also included in the cascade reaction in response to SI (Gu *et al*., 1998). The activation of ARC1 leads to a series of activations and to the inhibition of pollen germination, but the molecules involved later are not known up to now. ARC1 is a cytosolic protein that can be transported into the nucleus but in case of SI, its phosphorylation will lead to its transport to the endoplasmic reticulum (Stone *et al*., 2003). The current hypothesis is that the ARC1 protein, containing an E3 ubiquitin domain as well as a U-box domain, will ligate to specific substrates and by ubiquitination will induce their transport to the proteasome for degradation (Stone *et al*., 2003).

Two other candidate molecules involved in the cascade have been discovered recently: the *M*-locus protein kinase (MLPK) and the Exo70A1 (Murase *et al*., 2004; Samuel *et al*., 2009). MLPK, identified through a map-based cloning approach, has been found to be localized in the plasma membrane of the stigmatic papillae and to interact with SRK proteins (Kakita *et al*., 2007). However, even if the part of MLPK in the cascade

reaction is not known yet, the protein is an important component in the SI response as the gene coding MLPK is thought to be the same one as the one mutated in the *Brassica rapa* variety Yellow Sarson, causing the breakdown of its SI (Hinata *et al*., 1983). The other component recently discovered to interact with ARC1 is Exo70A1 (Samuel *et al*., 2009). This protein is a putative component of the exocyst complex, which has a function of polarized secretion in yeast and animals (Hsu *et al*., 2004; Munson and Novick, 2006). Synek *et al*. (2006) have shown that the protein Exo70A1 was required for a successful non-self-pollination in *Arabidopsis thaliana*, suggesting that the exocyst, in which Exo70A1 as a part, is involved in the secretion of molecules essential to the pollen germination/hydration.

With these findings, the latest hypothesis for the *Brassica* SI response (see Figure 1.2) is that the activation of SRK, by SCR from "self" pollen, in combination with MLPK will phosphorylate ARC1. ARC1 will then target Exo70A1 for ubiquitination, leading to its degradation by proteolysis and therefore preventing the secretion on the stigma surface of "compatible components" in favour of pollen germination (Samuel *et al*., 2009). However, with this model, the secretion of the "compatible components" have to be localized directly on the pollen-stigma contact point as compatible pollen located next to incompatible pollen can germinate normally.

This model is not complete and more research is being done in order to understand the complete SI mechanism. Moreover, not all scientists fully agree, as Kitashiba *et al*. (2011) have been unsuccessful to show the roles of Exo70A1 and MLPK. By overexpressing Exo70A1 and inactivating MLPK in transformed *Arabidopsis thaliana* (*A. thaliana* SRK-SCR), Kitashiba *et al*. (2011) were not able to reduce or suppress the SI

response, which is expected: an excess of Exo70A1 would have compensated for its degradation and MLPK is thought to be an essential component too.

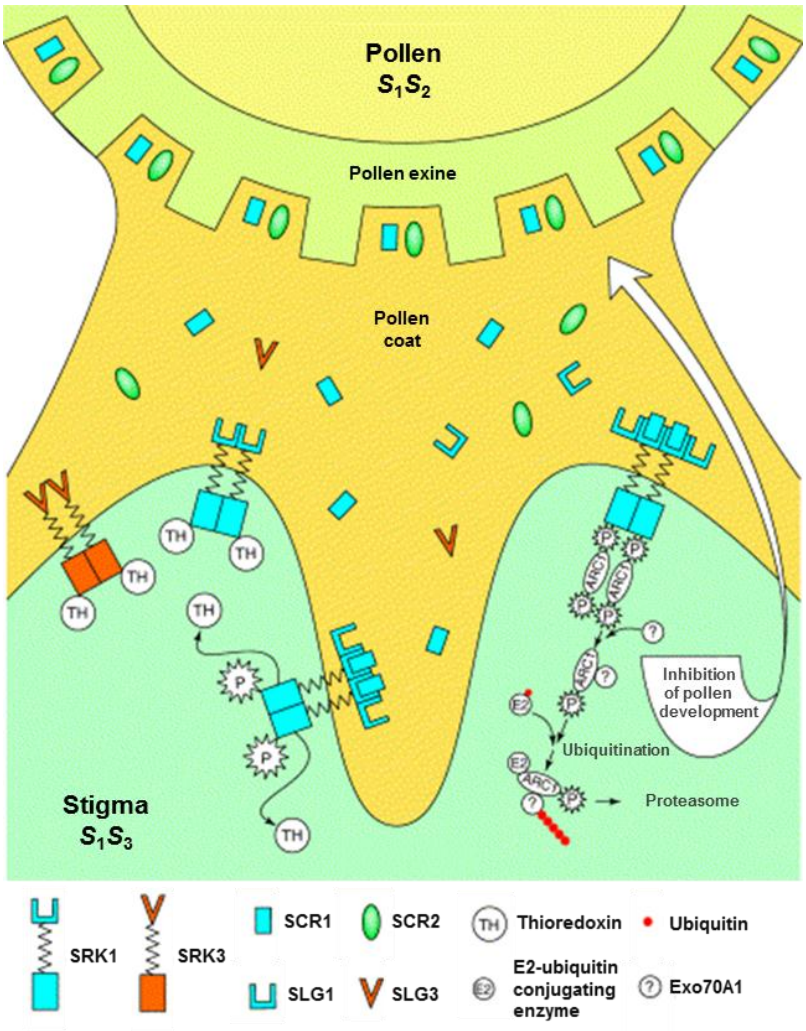


Figure 1.2: Model for the mechanism of sporophytic self-incompatibility (SSI) in Brassica. The female determinant of SSI is the *S*-locus receptor kinase (SRK); the male determinant is the *S*-locus cysteine-rich protein (SCR). Moreover, the stigma is also secreting the glycoprotein SLG, partially similar to SRK. In this model, the pollen haplotype is S_1S_2 and the female haplotype is S_1S_3 so with a common allele S_1 (represented in blue), the pollination will lead to a self-incompatibility response (both alleles have to be different for a successful pollination). SCR1 will interact with SRK1/SLG1, leading to the dissociation of THL1/2 (here annotated TH) and its phosphorylation, which is inhibited by THL1/2 in compatible pollination. The phosphoralyted SRK1 will interact with ARC1 (Armadillo-repeat motif-containing protein), which by ubiquitination will target Exo70A1 for its degradation in the proteasome. Exo70A1 is thought to be involved in the exocyst of factors enabling the germination of pollen; its degradation is therefore preventing the secretion of these factors, leading to the non-germination of "self" pollen. Figure adapted from Hiscock and McInnis (2003c).

1.3.3. Gametophytic self-incompatibility

GSI exists in major plant families such as Solanaceae, Rosaceae, Fabaceae, Papaveraceae and Poaceae. Two mechanisms of GSI have been described at the molecular level, the S-RNase mechanism in Solanaceae for example, and the S-glycoprotein mechanism in Papaveraceae.

1.3.3.1. S-RNase system

The S-RNase system is the most widespread GSI system, shared by many plants from the families Rosaceae, Solanaceae, Plantaginaceae and Rubiaceae. Most research on the GSI system is conducted on *Prunus* (Rosaceae) species such as pear, cherry and almond (Sassa *et al*., 1992; Sanzol *et al*., 2006; Tao *et al*., 1997; Tao *et al*., 1999), *Nicotiana tabacum* (tobacco; McClure *et al*., 1989), *Solanum tuberosum* (potato; Luu *et al*., 2000), *Lycopersicon esculentum* (tomato; Parry *et al*., 1998), *Antirrhinum* (dragon flowers, Xue *et al*., 1996) and *Petunia hybrid*a (garden petunia; Robbins *et al*., 2000). Among all these plants, the GSI system is similar, the S-RNase system, with some variations in the gene sequences but not in their functions. Many reviews of this GSI system are available today (Wang *et al*., 2003; Kao and Tsukamoto, 2004; Hua *et al*., 2008; Sims and Robbins, 2009; Meng *et al*., 2011).

This GSI system is characterized by the ability of both compatible and incompatible pollen to hydrate and undergo a normal germination on the surface of the stigma. It is only during the growth of the pollen tube, in the style, that the incompatibility response occurs, characterised by callose plugs developing along the tube growth (Ebert *et al.*, 1989). Therefore, compatible and incompatible pollen tubes at first appear

morphologically similar as the SI response is acting at a later stage, in the style, as has been reported in general for GSI (de Nettancourt, 2001).

The S-RNase system is not yet fully elucidated but the genes involved in its control have been mapped and well-studied. The self-incompatibility trait is genetically controlled by a single polymorphic *S*-locus (de Nettancourt, 1977). In this *S*-locus, at least two tightly linked genes are present, one *S* protein controlling the function expressed in the pistil and the other *S* gene expressed in the pollen. The pistil *S* component was first identified in *Nicotiana alata* (Anderson *et al*., 1986; McClure *et al*., 1989) as a cytosolic protein showing ribonuclease activity, the S-RNase. These S-RNases are basic glycoproteins of around 32 kDa, with a catalytic function which has been shown, by genetic engineering, to be essential for the rejection of the incompatible pollen (Murfett *et al*., 1994; Lee *et al*., 1994). The S-RNase gene has been sequenced in many species of the Solanaceae, and for a large number of *S*-alleles, shows variations between sequences with between 39% and 98% sequence identity (Anderson *et al*., 1989; Clark *et al*., 1990; Ai *et al*., 1990; Xu *et al*., 1990; Ioerger *et al*., 1990; Kheyr-Pour *et al*., 1990). The sequence of the S-RNase gene is composed of five conserved regions C1-C5, as well as two hypervariable regions HVa and HVb (Ioerger *et al*., 1991; Tsai *et al*., 1992). Deletions and insertions in these two hypervariable regions have been found to be linked with the *S*-allele diversity, leading to the conclusion that HVa and HVb were necessary but not sufficient to determine the *S*-allele specificity of the pistil (Kao and McCubbin, 1996; Zurek *et al*., 1997; Matton *et al*., 1997).

The male component of this GSI has been identified as an *S*-protein F-box gene named SLF or SFB in *Antirrhinum* of the Scrophulariaceae (Lai *et al*., 2002), *Prunus* (Entani *et al*., 2003; Ushijima *et al*., 2003; Yamane *et al*., 2003; Wang *et al*., 2004;

Sonneveld *et al*., 2005) and *Petunia* of the Solanaceae (Qiao *et al*., 2004; Sijacic *et al*., 2004). The gene coding for the PiSLF (*Petunia inflata* SLF) was found to closely map to the pistil-expressed S-RNase gene (Wang *et al*., 2004), with some hypervariable coding regions, responsible for the *S*-specificity of the SLF protein (Ushijima *et al*., 2004; Ikeda *et al*., 2004; Sonneveld *et al*., 2005). SLF is thought to be part of an SCF complex (Skp1-Cullin-F-box), which is a type of E3 ubiquitin ligase complex, linking a specific substrate to three ubiquitin enzymes (E1, E2 and E3) (Skowyra *et al*., 1997). The addition of poly-ubiquitin by the enzymes will target the substrate for degradation by the 26S proteasome (Hershko and Ciechanover, 1998). In this GSI system, in *Antirrhinum*, it is thought that the SLF is part of a conventional complex, containing four proteins: Cullin, Rbx1 (RING-HC finger protein), SLF and a novel Skp1 protein, called SSK1 (SLF-interacting SKP1-like 1) (Huang *et al*., 2006; Zhao *et al*., 2010). However, in *Petunia inflata*, no Skp1 or Rbx1 were found in the PiSCF complex, grouping PiSLF, PiCUL1-G (*Petunia inflata* Cullin 1) and PiSBP1 (*Petunia inflata* S-RNase Binding Protein 1, a RING-HC like protein) (Hua and Kao, 2006). Therefore, it seems that the SCF complex involved in the SI response could be different between species.

A model for the interaction between the pistil and the pollen *S*-components has been drawn (Meng *et al*., 2011): the S-RNases produced by the pistil are transported into the pollen tube where they are selectively targeted for degradation by the SLF (see Figure 1.3). In the case of an incompatible pollination between a pistil S_1S_2 and a pollen S_1, the S_1-RNase and S_2-RNase are transported into the pollen tube, where SLF-S_1 will recognize S_2-RNase as different, and therefore compatible, but S_1-RNase as incompatible; SLF-S_1 will target S_2-RNase for degradation, preventing degradation of pollen ribosomal RNA.

However, pollen rejection still occurs through the degradation of its RNA by in this example, the recognition of SLF-S_1 by S_1-RNase.

This model is supported by the observation of a competitive interaction between pollen *S*-alleles in diploid pollen (Brewbaker and Natarajan, 1960; Entani *et al*., 1999; de Nettancourt, 2001). In the case of a diploid pollen with two identical *S*-alleles (homoallelic pollen), only one type of SLF will be expressed in the pollen, therefore not disturbing the SI mechanism. However, in the case of a heteroallelic pollen (presenting two different *S*-alleles), two different SLF will be expressed, breaking down the SI mechanism. For example, the diploid pollen S_2S_3 will not recognize the self pistil S_2S_3 as SLF-S_2 will target S_3-RNase for degradation and SFL-S_3 will target S_2-RNase, leading to a normal pollination (see Figure 1.3 C).

Another finding supporting this model is the S-RNase localisation. S-RNases were found to be highly expressed in the pistil and the style, and in higher quantity in developing flowers (Cornish *et al*., 1987). The mechanism of transfer of the S-RNase through the membranes and into the pollen tube cytoplasm is not known yet but it has been shown that this transport is not S-haplotype-specific (Luu *et al*., 2000; Goldraij *et al*., 2006), ruling out the "gatekeeper" model described by Kao and Mc Cubbin (1996), where it was thought that the pollen *S*-component was a *S*-haplotype specific transporter, allowing similar *S*-haplotype S-RNase to enter the pollen tube, inhibiting its growth (hypothesis made when the pollen *S*-component was unknown). After transport, S-RNases have been found in the cytoplasm of the pollen tube (Luu *et al*., 2000), which is also the cellular localisation of SLF (Hua *et al*., 2007; Meng *et al*., 2009). However, Goldraij *et al*. (2006) have observed that the S-RNase were first compartmentalised in a vacuole before being released into the cytoplasm.

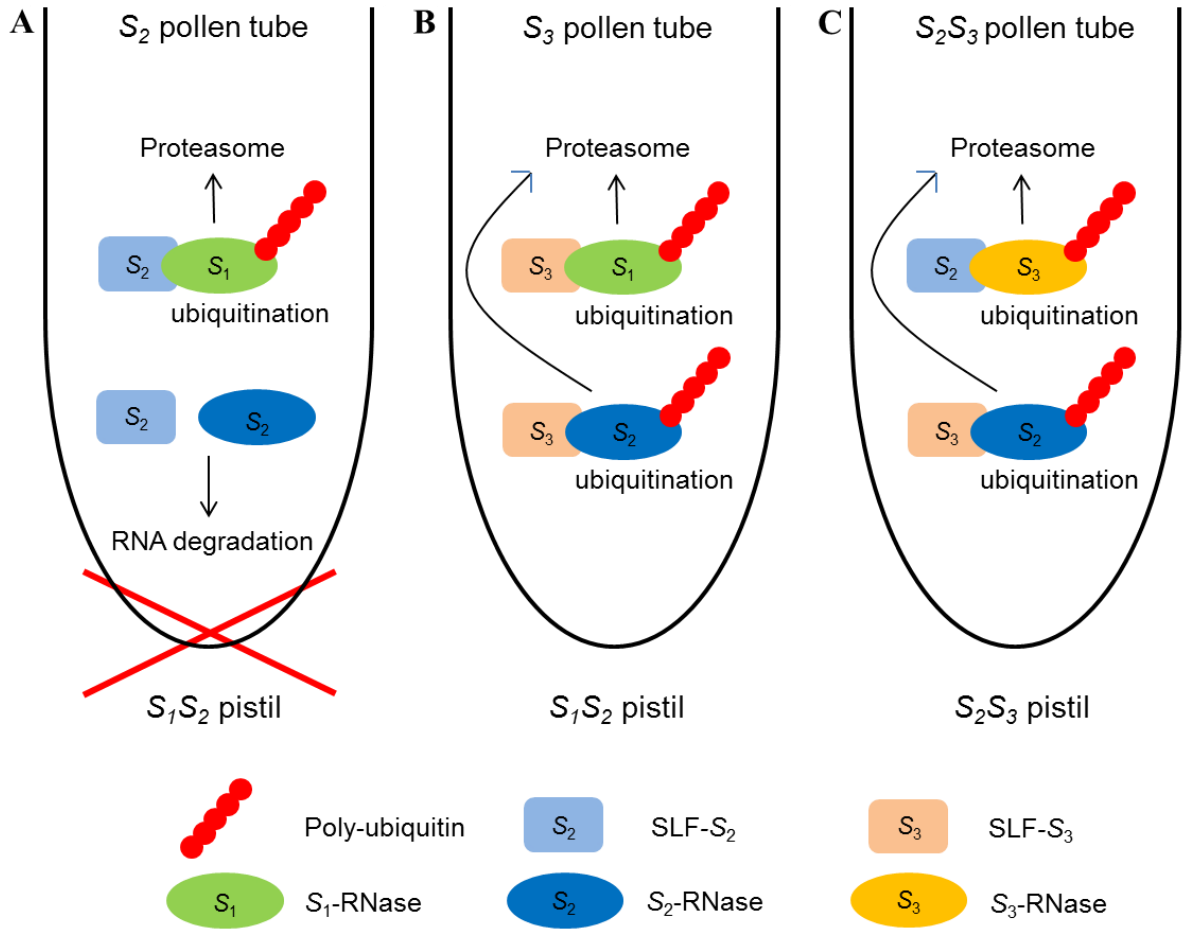


Figure 1.3: Model for the GSI S-RNase mechanism involved in self-pollen rejection in Rosaceae, Solanaceae and Scrophulariaceae. Allele specific S-RNases are produced by the pistil and transferred into the pollen tube where they are recognised by SLF as self or non-self. In the case of non-self-pollen, the SLF will recognise the S-RNase and ligate it to E1-E3 enzymes, resulting in the S-RNase ubiquitination, leading to its degradation in the proteasome. However, in the case of self-pollen, the S-RNase will not be ubiquitinated and therefore its toxicity will degrade the pollen RNA, leading to the pollen tube growth arrest. (A) Incompatible pollination as S_2-RNase is not degraded as the pollen contains SLF-S_2. (B) Compatible pollination as the pollen *S*-allele is not common with any pistil S alleles. Both S_1-RNase and S_2-RNase will be degraded and the tube will continue to grow normally. (C) Competitive interaction between pollen *S*-alleles in heteroallelic pollen. As pollen contains both S_2 and S_3 alleles, the pollen tube will contain SLF-S_2 and SLF-S_3 which will recognise S_3-RNase and S_2-RNase respectively, targeting them for degradation and leading to a compatible reaction even so pistil and pollen share common *S*-alleles. Figure modified from Meng *et al*. (2011).

More work is needed to fully understand this SI mechanism even if the main *S*-components are known. Not all the proteins involved in the cascade reaction leading to the

rejection of the "self" pollen are known and it seems that, within this mechanism, there are variations in the proteins involved. The S-RNase SI system expressed in these distantly related families (Solanaceae, Rosaceae and Scrophulariaceae) however has been found, through phylogenetic analysis, to have a common origin, an ancestor from which approximately 75% of the dicotyledons stem from (Igic and Kohn, 2001). This mechanism could therefore be present in many more families where SI is promoting outbreeding.

1.3.3.2. S-glycoprotein system: the Papaveraceae system

The S-glycoprotein system is another type of GSI, unique to the Papaveraceae, in which SI was first studied by Lawrence (1975) in *Papaver rhoeas*. The mechanisms involved in the self-pollen recognition and its rejection are well studied, with both female and male S-components isolated. The mechanism is governed by a single polyallelic *S*-locus, containing at least two genes, but in this system, unlike the S-RNase GSI, no dominance relationships have been observed between *S*-alleles (de Nettancourt, 2001). The system is reviewed by McClure and Franklin-Tong (2006) and Wheeler *et al*. (2010).

The pistil component was isolated from stigma by Foote *et al*. (1994) using isoelectric focusing gels to separate the stigmatic proteins. The pistil *S*-component was named *S* protein but was later renamed *PrsS* (*Papaver rhoeas* stigma *S*) (Wheeler *et al*., 2009). *PrsS* is a small protein (approximately 15 kDa) secreted by the stigmatic papilla cells (Foote *et al*., 1994). The protein coding genes have been identified and analysed for four different alleles, revealing high polymorphism: 40% to 46% of difference were observed between alleles (Walker *et al*., 1996). However, the protein secondary structure is highly conserved, with two hydrophilic regions identified to be essential in the biological

activity of the protein (Kakeda *et al*., 1998). *PrsS*, when discovered, was a novel protein so its function was unknown. It was however suspected that, because the protein was secreted and it has a signal peptide, *PrsS* was a type of signalling ligand. Today, *PrsS* is part of a protein family, SPH (*S*-Protein Homologue), specific to plants (Ride *et al*., 1999) and mainly expressed in floral tissues.

The male *S*-determinant was isolated by Wheeler *et al*. (2009) by sequencing the region close to the *PrsS*, as both female and male component of the SI mechanism have to be closely mapped to avoid recombination and SI breakdown. The gene *PrpS* (*Papaver rhoeas* pollen *S*) was found to code a small novel protein (approximately 20 kDa), highly hydrophobic with several predicted transmembrane passes (Wheeler *et al*., 2010) and localised in the pollen tube plasma membrane (Poulter, 2009; Wheeler *et al*., 2009). Several alleles of *PrpS* have been isolated, revealing a high polymorphism, similar to the one observed for *PrsS*, between alleles (40% to 50% of divergence) (Wheeler *et al*., 2009). The protein specificity was also demonstrated by an *in vitro* test in which pollen germination was induced and antisense oligodeoxynucleotide of specific *PrpS* alleles was added, inducing a self-incompatibility reaction in pollen grain with the same allele (Wheeler *et al*., 2009).

Early experiments showed that the concentration of intracellular cytosolic free calcium ($[Ca^{2+}]_i$) increased in incompatible pollination but did not change in a compatible reaction (Franklin-Tong *et al*., 1993). This increase in the $[Ca^{2+}]_i$ lead to the disappearance of the $[Ca^{2+}]_i$ gradient (Franklin-Tong *et al*., 1997), which is necessary for pollen tube growth as it is a directional messenger (Malho *et al*., 1994; Pierson *et al*., 1996). The increase in $[Ca^{2+}]_i$ is responsible for the depolymerization of F-actin which composes the cytoskeleton of the pollen tube, leading to a rapid alteration of the cytoskeleton, within the

first minute of the SI reaction (Geitmann *et al*., 2000; Snowman *et al*., 2002). Moreover, DNA fragmentation was observed in incompatible pollen specifically (Jordan *et al*., 2000b). DNA fragmentation is a common feature of programmed cell death (PCD), therefore it is thought that the increase of $[Ca^{2+}]_i$ is ultimately causing PCD (Thomas and Franklin-Tong, 2004; Bosch and Franklin-Tong, 2007). The DNA fragmentation in PCD is due to caspase, protein with a capsase-3-activity, but not previously found in plants (Woltering *et al*., 2002). However, good evidence for caspase-3-like activity has been found in plants (Lam and del Pozo, 2000; Schaller, 2004). Nevertheless, Thomas and Franklin-Tong (2004) have demonstrated, by treating incompatible pollen tube with caspase inhibitor, that DNA fragmentation in poppy pollen observed after an SI pollination was due to a caspase-3-like activity.

The model that is proposed for this GSI mechanism is that *PrsS*, secreted by the pistil will be recognized by the pollen tube, thanks to *PrpS*, a pollen-expressed receptor linked to a channel, leading to an increase in $[Ca^{2+}]_i$ (see Figure 1.4). The SI response is *S*-specific: *PrsS* will interact with *PrpS* if their *S*-haplotypes share one common *S*-allele, suggesting that *PrsS*-*PrpS* interaction is a ligand gating mechanism, *PrpS* being either an ion channel or a receptor coupled with a ion channel. *PrsS* has been shown to activate a calcium intake by the pollen but also a potassium intake, leading to the hypothesis that the ligand gating activated by *PrsS* is nonspecific cation conductance (Wu *et al*., 2011).

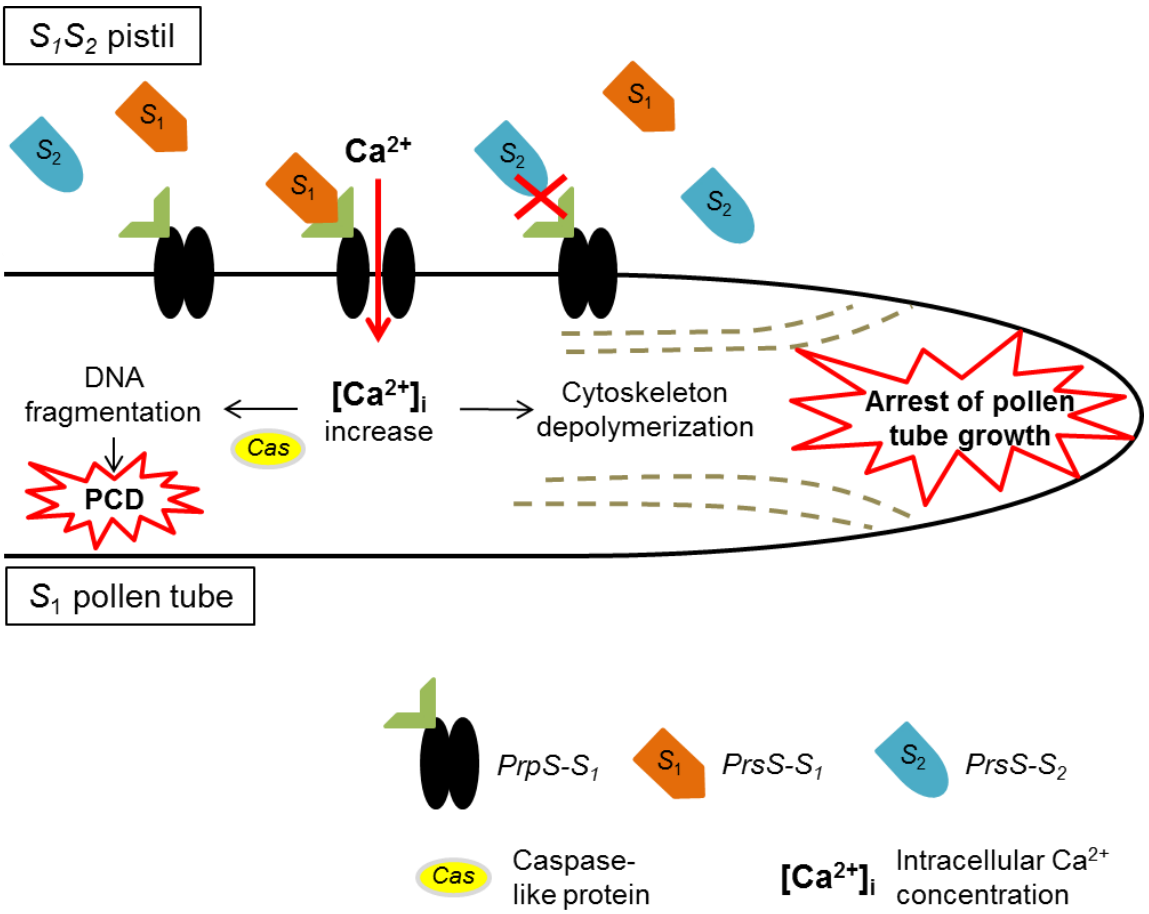


Figure 1.4: Model for the GSI S-glycoprotein system unique to the Papaveraceae. The SI system is a ligand gating type. The pistil *S*-specific component *PrsS*, secreted by the stigmatic cells into the extra cellular matrix, will be recognized by the pollen *S*-specific component, *PrpS* which is an ion channel or part of one. In a self-incompatible pollination, the complex *PrsS-PrpS* will lead to an intake of calcium by the pollen tube, resulting in an increase in $[Ca^{2+}]_i$ and therefore the loss of the gradient necessary to the pollen tube growth. Affected by the increase in $[Ca^{2+}]_i$ is the cytoskeleton with a depolymerisation of the F-actin and the DNA fragmentation, through caspase-3-like activity, leading to the PCD of the pollen tube. Figure adapted from Wheeler *et al.*, 2010.

This GSI system has been transferred to *Arabidopsis thaliana* in order to see if it was possible to genetically create a self-incompatible plant from a self-compatible species, even though *A. thaliana* is from the *Brassica* family, which is characterised by sporophytic SI. The transformation was done by expressing *PrpS* gene and by applying, *in vitro*, different *PrsS* alleles. The results showed that Papaver-type SI was successfully transferred to *A. thaliana*, and that only *PrsS* and *PrpS* were necessary for the SI to be functional,

leading to the conclusion that the proteins involved in the SI reaction, with the exception of *PrsS* and *PrpS,* are existing proteins, used in other mechanisms, that have been recruited to form a new signalling pathway in *Arabidopsis* (de Graaf *et al*., 2012). These findings have considerable practical implications for creating SI in other species as they show that the downstream pathways from the initial *Papaver* SI recognition system are present in other plant families and that SI can be induced simply by transferring the pollen- and stigma-part recognition genes to normally self-compatible species in a range of crops.

1.4. Self-incompatibility in grasses.

1.4.1. Introduction

The three major crops worldwide (rice, wheat and maize) are self-compatible species. However, self-incompatibility in grasses is a widespread mechanism, present in at least 16 genera (Connor, 1979) from subfamilies such as Pooideae, Panicoideae, Chloridoideae and Arundinoideae. Within subfamilies and even within tribes or genus, not all species are self-incompatible. In the case of the tribe Poeae, *Lolium perenne*, *L.multiforum, L. rigidum* as well as *Festuca pratensis* are self-incompatible but *Lolium temulentum* and *L. remotum* are self-compatible. SI is an important mechanism among grasses with economic importance, with species such as *Miscanthus*, *Secale cereale*, *Hordeum bulbosum* and *Phalaris coerulescens.*

The grass SI mechanism is gametophytically controlled but remains unique to the grasses. SI is genetically governed by a single multi-allelic *S*-locus in many SI systems (reviewed in Yang *et al*., 2008) but in the grass GSI system, it involves at least two multi-allelic independent loci, *S* and *Z* (Lundqvist, 1954; Hayman, 1956). This unique *S-Z*

system results in differences with other SI systems known so far. In autopolyploid grasses, the SI remains effective, as it has been observed in *Secale cereale, Festuca pratense* and *Dactylis glomerata* (Lundqvist, 1957, 1962 and 1968) and *Lolium perenne* (Fearon *et al*., 1984), unlike the Solanaceae GSI S-RNase system in tetraploid Prunus for eample, where the hetero-allelic pollen results in the breakdown of the SI system (Hauck *et al*., 2006).

1.4.2. Physiology of self-incompatibility in grasses

In a self-incompatible pollination, self-incompatible pollen is rejected during the pollination process, generally quickly after contact between stigma and pollen; five to ten minutes in *Phalaris tuberosa* (Knox and Heslop-Harrison, 1971); two minutes in *Secale cereale* (Shivanna *et al*., 1982) but in some cases the pollen tube can be stopped after penetration of the stigma cuticle like in *Hordeum bulbosum* (Heslop-Harrison, 1982) or even in the transmitting tracts of the stigma, as in *Alopecurus pratensis* (Shivanna *et al*., 1982). In *Lolium perenne*, when observed under microscope and with aniline blue staining, the pollen tube appears to be very small and to not enter the stigmatic cells. Incompatible pollen tube tips of grass species appear to be covered with a callose deposit. Callose along the pollen tube also appears in the S-RNase GSI system.

The grass SI system is classified as gametophytic SI as the haplotype of the pollen is recognized by the pistil. However, many of the characteristics of this SI system have features in common with SSI. For instance, grass pollen is tri-nucleate, with one vegetative cell producing the pollen tube and two gametes. Both other well-studied GSI, S-RNase and S-glycoprotein systems, have bi-nucleate pollen, with only one gamete (Yang *et al*., 2008), but tri-nucleate pollen are found in SSI systems. Moreover, the grass stigma surface is dry,

like the Papaveraceae but unlike the stigma in the S-RNase system, which have a wet surface, with secretion believed to facilitate pollen hydration.

Wehling *et al*. (1994) have observed in rye pollen protein phosphorylation as well as the involvement of Ca^{2+}-induced signal transduction in incompatible pollination. By treating stigmas with kinase inhibitors and Ca^{2+} antagonists, Wehling *et al*. (1994) were able to breakdown self-incompatibility, suggesting that a Ca^{2+} signalling cascade was involved in the SI response. A semi *in-vivo* assay was also realized in *L. perenne*; by treating stigmas with different Ca^{2+} inhibitors, SI was partially overcome and pollen tubes growing in the style were observed (Klaas *et al*., 2011). Ca^{2+} is known to be involved in SI signalling cascade in *Brassica* SSI and in *Papaver* GSI systems. The conclusions of Klaas *et al*. (2011) and Wehling *et al*. (1994) strongly suggest that for the grass SI response leads to a Ca^{2+} cascade mechanism, even though the initial *S*-components triggering this cascade are unknown. Another hypothesis for the grass SI response is the proteolysis pathway, involved in *Brassica* SSI and in the S-RNase GSI systems, as changes in phosphorylation activity were observed in self-incompatible pollen (Wehling *et al*., 1994).

1.4.3. Genetics of self-incompatibility in grasses

The grass SI system is mainly governed by two multi-allelic and independent loci *S* and *Z* (Lundqvist, 1954; Hayman, 1956). This *S*-*Z* system has been found to operate in *Secale cereale* (Lundqvist, 1954), *Phalaris coerulescens* (Hayman, 1956), *Lolium perenne* L. (Cornish *et al*., 1979) and many other grass species reviewed by Li *et al*. (1997). With this *S*-*Z* system, successful pollination will occur if at least one of the two alleles of *S* and *Z* from the pollen is different from the *S* and *Z* alleles in the stigma (see Figure 1.5).

Because this system is governed by two loci and because of the interactions between *S* and *Z*, the level of compatibility between two plants can vary from 0% to 100% with 50% and 75% compatibility observed (see Figure 1.5), unlike the only three type of compatibility (0%, 50% and 100%) observed for the single *S*-locus GSI systems (Baumann *et al*., 2000). Another unique characteristic of this SI is that the phenotype of pollination can vary according to the direction of the cross. If a female plant $S_1S_2Z_1Z_3$ is crossed with a plant $S_1S_1Z_1Z_2$ as a pollen donor, a compatibility of 50% will be observed, with compatible pollen grain been S_1Z_2 haplotype. However, in the reciprocal cross, 75% of the pollen will be compatible (S_1Z_3, S_2Z_1 and S_2Z_3).

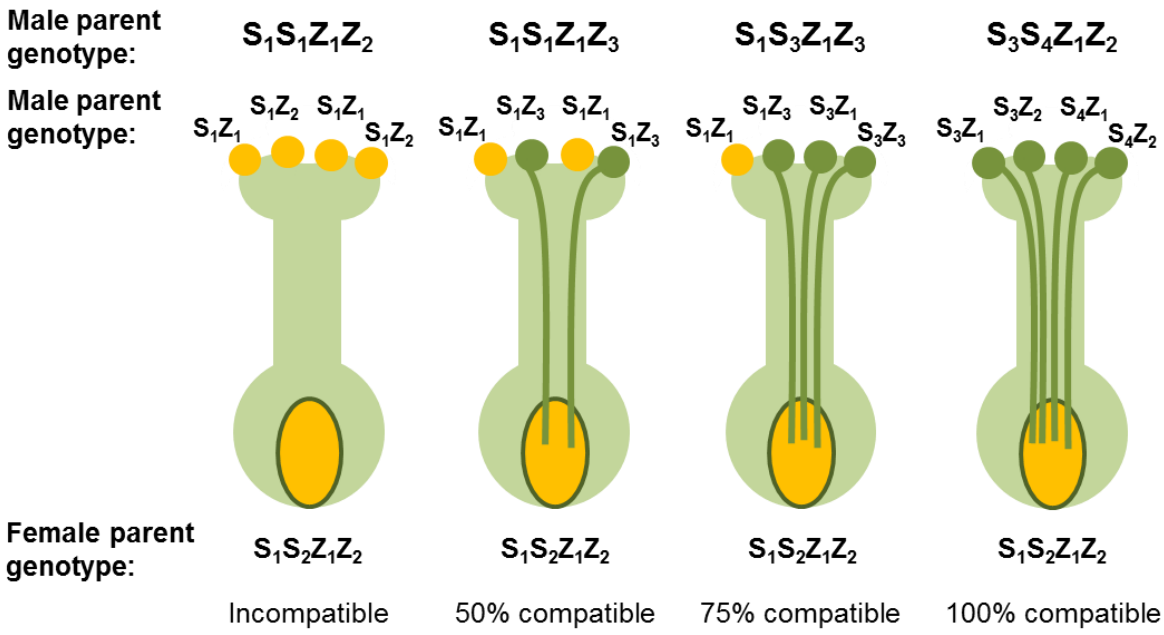


Figure 1.5: Genetic control of gametophytic self-incompatibility (GSI) by the two multiple-allelic loci system *S*-*Z*. A full incompatible reaction occurs when both *S* and *Z* alleles are common between the pollen and the pistil. But if at least one allele of *S* or *Z* is different between the two parents, there is a compatibility reaction with several degree of compatibility depending on the number of alleles that are not common: 50% of compatibility if one allele is different, either for *S* or *Z* and 75% if one allele is different for *S* and for *Z*.

Up to date, neither of the two loci have been identified but it is known that the *S* and *Z* loci are located on chromosome 1 and 2 respectively in *Secale* (Voylokov *et al*., 1998) and *Lolium perenne* (Thorogood *et al*., 2002). The *S*-locus was found to be linked to an isozyme phosphoglycoisomerase (PGI-2) and a leaf peroxidase *Prx*-7 genes in *Secale* (Gertz & Wricke, 1989) and in *Lolium perenne* (Cornish *et al*., 1980). Another gene, *Bm2*, was thought to be the *S*-gene pollen component, a protein with thioredoxin catalytic activity (Li *et al*., 1994) in *P. coerulescens* but it has now been mapped approximately 0.35cM away from the *S*-locus after more studies by Bian *et al*. (2004). Bian *et al*. (2004) delimited the *S*-locus in *P. coerulescens* to a region of 0.26cM on the short arm of chromosome 1. But this genetic distance is associated with a large physical region, the *S*-locus being in the subcentromeric region where recombination rates are expected to be reduced compared to more distal parts of the chromosome (Pedersen *et al*., 1995). Taking a comparative mapping approach i.e. using rice marker and sequence data to deduce the position of the *Lolium S*-locus can be problematic with this region because synteny between rice and Triticeae is not complete; chromosome 1 of wheat in this region is largely syntenic with rice chromosome 5 but with a small chromosome 10 insert (Sim *et al*., 2005).

Kakeda *et al*. (2008, 2009) have isolated, in *Hordeum bulbosum*, several pistil and anther-specific candidate genes using gene expression analysis, including some putative homologs of two rice genes being described with F-box motifs mapped in rice around 26 Mb on rice chromosome 5 (Os05g0532300 and Os05g0535200) and another highly polymorphic candidate gene around 6 Mb on the same chromosome in rice (Os05g0198050). A gene expression study has also been conducted in *Lolium perenne* L. by Yang *et al*. (2009) using the construction of several suppression subtractive hybridization (SSH) libraries. Eight expressed SI candidate genes for the *S*-locus have been

identified in SSH libraries and their expression was confirmed by real-time PCR. Among those candidate genes (Yang *et al*., 2009), four genes (*Can3, Can4, Can94* and *Can136*) have a protein kinase function which could imply that grass SI response is using a kinase pathway such as in the *Brassica* SSI system (SRK receptor on the stigma; Nasrallah, 2002) or downstream in the SI response such as in the *Papaver* GSI system (involvement of mitogen-activated protein kinase, MAPK; Rudd *et al*., 2003). One *S*-locus linked BAC clone *Lp*Bm2 was selected by Shinozuka *et al.* (2010) in *Lolium perenne* L. Its sequencing revealed two candidates genes syntenic with rice chromosome 5 Os05g0168800 (Spectrin repeat containing protein) and Os05g0169000 (Thioredoxin-like protein).

The mapping of the *Z*-locus is also well advanced. Bian *et al*. (2004), through fine-mapping, delimited the *Z*-locus in *Phalaris coerulescens* with the marker BCD266 mapped at 0.9 cM, which was later found to co-segregate with the *Z*-locus in *Hordeum bulbosum* (Kakeda *et al*., 2008). Hackauf and Wehling (2005) identified a putative ubiquitin-specific protease (UBP) gene in *Secale cereale* co-segregating with the *Z*-locus, and showing some pistil specific expression, but it was suggested later that the gene was not part of the *Z*-locus but could be involved in the SI cascade reaction (Yang *et al*., 2008). A proposed 1.5 cM *Z*-locus was isolated from the rice syntenic region on chromosome 4, corresponding to 125 kb with 12 predicted genes (Hackauf and Wehling, 2005). Ten candidate genes were identified in *Lolium perenne* by Yang *et al*. (2009) with function varying from extension of plant cell walls (Can 135), signalling process (Can 130) to putative channel activity (Can139). Three kinases have also been identified in *Lolium perenne* as pistil candidate genes by Van Daele *et al*. (2008). Recently, two more candidate genes have been identified in *Lolium perenne*: LpDUF247, a protein absent from the rice region syntenic to *Z*-locus

and with a domain DUF247 of unknown function, and LpTC116908, homologous to rice gene Os04g0647300 which is an ubiquitin-specific protease (Shinozuka *et al*., 2010).

Moreover, in addition to the *S* and *Z* components determining SI, other genetic factors are assumed to be involved in the signal cascade leading to the SI response. Additional loci *T* (Thorogood & Hayward, 1991; Thorogood *et al.*, 2005) and *F* (Thorogood *et al.*, 2002) have been identified. The *F*-locus was proposed, evidenced by segregation distortion on linkage group 3.While mapping a perennial ryegrass population: an association between the *S*-locus and the isoenzyme glutamate oxalo-acetate transminase GOT/3 mapped on LG3 was found where an *S*-allele/GOT/3 allele combination was transmitted at a greatly reduced frequency (Thorogood *et al.*, 2002). This region on LG3 is thought to be another locus involved in the SI mechanism in grass, named the *F*-locus, which is close to GOT/3 but has not been closely mapped yet. Despite the presence of this locus associated with SI, it was still possible to genotype plants based on *in-vitro* cross-pollination results for stigmatic inhibition according to a two locus system. This suggests that the *F*-locus is involved in another physiological aspect of the SI process and even that it could trigger a late-acting stylar inhibition of the pollen. This could explain the low seed set of self-compatible plants as observed by McCraw and Spoor (1983) based on stigmatic inhibition alone. Another region of segregation distortion was observed in an F2 perennial ryegrass mapping family on chromosome 5 (Thorogood and Hayward, 1991; Thorogood *et al.*, 2005), which indicates that another locus, named *T*-locus in ryegrass, is involved in SI in *L. perenne*. This self-compatible locus is probably equivalent to the *Secale* S5 (Fuong *et al*, 1993; Voylokov *et al*., 1998) and the *Phalaris T* locus (Hayman and Ritcher, 1992). The *T* locus acts gametophytically but unlike the *S*- and *Z*- *locus* has two allelic forms: one

leading to 0% compatibility and the other to 100%; in the heterozygous condition this would lead to 50% self-compatibility (Thorogood *et al.*, 2005).

The *S* and *Z* loci have not been characterized in any grass species. The candidate genes found in grasses all refer to homologues in the rice genome which are presumably inactivated forms as rice is a self-compatible species, or it is possible that the *S* and *Z* loci genes are simply not present in rice in any form (Kusaba *et al*., 2001).

In order to understand SI in ryegrass, one of the key elements is to identify the genes involved in this mechanism, in particular the *S*-locus in this project as the *Z*-locus is being extensively studied by Dr. Bruno Studer (ETH Zurich). My strategy for *S*-locus identification involves at first, the fine-mapping of the *S*-locus. This strategy serves to restrict the *S*-locus region size to a handful of candidate genes, identified by synteny in closely related species. Once the *S*-locus region has been narrowed sufficiently, our current knowledge of gene function and the characteristics of SI loci in other species can be used to predict the DNA sequence within this marker-flanked region that encodes for the *S*-locus and isolate candidate genes. In order to identify the *Lolium perenne* genes within the identified *S*-locus region, the genomic sequence had to be obtained. Using next generation sequencing, BAC clones covering the *S*-locus region would be sequenced and the *S*-locus genomic sequence built up. A complementary gene expression study of this region, with tissue specific RNA sequencing would enable to identify the expressed genes in the S-locus region as well as the differentially expressed genes in female or male tissue. Results from the *S*-locus study could be applied in breeding programs and part of this research project was to develop and test molecular markers in order to study the allelic diversity of SI in a breeding population. Finally, breakdown of SI naturally accuring is important in breeding and a mapping approach was taken in order to study two modified loci, *T* and *F*.

Chapter 2:

S-locus fine mapping

2.1. Introduction

2.1.1. Aims and strategies

Linkage maps (or genetic maps) have been developed for *Lolium perenne* (e.g. Armstead *et al.*, 2002; Jones *et al*., 2002; Studer *et al*., 2008) and the *S*-locus has been mapped on chromosome 1 in *Lolium perenne* (Thorogood *et al*., 2002), with molecular markers distant by 2cM (Yang *et al*., 2009). By comparing the available genetic maps and the molecular markers closely mapped to the *S*-locus, potential linked markers were selected to be mapped on our mapping populations, that were designed to exclude transmission through the pollen of one of the *S*-alleles and hence of linked markers, thus circumventing the need to phenotype plants for *S*. Once the *S*-locus region was isolated, specific markers were designed on *Lolium* transcript sequences, using the synteny between rice and *Lolium perenne*. The mapping was done by continually adding new markers in order to reduce the *S*-locus region size and increasing the population size as well as screening other related populations. In this study, over 80 markers have been tested on the *S*-population and over 10,000 plants have been genotyped with closely linked markers, leading to an *S*-locus region of approximately 0.12 cM. By synteny, the *S*-locus region was identified in rice and *Brachypodium*, 60Mb and 140 Mb in size respectively, and candidate genes were identified.

2.1.2. Linkage maps

A genetic map is essential to study the genome of any species and is usually one of the first steps in a map-based gene identification strategy. Molecular markers are anchored at a location in order to create the map, called a linkage map or genetic map. The genetic distances are based on genetic recombination rates, expressed in centiMorgan (cM). By comparing the genotypes of a mapping population for a number of molecular markers and

therefore the recombination rates between markers, an association between markers can be done and the genetic distances can be calculated.

Through his work, pioneering geneticist Gregor Mendel described the segregation of traits in sweet peas mainly, such as flower colour, seed shape and leaf size. From his observations, Mendel described in 1866 his Laws of Inheritance: the Law of Segregation stating that each trait has two factors (alleles), inherited from each of the parents, that are passed randomly to the progeny, and the Law of Independent Assortment, concluding that the segregation of two traits is independent from each other. However, this second law was proved not to be so universal when William Bateson *et al*. (1905) observed that some traits in sweet peas were often associated. Their conclusion was that some traits must be linked in some ways as their assortment (segregation) was not entirely independent. Later, Thomas Morgan, through his work on inheritance of observed body characteristics of fruit flies (*Drosophila melanogaster*) answered the question and in 1910, suggested the idea of a physical link between loci, the chromosome. If two traits are governed by two loci on the same chromosome, their segregation will be similar.

However, in his work of 1911, Morgan noticed that some traits, even if they are linked, do not always segregate together, leading to a break in the linkage: recombination between loci was discovered. Recombination events happen during meiosis, when cells are dividing in order to produce the gametes. During meiosis, the chromatin is condensed into dark staining bodies (chromosomes) that lie in the nucleus of the cell. Chromosomes occur in pairs, each so-called homologous chromosome being inherited from the two parents. The first division of the meiosis begins with the prophase, where homologous chromosomes are paired. During prophase, chromatids from the two homologous chromosomes may overlap and cross, forming a chiasma. During the separation of the

chromosomes in the anaphase, the arms of homologous chromosomes interchange around the chiasma, leading to crossing-over and recombination of the parental genomes in the progeny (see Figure 2.1). It is the recombination frequency, the frequency of occurrence of crossing-over, between factors that is measured in order to create a genetic map. The chances of recombination between factors are higher the further they are physically separated. Two factors are linked if segregation in the mapping population is not independent of each other and the recombination rate is under 50%. This recombination frequency is translated into centiMorgan (cM); 1cM is the equivalent to 1% recombination.

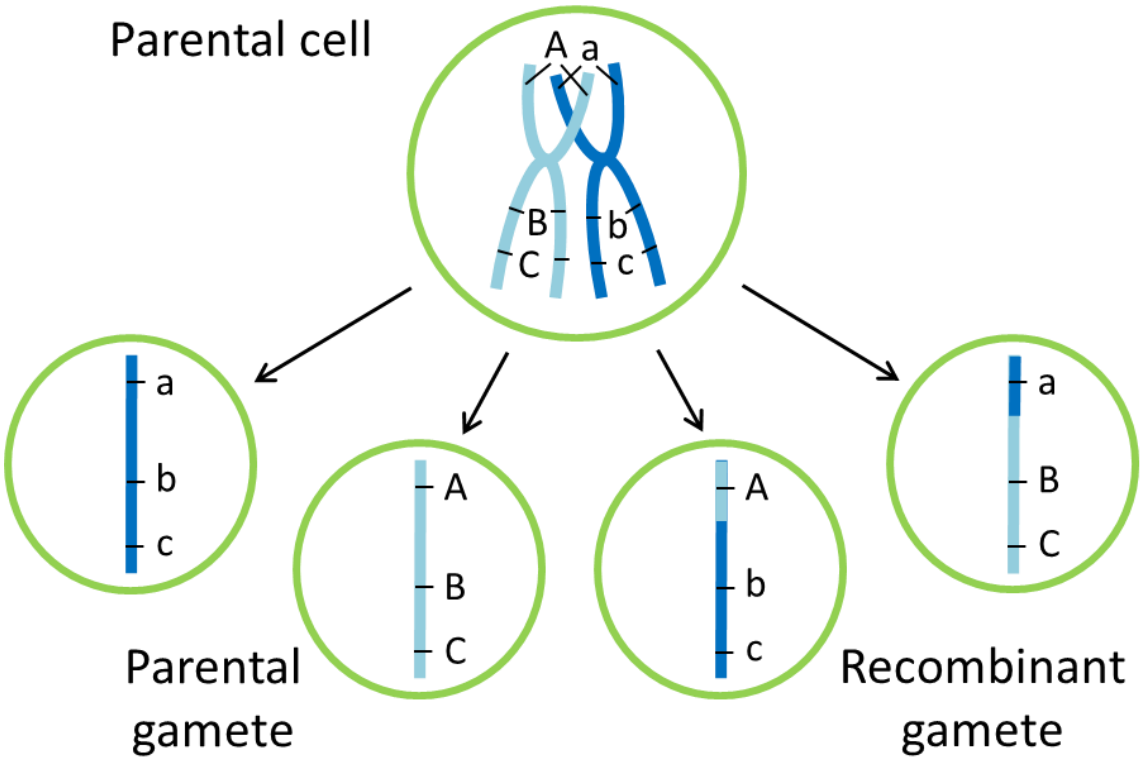


Figure 2.1: Graphic representation of a recombination due to a crossing over of the parental chromosome. Two chromosomes are represented with the blue bars, with three loci annotated, A, B and C; each locus is heterozygous in the parent. The parental cell can only lead to the creation of two gametes if the loci are linked: ABC and abc. But if a crossing-over occurs between the loci A and B, two new types of gametes will be created: Abc and aBC. The percentage of recombinants in the progeny will enable the calculation of the genetic distance between those loci A and B.

Originally, the recombination frequency was calculated using the formula:

Recombination frequency = 100 * (No. of recombinant / Total No. progeny)

However, this formula cannot be used in the creation of linkage maps as the recombination frequencies are not linear and the distance between A and another locus C for example, does not equal the distances A-B and B-C (Kearsey and Pooni, 1996). Haldane (1919) published a mapping function that is still used today, using recombination frequency (r) and map distance (m) in cM. The equation is as followed:

$$r = ½ (1-e^{-2m})$$

This formula assumes that there is no interference between crossing-overs. However, it has been shown that a crossing-over cannot occur at less than 15 cM from another crossing-over (Kearsey and Pooni, 1996). Kosambi (1944) developed another formula to link the recombination frequency (r) to genetic distance (m), assuming that there is interference between cross-over events:

$$r = ½ \tanh (2*m)$$

Nowadays, several computer packages are available for automating genetic mapping (reviewed in Semagn *et al.*, 2006).

The relation between a genetic map, based on recombination frequencies, and a physical map, based on number of base-pairs, is not the same for several reasons (see Figure 2.2). First, the genetic map is made using a certain mapping population, so the distance between markers is specific to this mapping population even if the distances between markers may vary only a little from one population to another. Another

complication is that the recombination frequency is not homogenous along the chromosome. It has been shown that the recombination frequency is reduced in the region close to the centromere of the chromosome (e.g. Hayward *et al*., 1998). Because of this, in rice, 1 cM can be a physical distance of 120 kb but also 1000 kb (Kurata *et al*., 1994).

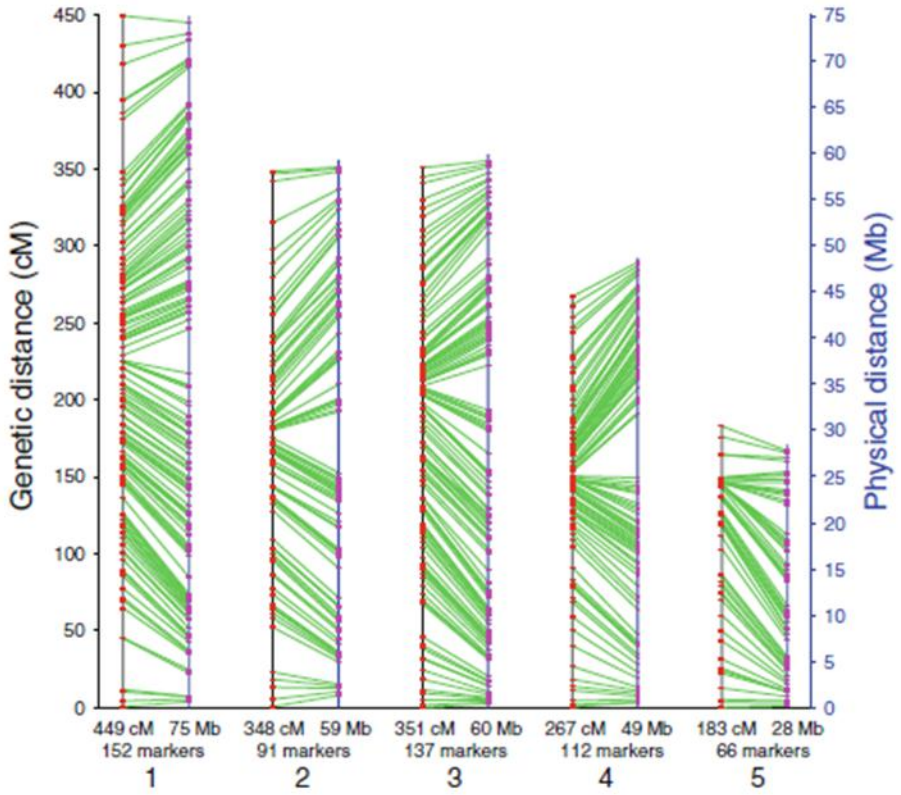


Figure 2.2: *Brachypodium* relation between genetic map and physical map. The black bars represent the linkage map (genetic map) whereas the blue bars represent the physical map. The coloured dots represent the molecular markers position. The green lines indicate the corresponding markers between the genetic and the physical maps. The gaps on the physical map indicate the non-uniformity of the recombination frequency, with less recombination in the centromeric region compared to the distal regions. Figure from Huo *et al*. (2011).

The first linkage was reported by Bateson *et al*. (1905) in sweet pea using morphological traits. But with the development of molecular markers, the visualisation of actual DNA variation that is the basis of observed inherited trait variation, ever-dense genetic maps have been created: from one marker every 10 cM approximately in human (Donis-Keller *et al*., 1987) to one marker every 0.002 cM (3.1 million markers over an estimated genome genetic size of 4782 cM; The international HapMap consortium, 2007) as techniques for marker development have evolved. Genetic maps in grasses are also well developed now; for example, a *Brachypodium distachyon* genetic map was made out of 558 markers (Huo *et al*., 2011) and there are many genetic maps of *Lolium perenne* (e.g. Tomaszewski *et al*., 2012; Studer *et al*., 2008; Jensen *et al*., 2005b).

Genetic maps are valuable tools in localizing quantitative trait loci (QTL) or genes. By associating the segregation of molecular markers with the phenotype studied, regions of the genome can be associated with phenotypes such as disease resistance. Once a gene or QTL has been located, the molecular markers located closely on the genetic map can be used for marker assisted selection (MAS) to accelerate breeding programmes.

But the segregation of traits and markers does not always follow the Mendelian segregation model. Some traits seem to be over-represented in some progeny, and this segregation cannot be explained using Mendel's laws. Mangelsdorf and Jones (1926) were the first to report a drift in some allele segregations in maize. It was then reported in many species and in particularly in the fruit fly by Sandler *et al*. (1959) who described a locus on a chromosome that would cause the over-representation of one allele over another. Since then, segregation distortion (SD) has been observed in many species such as rice (Harushima *et al*., 1996), barley (Konishi *et al*. 1992), tomato (Paterson *et al*., 1991), *Lolium perenne* (Jensen *et al*., 2005a) and human (Zöllner *et al*., 2004).

Segregation distortion can be the result of a single or a combination of mechanisms. The frequency of one allele over the other can occur during the creation of the gametes or spores but also at a later stage such as the pollen/spore germination, the seed development or even seed germination (Zamir and Tadmor, 1986; Lyttle, 1991). GSI can be a cause of SD as the alleles closely linked to the loci involved in SI are segregating with them. In *Lolium perenne*, the *S-Z* system of SI requires that, for a pollination to be successful, at least one of the loci has to be compatible, meaning at least one different allele between the two parents. Because of this mechanism, certain alleles are promoted over others, therefore creating a SD (Thorogood *et al*., 2002 and 2005). But because of SD, the segregation of markers linked to SD locus can be affected and therefore the genetic distance of linkage maps can be over or under estimated (Song *et al*., 2006). In *Lolium perenne*, many mapping projects have reported SD, with up to 60% of the markers showing SD (Jensen *et al*., 2005a). It has also been observed that SD can be different between mapping populations of the same species: 24% of SD in the population 'NA_6' compared to 15% in 'AU_6' population (Faville *et al*., 2004). This difference can be explained by the cross made to produce the mapping population; it seems that recombinant inbred line (RIL) populations have more SD than any type of population (Xu *et al*., 2007). One explanation is that RIL populations, because of the inbreeding, will accumulate more heterozygote genotypes due to the deleterious effect of some alleles when homozygous (Anhalt *et al*., 2008).

The effect that SD has on the creation of a linkage map has to be taken into account as it can impact on the markers' order or the genetic distances (Hackett and Broadfoot, 2003). But SD can also be used in mapping strategies: Harbord *et al*. (2000) have used SD to identify transgenes that are closely linked to the pollen *S*-locus of *Petunia*. This strategy

has been used in the mapping of the *S* and *Z* loci of the SI system in *Lolium perenne* (Yang *et al*., 2009).

2.1.3. Molecular markers

A genetic marker is a piece of DNA that has a unique location within the genome of an organism, with some variations (polymorphisms). These polymorphisms can be distinguished either directly by nucleotide sequencing or indirectly through their differing physical/structural properties that can be resolved by passing DNA samples through a conductive matrix to which an electric field has been applied (electrophoresis). The inheritance of these marker variants in populations in relation to other markers determines their position on a genetic linkage map. Semagn *et al*. (2006) have described over 30 different methods of revealing polymorphisms that have been developed but, because of ease of use and the amount of information they reveal, only a few of them are currently widely used (see Table 2.1).

There are basically two types of polymorphism in the genome: substitution (replacement of one or more nucleotides, either from the same nucleotide group [for example, the purine base A replaced by G] or from the other (pyrimidine) group [for example A replaced by G or C]) and insertion/deletion (InDel). From these two groups, a number of molecular markers have been described according to the number of polymorphic nucleotides involved and/or the location of the polymorphism in the genome.

For either group (substitution or InDel), one single nucleotide can be the source of polymorphism: Single Nucleotide Polymorphism (SNP). SNP represents the most common source of polymorphism in the genome (Syvänen, 2001); it is estimated that a SNP occurs

every 1000 bp in the human genome (Sachidanandam *et al*., 2001) and every 60-120 bp in maize (Ching *et al*., 2002). The variation in the SNPs distribution within a genome also varies (Barreiro *et al*., 2008); most of the SNPs are located in the non-coding regions. The effects of a SNP in the non-coding region are often not seen on the phenotype (silent mutation) unless it is in a part that regulates the gene expression. But they are excellent molecular markers as they provide a lot of polymorphism. If the SNP is located in a gene-coding region, it can be classified into one of two groups: the synonymous SNP where the mutation has no effect on the translated amino-acid chain due to the degenerate nature of the genetic code (one amino acid can be coded by several different codons, the third nucleotide often being unspecific) and the non-synonymous SNP where the amino-acid encoded is changed and the effects on the phenotype can vary from a complete loss (or gain) of a protein function or to a subtle structural difference in a protein that can alter its function or function-efficiency.

A lot of molecular methods that rely on SNP variation have been developed to discriminate allelic variants. One of the first methods used is the allele-specific oligonucleotide hybridization (ASO) (Wallace *et al*., 1979). Two allele specific probes are designed, overlapping the polymorphic base and are hybridized to the DNA; the perfect matching probe will stably hybridize to the target unlike the probe with a mismatch. In order to reveal the hybridization (or not), the most common method used is DNA chips containing the probes onto which the DNA samples are applied. Another method developed to reveal SNPs is to use primer extension, in which primers are designed to either end on the polymorphic base or flanking the SNP region. Using a polymerase, the target DNA will be extended and amplified. If the primer is designed to have the SNP on its 3' end, the extension with the polymerase will only happen if the primer perfectly

matches the target, and the results can be observed on an electrophoresis gel. In the other case, if the primers are designed to flank the SNP region, the primer extension will happen whatever the target is but as the polymorphism is only one base pair, the allele discrimination can only be based on size differences, using a fluorescently labelled primer and an electrophoresis by capillarity system in the case of an InDel SNP, or in the case of a substitution, methods such as sequencing, array, mass spectrometry can be used but are time consuming and expensive.

In some cases, a SNP can occur on a specific site, a restriction enzyme site, which is a specific DNA motif that a restriction enzyme (endonuclease) will recognize and digest. This will result in the cleavage of the DNA. There are three types of molecular markers used to reveal polymorphism on a restriction site: Cleaved Amplified Polymorphic sequence (CAPs), Restriction Fragment Length Polymorphism (RFLP) and Amplified Fragment Length Polymorphism (AFLP) markers. These molecular markers use the same principle: a restriction enzyme is used to cut the DNA into fragments (specific to the enzyme used) and the fragment sizes are then compared between samples. Restriction enzymes recognize a specific DNA motif (restriction site) of four to eight base pairs (bp). Because of the difference and specificity of the restriction site, some enzymes, with a small and common restriction site will cut more frequently in the genome and produce small fragments (frequent cutters) compared to the enzymes with a long restriction motif (rare cutters). The choice of the enzymes used in CAPs, RFLP and AFLP marker development is made according to the number of markers needed. Both CAPs and RFLP method uses more middle range enzymes (six bp cutter) to cut the double-stranded (ds) DNA. In the case of the AFLP markers, the DNA is cut using two enzymes most of the time, a rare cutter and a frequent cutter. The polymorphism (difference in size fragments) is then

revealed by electrophoresis using a radioactively labelled probe for the RFLP markers. For the other two markers (AFLP and CAPs), the fragments are then ligated to adaptors in order to amplify them using PCR before separating them by electrophoresis using silver staining or fluorescently labelled adaptors. However, the development of these two types of molecular markers (RFLP and AFLP) to genotype a large number of plants is time consuming and expensive. They were used to develop the first genetic maps.

The second type of polymorphism, the insertion/deletion (InDel) of one or more nucleic acids, has also led to the development of a large number of molecular markers. The most common types of InDel used as molecular markers are the ones occurring in a repetitive DNA sequence. Eukaryote genomes are composed of a large number of repetitive regions, where a motif of nucleotides is tandemly repeated (Tautz and Reuz, 1984; Vergnaud, 1989). The size of the repeated motif can vary from 1 up to 100s of nucleotides; according to the size of the motif, the regions are classified into microsatellites (or Simple Sequence Repeats, SSRs) for 2-6 bp repeats (Chambers and MacAvoy, 2000), minisatellites with a larger motif, from 10 to 60 bp approximately and satellites DNA which is a large tandem repeat (from 10 up to 1000s bp) from which the size and the sequence may vary (Richard *et al*., 2008). Those repetitive regions are known to be highly polymorphic, due to the ease of occurrence of the mechanism called "slipped-strand mispairing", which can be described as a misalignment of the DNA strands, due to the repetitiveness of the region, therefore leading to the insertion (or deletion) of one or more repeat units (Levinson and Gutman, 1987; Eisen, 1999). The variation of the number of repeats creates allelic diversity that can be used to develop molecular markers. SSRs are the most common used molecular markers to target repetitive regions as they are highly abundant in the genome: approximately one microsatellite every 50 kb in soybean for

example (Morgante and Olivieri, 1993). SSR markers are represented as a pair of primers flanking the microsatellite but specific to the region. By using PCR, the microsatellite is amplified and the size of it (therefore the number of repeats) can be compared with other samples by electrophoretic separation on agarose gels, polyacrylamide gels or by capillarity. The abundance and the ease of use of SSRs make them a marker of choice as they are highly polymorphic, co-dominant and highly transferrable between populations (Gupta *et al*., 1999). In rice, over 18,000 SSR sequences have been identified and can be used as molecular markers (The Rice Genome Mapping project, 2005). However, the development of such markers requires knowledge of the genome sequence: to use a repetitive region as a marker, they need to be found in the genome. In order to develop such markers, genomic libraries have to be constructed and the process is long and expensive. To that matter, the development of primers flanking repetitive regions can also be a problem as unique regions needs to be found for the primers to be designed on. Therefore, the use of SSRs markers in plant species of minor importance can be difficult as the genomic data are not always available and the cost to develop a library can be restricting for small research projects.

In order to exploit the available genomic data as well as the advantage of the repetitive region, SSRs are now being developed using EST (Expressed Sequence Tag) databases in cereals (Kota *et al*., 2001; Kantety *et al*., 2002). By using available data, the time and the cost of developing SSRs is much less but again, it requires some data that are not always available for species not widely studied. ESTs can also prove to be a good resource to develop other type of markers. ESTs are obtained by creating a cDNA library (a library of the expressed genes). Molecular markers based on ESTs can help to map specific genes or discover new ones. Any type of molecular marker can be developed from

them, as long as the EST is showing some polymorphism. For example, RFLPs have been developed from ESTs (EST-RFLP) and extensively used to produce genetic maps (e.g., Harushima *et al*., 1998 in rice) and physical maps (e.g., Kurata *et al*., 1997).

With the advance of sequencing technologies, more genetic data are available. By sequencing RFLPs and AFLPs for example, new types of markers can be developed, sequence tagged sites (STS). A STS is a short unique sequence that is amplified using a pair of primers and a PCR. Like EST-based markers, the allelic variation for such a marker can be due to any type of polymorphism, but STS can, unlike ESTs, be based on genomic non-coding regions, leading to more variations. STS markers have been first developed in the physical mapping of the human genome (Olsen *et al*., 1989) but were also used in the map development of crop species such as bread wheat (Talbert *et al*., 1994). STSs are good markers as they can be developed from an available resource and their use can be automated.

The number of methods to identify and use polymorphism can only increase with the years. Many more markers are available for use but none of them or the markers described previously can be generally used. The choice of the marker used is always dependent on the research project and purpose: number of markers required, research budget, time, genomic resources, population size, transferability of the research, etc. (see Table 2.1).

Table 2.1: Comparison between the most common used molecular markers

	Primer extension	**RFLP / AFLP/CAPs**	**Microsatellites**	**EST-based markers, STS**
Technique	Amplification using a primer pair (specific to the polymorphism or flanking it)	DNA Digestion using one or two enzymes, fragment length comparison	Amplification of repetitive region	Amplification of unique sequence
Type of polymorphism detected	SNPs, InDel	SNPs, InDel	Length polymorphism	SNPs, InDel
Method of detection	Electrophoresis (agarose, acrylamide or capillarity), sequencing	Electrophoresis (agarose, acrylamide or capillarity), radioactive probes (RFLP only)	Electrophoresis (agarose, acrylamide or capillarity)	Electrophoresis (agarose, acrylamide or capillarity), sequencing
Advantages	High abundance, easy to develop, easy to detect	High abundance, high reproducibility, no sequence data required	High abundance, High reproducibility, Co-dominant	High reproducibility, High abundance, Co-dominant
Disadvantages	Difficult reproducibility between populations, Dominant for some	Radioactivity, time consuming, labour intensive, subjectivity in scoring	High development cost	Need for sequence information

2.1.4. High Resolution Melting method

All molecular markers have to be revealed by a method, common and generally used for some of them (e.g. agarose gel, poly-acrylamide gel) but sometime more costly and difficult to access (e.g. electrophoresis by capillarity, sequencing). One new method that has been developed to reveal polymorphism and therefore any molecular markers based on PCR amplification is the High Resolution Melting curve analysis technique.

When a molecular marker based on PCR is used, analysis of the amplified product is different according to the type of marker used. The simplest and easiest method to do so is to use a gel electrophoresis separation method such as an agarose gel. However, this method is valid only if the marker is targeting a big length polymorphism and not just a SNP. An acrylamide gel, revealed using a silver staining is a much more discriminating electrophoresis method, enabling the genotyping of microsatellite markers. But this method is time consuming. Even still, SNP variation cannot be revealed. Other methods of electrophoresis such as electrophoresis by capillarity are available in order to identify SNP variations but this method involves special primers labelled with a fluorescent dye and is therefore a costly method to use to genotype a large number of plants.

High Resolution Melting (HRM) curve analysis is a powerful method of differentiating DNA polymorphism that can allow the genotyping of a large number of plants with a large number of markers, in a short time and for a small cost. It is regarded by some as a method that has high differentiation power as it is possible to differentiate both sequence and sequence length polymorphisms and also has the power to determine epigenetic variation caused by, for example, DNA methylation. The method, developed in 2003 from the collaboration between the University of Utah and Idaho Technology Inc. (Salt Lake City, UT, USA), uses the difference between DNA melting curves (i.e. the rate at which ds-DNA denatures to ss-DNA), that can be linked to any type of undefined polymorphism. The method is based on the use of a saturation dye that will interact and fluoresce with double-stranded DNA (ds-DNA) only. Ds-DNA is denatured from double-stranded to single-stranded with a temperature increase and the dissociation of the two strands occurs at a specific temperature, the melting temperature (Tm) depending on characteristics of the amplified sequence of DNA: its length, its GC content as well as its

heterozygosity (see Figure 2.3). The rate of dissociation is measured by the decreasing fluorescence as the temperature increases, and a melting profile is obtained and can be then compared with the profiles from different DNA samples. The method is divided into two steps: the amplification of a marker using a PCR with the saturation dye (e.g. LCGreen® Plus from Idaho Technology Inc., EvaGreen® from Biotium) included in the mix and the melting of the amplified marker (using the Light Scanner instrument (Idaho Technology Inc.) for example). The genotyping of the samples for a molecular marker is then done using software that will normalize the melting curve and classify the samples according to the curve shapes.

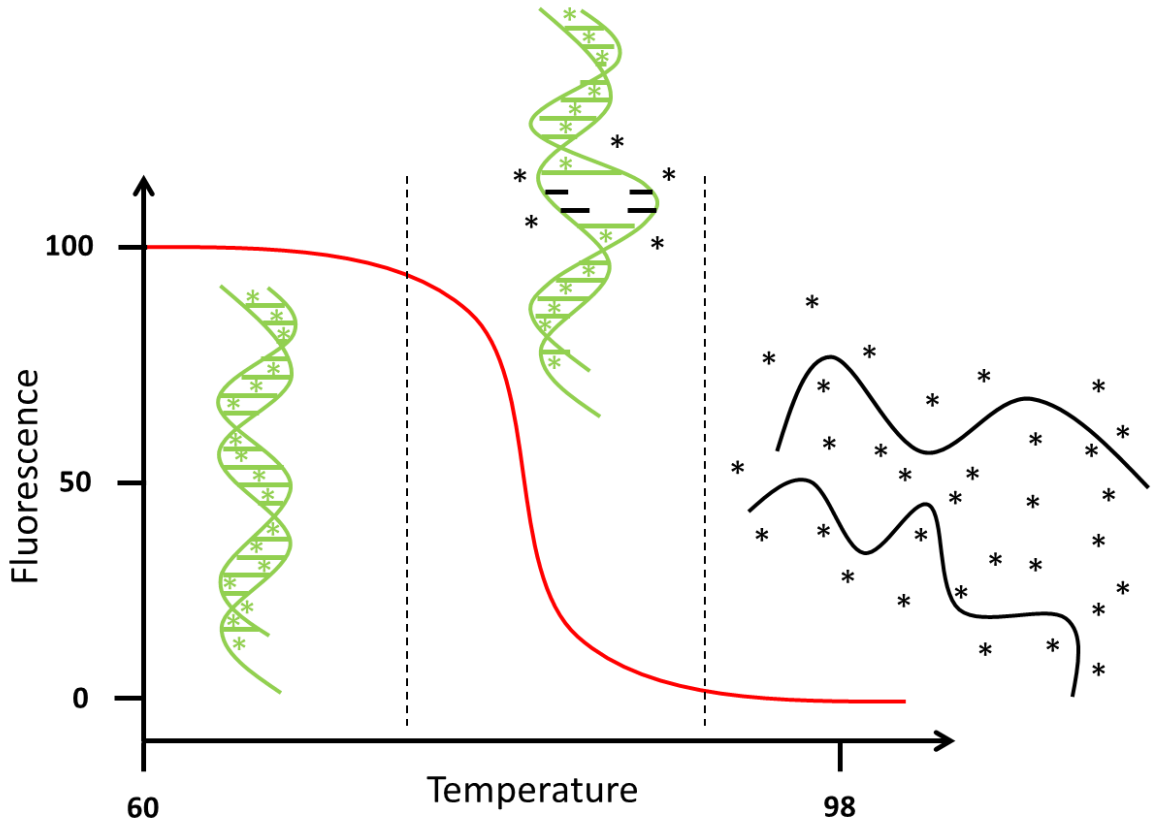


Figure 2.3: High Resolution Melting curve analysis principle. The graphic curve represents the decrease of the fluorescence as a function of the increase of the temperature. The state of the DNA (ds or ss) is represented for each stage of the melting: first ds, the DNA will melt and release progressively some fluorescence, leading to a ss DNA.

HRM curve analysis had been used in a large number of studies. Initially developed as a method for screening human clinical mutations (Wittwer *et al*., 2003), the method is now used widely to genotype a large number of molecular markers in a wide range of organisms. In almond, SNP markers have been specifically designed from ESTs to amplify small amplicons in order to map genes onto a linkage map. Wu *et al*. (2009) used HRM to genotype the mapping population. Other linkage maps have been developed using such an approach: SNPs mapping in apple (Chagné *et al*., 2008), STSs mapping in white lupin (Croxford *et al*., 2008). In rice for example, Li *et al.* (2011) have successfully mapped known STS and SSR markers using the HRM curve analysis technique and showed that it was faster and cheaper compared to other methods such as electrophoresis or Sanger sequencing. In *Lolium perenne*, Studer *et al.* (2009) proved that it is possible to map genes in a population without knowing the allele diversity of the mapping population. Using the parents and the grand-parents of their mapping population, the team was able to "blind-map" four genes by creating random molecular markers based on the gene sequences.

2.2. Plant populations

2.2.1. S-populations developed in IBERS

The S-populations used in the mapping of the *S*-locus in *Lolium perenne* were specifically designed in May 2007 by Dr. Daniel Thorogood at IBERS to investigate the *S*-locus and are derived from the first ever *Lolium perenne* mapping family produced by the International *Lolium* Genome Initiative (ILGI) collaboration. It is described by Jones *et al* (2002). In this study, three S-populations have been used: P235/59, P235/64 and P235/63. All three populations derived from crosses between siblings resulting from two crosses between ILGI population (P150/112) plants. A cross between the plant P150/112/129

($S_{12}Z_{13}$) as female and plant P150/112/132 ($S_{12}Z_{12}$) as pollinator resulted in an F1 population, P232/88. By the same method, a cross between the plant P150/112/41 ($S_{12}Z_{13}$) as female and plant P150/112/38 ($S_{13}Z_{13}$) as pollinator produced the F1 population P235/40. By crossing a plant from each of those two populations P232/88 and P235/40, the S-populations were obtained. The population P235/59 was obtained from a cross between the plants P235/40/13 ($S_{23}Z_{11}$) and P232/88/10 ($S_{22}Z_{12}$), the population P235/63 from a cross between the plants P235/40/14 ($S_{23}Z_{11}$) and P232/88/3 ($S_{22}Z_{12}$) and the population P235/64 from a cross between P235/40/17 ($S_{23}Z_{11}$) and P232/88/10 ($S_{22}Z_{12}$).

The crosses between the parents from the populations P235/40 and P232/88 were made in three separate pollen-proof isolation rooms using 5 clonal replicates for each parent. For each population, the parent P235/40 was used as the pollinator of the parent P232/88 ($S_{22}Z_{12}$ x $S_{23}Z_{11}$) as shown in Figure 2.4 with the example of the population P235/59. The progeny for the three S-populations were either $S_{23}Z_{11}$ or $S_{23}Z_{12}$ and therefore entirely heterozygous at the *S*-locus. *S*-locus-linked markers will similarly be heterozygous unless a recombination event has occurred between the *S*-locus and the marker, in which case marker genotypes will be homozygous for the S_2 allele-coupled marker allele. The frequency of recombinants can then be used to determine the recombination map distance between marker and *S*-locus.

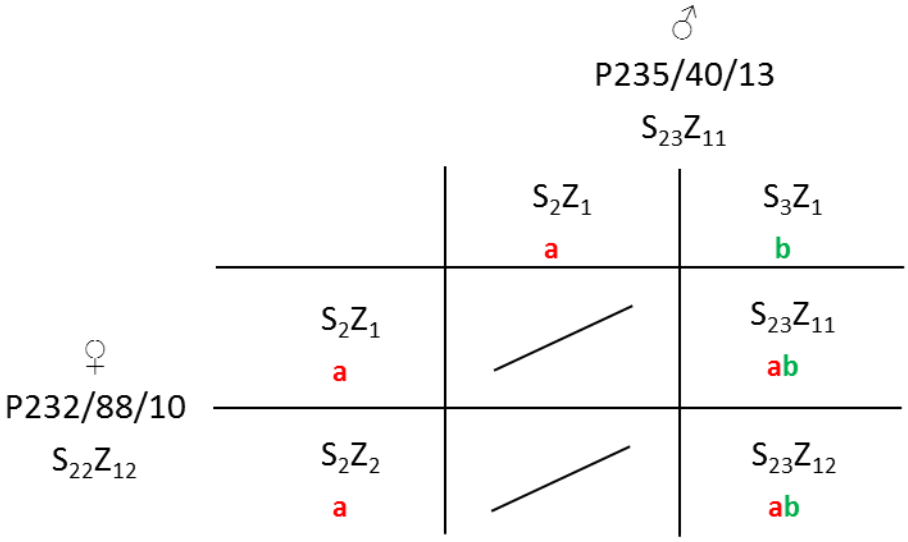


Figure 2.4: *S*-locus mapping population design for the population P235/59. In colour are the 2 alleles of an *S*-locus-linked marker. Only one type of pollen matches with the stigma, S_3Z_1, so all the progeny, where no recombination has occurred between marker and *S*-locus, is S_2S_3, allele ab for the marker. In a case of a recombination event between the *S*-locus and the marker, the plant will be homozygous for the marker, aa.

The seeds from the genotype P232/88 were harvested after the cross and represented the S-populations (i.e. from the plant P232/88/10 for the population P235/59). For the S-population P235/59, 3g of seeds were harvested, all seeds have been sown in a greenhouse throughout the experiment and 1466 plants from this population were used for the first fine-mapping of the *S*-locus.

The P235/59 population has not been previously used in mapping so no genetic maps were available for this population. Markers mapped on other populations of *Lolium* such as the ILGI population (Jones *et al*., 2002) were used to broadly locate the *S*-locus.

2.2.2. VrnA-S populations developed in The Danish Institute of Agricultural Sciences

The VrnA population was designed initially to map QTLs involved in vernalization response in *Lolium* (Jensen *et al*. 2005). The F1 population resulted from a cross between a Danish ecotype Falster and an Italian variety Veyo, selected for their contrasting vernalization requirements as Falster requires a high vernalization and Veyo requires no or low vernalization. Two plants from the F1 VrnA population were selected and crossed in order to produce the F2 VrnA population (NV#20/30-39): F1-30 and F1-39. Finally, some individual plants NV#20/30-39 were back-crossed with the parent F1-39 in order to create the VrnA-S populations.

In the large fine-mapping of the *S*-locus, four different VrnA-S populations have been used: S216, S267, S324 and S404. Each of these population resulted from a cross between a plant from the F2 VrnA population and the parent F1-39. The VrnA-S population S216 results from the cross between NV#20/30-39/216 and F1-39; S267 population from a cross between NV#20/30-39/264 and F1-39, S324 population from a cross between NV#20/30-39/324 and F1-39 and S404 population from a cross between NV#20/30-39/404 and F1-39.

2.3. Methods used for fine-mapping

2.3.1. DNA extraction protocols

Three different methods of DNA extraction have been used during this project in order to map and fine-map the *S*-locus: DNA extraction using the DNA 96 Plant kit (Qiagen, Hilden, Germany), DNA extraction using the robotic system Autogen

Autogenprep 740 (AutoGen Inc., Holliston, MA, USA) and the large scale DNA extraction (adapted from L. Westphal).

Whatever the DNA extraction protocol used, the quality and the quantity of the DNA were assessed on all or few random samples (approximately 10%) using a 1% agarose gel (1% agarose, 0.5X TBE) with 10% GelRed (Biotium, Hayward, CA, US). Some random DNA samples were quantified using a spectrophotometer (NanoDrop® ND-1000, Thermo Fisher Scientific Inc., Waltham, MA, US).

2.3.1.1. DNA extraction protocol using the DNA 96 Plant Kit from Qiagen

The screening was initiated with only 242 plants from the S-population P235/59 of which the DNA was extracted using the DNA 96 Plant kit (Qiagen, Hilden, Germany). The DNA extraction was done according to the QIAGEN protocol and using approximately 80mg of leaf sample. The samples were collected in collection microtubes and placed in liquid nitrogen. To each sample, 400µl of the working lysis buffer (Buffer AP1 preheated to 65°C with 0.25% RNase A (100mg/ml) and 0.25% of reagent DX) were added before being ground two times for 2 minutes at 20Hz using a Retsch-mill (Retsch, Haan, Germany). The collection racks were centrifuged quickly to remove any plant material from the caps and 130µl of the Buffer AP2 were added to each collection microtube. The racks were shaken vertically for 15 seconds and then centrifuged for 20 seconds at 3000rpm. The samples were incubated for 10 minutes at -20°C and then centrifuged for 5 minutes at 6000rpm. 400µl of the supernatant were transferred into a new rack of collection microtubes and 600µl of Buffer AP3/E were added to it. The rack was then

shaken for 15 seconds and centrifuged for 15 seconds at 3000rpm. A DNeasy plate was placed on top of a collection plate and 1ml of each sample were transferred into the DNeasy plate. The plate was centrifuged for 4 minutes at 6000rpm. 800µl of Buffer AW were added to each sample and the plate was then centrifuged again for 15 minutes at 6000rpm. The DNeasy plate was placed onto a new rack of Elution Microtubes and 100µl of Buffer AE was added to each sample. The plate was incubated for 1 minute at room temperature and centrifuge for 2 minutes at 6000rpm. This last step was repeated in order to obtain a final volume of 200µl of DNA.

2.3.1.2. DNA extraction protocol using Autogen automated system

After the initial screening was done, more plants from the S population P235/59 (approximately 1500 plants) were used for the first fine mapping. The DNA of these plants was extracted using the Autogen Autogenprep 740 (AutoGen Inc., Holliston, MA, USA) for automatic DNA extraction. Approximately 150mg of leaf were collected in Autogen tubes and placed in liquid nitrogen. The samples were ground using an Autogrinder (AutoGen Inc., Holliston, MA, USA) before being treated for the Autogen DNA extraction. To the ground samples, 430µl of warmed CTAB were added (100mM Tris-HCl pH 8.0, 20mM EDTA pH 8.0, 1.4M NaCl, 2% CTAB) before being incubated at 65°C for 45 minutes. The samples were then loaded in the Autogen Autogenprep 740 (AutoGen Inc., Holliston, MA, USA) for the automatic run of plant DNA extraction. The extraction program is divided into 7 steps: protein denaturation, removal of debris, precipitation of the DNA-CTAB complex, precipitation of DNA, DNA wash, drying of the DNA and

finally dissolving the DNA using 200µl of TE-RNase (10mM Tris-HCl pH 8.0, EDTA pH 8.0, RNase A 10mg/ml).

2.3.1.3. Large scale DNA extraction protocol

Large scale DNA extraction protocol aimed to extract DNA from a large number of plants quickly and at a low cost. Leaf samples from plants of the DTS and VrnA-S populations (approximately 15cm long) were collected in 96 well collection plates before being dried overnight in an oven at 60°C. Once the samples were dried, they were ground by placing a steel bead in each sample and using then the autogrinder for 35 seconds at 1500 strokes/min. The plates were spun for 30 seconds at 3700rpm in order to collect all the plant powder at the bottom and avoid contamination. Using a Rainin Liquidator 96 Pipetting System (Anachem Ltd., Luton, UK), 380µl of buffer extraction was added to each sample (0.2M NaCl, 0.05M Tris-HCl pH 8.0, 0.01M Na-EDTA, 0.01M DTT, 1M SDS, according to L. Westphal). Samples were vortexed and incubated in a water-bath at 60°C for 10 minutes. 50µl of 5M K-acetate was added to each sample; the samples were vortexed again and placed on ice to incubate for 20 minutes. The plates were centrifuged for 5 minutes at 3700rmp and 125µl of the supernatant was pipetted into a new 96-well plate. 100µl of isopropanol was added to each sample and the plate was mixed by inversion before being incubated for 30 minutes at -20°C. The plate was then centrifuged for 15 minutes at 3000rpm and the isopropanol (supernatant) was discarded by pipetting. The samples were dried at 60°C in a vacuum-centrifuge (Eppendorf, Hamburg, Germany) for approximately 20 minutes (to allow the isopropanol to evaporate). Finally, the DNA was dissolved into 20µl of sterile distilled water and left for at least 4 hours on ice before being used.

2.3.2. Marker selection

2.3.2.1. Selection of published markers around the *S*-locus

The closest marker in *Lolium* has been mapped at a distance of approximately 0.6 cM (Thorogood *et al*, 2002) and 0.1 cM (Yang *et al*., 2009) to the *S*-locus. This RFLP marker CDO98 (Jones *et al*, 2002) as well as the STS marker LOC_05g30800 (Yang *et al*, 2009) mapping 2 cM away from the *S*-locus were used to start the mapping. By comparing different public and non-public genetic maps of *Lolium perenne* (Jones *et al*. 2002; Jensen *et al*. 2005; O'Donoughue *et al*. 1992; Studer *et al*. 2008; Yang *et al*. 2009), twenty pairs of primers were selected to start the first screening of a sample of the S population P235/59 (see Table 2.2).

Table 2.2: Available markers used for the initial mapping of the *S*-locus

Oligoname	Accession number	Forward sequence	Reverse sequence	Screening
As_CDO1173	AA231926/ AA231927	CATGAACGACCGGGATACTT	GGCGTTCTTCATCTTCTTCAA	Polymorphic
G01_031	ES699945	ATGAACACCCAGGATTGGAA	TGTATGCAGCTCAGGGTTTG	No amplification
G01_063	EY457990	GTGGTGCAGTTTGGTCAATG	GACGCAACAAAACTTGTGGA	No amplification
G02_017	ES700234	CCTCATCCCAAACCCTAACC	AGCTCCTTCTCCTCCCTGAC	Polymorphic
G05_081	ES699712	CCAACCAGAAGAGTCGGAAG	AGGGTCTCGGAGATGCTG	No amplification
G05_127	ES700228	AAATGTTGTCCGGTGAGAGG	GATGAGCGGTCTCTTCTTGC	No amplification
Hv_BCD207_SSR_B2-H5	BQ548066/BQ548067	AACAACAGTTTTCGATGACCTC	TCTCTACTGAAAGTGTCATGGTCA	Not polymorphic
Hv_BCD921_WHE01D12	BE438643	CCCTTACCATAGCTGCCTCA	AGCTGGGGATACAGTGCTTTT	Polymorphic
Hv_BCD921_WHE1D0122	BE438796	CTCTCTCCCGCCTCCTCT	GAACACGCACTACGCAGAGA	Not specific
LOC_Os05g25850	-	GCTTTGGATAAAGAGGCCAA	TGCTCCCAGACATCAATTCC	No amplification
LOC_Os05g30800	-	CCTCAGACCATGATGCTTGG	TCCTCTTGCACTGCCTTGTT	Not polymorphic
LOC_Os05g43480	-	CGGCCTGAGGTATTGGAGGG	TGCTTCCGAGCTGAGGTGAG	No amplification
LOC_Os05g46550	-	TTGGCCAGATTTGAGAATGA	ACTTCAATGGGTTCTGCTGC	No amplification
LpBCD1072	AA231687 /AA231688	TGTTACGCCACTGTCTCTGG	TGACCAAGCTGCATCACTTC	Not specific
LpCDO98	AA231728/AA231729	GGTAGGATGGGCAGTTCAGC	AGGCGAGTCTGTTTGTACGC	Not polymorphic
LpHSLF1	AB511859	TGCCTGTGTGTCTCGTGAAT	CTTCACACCAATGCCTGTGT	Not polymorphic
LpHSLF2	AB511862	TTGTTTTGGAGGAGCTGGAT	ACACACTGTGCAAAGACTCG	Not polymorphic
LpSSR057	AY919052	TAGCCTCCAGAAACAAAGTC	CATAGCAGTACAGCCAGTCA	Not specific
Pc_Trx_ext	AF159388/AF159389	GAGGAGCAACCTACGACCAG	TGCCATCGCCAATAGCTT	No amplification
Ta_PSR168_SSR	AJ440601/AJ440602	ATGCGGAAGGTTCAGAGAAA	GCCTCTCCTCAGCTGACATC	Not specific

2.3.2.2. Primers based on *Lolium perenne* reads

In order to develop more compatible markers for the S-population P235/59, primers were designed on *Lolium perenne* sequences. Using CLC Genomics Workbench software (CLC bio, Aarhus, Denmark), an alignment between the rice sequence and some 454 cDNA reads was done to look for conserved region to design primers on. The 454 cDNA read information was given by Dr. Bruno Studer (ETH Zürich, Switzerland) and are reads from RNA extracted from various tissues from three different genotypes of *Lolium perenne*: F1_30 and F1_39, the mapping parents of the VrnA mapping family, as well as another inbred genotype. Approximately 802,156 high quality cDNA reads were available. The rice sequence selected was obtain as a fasta file from NCBI (http://www.ncbi.nlm.nih.gov/mapview/), using the RAP Build 3 of *Oriza sativa japonica*. By doing a reference assembly using the CLC Genomics Workbench software (using default parameters), the cDNA *Lolium* reads were aligned with the rice sequence of the region of interest. For each match between rice sequence and *Lolium* reads, a comparison between the different *Lolium* genotypes and *Lolium* reads was made in order to look for polymorphism such as SNPs and length polymorphisms. Using the software Primer3 (http://frodo.wi.mit.edu/primer3/, Rozen and Skaletsky, 2000), primers were designed on each side of the SNPs, with a PCR product size of between 80 to 150bp. The primer names were given according to the position of the *Lolium* reads (used to design the primers) on the rice sequence in the CLC alignment (i.e. the primer 05_01603 was design from the rice region around 1.603Mb).

To start the fine-mapping, markers were design on the rice region 1.5 to 5Mb on chromosome 5 every 250Kb. The region of interest was narrowed down and further markers were designed. During this fine-mapping, 82 primer pairs have been designed but

only 40 of them have been used to screen the S-population P235/59; they are shown in Table 2.3.

Table 2.3: Fine-mapping markers design on *Lolium perenne* cDNA reads aligned with the rice region 1.5 to 5Mb. For each marker, the table contains the forward and reverse sequence of the primer, the annealing temperature (Ta), the PCR program used (see description of the PCR program in the section 2.3.3) and if the marker is specific and polymorphic.

Name	Forward sequence	Reverse sequence	Ta	PCR program	Polymorphism
05_01603	ATGCTGTCCCAGCTCCAC	GAGGTTGAAGGAGCTGTTCG	60	HiRes63	Yes
05_01953	GTGGGCCACTGCTCGTC	GATCGCCTTGGGGTCCT	61	HiRes63	Not Specific
05_02325	CATTGACCACCCCAATCAAG	TGAGGGGGACTTGCTACCTA	60	HiRes63	Yes
05_02489	ACAACCCTGCCTGGATAACT	GACCAGCAGGCTGAAGACC	60	HiRes63	No
05_02672	GGTACTTGCGAACGATGTCA	CTTCTTCCCGATCCAAAACA	60	HiRes63	Yes
05_02720	AGCCCAACAGCTATTCCAAG	TCTGGTTTTTGGTGAAGGTG	59	HiRes63	Yes
05_02774	CCAGCTGGCTCTTTGGTAAG	GTCGGATGATGATGTTGTGC	60	HiRes63	No
05_02790	CATCGCCAGCATGCTTATAG	CCACTTGCTCTTCCTCTTCC	59	HiRes63	Yes
05_02802	AAATTCTCCGGAGTTGAGGAA	AAATTTGGAGTGGATCCTTCA	59	HiRes63	Yes
05_02823	CTTGTTCCGTTTCGATTGGT	TTGTCTTTCCTGGCCTGTCT	60	HiRes63	No
05_02826	ATTCAGACCAGGCAGCCATA	ACCTGCTGCAAGAAACTGCT	60	HiRes63	No
05_02827	GCAGCAAGTGATGATGAAGG	AGCTGCTTCAACTTGCCATC	60	HiRes63	Yes
05_02833	TGCAGCCAGAGAGGATGTC	ATGCTGGTTTCAGTGCCATT	58	HiRes63	Yes
05_02834	GGCTCGGTTAATTGATCCAAA	AATGCTGCCAGAGCAACAC	61	HiRes63	Not Specific
05_02843	GCTCCTGCACGTAGTTCTCG	CTCCGGTCGGTGCTGTC	61	HiRes65	No
05_02852	GAGGAGTCTAGGGAGGAGGA	GTCATCGCTTCTTCGTCGT	58	HiRes65	No
05_02863	GGAGCAGCTTCAGGAAAACA	TCGGATGGTTCCAACTATGG	60	HiRes63	Yes
05_02867	GCTTCAGCACCTGAACAAAA	TGGTCGGTGAGTCTGAAATG	59	HiRes63	No
05_02869_1	TCCAACATTCCATCAAGCAA	CGAGAATTCTACCAATGCGACT	60	HiRes63	No
05_02869_2	CATTGGCTCATATACTCCAGCA	CCCGTTATGTTGCCATGAAT	60	HiRes63	No

Name	Forward sequence	Reverse sequence	Ta	PCR program for HRM	Polymorphic
05_02870	GCAGCATCAGTACAGCAGAAG	CCACAAGCTACTGGGGAGAA	59	HiRes63	No
05_02871	GCATCTCAAGACGATCAGCA	ACCGTCCAAACTTCACTTGC	60	HiRes65	Yes
05_02876	TAAGGGCTCATCATCCTTGG	TTCACCGAGCTACAATCCTTC	60	HiRes63	No
05_02878	ACCTCTGCCTCAAACACATT	AACTGGATGATGAGAACTGGA	57	HiRes63	No
05_02889	GGACGTCAAGACAACCACTG	GGAGATTGTCTGTGCTTCGAG	60	HiRes65	Yes
05_02904	AGGAAGGACTCGCAGACGTA	GAGGGGTCGGCGTTGAG	61	HiRes65	Yes
05_02907	TCCAACATTCCATCAAGCAA	TGAGCCAATGGAACCAGATT	60	HiRes63	No
05_02911	AAGCGGGAGACGGTTGAG	ACCCGCTACGAGCCTGTG	62	HiRes63	Yes
05_02915	CCGACAATTTCCCGTACACT	TTAGGAAAGATTGGGGTTGG	59	HiRes63	Yes
05_03010	AGGATCGCCGGGAAGAC	GTCATCACGGGCCACAG	60	HiRes63	Not Specific
05_03129	GGTTCTCTGGTTGAGGTGGA	TGGCAACATCAACGAACAGT	60	HiRes63	Yes
05_03170	AACATTGCTGTGGTTGTTGG	TCAGTGAACCTCAGGGCAGT	60	HiRes63	Yes
05_03191	CCCTCGGCATCATCAATC	TGGAAGGCCTTGAGAACATC	60	HiRes63	No
05_03283	AGTCCGATGTTCAGGAAGGA	GAGCTGGAACTGGCGATCT	60	HiRes63	No
05_03433	GTCGCAGAGGTCGCTCAG	GGAGGAGGGTGAGCATGAC	60	HiRes63	Yes
05_03637	CACCCGATCCAGAATTCCTC	GCCAAGCATGCACGTGAG	62	HiRes63	Not Specific
05_03702	CCCACACGGCTTCTCACC	CCTTGTACACCCTCCAGCTC	61	HiRes63	Not Specific
05_04047	CAGCACGGGTACCTAGCAA	CTCCATCCTGTGGGGTAAGA	60	HiRes63	No
05_04731	TCAGAATGACGTGGCCAAT	TGCGTCCAACTCTTTGCTC	60	HiRes63	Yes
05_04820	AGATGCTGAGCGCCAATC	CTTCTCCAGTGGCCCTTTG	60	HiRes63	Not Specific

2.3.2.3. Primers based on rice and *Brachypodium* genes

Using the results of the fine-mapping with the primers designed from *Lolium* reads the syntenic rice region was narrowed down to eight genes. The syntenic *Brachypodium* region was identified with only eleven genes. Primers were designed using the rice and *Brachypodium* gene sequences (http://www.ncbi.nlm.nih.gov/, http://www.gramene.org/), and when possible, the primers were designed from the consensus sequence of the two species genes. The *Lolium* reads used to design the primers in the previous section were also aligned to some rice genes, in order to look for polymorphism between the different ryegrass genotypes.

As described previously, the software Primer3 (http://frodo.wi.mit.edu/primer3/) was used to design the primer pairs, either targeting a gene intron, a polymorphism between the two species sequences or a polymorphism between the *Lolium* genotypes. Primers were designed to have an annealing temperature around 60°C and the PCR product size between 100 to 220bp. The names of the primers were given according to the sequence they were designed from: R is from *Lolium* reads aligned to the rice gene, RB from *Lolium* reads aligned to the consensus gene sequence between rice and *Brachypodium*, B from *Lolium* reads aligned to the *Brachypodium* gene and without letter, from the rice gene itself, without *Lolium* sequence data; the number in the marker name is according to the rice or *Brachypodium* gene it is based on. A total of 16 primer pairs were designed and screened on all S-populations (see Table 2.4).

Table 2.4: Fine-mapping markers design from rice and *Brachypodium* genes. The name of the primer was given according to the gene it was targeting: R for rice, B for *Brachypodium* and RB for the consensus sequence between the two species. The number after the letter is matching the name of the targeted gene.

Name	Forward sequence	Reverse sequence	Ta	PCR program for HRM	Polymorphic	Gene targeted
05_R0149900_1	CGTGCTTATCTAGCGCTCTCT	CCTTCTCTCTCCCACTGGAAC	60	HiRes65	Not specific	Os05g0149900
05_R0150000_1	GCGGTGAGCAAGACGAAG	GAGGCCTTGTCGATGAGC	60	HiRes65	No	Os05g0150000
05_R0150100_1	TGTTTAGGCAAAGGGATGGA	GGGGAGCCAGAAGGAGTAGT	60	HiRes65	No	Os05g0150100
05_R0150300_1	GGCTCTTGGACATCTTGGAA	CTGCCCTAGTCGAAAGCAAG	60	HiRes63	Yes	Os05g0150300
05_RB0149900_2	TTCAGGGGAGCTCAAGAAC	TGATGCATAACCTTCATTGCTT	59	HiRes63	No	Os05g0149900
05_RB0150000_1	CTCCGGTCGGTGCTCTC	GGACGTAGTTCTCCCCGAAG	61	HiRes65	No	Os05g0150000
05_RB0150300_1	TTGTTCAGCAGCTTCATAAGG	GGTAAACCTTTCTCAACATCAGA	58	HiRes63	Yes	Os05g0150300
05_B35790.1_1	GTCTTCCTCCCCGAGAAATC	GACCTTCCGGTAGAACCACA	60	HiRes63	No	Bradi2g35790.1
05_B35780.1_1	ATTTCATGCCATTCCGAAAC	AGTCCTGGGCACCCATATC	60	HiRes63	Yes	Bradi2g35780.1
05_B35770.1_1	TGTCACTGCCTCCTCTTCCT	ATTTTGGAACGCCAAAACTG	60	HiRes65	No	Bradi2g35770.1
05_B35750.1_2	TCCAGAAGTAACGGGGGTAA	TATGGCCACTGCTCTGTGAG	60	HiRes63	Not specific	Bradi2g35750.1
05_B35750.1_3	AAGGTGGCACCAACTACTACG	TTGTATGGTGTTCATCCTGGTT	59	HiRes63	Not specific	Bradi2g35750.1
05_0149900_1	GGAGAAGCAAATGAAGGGAAT	CATCATCTTGGTCACCATCG	60	HiRes63	No	Os05g0149900
05_0149900_2	CGATGGTGACCAAGATGATG	CCATCCGCACCACATAGAG	60	HiRes63	No	Os05g0149900
05_0150000_1	GGGATTAGCAAAACATGTCCA	AAATTCTCCGGAGTTGAGGAA	60	HiRes63	No	Os05g0150000
05_0150000_2	TGGTGCAAGAGAATACCCAAA	AAACGGAAGGTTGGACATTG	60	HiRes63	No	Os05g0150000

2.3.3. Screening method

The initial screening of the S population P235/59 with publically available markers was done using a subset of 16 randomly chosen family member plants in order to identify polymorphic markers showing segregation distortion (i.e. an excess of heterozygous plants). Any homozygotes isolated from this first screening were then used to test the new primers as they are either recombinants or self (a plant is considered as resulting from a self-pollination if it is genotyped homozygote for both flanking markers). Any new primer was tested on a subset of DNA made out of recombinants previously identified and the parent P232/88/10 as well as two plants (P235/40/14 and P235/40/17) related to the parent P235/40/13.

DNA was amplified by Polymerase Chain Reaction (PCR) in a 10µl PCR reaction using 1X LightScanner Master Mix (Idaho Technology Inc., Salt Lake City, UT, US), 0.3µM of each forward and reverse primer and 1µl of DNA (approximately 20ng/µl). An additional 14µl of mineral oil (Sigma-Aldrich, St. Louis, MO, US) were added before the PCR to avoid sample evaporation during the melting of the plate. The PCR program used are shown in Table 2.5. The melting of the plate was done between 60 and 98°C using the 96-well LightScanner (Idaho Technology Inc., Salt Lake City, UT, USA) and the genotyping was done using the software provided, LightScanner® and Call-IT® (Idaho Technology Inc., Salt Lake City, UT, USA).

Table 2.5: PCR program for the screening of the markers using HRM curve analysis.

HiRes63 Temperature	HiRes65 Temperature	Time	
95	95	02:00	
94	94	00:30	40 cycles
63	65	00:30	
72	72	00:30	
72	72	02:00	
94	94	00:30	
25	25	00:30	
4	4	∞	

After the initial screening with existing markers and the first screening with new markers designed on *Lolium* reads, two flanking markers (i.e. markers that showed the lowest levels of recombination) were identified: 05_02720 and 05_02915. These two markers were used throughout the fine-mapping process as their genotyping was easy, the difference between the curve shapes of homozygote and heterozygote samples was clear (see Figure 2.5). Any recombinant for those two flanking markers was then screened using any polymorphic markers mapping within the *S*-region, bounded by the two recombinant flanking markers.

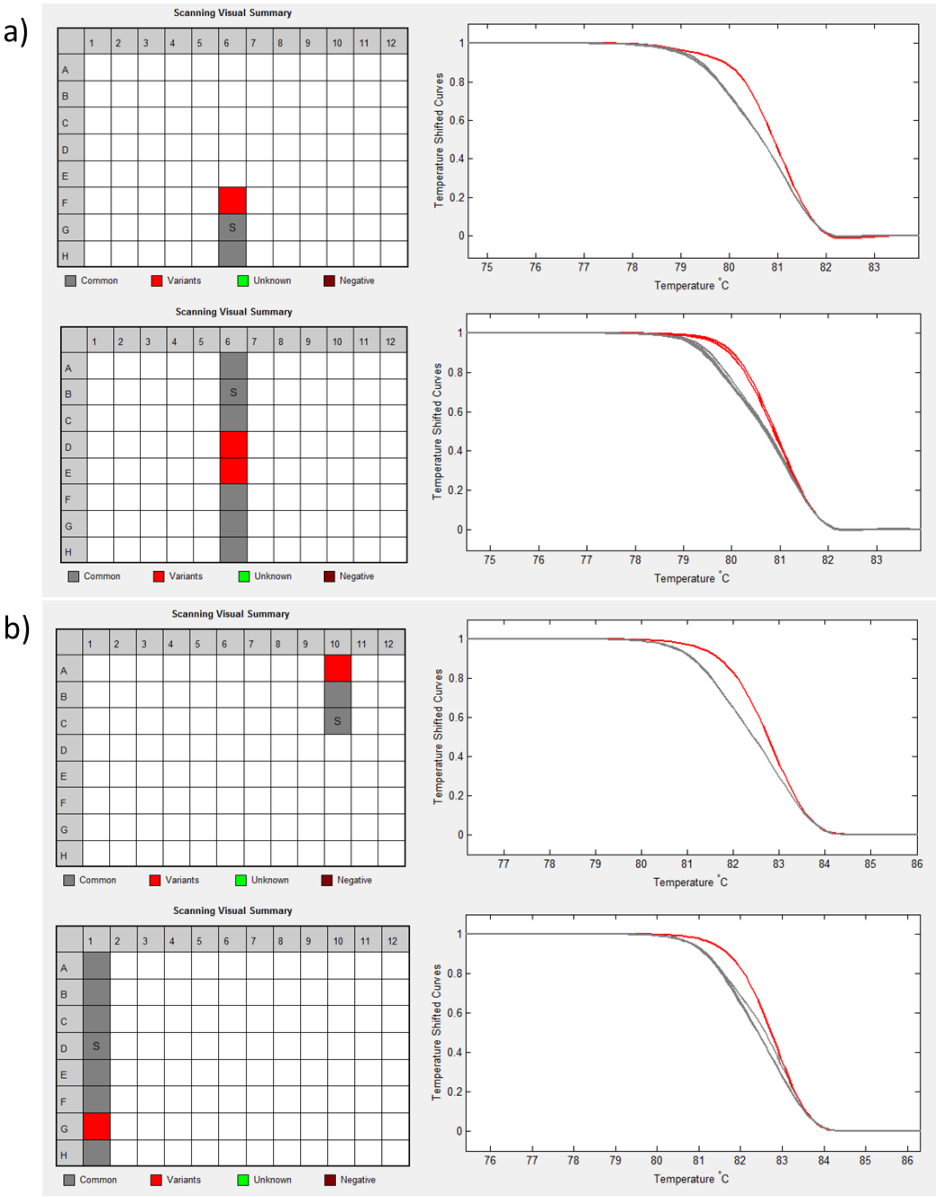


Figure 2.5: HRM profiles of the two flanking markers (a) 05_02720 and (b) 05_02915. For each marker, the top table and graphic represent the parents' genotype and melting profile curves, the bottom table and graphic, an S-subset made with recombinant plants from the 242 plants of the P235/59 S-population. The parents are P232/88/10 (S_{22}), P235/40/14 (S_{23}) and P235/40/17 (S_{23}) from top to bottom in the table representation. For both markers, the parent P232/88/10 (in red) shows a different curve shape as its genotype for the S-locus is homozygous. In the S-subset, the red samples have the same genotype as the parent P232/88/10, which mean that they represent recombinants between the markers and the S-locus: 4th and 5th plants (D6 and E6) for marker 05_02720 and the 7th (G1) for 05_02915. If a plant was showing a homozygote genotype for both markers, it would be classified as a self and not a recombinant.

For the large scale screening, in order to save time and money, the protocol was slightly different for the genotyping. DNA was amplified by PCR in a 6µl volume reaction containing 1X LightScanner Master Mix (Idaho Technology Inc., Salt Lake City, UT, US) and 0.3µM of each forward and reverse primer. In order to make the screening more efficient, the DNA was transferred to the PCR mix using a stamp tool. The clean tool was sterilized tree times in 90% ethanol before being flamed. Once the stamp tool was cool enough, it was dipped into the DNA (undiluted after the DNA extraction therefore highly concentrated) and then transferred to the PCR plate, by stirring the stamp tool into the PCR mix. The amplification was carried as described previously in Table 2.5.

2.4. Results

2.4.1. Initial screening using available markers

The initial screening of the S-population P235/59 with these primers was done using a subset of 16 plants. Out of the twenty initial markers, only three were polymorphic and specific to the *S*-locus region on LG 1 (see Table 2.2, section 2.3.2.1). Two markers, As_CDO1173 and Hv_BCD921_WHE01D12 were selected to screen all 242 plants of the S population P235/59. Only six recombinants out of 242 plants were found for the marker As_CDO1173 but the marker Hv_BCD921_WHE01D12 had too many recombinants (19 out of 96), leading to the conclusion that it was too far away from the *S*-locus. Unfortunately, both markers mapped around 2cM from the *S*-locus, LpCDO98 and LOC_Os05g30800 were not polymorphic for the S-population P235/59 so they were not used in the fine-mapping.

By querying BLAST researches of the markers sequences, the rice region 1.5 to 5Mb on chromosome 5 was selected to design new markers, using *Lolium perenne* reads.

2.4.2. Screening of the whole S-population P235/59 with primers designed from *Lolium perenne* reads and rice and *Brachypodium* gene sequence

The initial screening of the 242 plants of the S-population P235/59 with existing (published and non-published) markers has led to the selection of a rice region of 3.5Mb (between 1.5 and 5Mb on chromosome 5). From this region, markers were designed using *Lolium* reads aligning on the rice region of interest. Because of the design of this S-population, the genotyping of the plants was made easy: any homozygote was deemed to be a recombinant or a self.

The initial screening enabled selection of recombinants for the marker As_CDO1173 only, which is on the top of the *S*-locus region in *Lolium perenne*. In order to detect a flanking marker at the other end of the *S*-locus region, four markers designed at the end of the rice region of interest have been screened on the 242 plants: 05_03702, 05_04047, 05_04731 and 05_04820. Only the marker 05_04731 was polymorphic. Using the two flanking markers, As_CDO1173 and 05_04731, an *S*-subset of 14 recombinants has been made to test the new markers for polymorphism. This subset also included one parent: P232/88/10, homozygous for the *S*-locus and the plant P235/40/17 which is heterozygous for the *S*-locus (similar to the parent P235/40/13).

Using this *S*-subset, markers designed in the *S*-region were screened for polymorphism (see Table 2.3, section 2.3.2.2). According to the number of recombinants, two flanking markers were selected (05_02720 and 05_02915) to screen any new DNA

extracted from the S-population P235/59. The region of interest was narrowed down by taking the number of the recombinants into account. More primers were designed along the screening, and used to genotype any recombinants for the flanking markers.

From this P235/59 population, a total of 1393 plants have been screened with the two flanking markers: 05_02720 and 05_02915. All the recombinants have then been screened with eight polymorphic markers mapped between the two flanking markers. The results are shown in Table 2.6.

Table 2.6: Results from the screening of the S-population P235/59 with ten markers using the HRM curve analysis. The markers are linearly organized according to their rice position.

Marker name	**Number of recombinants out of 1393 plants**
05_02720	27
05_02790	3
05_02827	0
05_02833	0
05_02863	0
05_02871	0
05_02889	1
05_02904	2
05_02911	3
05_02915	21

The region of interest (between markers 05_02790 and 05_02889) was queried by BLAST to the *Oryza sativa* genome (using megaBLAST, http://www.ncbi.nlm.nih.gov/ genome/seq/BlastGen/ BlastGen.cgi?taxid=4530) and to the *Brachypodium distachyon*

genome (using BLASTN, http://www.gramene.org/Multi/blastview?species=Brachypodium_distachyon).

In the region of interest in rice, eight genes are annotated on the 60Kb region with four of them having *Lolium* markers on them. According to the *Brachypodium* genome, eleven genes are annotated in the same region, which represents approximately 140Kb. The description of the genes as well as the synteny between those rice and *Brachypodium* genes can be found in Table 2.7.

Table 2.7: Synteny between *Brachypodium* and rice genes in the *S*-locus region. The function of those genes has been found on http://www.ncbi.nlm.nih.gov/gene/ and the BLAST queries using http://blast.ncbi.nlm.nih.gov/ and http://www.brachybase.org/blast/.

Brachypodium genes	**Hits on Rice genome**	**E value**	**Query coverage**	**Rice protein information (NCBI)**
Bradi2g35707.1	Os05g0150600	0.0	23%	ATP-dependent DNA helicase RecQ family protein
Bradi2g35720.1	Os05g0150500	0.0	37%	Cyclin-like F-box domain containing protein, leucine rich, transport inhibitor response 1 (TIR1)
Bradi2g35730.1	Os05g0150400	1E-85	13%	Double-stranded RNA binding domain containing protein
Bradi2g35740.1	Os05g0150300	0.0	42%	Similar to possible global transcription activator SNF2L1 (SWI/SNF related matrix associated actin dependent regulator of chromatin subfamily A member 1)
Bradi2g35750.1	Os05g0198000	0.0.	60%	Protein of unknown function DUF247 plant family protein
Bradi2g35760.1	Os05g0150000	1E-89	27%	Protein of unknown function UPF0001 family protein
N/A	Os05g0150100	-	-	Protein of unknown function DUF295 family protein
Bradi2g35767.1	Os07g0141700	1E-61	6%	NB-ARC domain containing protein, unknown function
Bradi2g35780.1	Os05g0163700	7E-118	15%	Similar to Acyl-coenzyme A oxidase 4, peroxisomal
Bradi2g35790.1	Os04g0162600	2E-109	73%	Protein of unknown function DUF295 family protein
Bradi2g35800.1	Os05g0149900	4E-118	26%	Tetratricopeptide-like (TPR-like) helical domain containing protein
Bradi2g35807.1	Os05g0149800	6E-85	25%	Calcium-binding EF-hand domain containing protein

2.4.3. Large scale fine-mapping of the *S*-locus using different *S*-populations

In order to increase the number of plants for the *S*-locus mapping, more *S*-populations had to be used as all the plants from the population P235/59 (1393 plants) were used in the first fine-mapping of the *S*-locus. Two populations derived from the same grand-parents as P235/59 were used: P235/63 and P235/64 (described in section 2.2.1.), and named DTS for this project. Moreover, four populations unrelated to the S-population P235/59, but related between each other were added for the fine-mapping: S216, S267, S324 and S404 (described in section 2.2.2). Those new VrnA-S populations are not related to the first population used and therefore, the *S*-locus should have different *S*-alleles and could as well as have different polymorphisms for the markers used, but the genotype of all the plants should be heterozygote at *S* unless recombinant.

The DNA was extracted using a different method, faster and cheaper, described in the section 2.3.1.3. The DNA was amplified using two markers flanking the *S*-locus: As_CDO1173 and 05_03433, and not 05_02720 and 05_02915 as their genotyping were impossible (monomorphic marker) or not as easy with the VrnA-S populations. The Table 2.8 shows the results of this genotyping with the two flanking markers selected. This large-scale fine-mapping was done using 8,784 plants in total, spread between six populations (see Table 2.8 for the detailed of the plant number).

Table 2.8: Large-scale fine-mapping of the *S*-locus using two flanking markers and five large mapping populations. The table is representing the number of plants screened per populations as well as the number of potential recombinants and selfs for both flanking markers As_CDO_1173 and 05_03433.

	Number of recombinants		**Number of plants resulting from a self-pollination (%)**	**Size of the population**
	Primer 6 (As_CDO_1173)	**Primer 05_03433**		
VrnA_S216	26	1	29 (2%)	1425
VrnA_S267	8	7	20 (2%)	964
VrnA_S324	11	0	8 (1%)	746
VrnA_S404	13	3	2 (0.2%)	994
DTS	40	19	4 (0.08%)	4655

The genotyping of these plants led to the selection of 46 plants, homozygous for either flanking markers or both (plant resulting from a self-pollination), to be screened by the other *S*-linked markers. The results of this final screening are shown in Table 2.9. Not all the *S*-linked markers were successfully screened. Because the marker 05_02827 is closely related to the marker 05_02833, only the marker 05_02833 was used for the screening as the scoring was easier with this marker. The same was done with the marker 05_02863, used to screen any recombinants but not the marker 05_02871.

Table 2.9: Genotyping using *S*-linked markers of the 46 recombinants isolated with the large screening using the flanking markers As_CDO_1173 and 05_03433. All the *S*-linked markers previously identified were screened on all the recombinants but not all were polymorphic; therefore, the number of recombinants does not always represent the exact genetic distance of the *S*-locus (* marked the number of recombinants out of 4655 plants from the DTS population only).

Marker name	**Number of recombinants out of 8,784 plants**	**Polymorphism for DTS populations**	**Polymorphism for VrnA-S populations**
As_CDO_1173	98	Polymorphic	Polymorphic
05_02790	6	Polymorphic	Polymorphic
05_02833	0	Polymorphic	Polymorphic
05_02863	0	Polymorphic	Polymorphic
05_02889	0*	Polymorphic	Monomorphic
05_02904	0*	Polymorphic	Monomorphic
05_02915	6*	Polymorphic	Monomorphic
05_03433	30	Polymorphic	Polymorphic

The recombinant plants were re-potted in individual 3.5inch pots in order to let them grow; the rest of the plants were discarded in order to save space and time.

2.4.4. Genotyping of the recombinants with the gene specific markers

Using the synteny between *Lolium* and rice and *Brachypodium*, new markers were designed from the gene sequences identified in the *S*-region (see section 2.3.2.3.). These markers were used to genotype the 46 plants isolated from the large-scale fine-mapping, homozygous for either flanking markers or both (plant resulting from a self-pollination), as well as the recombinants isolated from the S-population P235/59. Unfortunately, only three

markers were showing some polymorphism: 05_R0150300_1, 05_RB0150300_1 and 05_B35780.1_1. Out of those three markers, the markers 05_R0150300_1 and 05_RB0150300_1 were showing no recombinants. The marker 05_B35780.1_1 however was showing some good polymorphism and was easy to score; the recombination rate was 29 out of 10,177 plants.

2.5. Discussion

The strategy in this project to identify and map the *S*-locus was to use a recombination-based fine-mapping approach, which is a common method to identify genes. The *S*-locus mapping has been conducted in different grass species, but so far, none of them have succeeded: in *Lolium perenne*, the closest marker CDO98 is mapped at 0.1 cM from the *S*-locus (Yang *et al*., 2009), in *Phalaris coerulescens*, the closest markers (Wg811 and Psr168) were mapped at 0.13 cM (Bian *et al*., 2004) and in *Hordeum bulbosum*, marker HTL showed complete linkage (0 cM) to *S*-locus as no recombination were found with a population of 662 plants (Kakeda *et al*., 20008). Using markers available and mapped in the region close to the *S*-locus (Jones *et al*. 2002; Jensen *et al*. 2005; O'Donoughue *et al*. 1992; Studer *et al*. 2008; Yang *et al*. 2009), the mapping of the *S*-locus was done initially on a small S-population of 242 plants. New markers were designed once the *S*-region was delimitated on the rice chromosome 5, using *Lolium* reads in order to have a higher chance of polymorphism in our population. As the markers were designed and screened for recombination, the S-population was increased to 1393 plants at first, and by using other unrelated or not populations, the large-scale fine-mapping was done by genotyping 10,177 plants.

With this fine-mapping, the *S*-locus has been delimitated to a region less than 0.11 cM with nine recombinants out of 10,177 plants screened for the flanking marker 05_02790 and one recombinant for the other flanking marker, 05_02889, out of 6048 plants (DTS and P235/59 as the marker is monomorphic for the VrnA-S populations). The mapping was carried out by adding markers and plants along the screening in order to reduce the size of the *S*-locus region. In order to have a large number of plants, several populations had to be used.

However, it came with a drawback: not all the markers were polymorphic for every population. This resulted in the impossibility to calculate the exact genetic distance for the lower flanking marker (05_02889) but did not prevent to reduce the *S*-locus region as all the non-recombinant markers, previously identified with the first screening of the P235/59 S-population, were polymorphic for all population.

The first fine-mapping described in this study led to an *S*-locus region with a genetic size of 0.30 cM (using 1393 plants). The large-scale fine-mapping conducted afterward using several populations reduced the genetic size of the *S*-locus down to 0.11 cM. Unfortunately, it did not reduce the physical size of the *S*-locus as no recombinants were found for any of the non-recombinant markers isolated in the fine-mapping using only the P235/59 population. In this case, increasing the number of plants proved not to be helpful for the mapping. However, because several unrelated populations were used in the mapping, we can be sure that the segregation distortion observed in this region is not due to the population used in the mapping but to the *S*-locus itself. Moreover, because no recombinants were found for the markers 05_02833 and 05_02863 out of over 10,000 plants, the feeling that these markers are linked and maybe mapped onto the *S*-genes is strong.

Another observation from that large-scale fine-mapping is that the number of recombinants as well as the number of selfs is different according to the population, but the variation of recombinants is not significant between populations. However, the number of self varies from 2% (VrnA-S216) to less than 0.1% (DTS). These variations can be explained by the way populations were created: by cross-pollination in an isolation chamber for the DTS and P235/59 populations and by cross-pollination using a bag for the VrnA-S populations. The use of bags for cross-pollination can increase the number of self-pollination as self-incompatibility can be modulated by changes in temperature and the temperature in bags is often higher.

The results of the fine-mapping have been summarised in the Figure 2.6 that shows the *S*-locus region with the *S*-linked markers as well as the synteny between *Lolium* and two genetically close species: rice and *Brachypodium*. The *S*-locus region, by synteny, includes eight genes in rice on chromosome 5, from 2.829 to 2.889 Mb but correspond to a larger region in *Brachypodium* on chromosome Bd2: from 36.110 to 36.248 Mb (138 Mb), with eleven genes annotated. However, not all these genes can be considered as candidates genes. The two flanking genes of the *S*-locus region in rice and *Brachypodium*, Os05g0149800 and Os05g0150600, Bradi2g35807.1 and Bradi2g35707.1 respectively are not considered as candidates genes as both flanking markers 05_02790 and 05_02889 map onto them. The number of recombinants for the marker 05_02790 is higher than for the marker 05_02889 (9 against 1) which would mean in theory that 05_02790 is further away from the *S*-locus than 05_02889 (0.088 cM against 0.016 cM). However, because of the effect of segregation distortion on the linkage map, this result has to be considered carefully and every gene in this region should be considered as a candidate gene, with some degree of preference maybe.

But even if the gene Os05g0149800, syntenic to the *Brachypodium* gene Bradi2g35807.1 cannot be an S-gene, it does not necessarily means it is not involved in the SI mechanism and the cascade resulting from the recognition of the self-pollen. Indeed, the gene Os05g0149800 is annotated as a protein containing a calcium-binding EF-hand domain. The GSI system of *Papaver* is rejecting the self-pollen through an intracellular signalling pathway, involving rapid calcium cytosolic concentration increase, leading to the activation of a Ca^{2+}-dependent protein kinase (CDPK) (Franklin-Tong, 1999; Holdaway-Clarke and Hepler, 2003; Ge *et al*., 2007). Moreover, in *Brassica rapa*, an SSI system, a Ca^{2+}-signalling-related actin reorganization has been reported by Iwano *et al*. (2007) to occur after a self-pollination. In grasses SI, calcium could also play a part in the chain reaction as Wehling *et al*. (2004) have shown in *Secale cereale* that SI could be suppressed by the application of Ca^{2+} channel blockers. The same effects were observed in *Lolium perenne* (Yang *et al*., 2009; Klaas *et al*., 2010). Therefore, even if the gene Os05g0149800 is not involved in the initial recognition and interaction between the pollen and the stigma, it can be part of the SI mechanism and be involved in the self-pollen rejection.

Another gene that is not considered as a candidate gene for *S* is the *Brachypodium* gene Bradi2g35780.1. Indeed, the marker 05_B35780.1_1, designed by aligning *Lolium* cDNA reads to the gene sequence of Bradi2g35780.1, has shown some recombination events: 29 recombinants out of 10,177 plants, so more than the marker 05_02889, which is the flanking marker , and nearly as many as the marker 05_02915. By synteny, the gene Bradi2g35780.1 corresponds to the rice gene Os05g0163700, which is mapped in rice around 3,710 Kb. The marker 05_03433 has been initially designed around the rice region 3,433 Kb and mapping to the gene Os05g0159200 so it seems that for this gene, the

synteny between rice and *Lolium* is better than the synteny between *Brachypodium* and *Lolium*, even if the marker 05_B35780.1_1 is mapped above the marker 05_03433, which is not the order they are mapped in rice (see Figure 2.6).

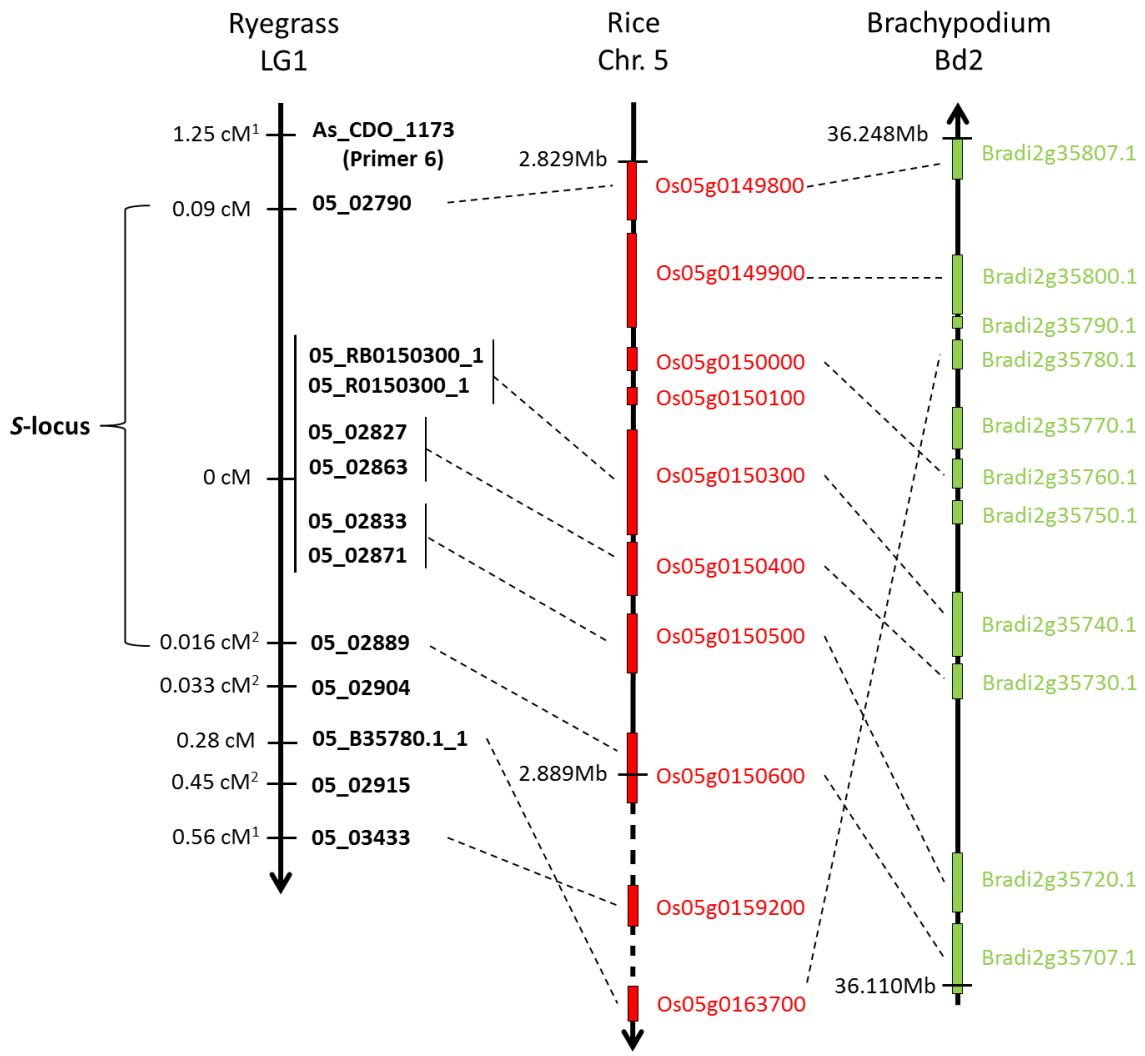


Figure 2.6: Syntenic map between *Lolium perenne*, *Oryza sativa* and *Brachypodium distachyon* in the *S*-locus region of *Lolium*. The ryegrass linkage group 1 is represented with the markers used in the fine-mapping. The genetic distances have been calculated from the recombination frequency out of 10,177 plants, for the exception of ([1]) calculated out of 6048 plants (DTS and P235/59) and ([2]) out of 8784 plants (DTS and VrnA-S populations). The rice chromosome 5 and *Brachypodium* Bd2 are represented with the genes present in this region of interest, data taken from the NCBI Genes_Sequence Map (http://www.ncbi.nlm.nih.gov/projects/mapview/maps.cgi?taxid=4530&build=current&chr=5&beg=2790000&end=2900000) and the *Brachypodium distachyon* JGIv1.0 Gene annotation respectively (http://www.gramene.org/Brachypodium_distachyon/Location/Genome; The International *Brachypodium* Initiative, 2005). The distances on those linkage groups are physical distances and the chromosome orientation is shown by the arrow on the bar.

The remaining six genes mapped in the *S*-locus rice region are corresponding to five of the nine remaining genes in *Brachypodium*. By querying BLAST, the other four genes included in the *Brachypodium S*-region can also be found in rice, but on different chromosomes (Os4 and Os7). The synteny between the *Brachypodium* chromosome 2 (Bd2) and the rice chromosome 5 (Os5) is good (see Figure 2.7, The International *Brachypodium* Initiative, 2010) even if Os5 is a mirror image of Bd2 (orientated in the opposite direction). However, it is not unusual to have loci homologous to other rice regions as represented on Figure 2.7, with small coloured stripes all along *Brachypodium* chromosomes.

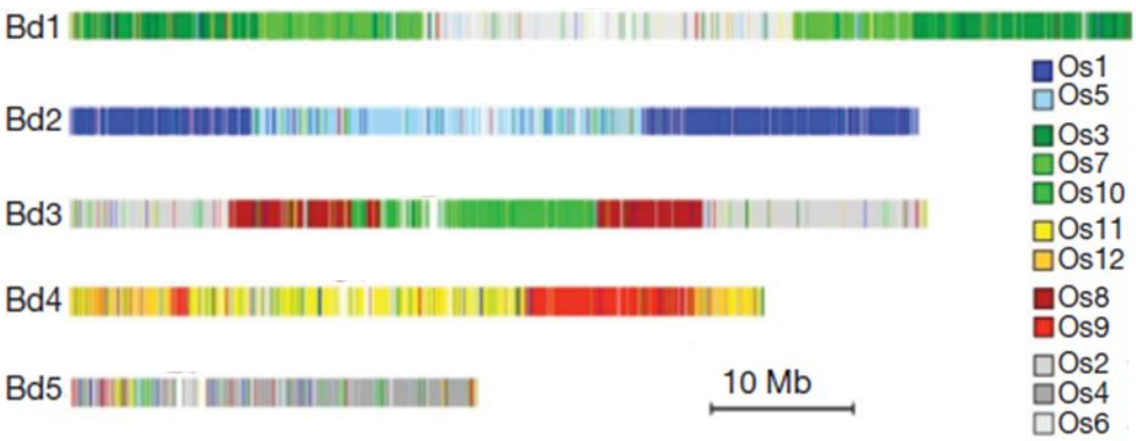


Figure 2.7: Synteny between the rice and *Brachypodium* genomes. The five *Brachypodium* chromosomes are represented in different colour, according to their synteny with rice chromosomes. Figure modified from The International *Brachypodium* Initiative (2010).

Within the five genes that have a good synteny between Os5 and Bd2, the gene Os05g0150000 (homologue to Bradi2g35760.1) is annotated as a protein coding gene but with unknown function. The coding sequence in rice is short for this gene, 642 bp, and the start and stop codon are not identified. However, the gene Bradi2g35760.1 is bigger, 4500

bp with an mRNA of 1176 bp, and even if its function is unknown, the protein it is coding is similar to a pyridoxal phosphate enzyme. This enzyme is involved in amino-acid biosynthesis, including deamination and transamination (Hayashi, 1995). This function makes the gene Bradi2g35760.1 unlikely to be a good candidate gene as an *S*-gene as it is not involved any recognition mechanism.

Among the rice and *Brachypodium* genes mapped in the *S*-region, the gene Os05g0150100, which is not found in *Brachypodium*, and the gene Bradi2g35790.1, syntenic to the rice gene Os04g0162600, are annotated as protein coding genes with unknown function. However, those genes have a common domain in their protein: DUF 295. A DUF domain is a Domain of Unknown Function; these domains are classified by number according to their similarity and represents around 20% of the protein family known so far (Bateman *et al*., 2010). This DUF 295 can be found in different species, such as *Arabidopsis thaliana*, maize, *Sorghum bicolor* and *Hordeum vulgare*. It has been reported over 300 times to be associated with an F-box domain, cyclin-like domain (http://www.ebi.ac.uk, interpro domain IPR005174). The gene Os05g0150500 (syntenic to Bradi2g35720.1) in the *S*-region is also annotated as a cyclin-like F-box domain containing, transport inhibitor response 1 protein. F-box domains were first described by Bai *et al*. (1996) as a motif required in the protein-protein interaction and the first function of F-box protein described was as a component of SCF (Skp1-Cullin-F-box) ubiquitin-ligase (E3) complex (Skowyra *et al*., 1997), conferring the specificity of the SCF complex to recognize and target certain substrates (Kipreos and Pagano, 2000). SCF complexes are made out of four components: Skp1 (S-phase kinase-associated protein), a cullin, Rbx1 and an F-box protein (Deshaies, 1999; Craig and Tyers, 1999) and are creating the link between substrates and ubiquitin-conjugating enzymes, which will lead to the transfer of

ubiquitin to the substrate, marking it for proteolysis (Hershko and Ciechanover, 1998). SCF complexes are known to be involved in SI and more specifically the GSI *S*-RNases systems found in Solanaceae, Scrophulariaceae and Rosaceae, where the male component of the *S*-locus is an F-box (SLF or SFB) (Sijacic *et al*., 2004; Lai *et al*., 2002; Ushijima *et al*., 2003). Therefore, because of the F-box part in the self-pollen rejection in GSI *S*-RNase system, as well as the specificity of recognition it confers, genes like Os05g0150500 are good candidate genes.

Another gene in this *S*-region annotated as coding for an unknown protein is the *Brachypodium* gene Bradi2g35750.1, syntenic to the rice gene Os05g0198000, mapped around 6,030 Kb on LG5. Kakeda *et al*. (2008) have identified four linked markers to the *S*-locus in *Hordeum bulbosum*: LC34, HTL, HPS10 and HAS175. The marker HPS10 is mapping on the rice gene Os05g0198050, which is located next to the candidate gene Os05g0198000 (Bradi2g35750.1) found in this study. Moreover, the marker HPS10 has been shown to be specifically expressed in pistil, with an expression level increasing with flower maturity (Kakeda *et al*., 2008). So even if the functions of the genes are unknown, the gene Bradi2g35750.1 (Os05g0198000) remains a possible good candidate gene.

Another gene included in the *S*-locus region is the *Brachypodium* gene Bradi2g35800.1, syntenic with the rice Os05g0149900, annotated as a protein containing a tetratricopeptide-like (TPR-like) helical domain. TPR repeats are structural motifs that lead to the folding of the protein, giving it a function such as mitochondrial or peroxisomal transport for example. They have a highly specific binding site targeting peptide, and their basic function is to mediate protein-protein interactions (D'Andrea and Regan, 2003). Therefore, because of their specificity, Bradi2g35800.1 can be a good candidate gene even if it has not been reported previously in any SI systems.

The gene Bradi2g35730.1, syntenic with the rice gene Os05g0150400, is annotated as a protein containing a double-stranded RNA binding domain (dsRBD). Double-stranded RNA binding proteins (dsRBP) are widely abundant protein domain with key roles in post-transcriptional regulation of gene expression (Burd and Dreyfuss, 1994). These proteins, through their specific dsRBD, are interacting with ds-RNA. However, the interaction of these protein are not specific to the ds-RNA sequences they interact with and therefore, this gene is unlikely to be an *S*-gene.

Another gene with a transcription function is the gene Bradi2g35740.1, syntenic to rice Os05g0150300 and on which the markers 05_02833 and 05_02871 are mapped. This gene is coding for a global transcription factor activator similar SNF2L1, and this type of protein are known to be involved in processes such as transcription regulation, DNA repair, DNA recombination and chromatin decondensation. Therefore, again, it is unlikely to be an *S*-gene but could still be involved in the SI response.

Finally, the last annotated gene in the *S*-locus region is the *Brachypodium* gene Bradi2g35767.1, syntenic to rice Os07g0141700, mapped on LG7. This gene is annotated as a protein containing an NB-ARC domain (Nucleotide-Binding – APAF-1, R protein and CED-4), of which function is thought to be the regulation of the R genes involved in disease resistance in plants (Van Ooijen *et al*., 2008).

All genes detailed previously have been mapped by synteny to the *S*-locus *Lolium* region. However, because rice and *Brachypodium* are not self-incompatible, it is possible that the *Lolium S*-region contains some genes not annotated or not mapped in the syntenic regions of rice and *Brachypodium*. In order to investigate the genes present in the *Lolium S*-locus, the *S*-region has to be sequenced which is the topic of the following chapter.

Chapter 3:

S-locus region study

3.1. Introduction

3.1.1. Aims and strategy

After detailed definition of the *S*-locus region by fine-mapping, the aim of the project was to study the region at molecular level. The aim was to sequence the *S*-locus region by identifying and sequencing the BAC clones that span the region. The sequencing was done using next generation sequencing tool: 454 pyrosequencing (Roche). Once the nucleic acid sequence of the *S*-locus was built, the annotation of the region, meaning the identification of the genes, was done by finding synteny with other closely related plants such as *Brachypodium distachyon*, *Hordeum bulbosum* and *Oryza sativa*.

Finally, after identifying the genes in the *S*-locus region, the strategy to narrow down the number of candidate genes was to look at the expression of these genes in different tissues where the genes were likely to function. Because it is likely that the *S*-locus has at least two components, a female part and a male part, the gene expression study was done on stigma and pollen samples, as well as stigma samples pollinated with pollen giving self-incompatible and cross-compatible pollinations. From those samples, RNA was extracted and sequenced, using a different tool: Illumina® sequencing. Once assembled, the RNA reads were compared between samples in order to identify genes that were tissue-specific and overexpressed, therefore good *S*-gene candidates.

3.1.2. Next generation sequencing: methods, tools and advantages

The first sequencing system developed was in 1977 by Frederick Sanger. The sequencing was based on the chain-termination method which is the production of a complementary sequence of target sequence, by adding modified nucleotides one at a time.

The Sanger sequencing was then automated by Applied Biosystems in 1987 and current models are still in use today, e.g. for projects such as sequencing PCR amplified product and SNP genotyping. A number of sequencing platforms, termed 'next generation sequencing' (NGS) (Kircher and Kelso, 2010) are available today, which allow high-throughput sequencing in parallel, with reduced costs thanks to technology advances: from Roche (454; Margulies *et al.,* 2005) and Life Technologies (Ion torrent; Rothberg *et al.,* 2011), Illumina (HiSeq and MiSeq; Bentley *et al.,* 2008) and Pacific Biosystems (PacBio; Eid *et al.,* 2009). These platforms use different methods to achieve the large scale sequencing output they deliver; for example, both 454 and ion torrent use pyro-sequencing but are detected by fluorescence or pH differences respectively (Margulies *et al.,* 2005; Bentley *et al.,* 2008). NGS produce thousands-millions of nucleotide sequences in parallel, whether it is in predefined wells (e.g. 454 and ion torrent) or on an array (e.g. Illumina). The main advantages for using these platforms are the reduced cost, reduced time and high throughput (see Table 3.1). Advances have been such that the human genome sequencing project, sequenced using Sanger sequencing, was completed in 2003 (The Human Genome Project) after approximately 13 years and at a cost of around £1.8b. Using NGS technologies, five human genomes can be sequenced today in one sequencing run, with a cost of £3000 per genome, all in one week.

All of these advances have given rise to a number of novel applications such as RNA-seq (transcriptome profiling) and CHIP-seq (transcription factor binding site identification), to name just two (Kircher and Kelso, 2010; Morozova & Marra, 2008). However, despite their increased sequencing output, assembly of higher eukaryote genomes still remains difficult given their inherent complexity (typically due to repetitive regions) (Mortazavi *et al.,* 2008).

Table 3.1: Comparison table between the two most popular NGS, 454 and Illumina, with Sanger sequencing. Table adapted from Liu *et al*. (2012).

	454 GS FLX + (Roche)	**HiSeq 2000 (Illumina)**	**Sanger ABI3730 (Applied Biosystems)**
Read length	700 bp	50 bp, 50 bp PE and 101 bp PE	400 - 900 bp
Throughput	0.7 Gb	600 Gb	1.9 - 84 Kb
Time/run	24 hours	3 - 10 days	20 mins to 3 hours
Cost/million bases	£6.50	£0.04	£1,500
Advantages	Read length, fast	High throughput	High quality, long read length
Disadvantages	Error rate when polybase more than 6 (polyA tail for example), high cost, low throughput	Short read assembly	High cost, low throughput

The two most popular NGS technologies used today are the 454 and the HiSeq (Illumina). These platforms use different principles for sequencing and sample preparation. The Figure 3.1 and 3.2 compare and contrast these two technologies.

454 was the first NGS platform to be available and was developed in 2005 by Jonathan Rothberg and his team, who later developed the Ion torrent platform (Life technologies). Rothberg *et al.* (2011) recognised that to overcome single tube Sanger sequencing reactions (one sample sequenced at a time), a number of innovations were required, such as reducing the volume of the sequencing reaction from micro-litres (10^{-6}l) to pico-litres (10^{-9}l). To reduce the sequencing reaction volume, emulsion PCR was developed: by mixing oil with water, pico sized droplets of water are created, where each droplet contains a single DNA strand (see Figure 3.1). 454 uses pyro-sequencing which

uses a complementary enzyme reaction: an Adenosine Tri-Phosphate (ATP) molecule is created upon nucleotide incorporation, the ATP is then converted to light by luciferase enzyme, which is detected (Ronaghi *et al.*, 2000). Critically, for sequence determination, single species of nucleotide are added sequentially. For example, cytosine nucleotides are added to the reaction which are complementary to the strand of DNA being amplified therefore cytosine is incorporated; ATP is then created and used to emit light. Should the complementary strand contain three guanine nucleotides then three cytosine will be incorporated, producing more light than a single or double incorporation, which can be resolved by quantifying the light emission. Nucleotides are added to the reaction one by one, so if the nucleotide added to the reaction is not complementary to the DNA sequence, the chain extension is prevented until the correct complementary nucleotide is added (see Figure 3.2). Although this works well for short homo-polymers, long homo-polymer stretches (such as poly A tails on mRNA) can cause problems (Quail *et al.*, 2012). 454 sequencing occurs on a plate which contains in excess of 1 million wells which will contain a single pico-litre reaction volume from the emulsion described above. Prior to sequencing, each DNA molecule is amplified over a thousand times to allow the amount of light emitted by nucleotide incorporation to be detectable. Given that the composition of each DNA strand will affect the length of DNA sequenced means the final amount of sequencing will vary from each reaction. Despite the differences, the overall output is around 450bp for FLX platform and 700bp for the FLX+ which are considered as "long reads".

The Illumina platform results from collaboration between several microarray based companies and therefore shares many similarities with microarray technology. Like microarrays, DNA is hybridised to an array via adapters ligated to the DNA (see Figure

3.1). The DNA is then amplified locally in clusters where 1000 molecules of each strand of DNA are replicated, like 454, to become detectable at the sequencing stage. Both ends of a DNA molecule have adapters which hybridise to the array and those adapters are the basis for the bridge technology which is essential for PCR amplification but also allows each strand of DNA to be sequenced in both ways, thus paired end sequencing is possible. The sequencing is done by what is called reversible chain-termination (Bentley *et al.,* 2008). Essentially, a modified polymerase incorporates a modified nucleotide (dNTPs) which has a fluorescent moiety attached to them (see Figure 3.2). The incorporation of these modified dNTPs prevents further chain extension, which terminates chain extension but leaves a fluorescent tag on the incorporated dNTP (Wang *et al.,* 2009). This is then detected by laser excitation. Once the sequences are detected, the fluorescent moiety is removed and a subsequent round of sequencing proceeds. The Illumina platform has dominated over other NGS technologies chiefly due to its large output. Each flow cell on a HiSeq200 has eight channels, each of which can yield around 180 million paired end reads at 100 bases each end and this is set to rise to 150bp (Quail *et al.,* 2012). Moreover for RNA-seq, a single channel provides enough coverage for multiple samples. Although longer reads can be advantageous for genome sequencing, a variant library preparation called mate-pair is at least as effective at spanning repetitive regions. Mate-pair libraries are made by fragmenting the nucleic sequence into larger fragments, which are selected for a particular size (from 2 to 5Kb generally) using electrophoresis. Using mate pair libraries, the distance between paired ends can be increased to 40kbp (Lucigen, NxSeq® 40 kb Mate-Pair Cloning Kit), which is extremely useful for over-coming one of the complexities of higher eukaryote genomes (van Bakel *et al.,* 2011).

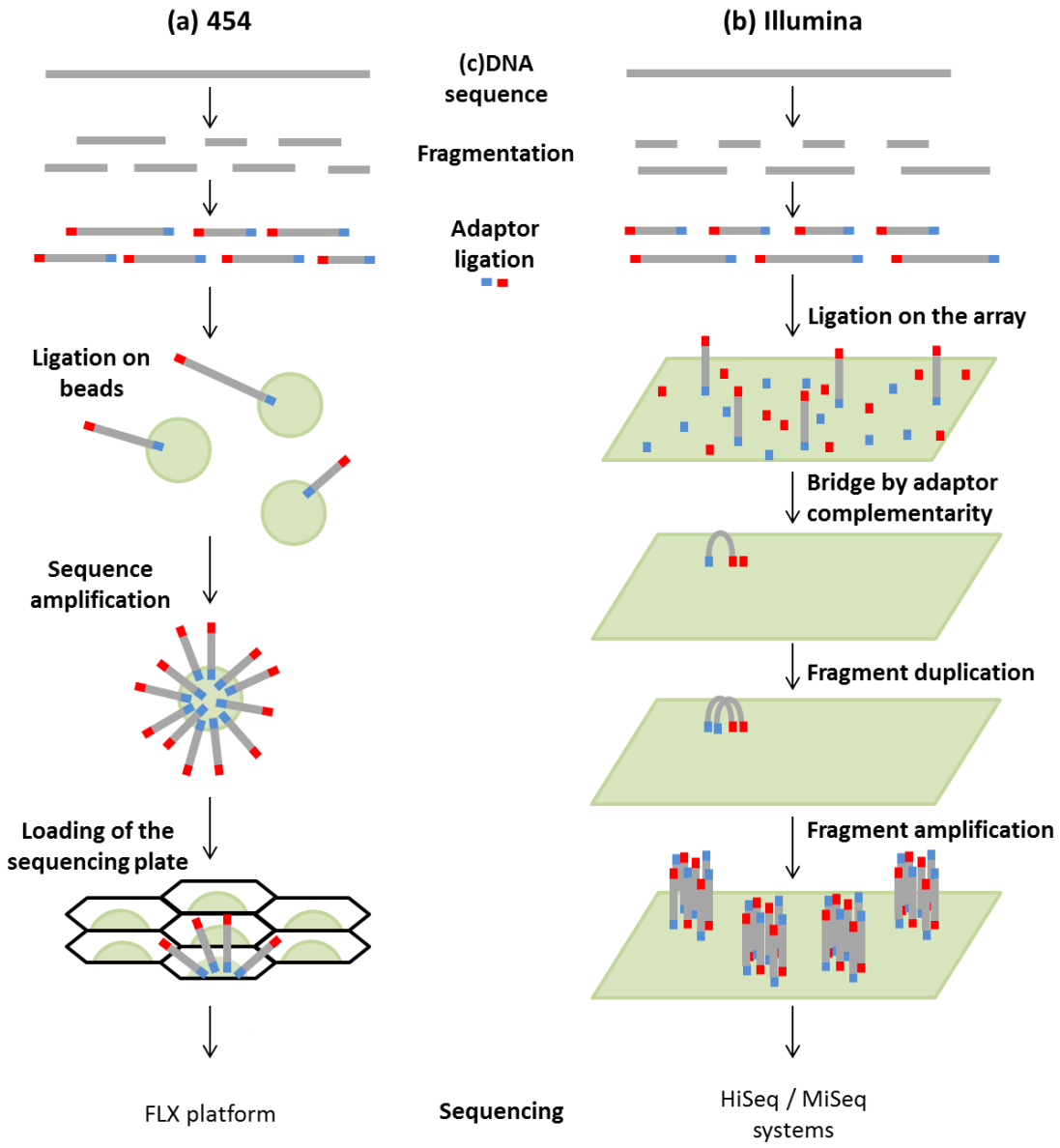


Figure 3.1: Graphic representation comparing the preparation methods for (a) 454 and (b) Illumina sequencing. Both sequencing methods require samples to be fragmented and specific adaptors ligated to each fragments. Fragments are then ligated to specific supports: (a) beads for 454 and (b) a sequencing array/chip for Illumina.

(a) 454: After fragmentation, an emulsion is created with beads, where each fragment is ligated to one bead. The fragments are then amplified on the beads before the solution is loaded onto the sequencing plate; each bead will be loaded into one well.

(b) Illumina: After fragmentation and adaptor ligation, the fragments are loaded onto the Illumina array. The array contains adaptors that will enable the fragment to form a bridge. Once replicated, each fragment will have copies of itself, oriented in both directions (paired-end).

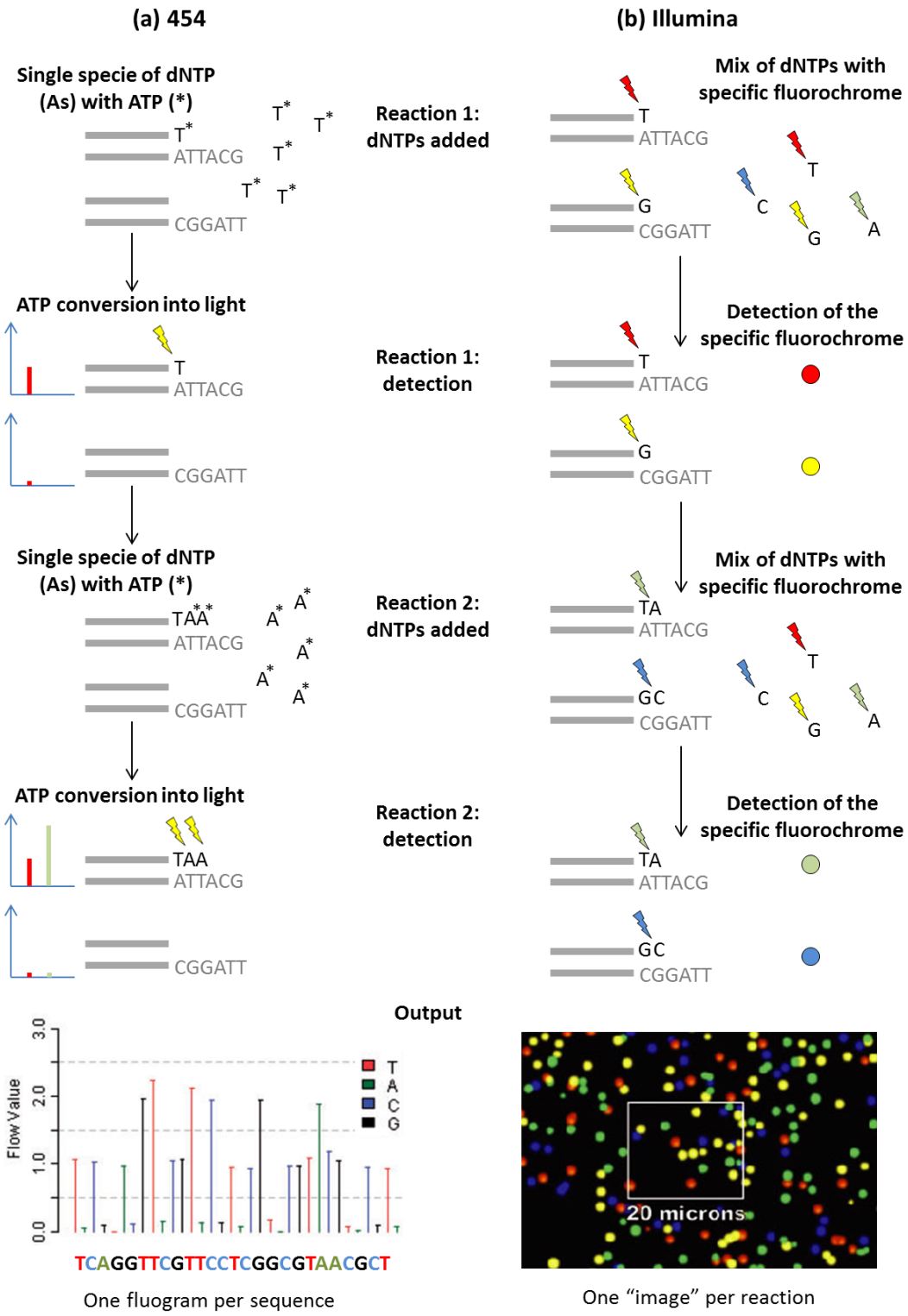


Figure 3.2: Graphic representation comparing the detection methods for (a) 454 and (b) Illumina sequencing.

(a) 454: The sequencing is done by copying the sequences, adding one species of dNTP at a time (per reaction), following the same order (T,A,C and G). If the nucleotide is complementary, ATP is produced and transformed into light. The light intensity

is proportional to the number of nucleotides incorporated. The reaction is then washed and the next dNTP is added for the next reaction.

(b) Illumina: Unlike with 454 sequencing, the dNTPs are added as a mix as they have a specific fluorochrome. One nucleotide is incorporated at a time to form the complementary sequence of the target. Once the nucleotide is incorporated, the fluorescence is detected using a camera, creating one image of the all array per reaction. The fluorochrome is then removed so another nucleotide can be added to elongate the sequence in the next reaction. The sequence is determined by the succession of colour at one single location (one read).

Given the large output of data from NGS platforms, specialist software tools have been developed to analyse the results. These tools can be broadly divided into the following categories: short read *de novo* assemblers, long read *de novo* assemblers, read aligners and downstream analysis tools. Short reads from the Illumina typically use de Bruijn graphs for assemblers and have been found not to be effective for long reads (de Bruijn, 1946; Garg *et al.,* 2011). However the assembled contigs from short reads are amenable to long read assemblers, which typically rely on overlapping reads. But long read assemblers are not suitable for short reads directly. Three software tools are widely used for long read assembly which include CAP3, Newbler and MIRA. CAP3 was created pre-NGS and has been used extensively for merging sequencing data; simply, it uses similarity and overlap to join sequences into contiguous sequences (Huang & Madan, 1999). The Newbler assembler was made specifically for 454 sequencing data and is freely available from Roche (Margulies *et al.,* 2005). Short read aligners typically use simple scoring for read mapping, based on match and mis-match (Langmead & Salzberg, 2012; Li *et al.,* 2008; Li *et al.,* 2009b). It may seem intuitive to use BLAST based methods for read aligning but the short length is not amenable for BLAST based alignment methods (Altschul *et al.,* 1990). The tools described above represent only part of the software required. Other software required include packages that process the output file from

programs such as SAMtools (Sequence Alignment/Map; Li *et al*, 2009a), MAKER 2 for sequence annotation (Holt and Yandell, 2011) and EdgeR for the identification of differentially expressed genes (Robinson *et al*., 2011) for example. Down-stream analysis tools encompass a large variety of algorithms particularly tailored for a required task and are statistically based.

3.2. Material

3.2.1. BAC libraries used for the S-locus sequencing

Two *Lolium perenne* L. BAC libraries have been used in this project, both constructed as part of the EU Framework 5 project (www.grasp-euv.dk). The libraries were made by…. from two different *Lolium perenne* L. genotypes (Farrar *et al*., 2007): LTS18, a progeny plant of the WSC mapping family (Armstead *et al*., 2002; Turner *et al*., 2006) and NV#20F1-30, one of the parents of the VrnA population (Jensen *et al*., 2005). The average insert size of both libraries is approximately 100Kb and taken together, the two libraries represent almost ten genome equivalents.

3.2.2. Plant selection and sampling for the tissue specific transcriptome analysis

Because the aim of this study was to look at gene expression in the *S*-locus region, the sampling was designed in order to have a variety of characterised genotypes at the *S*-locus as well as a tissue-specific expression.

Pollen and stigmas were sampled from different plants according to their *S* allele combinations. The pollen was collected into clear bags and transferred into microcentrifuge tubes; the pollen volume was approximately 0.1 ml per sample. The stigmas were collected from unopened florets with non-dehiscent anthers. Non-pollinated pistils at different stages of maturity were sampled and the ovary was removed using a razor blade. Approximately 50 pistils were sampled to create one stigma sample. All samples were placed in liquid nitrogen before being stored in -80°C.

For the pollen and stigma samples, one plant from the P235/59 population (P235/59/3) was selected. By using flanking markers around the *S*-locus, the genotype of this plant was confirmed to be heterozygous, S_2S_3. In order to have a homozygous plant at the *S*-locus, and preferably the same allele as the plant P235/59/3, a plant from the related population P235/58 was used. The population P235/58 was designed by Dr. Daniel Thorogood and is the reciprocal cross progeny to the P235/59 population; therefore the population is either homozygous S_2S_2 or heterozygous S_2S_3. The chosen plant P235/58/3 was genotyped for the *S*-locus by using the *S* flanking markers and was predicted to be homozygous S_2S_2. Finally, in order to have more *S*-allele diversity, another genotype was sampled, the plant F1-30, one of the parents of the S-VrnA populations. This plant has been genotyped as heterozygous, S_4S_5 (the allele numbers are allocated at random).

Finally, in order to have a complete expression pattern study, the samples of self-pollination (therefore self-incompatible) and cross-pollination (successful pollination) were taken. Semi *in-vivo* pollinations were done as described in chapter 5 section 5.3.1. The self-pollinations were done using the F1-30 plant. The cross pollinations were done using two unrelated plants: F1-30 and Foxtrot (DLF-Trifolium, Store Heddinge, Denmark). In

both cases, pistils were collected approximately 2 hours after pollination, the ovaries were removed and the stigma samples were placed in liquid nitrogen.

For a better comparison of the results, three replicates (three sampling) were done for the stigma (F1-30_1/2/3), pollen (Male_1/2/3) and self-pollination (SI_1/2/3) from the plant F1-30 as well as for the cross pollination (compatible reaction, named SC_1/2/3) between plant F1-30 and the plant Foxtrot. The Table 3.2 summarizes all the samples and the *S*-allele diversity sampled in order to study the gene expression in the *S*-locus region.

Table 3.2: Summary of the samples used for the gene expression study in the *S*-locus region. For each sample, its name, the tissue, the plant genotype and the relative *S* allele composition is given. The sample 59_21 P was extracted from a plant of the P235/59 population, therefore heterozygous for *S* but was selected to be part of the analysis for the *Z* allele as it is heterozygous for *Z* (unlike the plant 59_3). Only the pollen sample from this plant was part of the study as the library preparation from the stigma sample failed.

Sample name	**Tissue**	**Genotype**	***S* alleles composition**
F1_30_1	stigma	F1_30	S_1/S_2
F1_30_2	stigma	F1_30	S_1/S_2
F1_30_3	stigma	F1_30	S_1/S_2
Male_1	pollen	F1_30	S_1/S_2
Male_2	pollen	F1_30	S_1/S_2
Male_3	pollen	F1_30	S_1/S_2
SI_1	stigma/pollen	F1_30	S_1/S_2
SI_2	stigma/pollen	F1_30	S_1/S_2
SI_3	stigma/pollen	F1_30	S_1/S_2
SC_1	stigma/pollen	F1_30*Foxtrot	$S_1/S_2/S_3/S_4$
SC_2	stigma/pollen	F1_30*Foxtrot	$S_1/S_2/S_3/S_4$
SC_3	stigma/pollen	F1_30*Foxtrot	$S_1/S_2/S_3/S_4$
58_3 S	stigma	P235/58/3	S_5/S_5
58_3 P	pollen	P235/58/3	S_5/S_5
59_3 S	stigma	P235/59/3	S_5/S_6
59_3 P	pollen	P235/59/3	S_5/S_6
59_21 P	pollen	P235/59/21	S_5/S_6

3.3. Methods

3.3.1. Methods for the sequencing of the *S*-locus BAC clones

The study was done in five main steps: isolation of the BAC clones covering the *S*-locus region, sequencing the selected BAC clones using 454 sequencing, filtering the nucleotides obtained from sequencing, assembling the sequences into scaffolds and annotating the scaffolds in order to identify genes in the *S*-locus region of *Lolium perenne*.

3.3.1.1. Identification of the BAC clones covering the *S*-locus

Both BAC libraries were screened using markers around the *S*-locus region. The two flanking markers 05_02790 and 05_02889 as well as the markers showing no recombination (05_02827, 05_02833, 05_02863 and 05_02871) were used to fish the BAC clones.

Once the BAC clones had been isolated, DNA was extracted from the BAC clones in order to order the BAC clones within the *S*-locus. In order to do so, the ends of the BAC clones were sequenced and two pairs of primers were designed for each of them, for each end of the BAC clones. Primers were screened on the BAC clones' DNA in order to align them and to select the BAC clones for next generation sequencing.

3.3.1.2. BAC library screening

The screening of the superpool and the clones plates was done using a PCR amplification volume of 10µl containing 0.25 unit DreamTaq DNA Polymerase (Fermentas, St. Leon-Rot, Germany), 1X DreamTaq Buffer, 200µM dNTPs (Roche Applied Biosciences, Penzberg, Germany), 0.2µM of each forward and reverse primer and

1µl of the BAC DNA or bacteria clones. The amplification was done at 95°C for 4 min, followed by 30 cycles of 30 s at 94°C, 30 s at a primer-specific temperature (see Table 2.3 from chapter 2 section 2.3.2) and 1 min at 72°C, and finally 10 min at 72°C using a thermocycler (GeneAmp PCR System 9700, Applied Biosystems, Foster City, CA, USA). The PCR products were then analysed on 1% agarose gels (1% agarose, 0.5X TAE) using 10% GelRed (Biotium, Hayward, CA, USA).

The identification of the BAC clones covering the *S*-locus region was done by screening at first the BAC libraries superpools. For each BAC libraries, the superpools have been organised in a tri-dimensional way, enabling to identify the corresponding BAC clone with only two PCR amplifications. Indeed, the PCR screen of the superpool generates three positive amplifications per positive BAC colony, identifying the BAC library plate to screen after in order to identify the positive BAC clone.

3.3.1.2.1. Extraction of the BAC clone DNA

Prior to the DNA extraction, the BAC clones were grown individually overnight at 37°C in petri dishes on chloramphenicol (25ng/ml) LB-agar media (0.1 w/v triptone, 0.05w/v yeast extract, 0.05 w/v NaCl, 1.5% agar). Three independent colonies from each BAC clone were then grown overnight at 37°C in 2ml individual tubes containing chloramphenicol (25ng/ml) LB (0.1 w/v tryptone, 0.05w/v yeast extract, 0.05 w/v NaCl). The BAC clones were then prepared by extracting the DNA using the BAC clone miniprep (Qiagen, Hilden, Germany) protocol (see Table 3.3 for the composition of each buffer). The tubes were centrifuged at 4,000 rpm for 5 minutes and the supernatant was discarded. The pellet was resuspended in 300µl ice-cold P1 buffer and 300µl P2 buffer. Tubes were inverted to mix and incubated at room temperature for 5 minutes before 300µl of ice-cold

N3 buffer were added. Tubes were inverted and incubated on ice for 10 minutes before being centrifuged at 13,000 rpm for 10 minutes. The supernatant was transferred into fresh 1.5ml tubes and centrifuged again at 13,000 rpm for 10 minutes. The supernatant was transferred into a fresh 1.5ml tube and 650µl of isopropanol were added to it. Tubes were mixed by inversion, incubated at room temperature for 2 minutes and then centrifuged at 14,000 rpm for 20 minutes. The isopropanol was removed and the pellet was washed with 500µl of 70% ethanol. Tubes were incubated at room temperature for 2 minutes, centrifuged at 14,000 rpm for 2 minutes before the ethanol was removed. The pellets were leaved to dry at room temperature (30 minutes approximately) before been resuspended in 40µl of 0.5X TE buffer.

Table 3.3: Composition of buffers used in the BAC clone miniprep protocol.

Buffer	Composition
P1 Resuspension buffer	50mM Tris-HCl pH 8.0 10mM EDTA 100µg/µl RNase
P2 Lysis buffer	200mM NaOH 1% SDS
N3 Neutralisation buffer	3M potassium acetate pH 5.5

3.3.1.2.2. BAC-end sequencing and primer design

For the BAC-end sequencing, 12µl of the BAC clone DNA were mixed with 2µl of the primer, forward or reverse as required. Each BAC clone DNA was sent for BAC-end sequencing using the forward and reverse primer separately. The primers used for the

BAC-end sequencing were as followed: forward, GTTTTCCCAGTCACGAC and reverse, CGGATAACAATTTCACACAGGA. The sequencing was done using the ABI3730 DNA Analyser (Applied Biosystems, Foster City, CA, USA).

For each sequence obtained by sequencing the ends of the BAC clone inserts, a pair of primers was designed using the software primer3 (http://frodo.wi.mit.edu/primer3/; Rozen and Skaletsky, 2000). Because three individual BAC colonies were sequenced for each BAC clone, the primers were designed on the consensus sequence of the three sequences. Primers were designed for each BAC clone ends (see Table 3.4) but initially, only primers from the LTS18 BAC library sequences were used to align the BAC clones from the BAC library LTS18.

Table 3.4: BAC-end sequence primers designed from *Lolium perenne* BAC clone end-sequences from the BAC library GRASP LTS18 and GRASP NV#20F1-30.

Oligoname	BAC clone	Forward	Reverse	Product size	Annealing temperature (Ta)
SBAC-174-F	P174C12C	AGGCAACTGGAGGTTGAACT	CAGGGGCACAAGTACACAAG	233	59
SBAC-175-f	P175C6E	CGGATAATCTGGCCTCAAGT	GACCTGAGGAATTGCTGCTC	212	58
SBAC-175-r		TGCAGCATACAACACCAACA	GAGCTTGGGATTGCAAAGAG	245	60
SBAC-169-F	P169C11B	TAAGCAAAAGAAGGGGCAAG	CACCGTACTAACGGCAGAGC	180	59
SBAC-169-R		TCCCCTTCTTCATCGACAAC	AGCAAACCGACTCTCCAAGA	100	57
SBAC-223-F	P223	GACGGTGCAACTCCGATAAT	GCAGACTAGGCTGGTTGTCC	199	59
SBAC-223-R		TCAGTGGCCTCCTCTCTGTT	GCTCAAGTCCATCACCCCTA	229	59
SBAC-29-F	P029C5F	TGCACTTCGGCAAGCTATTA	AAAACGCATATCACCCCTTG	122	57
SBAC-29-R		GAAACTGCCAAGCACTCCTC	TCCTCGCCTCAGTCTCCTTA	121	59
SBAC-227-F	P227	GAGCAATGCCCAGAAGTGTT	TCCTCCACACCAAACTCAAA	106	57
SBAC-227-R		GATTTCGTCCAATTCCGAGA	ACTAAGCTTGGGGATGCTGA	224	56
SBAC-180-F	P180C7C	AATCCGTTCGAGGCATACAC	CCCCTTTGGTCCCTAGAAAA	192	57
SBAC-180-R		TATGGGAGGGATCCACAAAA	CTTCAGCTTTGGGAATGGAG	250	56
SBAC-202-F	P202C7E	CGAACTTTTCCTGCTTTTGC	ACTGCAGATTTGCGAGAACC	212	59
SBAC-202-R		ATGGTCTGCAACACGTACCA	GGTTACAAGCCGAACCAAGA	248	57
SBAC-714-F	714-6C	TCTGTATCCGGTACCCAAGC	ACGGCTTTCGAAAACTGAGA	159	58
SBAC-714-R		AGTGACCGAGCCTAGGGTTT	GGACGCCAAAATGAAGAAAA	210	57
SBAC-573-F	573-8A	GCCTCCACCTCCTAGCTTCT	AAGGATGAAGGCATGACACC	215	61
SBAC-573-R		GAGGTGATGAACGGGTTGAC	CTGGCTTGGTTCCCTTTACA	170	59

Oligoname	BAC clone	Forward	Reverse	Product size	Annealing temperature (Ta)
SBAC-764-F	764-6D	GCTCTGGGAACCAAATGAAA	AAGGAAATGGGATTGGTTCA	199	61
SBAC-764-R		CCTTCTGCTCCACCACTAGC	GCCGATGCTCCAACTGTATT	188	61
SBAC-716-F	716-4B	GAATGCAGCGATCAAAACAA	GAAAGCAACCGCTGAAACTT	189	58
SBAC-716-R		GCCCCTATAGGATTGGAAGC	CCGCCTATGATCCTCAATGT	159	57
SBAC-520-10h-F	520-10H	TGGATCCGATCAAGAAGGTC	AGCTTGGGGGAGATTGAAGT	186	61
SBAC-520-10h-R		AACGTCCAAAGGTGTTCCAG	AAACAAACGGCAGAACGAAG	241	58
SBAC-681-F	681-10B	GCGAATTCCCACGTTCAC	AAGATCGGTAAGTGCGGATG	120	57
SBAC-681-R		CGGTAAACCACCAACAATCC	TCAAGGGGCAAGTTATGGAC	100	57
SBAC-520-7a-F	520-7A	CTCGGTTTGACCACCAGTTT	TCTGATCGAGGAAGCCTGTT	114	57
SBAC-520-7a-R		AGTTTTGCAGGCTTTGCATT	CGTGCTAGGTGGGAAAAAGA	173	59

3.3.1.3. Screening of the BAC clones with BAC-end primers

In order to align the BAC clones, PCRs were conducted using the BAC clone DNA extracted previously from 520-7A, 520-10H, 573-8A, 681-10B, 714-6C, 716-4B and 764-6D. Each BAC clone DNA was amplified by PCR using every pair of primers designed from the BAC end sequences. The amplification was done with a first step at 95°C for 4 min, followed by 40 cycles of 30 s at 94°C, 30 s at a primer-specific temperature (see Table 3.4for Ta) and 2 min at 72°C, and finally 7 min at 72°C. The PCRs were carried out using a thermocycler GeneAmp PCR System 9700 (Applied Biosystems, Foster City, CA, USA) and in a 10µl volume reaction made as described previously (chapter 3 section 3.3.1.2). The PCR products were then analysed on 1% agarose gels (1% agarose, 0.5X TAE) using 10% GelRed (Biotium, Hayward, CA, USA). Finally, the BAC clone order was constructed by analysing the agarose gel and a contig of the *S*-locus region was built.

3.3.1.4. Preparation for sequencing: library preparation method

The library preparation was done using the GS FLX Titanium Rapid Library Preparation Method (Roche Applied Biosciences, Penzberg, Germany), that contains most of the buffers and kit used in this protocol, and was divided into four main steps: DNA nebulization, fragment-end repair, small fragment removal and the quality and quantity assessment of the library. The library as well as the sequencing were done in Aberystwyth University (UK) by Dr Susan Girdwood (Method detailed in the supplementary file: Method for the 454 library preparation). The shotgun sequencing was done using the 454 sequencer GS FLX (Roche Applied Biosciences, Penzberg, Germany).

3.3.1.5. Assembly of the BAC clones sequences and filtering

The assembly of all the sequences covering the *S*-locus region was done in two steps. First the sequences were assembled for each BAC clone separately, leading to a number of contigs per BAC clone. Then, the contigs were all assembled together into scaffolds, grouping both BAC libraries.

The assembly of the reads into a contig was done using the software GS De Novo Assembler version 2.3 (also called Newbler), provided by Roche; the assembly was done by Dr Susan Girdwood. The software found overlaps between reads and assembled them if the overlap were over 40bp with a minimum of 90% of identity for the overlap between the two reads. The software then created a consensus sequence from all the alignments that were possible, that made up the contig.

Using CLC Genomics Workbench software (CLC bio, Aarhus, Denmark) and the tool Trim Sequences, the sequences were trimmed and filtered from any vector sequences (*Hind*III-digested pIndigoBAC-5 vector (Epicentre Biotechnologies, Madison, WI, USA), sequencing adapters left, and bacterial genome sequences.

The assembly of the BAC clone sequences into scaffolds was done by Dr. Stephen Byrne from Aarhus University (Denmark), using four different paired-end and mate pair *Lolium* libraries (540 bp paired-end, ~ 1.8 Kb mate pair, ~ 3.5 Kb mate pair and ~ 8.4 Kb mate pair library; Aarhus University, Denmark). The software Bowtie 2.0.0-beta 2 was used for the alignment (Langmead *et al*., 2009), as it has been developed to align short DNA reads to a scaffold genome. Along with Bowtie, the software SSPACE Premium V1.0 (Boetzer *et al*., 2011) was used to assemble the aligned contigs into scaffolds.

3.3.1.6. Gene annotation

The annotation of the scaffolds of the *S*-locus was done by Dr. Stephen Byrne (Aarhus University, Denmark) as it requires Linux operating system programing experience. The MAKER 2 pipeline was used for the scaffold annotations (Holt and Yandell, 2011). MAKER annotates sequences with structural and functional annotations, by aligning the sequences to an EST database, creating a training file for the *ab initio* gene predictor (SNAP).

In order to visualize the results from the gene prediction and annotation (a gff3 file), the software Apollo (Lewis *et al*., 2002) was used. BLAST searches were carried out using the Apollo software function, searching the NCBI database in order to identify genes.

3.3.2. Methods for the tissue specific transcriptome sequencing and analysis using Illumina sequencing.

3.3.2.1. RNA extraction protocols

The RNA extraction protocols used in this study were different because the samples were from different tissues and the plants were from different research institutes. The pollen RNA extracted from the plant F1-30, as well as the RNA extracted from the SC and SI samples, was done using the RNeasy Plant Mini Kit (Qiagen) and the stigma RNA from F1-30 using the PicoPure RNA Isolation Kit (Arcturus Bioscience, Inc.). Both types of extractions were conducted by Dr. Bruno Studer (ETH Zürich, Switzerland) in The Danish Institute of Agricultural Sciences (Aarhus University, Denmark). The samples from the population P235/59 and P235/58 were extracted by Dr. Matthew Hegarty from IBERS

(Aberystwyth University, UK) using a Tri Reagent® (Sigma-Aldrich, St. Louis, MO, USA) extraction method.

3.3.2.1.1. Pollen extraction using RNeasy Plant Mini Kit (Qiagen)

The pollen total RNA extraction was done using the RNeasy Plant Mini Kit (Qiagen) and following the manufacturer's instructions. Because the amount and the nature of the sample, the pollen grinding was done directly in the collection 1.5ml tube in liquid nitrogen, using a pellet pestle. Without letting the sample thaw, 450 µl of the buffer RLT was added and the tube was vortexed; this buffer contains β-mercaptoethanol (1%) for cell-wall lysis. The lysate was transferred to a QIAshredder spin column and centrifuged for 2 minutes at 13,000 rpm. The supernatant of the flow-through was transferred to a new micro-centrifuge tube, 250 µl of 96% ethanol was added and mixed by pipetting. The sample was then transferred to an RNeasy spin column and centrifuged for 15 seconds at 13,000 rpm. At this stage, the total RNA is attached to the column and the flow-through is discarded. The RNeasy column is then rinsed with 700 µl and centrifuged for 15 seconds at 13,000 rpm. The flow-through is discarded again and the column is washed with 500 µl of buffer RPE and centrifuged for 15 seconds at 13,000 rpm. This last step is repeated one more time before the 1.5 ml collection tube is replaced with a clean one. A volume of 50 µl of RNase-free water was added directly onto the column membrane to dissolve the total RNA and the tubes were centrifuge one last time for 1 minute at 13,000 rpm.

The total RNA was then treated to eliminate any contaminating DNA. The DNase treatment was carried out using the RNase-Free DNase Set (Qiagen), following the

manufacturer's protocol. To each sample (with a volume of 50 µl), 10 µl of buffer RDD and 2.5 µl of DNase I (stock solution) were added and the volume was made up to 100 µl using RNase-free water. Samples were incubated for 10 minutes at room temperature before going back to the RNeasy Plant Mini Kit clean-up protocol. 350 µl of Buffer RLT was added to each sample and mixed, followed by 250 µl of 96% ethanol. Samples were mixed by pipetting and 700 µl of each was transferred to an RNeasy Mini spin column and centrifuged for 15 seconds at 13,000 rpm. The flow-through was discarded and the column was rinsed using 500 µl of buffer RPE and centrifuged for 15 seconds at 13.000 rpm. This step was repeated after the flow-through was discarded, but the centrifugation was done for 2 minutes at 13.000 rpm. The column was placed into a new collection tube and 30 µl of RNase-free water was added onto the membrane. Samples were centrifuged for 1 minute at 13.000 rpm in order to collect the dissolved DNA-free RNA.

The total RNA concentrations were quantified using the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA).

3.3.2.1.2. Stigma extraction using PicoPure RNA Isolation Kit (Arcturus Bioscience, Inc)

The PicoPure RNA Isolation kit (Arcturus Bioscience, Inc., CA, USA) was used to extracted the RNA from the stigma samples containing only 20 stigmas. This extraction method was chosen for these small samples as it is normally used to extract RNA from a sample made out of 10 cells or less. A step of grinding was added to the protocol: the samples were ground in the tube in liquid nitrogen using a pellet pestle. Then, following the manufacturer's instructions, 50 µl of the extraction buffer (XB) was added, mixed by

inversion and incubated at 42°C for 30 minutes. The tube was then centrifuged for 2 minutes at 2,500 rpm. The RNA Purification column was pre-conditioned by adding 250 µl of the conditioning buffer (CB) onto its membrane, incubated at room temperature for 5 min and centrifuged for 1 minute at 13,000 rpm. 50 µl of 70% ethanol was added to the cell extract and mixed by pipetting before the whole extract was transferred in the pre-conditioned purification column. The column was then centrifuged for 2 minutes at 1,000 rpm immediately followed by a centrifugation for 30 seconds at 13,000 rpm. 100 µl of the wash buffer (W1) was added to the column and centrifuge for 1 minute at 10,000 rpm. At this stage, a DNase treatment was applied using the RNase-free DNase set (Qiagen) by adding 40 µl of the DNase incubation mix (10 µl DNase I stock solution, 30 µl buffer RDD) onto the column membrane which was incubated for 15 minutes at room temperature. The column was then centrifuged for 15 seconds at 10,000rpm. The flow-through was discarded and 100 µl of the wash buffer 2 (W2) was added before centrifugation for 1 minute at 10,000rpm. Another 100 µl of W2 was added and the column was centrifuged for 2 minutes at 13,000 rpm. The column was transferred to a new 0.5 ml micro-centrifuge tube before 30 µl of elution buffer (EB) was added. The column was incubated for one minute at room temperature and then centrifuged for one minute at 3,000 rpm followed by one minute at 13,000 rpm.

The total RNA was quantified using the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA).

3.3.2.1.3. Tri Reagent® RNA extraction protocol

The first step of the protocol was to grind the samples with 1ml of Tri Reagent® (Sigma-Aldrich), using the Retsch-mill (Retsch, Haan, Germany) with a frequency of 30

Hz for 90 seconds. The samples were left to incubate at room temperature for 10 minutes before 160 µl of ice cold chloroform was added to each samples. The tubes were mixed by inversion and incubated at room temperature for a further three minutes. After a centrifugation for 15 minutes at 4°C and 13,000 rpm, the supernatant was transferred to a new tube and the step was repeated by adding 160 µl of chloroform. After recovering the supernatant again, 0.5 µl of glycogen and 500 µl of isopropanol were added. The samples were mixed by inversion and incubated for 10 minutes at room temperature before been centrifuged for 10 minutes at 4°C and 13,000 rpm. The supernatant was discarded without disturbing the pellet and 1 ml of 75% ethanol was added. Tubes were vortexed for a few seconds before been centrifuged for 5 minutes at 4°C and 13,000 rpm. The ethanol was pipetted off and the tubes were left open for 5 minutes in order for all the ethanol to evaporate. Finally, the pellet was dissolved in 20 µl of nuclease free water.

3.3.2.2. Preparation for Next Generation Sequencing RNAseq with HiSeq 2000 Illumina sequencing

Two libraries were prepared with the RNA samples: one by Dr. Matthew Hegarty using the RNA extracted using the Tri Reagent® protocol (59_3_S, 59_3_P, 59_21_P, 58_3_S and 58_3_P), the other by the company doing the sequencing, Aros Applied Biotechnology A/S (Aarhus, Denmark) but both libraries were prepared using the same method: TruSeq® RNA Sample Preparation Kit (Illumina, Inc., San Diego, CA, USA). The library preparation was divided into seven steps: purification of the RNA, first and second strand cDNA synthesis, repaired of the end fragment, adenylate 3' end, ligation of

the adapters and PCR amplification of the library (Method detailed in the supplementary file: Method for the Illumina library preparation).

In total, 17 cDNA libraries were sequenced using Illumina sequencing (subcontracted to AROS Applied Biotechnology, Aarhus, Denmark) and 76 bp paired-end (PE) reads were generated for each of them. All libraries were pooled into two groups and each group was distributed over two lanes, with an expectancy minimum of 51M reads/sample.

3.3.2.3. *De novo* transcriptome assembly method

The transcriptomes assembly was conducted by Dr Stephen Byrne (Aarhus University, Denmark). The program FastQC (Babraham Institute, Cambridge, UK) was used to visualize the read quality and length of the Illumina HiSeq2000 raw reads. The sequences were trimmed using the software Homer-Tools (Heinz *et al*., 2010) to remove Illumina adaptors at the 3´ end of the reads, and quality trimmed using the software Sickle (stringent cut-off of Q20 and at least 40bp length). *De novo* transcriptome assembly was performed using the software Trinity (minimum contig length 200bp) (Grabherr *et al*., 2011).

3.3.2.4. Transcripts alignment and differential expression analysis

The alignment of the transcripts to the *S*-locus genomic sequence (scaffold of the BAC clones) was done by Dr Stephen Byrne, using BOWTIE (Langmead *et al*., 2009).

The results of the alignment were visualized using the software IGV 2.1 (The Broad Institute; Robinson *et al*., 2011).

In order to compare the level of expression between stigma and pollen tissues and between self-pollination and cross-pollination, a count of the raw reads for each annotation was done by Dr Steven Yates (Essex University), using R (R Development Core Team, 2008) and Rsamtools (part of Bioconducteur, Gentleman *et al*., 2004) to extract the read counts. Raw reads counts were then computed by EdgeR (Robinson *et al*., 2011) to calculate the biological coefficient of variation (BCV) in order to assess the variation between replicates. EdgeR was also used to calculate differential expression between samples for each gene. For each gene annotation, the logFC (log_2 Fold Change; difference between tissues expression), the logCPM (log_2 Count Per Million; normalised average between all samples) and P Value (exact test for the negative binomial distribution, similar to Fisher's exact test) were obtained. A normalization of the gene expression was also calculated: RPKM (Reads per Kb of exon model; Mortazavi *et al*., 2008).

3.3.3. Methods for assessing allelic diversity of the gene Osg0150500-like in *Lolium perenne*

The *Lolium* region annotated as Os05g0150500-like can be found in scaffold 1 between 3,011 and 7,497 bp, scaffold 11 between 202 and 4,854 bp and in scaffold 15, between 5,004 and 10,798 bp.

All the sequences alignments were done using the multiple sequence alignment tool MAFFT (http://mafft.cbrc.jp/alignment/server/), which is used in any type of alignment (nucleotides or amino-acids). After selection of the scaffold region, the gene structure from

the scaffolds was predicted using two software programs: Augustus (http://bioinf.uni-greifswald.de/augustus/, Stanke *et al*., 2004) and FGENESH (http://linux1.softberry.com/berry.phtml?topic=fgenesh&group=programs&subgroup=gfind, Softberry). Two software programs were used in order to compare the prediction to the coding sequence (CDS) of *Brachypodium* and rice, by aligning them using MAFFT tool. From this alignment, a predicted CDS for *Lolium* was identified.

Using the tool Map Reads to a Reference Sequence from CLC Genomics Workbench software, the transcripts from stigma tissue covering the *Lolium* gene sequence were pooled out for each transcriptome. Using Spidey software which allows alignment of cDNA (transcripts) to genomic DNA (http://www.ncbi.nlm.nih.gov/spidey/), the pooled out transcripts were mapped to the scaffold sequence to confirm their alignment. Then, using MAFFT, the transcripts were aligned to the *Lolium* CDS as well as the *Brachypodium* sequence in order to look for polymorphism.

Finally, for each transcript, the corresponding protein sequence was predicted using the translate tool from ExPASy (Swiss Institute of Bioinformatics, http://web.expasy.org/translate/). As none of the transcript sequences were covering the entire gene, the protein sequence from the three possible reading frames was assessed by aligning them to the *Lolium* protein from the predicted *Lolium* CDS. Once all the protein sequences were obtained, an alignment between the rice Os05g0150500 protein, the *Brachypodium* Bradi2g35720.1 protein, the predicted *Lolium* gene and the different *Lolium* genotypes was done using MAFFT in order to assess the variability of the protein between species and between genotypes.

3.4. Results

3.4.1. *S*-locus genomic sequencing

3.4.1.1. Identification of the BAC clones covering the *S*-locus and their alignment

The screening of the two BAC libraries, LTS18 and NV#20F1-30, with the flanking markers as well as the non-recombinant markers had led to the isolation of 10 and 11 clones, respectively (see Tables 3.5 and 3.6).

Table 3.5: Results of the GRASP LTS18 BAC library screening. The clone identifications are annotated between brackets.

Markers name	Number of clones
05_02790	2 (520-7A, 520-10H)
05_02827	2 (714-6C, 573-8A)
05_02833	2 (714-6C, 764-6D)
05_02863	2 (714-6C, 573-8A)
05_02871	2 (714-6C, 764-6D)
05_02889	5 (681-10B, 716-4B, 946, 954, 959)

Table 3.6: Results of the GRASP NV#20F1-30 BAC library screening. The clone identifications are annotated between brackets.

Markers name	**Number of clones**
05_02802	5 (P029C5F, P041C5B, P180C7C, P223, P227)
05_02827	3 (P169C11B, P174C12C, P202C7E)
05_02833	3 (P175C6E, P152C2B, P061C12E)
05_02863	3 (P169C11B, P174C12C, P202C7E)
05_02871	3 (P175C6E, P152C2B, P061C12E)
05_02889	3 (P175C6E, P152C2B, P061C12E)

For the screening of the NV#20F1-30 BAC library, the marker 05_02802 was used because the marker 05_02790 did not amplify in this BAC library superpool. This marker is mapped in rice between the markers 05_02790 and 05_02827 so by screening it on the BAC library instead of the top flanking marker 05_02790, a part of the *S*-locus region is missing from the BAC clones isolated from this BAC library. Moreover in this BAC library, no contig can be built between the markers 05_02827 and 05_02863 and the markers 05_02833 and 05_02871 as there are no identical clones between the two groups.

However, with the BAC library LTS18, it was possible to construct a contig between those four markers as they all have one BAC clone in common, 714-6C. In the case of this BAC library, there are two gaps: one between the non-recombinant markers (05_02827, 05_02833, 05_02863 and 05_02871) and the top flanking marker 05_02790, the other one between the non-recombinant markers and the bottom flanking marker,

05_02889. The gaps in the alignment of the BAC clones might need to be filled by designing more markers based on the BAC clones end sequences.

From the identified BAC clones from the LTS18 BAC library, 7 BAC clones were used for the BAC-end sequencing: 520-7A, 520-10H, 573-8A (named 573), 714-6C (named 714), 764-6D (named 764), 716-4B (named 716) and 681-10B (named 681). For the NV#20F1-30 BAC library, all the BAC clones were isolated in order to sequence the ends of the insert.

After screening the BAC clone DNA with the BAC-end sequence primers, the BAC clones from the LTS18 BAC library were ordered within the *S*-contig (see Figure 3.3 and 3.4). However, the BAC-end marker SBAC-174-F was not used for the alignment of the BAC clones as it was amplifying most of the BAC clone DNA, suggesting that the marker was designed in a highly repetitive region. The same happened with most of the markers designed from the BAC-end sequences of the NV#20F1-30 BAC library. Only the markers SBAC-175-r and SBAC-202-F were useful in the confirmation of the alignment of the BAC clones, as the rest of the markers were showing multiple hits, suggesting that they were not specific.

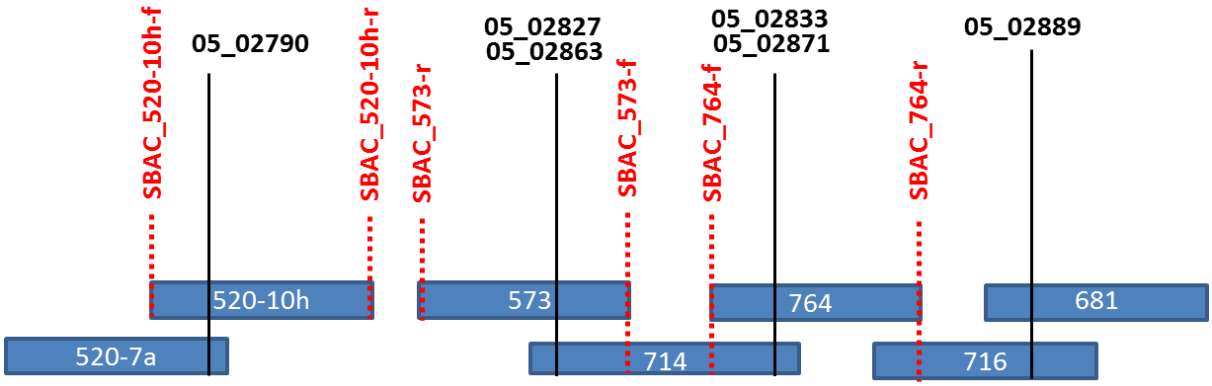


Figure 3.3: Schematic of the alignment of the BAC clones from LTS18 BAC library by using BAC-end sequence primers. The BAC clones are represented with the blue bars and the BAC-end sequence primers in red. Also annotated on the figure are the HRM primers used to fish the BAC clones.

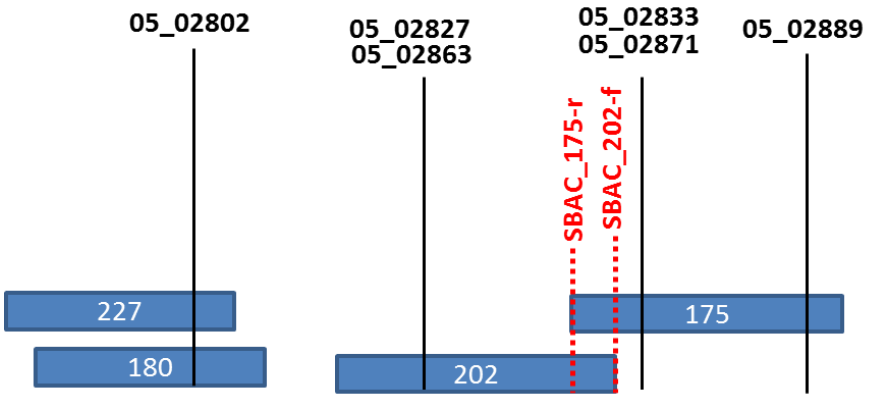


Figure 3.4: Schematic of the alignment of the BAC clones from NV#20F1-30 BAC library by using BAC-end sequence primers. The BAC clones are represented with the blue bars and the BAC-end sequence primers in red. Also annotated on the figure are the HRM primers used to fish the BAC clones.

For the sequencing of the BAC clones from the LTS18 BAC library, only the BAC clones 520_10H, 520_7A, 573, 714 and 764 were selected. Indeed, the BAC-end primers have been used to screen the DNA from the recombinants isolated from the fine-mapping described in Chapter 2. The marker SBAC-764-f was found to have the same recombinants

as the flanking marker 05_02889, therefore becoming the new flanking marker (at the bottom of the *S*-locus) and meaning that it was not necessary to sequence the BAC clones 716 and 681 from this library.

For the library NV#20F1-30, the BAC clones 175, 180, 202 and 227 were selected for sequencing as they cover the entire *S*-locus region except the top flanking marker 05_02790.

3.4.1.2. Sequencing results

The shotgun sequencing of the nine BAC clones was done using pyrosequencing (454 sequencing). The nine samples were pooled together into two sections of a run (one run of sequencing is made out of eight lanes). The estimated number of reads out of one run using the 454 sequencer GS FLX Titanium XLR70 is 1,000,000 so the estimation of our sequencing should be around 250,000 reads (1/4 of a run). Knowing that the average length of the reads is around 450bp and that the average size of the BAC clone is 100kb, the sequencing is estimated to cover each BAC clone 120 times.

The reads from the sequencing were automatically passed through a quality filter: GS De Novo Assembler (454 Life Sciences, Roche Diagnostics, Basel, Switzerland). The filter removed approximately 30% of the reads, from 383,987 reads to 101,479 reads, but still retaining 38,781,541 bp which correspond to 43 time coverage for each BAC clone.

The reads were also sorted according to their MID sequences (and therefore according to the BAC clone they belong to) using the same software, GS De Novo

Assembler. After sorting the reads, the MID sequence vectors were trimmed to avoid misalignment and misassembly.

3.4.1.3. Results of the assembly of the sequences into contigs

At first, the sequences obtained from the sequencing of the BAC clones were assembled into contigs for each BAC clone separately. The assembly of the reads lead to a large number of contigs for each BAC clones, with up to 213 contigs for the BAC clone 520_7A. The results of the assembly are presented in Table 3.7.

However, the reads were not cleaned up of any contaminating DNA. This step was conducted after the assembly into contigs, using CLC Genomics Workbench software, by removing vector and bacterial sequences out of the contigs. The results of this trimming, presented in Table 3.8, show that the number of contigs per BAC clone was considerably reduced, with only one contig left for the BAC clone 520_7A (instead of 213), but the number of bases covering the BAC clones was not reduced much.

Table 3.7: Table presenting the statistic of the assembly of the BAC clones DNA using Newbler. For each BAC clone, the table presents the number of reads before and after assembly, as well as the contig resulting from the assembly with a size over 500 bp and the non-assembled reads.

BAC clone	Total number of reads	Number of assembled reads	Contig > 500 bp				Non-assembled reads			
			Number of contig	Average contig size (bp)	N50 contig size (bp)	Largest contig size (bp)	Number of partially assembled reads	Number of singletons	Number of repeats	Number of short reads
175	12,975	6,224 (48%)	169	1,114	983	30,813	606	5,186	24	585
180	7,731	4,059 (52.5%)	29	2,679	11,837	18,047	165	2,944	1	496
202	6,041	2,591 (43%)	44	2,343	5,443	16,261	140	2,741	2	517
227	2,815	1,134 (40.3%)	46	1,374	2,073	5,457	69	1,415	6	40
573	20,971	15,781 (75.3%)	91	1,756	5,671	32,949	569	3,638	146	684
714	17,628	12,501 (71%)	84	1,831	9,386	45,801	489	3,804	27	641
764	4,234	2,080 (49%)	46	2,071	3,130	9,677	108	1,726	17	264
520_7A	10,562	3,635 (34.4%)	213	719	640	16,121	672	5,183	0	651
520_10H	673	247 (36.7%)	31	781	760	3,237	60	274	4	39

Table 3.8: Statistics from the BAC clones reads trimming. For each BAC clones, the number of contig before and after trimming (removal of any vector, bacterial or adapter sequences) and the total size (in bp) represented by the contig.

BAC clone	Before trimming		After trimming	
	Number of contig (>500 bp)	Total size (bp)	Number of contig (>500 bp)	Total size (bp)
175	169	188293	10	86958
180	29	77692	8	63483
202	44	103127	20	88856
227	46	63247	40	57645
573	91	159831	23	116001
714	84	153833	13	110191
764	46	95297	38	89745
520_7A	213	153246	1	16121
520_10H	31	24220	29	23022

3.4.1.4. Results of the assembly of the contigs into scaffolds

From this assembly, the contigs from all the BAC clones (182 contigs with an average size of 3488 bp) have been assembled into 70 scaffolds, with an average size of 10232 bp. The assembly was done progressively by both software (Bowtie 2.0.0-beta 2 and SSPACE Premium V1.0), starting with the first scaffolding using the 540 bp paired-end library, followed by the second scaffolding with the 1.8 Kb mate pair library, a third scaffolding with the 3.5 Kb mate pair library and the final scaffolding with the 8.4 Kb mate pair library. The contribution by the different libraries to the scaffolding is shown in Table 3.9.

The assembly has resulted in several contiguous sequences over the *S*-locus. However, within each scaffold, some sequences are represented by unknown nucleotides

(N). Those Ns most likely correspond to repetitive, low complexity regions that could not be assembled with the BAC reads and the paired-end and mate pair libraries.

Table 3.9: Summary of the assembly of the *S*-locus contigs into scaffolds using four different paired-end and mate pair *Lolium* libraries. The first scaffolding was made with the 540 bp paired-end library, the 2nd scaffolding using the 1.8 Kb mate pair library, the 3rd scaffolding with the 3.5 Kb mate pair library and the 4th scaffolding with the 8.4 Kb mate pair library.

	Number of sequences	Number of bp	Maximum size	Minimum size	Average size	N50
Contigs	182	662833	45801	139	3488	7968
1st scaffolding	151	665241	45801	139	4405	9629
2nd scaffolding	94	688779	50090	515	7327	13774
3rd scaffolding	78	704453	50090	515	9031	25832
4th scaffolding	70	716307	83770	515	10232	47009

3.4.1.5. *S*-locus region annotation

Using the software Apollo, the annotations of the scaffolds was visualized and recorded. For each scaffold, the software presented the nucleic sequence with the *Lolium* ESTs matching it. The Figure 3.5 represents an example of the visualization using Apollo.

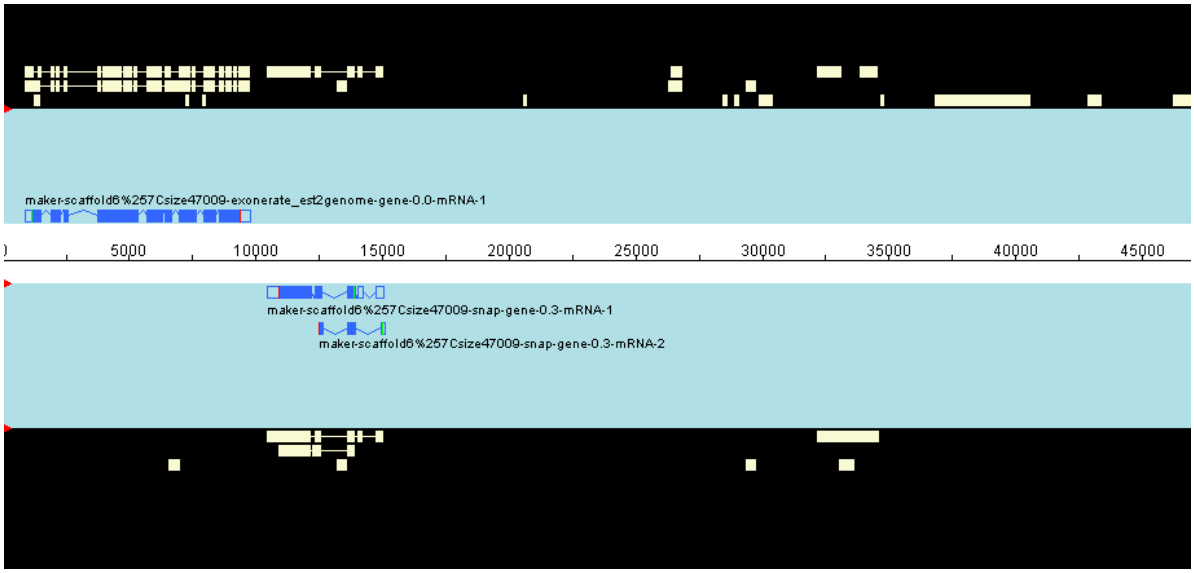


Figure 3.5: Visualization of the gene annotation for the scaffold 6. The scaffold is represented by the ruler (scaffold 6 size is about 45000 bp) and the bars in light colours represent the *Lolium* ESTs matching the scaffold. From those EST, the genes are predicted, represented in blue. Not all the ESTs are associated with gene prediction so annotations have to be checked manually for possible extra genes.

From the annotation of the 70 scaffolds, only 16 scaffolds were annotated with coding sequences (see Table 3.10). Nine genes from NCBI database were predicted within the *S*-locus region; however, a few annotations were not identified so there could be potentially unidentified genes. However retrotransposons were identified regularly during the annotation process. Moreover, several annotations are common between scaffolds, with up to three different annotations syntenic to the rice gene Os05g0150500 and Os05g0150400.

Table 3.10: Annotation of the scaffolds covering the *S*-locus region. Only the scaffolds with annotations are detailed in this table. The table include for each scaffold the starting and ending position of the annotation as well as the results from a BLAST search.

Scaffold number	Annotation		BLAST results		
	Start position (bp)	End position (bp)	*Oryza sativa*	*Brachypodium distachyon*	Other grasses
1	3,011	7,497	Os05g0150500 / TIR1	Bradi2g35720.1	
	45,508	47,943	No BLAST hit	No BLAST hit	*Hordeum vulgare* predicted protein (GI: 326532265)
2	905	9,730	Os05g0150300	Bradi2g35740.1	
	10,424	14,965	Os05g0150400	Bradi2g35730.1	
3	865	5,878	Os05g0150400	Bradi2g35730.1	
	31,205	38,104	No BLAST hit	No BLAST hit	*Lolium perenne* retrotransposon (GI: 359301480)
4	665	1,561	No BLAST hit	No BLAST hit	No BLAST hit
	45,398	48,007	No BLAST hit	No BLAST hit	No BLAST hit
	70,205	79,918	No BLAST hit	No BLAST hit	*L. perenne* retrotransposon (GI: 359301480)
6	905	9,736	Os05g0150300	Bradi2g35740.1	
	10,431	14,972	Os05g0150400	Bradi2g35730.1	
	36,800	40,550	No BLAST hit	No BLAST hit	No BLAST hit
7	6,657	7,239	No BLAST hit	No BLAST hit	No BLAST hit
	11,027	19,548	Os05g0150600	Bradi2g35707.1	
8	1,000	11,000	Os04g0142400		
	11,000	25,000	No BLAST hit	No BLAST hit	*L. perenne* cinnamoyl-CoA reductase (CCR1) gene, promoter region GI:17978547
11	202	4,854	Os05g0150500 / TIR1	Bradi2g35720.1	
	7,175	11,809	No BLAST hit	No BLAST hit	*L. perenne* retrotransposon (GI: 359301480)
13	6,062	10,848	No BLAST hit	No BLAST hit	No BLAST hit

Scaffold number	Annotation		BLAST results		
	Start position (bp)	End position (bp)	*Oryza sativa*	*Brachypodium distachyon*	Other
15	5,894	10,587	Os05g0150500 / TIR1	Bradi2g35720.1	
17	2,518	3,365	Partial Os05g0150000	Partial Bradi2g35760.1	
	6,053	7,230	Os05g0149900	Bradi2g35800.1	
23	190	604	No BLAST hit	No BLAST hit	No BLAST hit
42	1,590	2,067	No BLAST hit	Partial Bradi2g35767.1	*L. perenne* NBS-LRR disease resistance protein-like gene (GI:156106257)
47	819	1,205	No BLAST hit	No BLAST hit	*L. perenne* retrotransposon (GI: 359301480)
52	66	432	No BLAST hit	No BLAST hit	*L. perenne* retrotransposon (GI: 359301480)
56	225	1,043	No BLAST hit	Partial Bradi2g35767.1	*L. perenne* NBS-LRR disease resistance protein-like gene (GI:156106257)

3.4.2. Tissue specific transcriptome analysis

3.4.2.1. Results from the sequencing

After sequencing, the reads were filtered and separated using their index. The Table 3.11 shows the number of paired-end reads obtained for each library after filtering out short reads and bad quality reads.

Table 3.11: Number of paired-end (PE) reads after sequencing using Illumina sequencing.

Sample Name	Number of PE reads (in millions)
58_3_S	95.4
58_3_P	89.8
59_3_S	60.5
59_3_P	90.0
59_21_P	96.8
SI_1	100.0
SI_2	84.3
SI_3_	90.2
SC_1	88.5
SC_2	89.9
SC_3	29.1
F1_30_1	73.6
F1_30_2	66.0
F1_30_3	66.0
Male_1	69.0
Male_2	64.1
Male_3	59.1

The number of reads per samples has exceeded the expectancy with an average of 79.7 M reads per sample after filtering. Only the sample SC_3 had a low number of reads, with 29.1 M, and was not used in the expression study.

3.4.2.2. *De* novo assembly results

The transcripts were assembled for each sample, therefore creating a transcriptome for each plant genotype and tissue. The transcriptome for each sample is composed of a large number of reads with a minimum size set at 200 bp. After assembly, the samples 59_3_P has the lower number of reads (57743) constituting its transcriptome. The N50 was calculated for each transcriptome; when reads are arranged decreasingly according to their size, N50 represents the middle point, from which half of the nucleotides are represented; therefore the sum of all reads with read length above N50 represents 50% of the bases sequenced. The greater the N50, the better the assembly and in this study, N50 is around 650 bp for the sample 59_3_S but much greater for sample 58_3_S with 1134 bp. The Table 3.12 represents all the data from the assembly for each transcriptome from samples from Aberystwyth University (UK) only.

Table 3.12: Transcriptome details after assembly of the 50 bp reads. For each sample, the table gives the number of transcripts the transcriptome is made of, the maximum size of read, the size mean and the N50.

Sample names	**Number of reads**	**Mean (bp)**	**N50 (bp)**	**Maximum size (bp)**
58_3_P	102,381	676	1,034	13,352
58_3_S	139,266	701	1,134	13,348
59_3_P	81,250	690	1,068	13,285
59_3_S	57,743	521	669	8,068
59_21_P	163,538	558	717	25,478

3.4.2.3. Gene expression of the *S*-locus region

Using the software IGV, the expression in the *S*-locus region for each scaffold was assessed. A large number of the scaffolds showed no expression. Most of the scaffolds with no expression in pollen and pistil were not annotated and therefore did not show any expression. However, few scaffolds with no annotation but showing expression were investigated. The expressed regions were showing similarities with *Lolium perenne* retrotransposons Camilla (Asp *et al*., 2011; GenBank GI: 359301480), and the expression level was not specific to a tissue or a genotypes. Other unannotated regions showed an expression pattern that could be associated with a tissue, such as scaffold 11 shown in Figure 3.6.

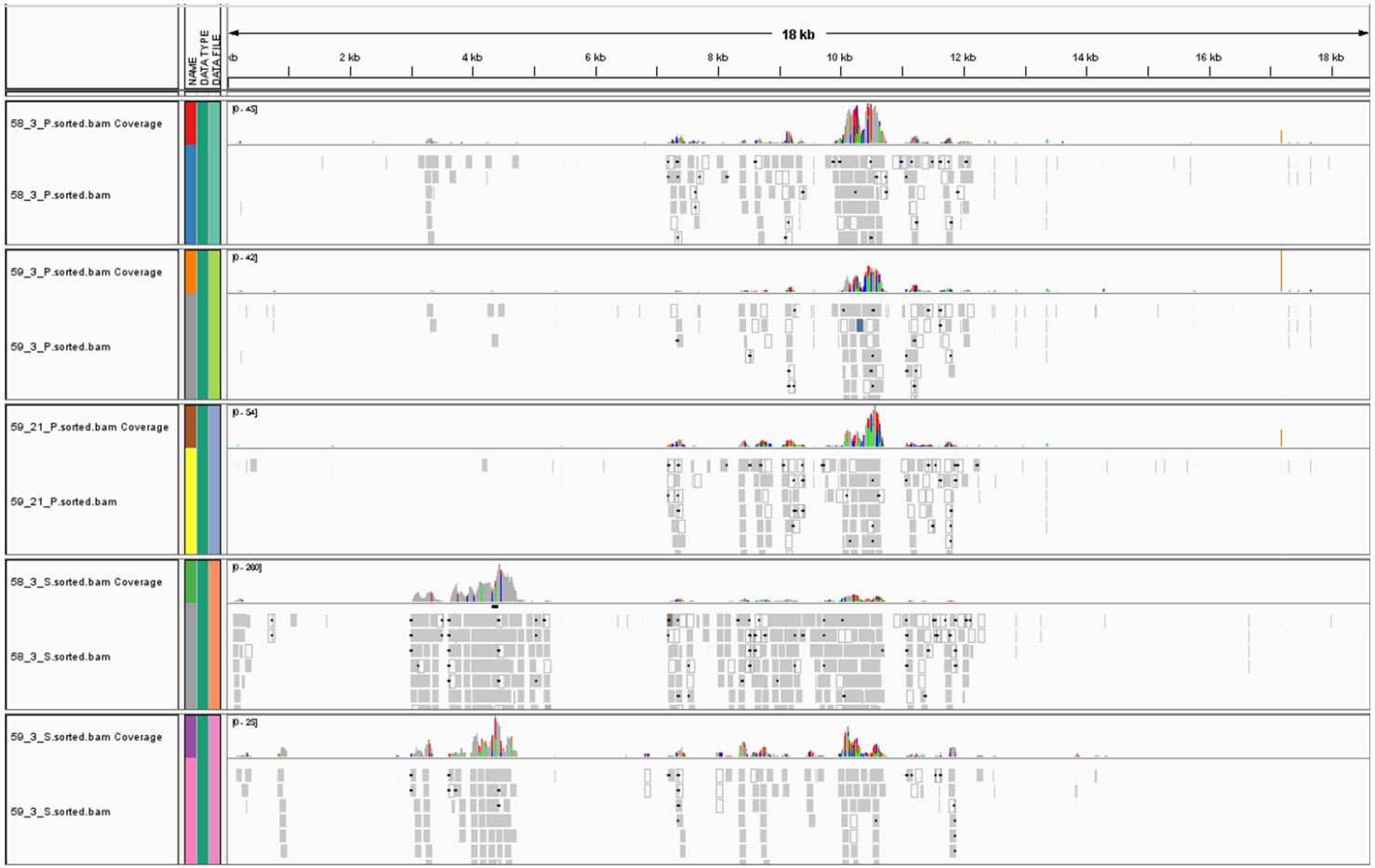


Figure 3.6: Print screen of the alignment between transcriptome and genomic DNA (scaffold 11). The expression level can be analysed by looking at the number of reads aligned for each sample to a specific part of the scaffold. The reads are represented with the grey bars, with the level of expression (coverage) represented by the coloured graphs. The samples are from top to bottom 58_3_P, 59_3_P, 59_21_P, 58_3_S and 58_3_S. The scaffold region between 2 and 5 kb seems to be specifically expressed in stigma tissue whereas the region from 7 to 12 kb is represented in all five samples.

In order to compare the gene expression level between tissues and between cross-pollination (compatible) and self-pollination (incompatible), the raw reads mapped to the annotation regions as well as the non-annotated regions showing expression pattern were extracted using Rsamtools. The variation between replicates was assessed by calculating the biological coefficient of variation (BCV). For the tissue expression analysis, the samples were analysed into two separate groups: the three biological repeats of stigma compared to pollen (F1_30_1, 2 and 3 compared with Male_1, 2 and 3) and the genotype samples, comparing stigma against pollen (58_3_S and 59_3_S against 58_3_P, 59_3_P and 59_21_P). The BCV for these two comparisons was calculated at 25.6% for the biological tissue replicate and 43.5% for the genotypes tissue samples. The pollination comparison (compatible against incompatible) gave a BCV of 29.3%.

The gene expression study was done for each annotation by comparing the level of expression between stigma and pollen and between incompatible pollination (SI) and compatible pollination (SC) (see Appendices 1 and 2 for the detailed results). From this, it was observed that approximately 80% of the reads mapped to the annotations from scaffold 2 (905 to 9730 bp) and scaffold 4 (905 to 9736 bp), both syntenic to the rice gene Os05g0150300 and *Brachypodium* gene Bradi2g35740.1. These two annotations are over expressed in stigma, 4.9 fold compared to pollen. However, in order to have a better representation of the level of expression of the other annotations, these two annotations were removed and another analysis was done for both tissue and pollination comparison. The BCV for the tissue samples was calculated again: 24% for the F1_30/Male samples and 55.54% for the genotype samples, which shows less variation for the first tissue samples group but more for the second. All the results from the expression studies for both tissue and pollination samples are shown in Appendices 1 and 2 respectively.

The simplified results of the tissue-specific expression analysis are presented in Table 3.13. Out of the initial 27 annotations identified in the *S*-locus region, twenty annotations were assessed in the expression study (five annotations were removed because they were duplicated and two annotations were removed because they were representing 80% of the gene expression level). Out of the syntenic *Lolium* annotations with rice and *Brachypodium*, four genes are up-regulated in stigma tissue: Os0g0149900/Bradi2g35800.1 by 3-fold, Os0g0150000/ Bradi2g35760.1 by 11.3-fold, Os0g0150400/Bradi2g35730.1 by 13-fold, and Os05g0150500/Bradi2g35720.1 by 5.6-fold on average. The other gene being more highly expressed 2.6-fold in stigma is the partial Bradi2g35767.1 gene, syntenic also to the *L. perenne* NBS-LRR disease resistance protein-like gene. However, the annotation from the scaffold 56 is also syntenic to these genes and is highly expressed in pollen, by the same amount approximately (3 fold). As for the *Lolium* annotations over expressed in pollen, only three annotations have a synteny with a grass species: the gene Os05g0150600/ Bradi2g35707.1 over expressed 4-fold, the *L. perenne* promoter CCR1 gene over expressed 1.9-fold and the *Hordeum vulgare* predicted protein over expressed 2.2-fold. Also over expressed in pollen are three annotations where no synteny was found with other plants, with an increase of 5.3-fold approximately. The *Lolium* retrotransposon annotations showed high expression levels but not differences in the level of expression between pollen and stigma.

As for the gene expression comparison between self-pollination and cross-pollination in the *S*-locus region, only one annotation was differentially expressed: scaffold 4 annotation starting at 665 bp (see Appendix 2). The annotation expression was higher in the compatible pollination samples but by 2.1-fold. Considering that only 13 reads on average are covering the annotation in a self-incompatible pollination and 23 in a compatible reaction, this hardly represents a significant difference. Once again, in the pollination expression level, the

annotations from scaffold 2 and 4 - 905bp (Os05g0150300/Bradi2g35740.1) were predominantly expressed, but at the same level between the two types of pollination.

Table 3.13: Comparison of the gene expression level between two tissues: pollen and stigma. For each annotation, the comparison has been made using three samples repeated (F1_30 for stigma and Male for pollen) and using different genotypes as repeats (IBERS samples: 58_3_S/P, 59_3_S/P, 59_21_P). For each comparison, the logFC (log base 2) and the P Value are given and the significant changes are highlighted in yellow if the expression is higher in stigma compared to pollen, and green is the expression level is higher in pollen (the absence of colour signifying the non-significance of the expression difference between pollen and stigma).

	Annotations			**Biological replicates (F1-30/Male)**		**Genotype samples (IBERS)**	
Scaffold number	**Start position (bp)**	**End position (bp)**	**Syntenic gene**	**logFC**	**P Value**	**logFC**	**P Value**
1	3,011	7,497	Os05g0150500/Bradi2g35720.1	-2.594	4.62E-18	-4.249	2.49E-05
	45,508	47,943	*Hordeum vulgare* predicted protein	1.125	8.55E-04	-0.786	0.42
2	10,424	14,965	Os05g0150400 / Bradi2g35730.1	-3.623	1.34E-30	-2.45	3.90E-03
3	31,205	38,104	*L. perenne* retrotransposon	0.973	1.49E-04	0.618	0.44
4	665	1,561	No synteny found	2.762	1.25E-07	1.182	0.096
	45,398	48,007	No synteny found	0.419	0.21	-3.178	2.30E-03
6	10,431	14,972	Os05g0150400 / Bradi2g35730.1	-3.858	1.58E-34	-2.754	1.10E-03
	36,800	40,550	No synteny found	2.014	1.77E-14	1.571	0.04
7	6,657	7,239	No synteny found	2.319	9.40E-08	0.827	0.54
	11,027	19,548	Os05g0150600 / Bradi2g35707.1	2.031	2.23E-09	1.369	0.033
8	1,000	11,000	Os04g0142400	0.773	9.16E-03	-1.569	0.043
	11,000	25,000	*L. perenne* promoter CCR1 gene	0.936	1.44E-03	2.403	4.40E-04
11	202	4,854	Os05g0150500 / Bradi2g35720.1	-3.518	5.32E-29	-4.236	7.11E-06
15	5,894	10,587	Os05g0150500 / Bradi2g35720.1	-1.489	4.81E-07	-4.114	2.95E-05
17	2,518	3,365	Partial Os05g0150000 / Bradi2g35760.1	-3.541	1.99E-21	-0.919	0.32
	6,053	7,230	Os05g0149900 / Bradi2g35800.1	-1.604	2.47E-05	-0.269	0.92
23	190	604	No synteny found	1.108	0.21	-0.024	0.9
42	1,590	2,067	Partial Bradi2g35767.1 / *L. perenne* NBS-LRR disease resistance protein-like gene	-1.372	1.99E-03	1.956	0.13
47	819	1,205	*L. perenne* retrotransposon	0.778	0.026	1.209	0.11
56	225	1,043	Partial Bradi2g35767.1 / L. perenne NBS-LRR disease resistance protein-like gene	-1.604	2.11E-04	1.121	0.63

3.4.3. Allelic diversity of the gene Osg0150500-like in *Lolium perenne*

As shown previously in Figure 3.6 and in Table 3.13, the *Lolium* region annotated as Os05g0150500-like showed some differences in expression between pollen and stigma. This *Lolium* gene can be found in scaffold 1 between 3,011 and 7,497 bp, scaffold 11 between 202 and 4,854 bp and in scaffold 15, between 5,004 and 10,798 bp.

3.4.3.1. Isolation of the *Lolium* coding sequence and protein sequence

By aligning the scaffold 1, 11 and 15 to the rice and *Brachypodium* genes (see Appendix 3), scaffold 15 was selected as the scaffold to use for the genomic sequence of the annotated gene. Indeed, the alignment revealed that scaffold 1 was not suitable for the study of the gene as its alignment did not match well with the other two scaffolds, probably due to an error during the scaffold assembly. The aligned sequence of scaffold 11 was found not to align fully to the *Brachypodium* genes, unlike scaffold 15, which was therefore used in the study to predict the coding sequence and the protein sequence.

Using the online program Augustus, the *Lolium* CDS of Os05g0150500-like gene was predicted from scaffold 15 (see Appendix 4). The prediction revealed that the gene was made out of four exons and three introns, as represented on Figure 3.7. However, the prediction from the software FGENESH was different as the predicted gene is composed of three exons and not four, as shown with the Figure 3.8 (see Appendix 5 for the predicted CDS and protein sequences from FGENESH). Both predicted CDS were aligned to the *Brachypodium* CDS using MAFFT; the results are presented in Figure 3.9. From this alignment, the predicted CDS and protein from FGENESH were selected for the rest of the

study as the Augustus-predicted CDS was missing a part of the sequence. The rice Os05g0150500 CDS was not used in this alignment as it did not align to the sequence as well as the *Brachypodium* CDS.

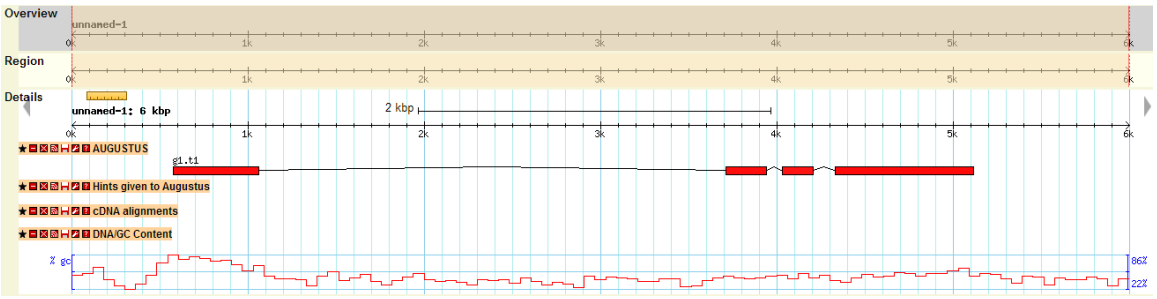
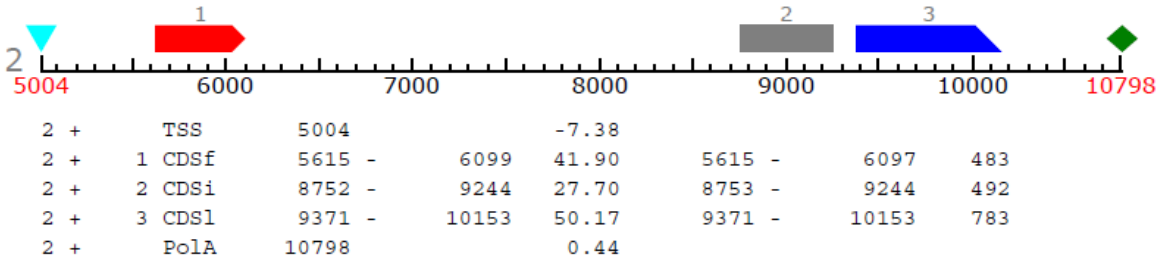


Figure 3.7: Prediction of the Os05g0150500-like gene in *Lolium perenne* using Augustus. The prediction was made from the *Lolium* scaffold 15. The red bars are representing the exons and the thin grey lines are representing the introns. The red graph at the bottom represents the GC content of the sequence.



2	+		TSS	5004		-7.38		
2	+	1	CDSf	5615	- 6099	41.90	5615 - 6097	483
2	+	2	CDSi	8752	- 9244	27.70	8753 - 9244	492
2	+	3	CDSl	9371	- 10153	50.17	9371 - 10153	783
2	+		PolA	10798		0.44		

Figure 3.8: Prediction of the Os05g0150500-like gene in *Lolium perenne* using FGENESH. The prediction was made from the *Lolium* scaffold 15. The red, grey and blue bars represent the three exons, the light blue triangle represents the TATA-box and the green diamond represents the polyA signal. In red letters are the start and end positions (bp) of the gene in scaffold 15.

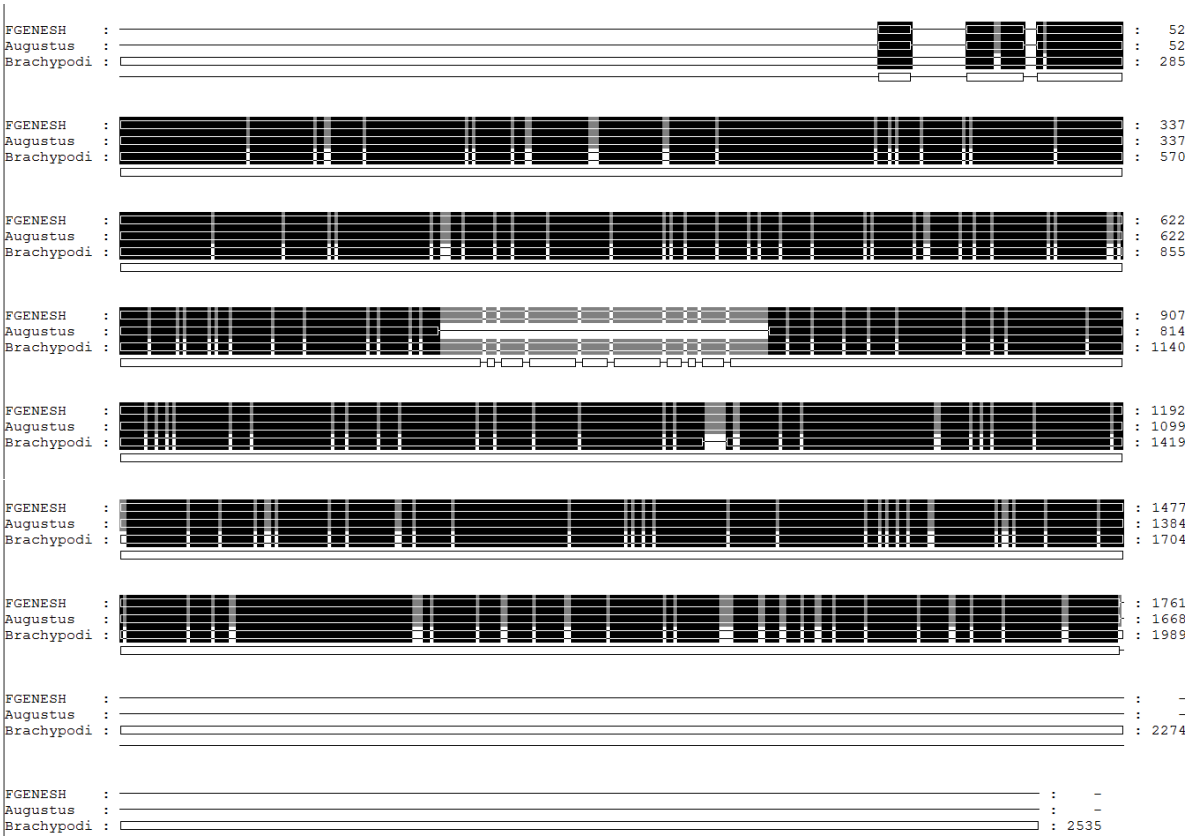


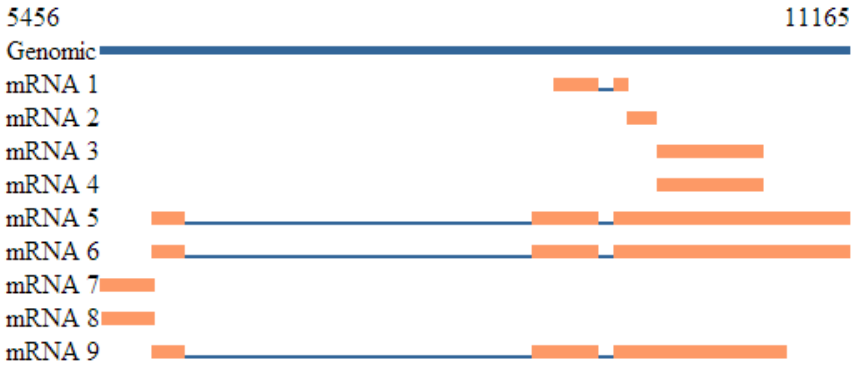
Figure 3.9: Graphic representation of the alignment between the CDS from *Lolium perenne* (predicted from FGENESH and Augustus) and *Brachypodium* Bradi2g35720.1. In black are represented the nucleotides common between the three sequences and in grey, nucleotides common between two sequences.

3.4.3.2. Alignment of the *Lolium* transcripts to the *Lolium* predicted CDS

Using the CLC Genomics Workbench software, the *Lolium* cDNA transcripts from stigma samples (genotype 58_3, 59_3 and F1_30) covering the *Lolium* Os05g0150500-like gene from scaffold 15 were extracted. Only transcripts from stigma samples were pooled out as only few small transcripts were found in pollen samples. For confirmation, the selected transcripts were aligned to the scaffold 15 (see Figure 3.10). An alignment of the transcripts against the predicted *Lolium* CDS is also presented in Figure 3.10 (b), showing

that the sequence from the genotype 59_3 is not as good as the other two, as the sequence is missing the beginning of the CDS. This alignment also showed that the genotype 58_3 reads are shorter than the actual predicted CDS, as they only start at the nucleotide 241, unlike for the genotype F1_30, which is covering the whole CDS. Moreover, for each genotype, two transcripts map to the same location. This is due to the fact that, if any difference was found when the reads were assembled, two sets of transcripts were built, showing that the sample is heterozygous. The details of the selected transcripts are presented in Table 3.14.

A



B

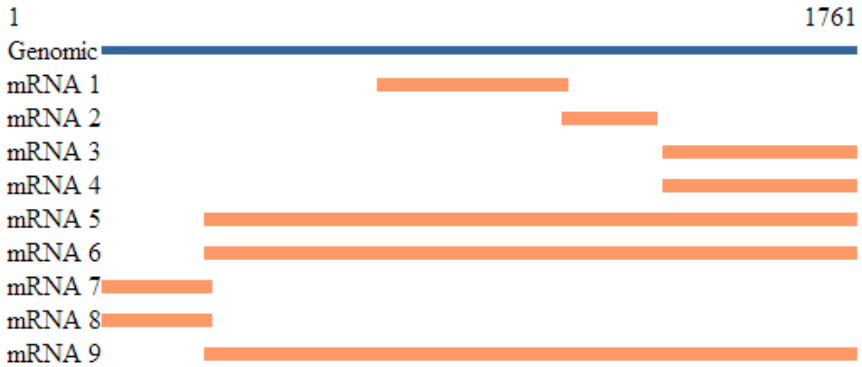


Figure 3.10: Graphic representation of the *Lolium* transcripts mapped to the *Lolium* scaffold 15 (A) and predicted CDS (B). The transcripts represented by orange bars are from three different genotypes, from stigma samples: 59_3 (mRNA1, mRNA2, mRNA 3and mRNA 4), 58_3 (mRNA 5 and mRNA 6) and F1_30 (mRNA 7, mRNA 8 and mRNA 9).

Table 3.14: Table of the correspondence between genotype, read names and mRNA number. For each genotype, several reads are needed to cover the whole *Lolium* CDS. The mRNA number corresponds to the number in Figure 3.10.

Genotype	Transcript name	mRNA number
F1_30	F1_30_S_11	mRNA 7
	F1_30_S_12	mRNA 8
	F1_30_S_2	mRNA 9
58_3	58_3S_11	mRNA 5
	58_3S_11	mRNA 6
59_3	59_3S_1	mRNA 1
	59_3S_2	mRNA 2
	59_3S_31	mRNA 3
	59_3S_32	mRNA 4

The *Lolium* transcripts (cDNA from stigma) were aligned to the *Lolium* predicted CDS as well as the *Brachypodium* Bradi2g35720.1 CDS in order to look at the polymorphism between *Lolium* transcript and the conservation of the gene between species (presented in Appendix 6). This alignment highlighted few polymorphisms between *Lolium* and *Brachypodium* sequences, mainly of one nucleotide. Two large insertion/deletions (InDel) occured between the two species, around the *Brachypodium* position 10 and 1090 bp, but the InDels are of 18 and 6 bp respectively, so the reading frame for the translation into protein would not be disrupted. Only a few polymorphisms between *Lolium* genotypes were highlighted, and could not be associated with any genotype of the *S*-allele.

3.4.3.3. Alignment of the predicted *Lolium* protein sequence for each genotype

The comparison of the protein sequence is important as it determines if the Os05g0150500-like *Lolium* protein is highly polymorphic and if the polymorphism can be associated to the *S*-allele of the different genotype, which would indicate that this gene is potentially the stigma *S*-gene. The alignment of the protein sequences from rice, *Brachypodium* and *Lolium* is presented in Figure 3.11. This alignment revealed that the protein in rice is shorter than in *Brachypodium* and *Lolium*, but the rest of the sequences were well conserved, with only few variations of amino-acids.

The alignment of the protein sequences of the predicted *Lolium* gene and the different genotypes is presented in Figure 3.12. No polymorphism between the different genotypes was found, to the exception of one AA for the genotype F1-30 (an M instead of an I).

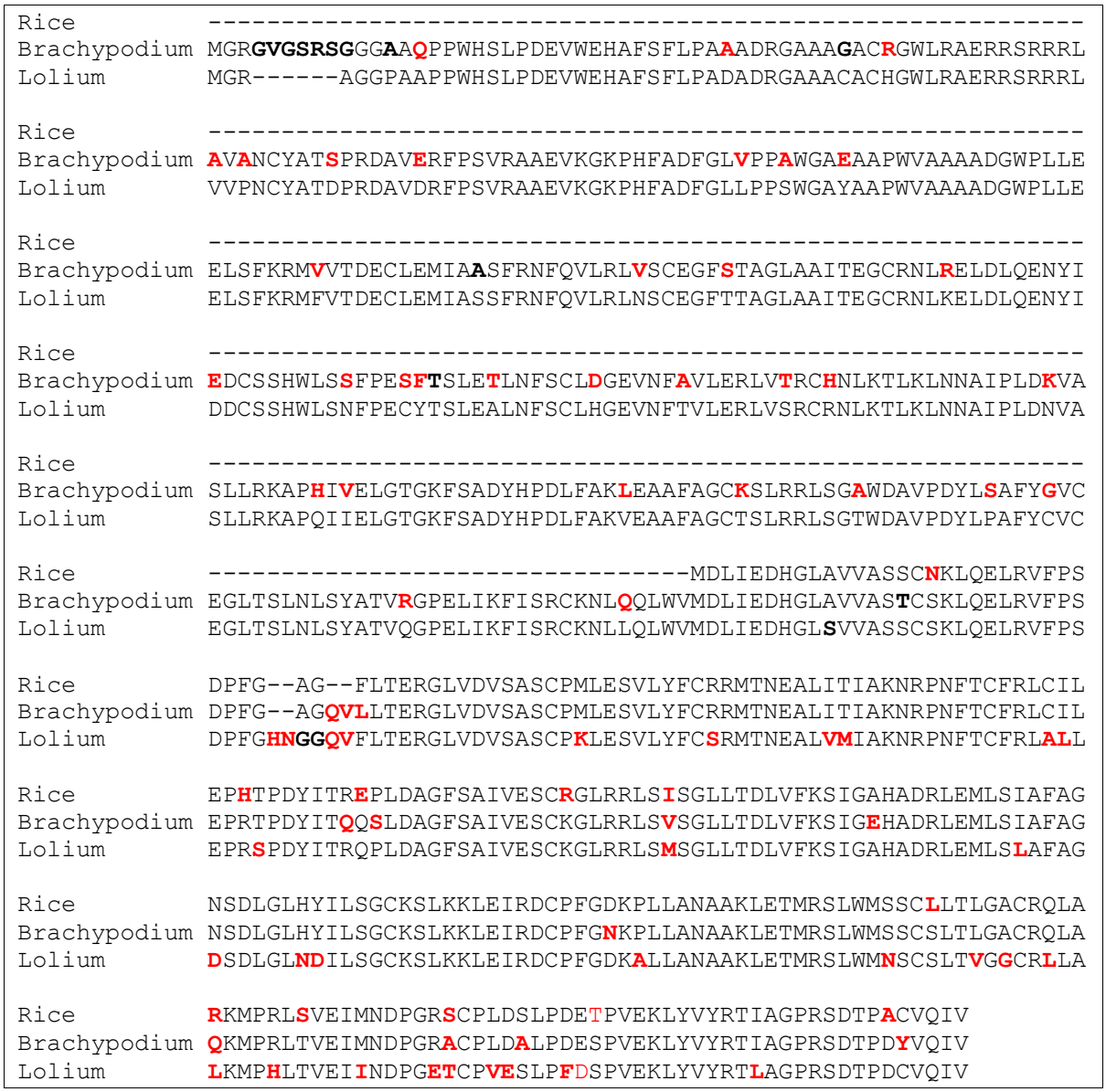


Figure 3.11: Alignment of the rice Os05g0150500 protein, the *Brachypodium* Bradi2g35720.1 protein and the predicted *Lolium* protein sequences. In red are represented the differences between the three species.

```
Predicted   MGRAGGPAAPPWHSLPDEVWEHAFSFLPADADRGAAACACHGWLRAERRSRRRLVVPNCYATDPRDA
F1_30_S_11  MGRAGGPAAPPWHSLPDEVWEHAFSFLPADADRGAAACACHGWLRAERRSRRRLVVPNCYATDPRDA
F1_30_S_12  MGRAGGPAAPPWHSLPDEVWEHAFSFLPADADRGAAACACHGWLRAERRSRRRLVVPNCYATDPRDA

Predicted   VDRFPSVRAAEVKGKPHFADFGLLPPSWGAYAAPWVAAAADGWPLLEELSFKRMFVTDECLEMIASS
F1_30_S_11  VDRFPSVRAAEVKGKPHF----------------------------------------------------
F1_30_S_12  VDRFPSVRAAEVKGKPHF----------------------------------------------------
F1_30_S_2   -------------GKPHFADFGLLPPSWGAYAAPWVAAAADGWPLLEELSFKRMFVTDECLEMIASS
58_3S_11    -------------GKPHFADFGLLPPSWGAYAAPWVAAAADGWPLLEELSFKRMFVTDECLEMIASS
58_3S_12    -------------GKPHFADFGLLPPSWGAYAAPWVAAAADGWPLLEELSFKRMFVTDECLEMIASS


Predicted   FRNFQVLRLNSCEGFTTAGLAAITEGCRNLKELDLQENYIDDCSSHWLSNFPECYTSLEALNFSCLH
F1_30_S_2   FRNFQVLRLNSCEGFTTAGLAAITEGCRNLKELDLQENYIDDCSSHWLSNFPECYTSLEALNFSCLH
58_3S_11    FRNFQVLRLNSCEGFTTAGLAAITEGCRNLKELDLQENYIDDCSSHWLSNFPECYTSLEALNFSCLH
58_3S_12    FRNFQVLRLNSCEGFTTAGLAAITEGCRNLKELDLQENYIDDCSSHWLSNFPECYTSLEALNFSCLH

Predicted   GEVNFTVLERLVSRCRNLKTLKLNNAIPLDNVASLLRKAPQIIELGTGKFSADYHPDLFAKVEAAFA
F1_30_S_2   GEVNFTVLERLVSRCRNLKTLKLNNAIPLDNVASLLRKAPQIIELGTGKFSADYHPDLFAKVEAAFA
58_3S_11    GEVNFTVLERLVSRCRNLKTLKLNNAIPLDNVASLLRKAPQIIELGTGKFSADYHPDLFAKVEAAFA
58_3S_12    GEVNFTVLERLVSRCRNLKTLKLNNAIPLDNVASLLRKAPQIIELGTGKFSADYHPDLFAKVEAAFA
59_3S_1     --------------CRNLKTLKLNNAIPLDNVASLLRKAPQIIELGTGKFSADYHPDLFAKVEAAFA

Predicted   GCTSLRRLSGTWDAVPDYLPAFYCVCEGLTSLNLSYATVQGPELIKFISRCKNLLQLWVMDLIEDHG
F1_30_S_2   GCTSLRRLSGTWDAVPDYLPAFYCVCEGLTSLNLSYATVQGPELIKFISRCKNLLQLWVMDLIEDHG
58_3S_11    GCTSLRRLSGTWDAVPDYLPAFYCVCEGLTSLNLSYATVQGPELIKFISRCKNLLQLWVMDLIEDHG
58_3S_12    GCTSLRRLSGTWDAVPDYLPAFYCVCEGLTSLNLSYATVQGPELIKFISRCKNLLQLWVMDLIEDHG
59_3S_1     GCTSLRRLSGTWDAVPDYLPAFYCVCEGLTSLNLSYATVQGPELIKFISRCKNLLQLWVMDLIEDHG

Predicted   LSVVASSCSKLQELRVFPSDPFGHNGGQVFLTERGLVDVSASCPKLESVLYFCSRMTNEALVMIAKN
F1_30_S_2   LSVVASSCSKLQELRVFPSDPFGHNGGQVFLTERGLVDVSASCPKLESVLYFCSRMTNEALVMIAKN
58_3S_11    LSVVASSCSKLQELRVFPSDPFGHNGGQVFLTERGLVDVSASCPKLESVLYFCSRMTNEALVMIAKN
58_3S_12    LSVVASSCSKLQELRVFPSDPFGHNGGQVFLTERGLVDVSASCPKLESVLYFCSRMTNEALVMIAKN
59_3S_1     LSVVASSCSKLQELRVFPSDPFGHNG------------------------------------------
59_3S_2     ------------------------NGGQVFLTERGLVDVSASCPKLESVLYFCSRMTNEALVMIAKN

Predicted   RPNFTCFRLALLEPRSPDYITRQPLDAGFSAIVESCKGLRRLSMSGLLTDLVFKSIGAHADRLEMLS
F1_30_S_2   RPNFTCFRLALLEPRSPDYMTRQPLDAGFSAIVESCKGLRRLSMSGLLTDLVFKSIGAHADRLEMLS
58_3S_11    RPNFTCFRLALLEPRSPDYITRQPLDAGFSAIVESCKGLRRLSMSGLLTDLVFKSIGAHADRLEMLS
58_3S_12    RPNFTCFRLALLEPRSPDYITRQPLDAGFSAIVESCKGLRRLSMSGLLTDLVFKSIGAHADRLEMLS
59_3S_2     RPNFTCFRLALLEPRSPDYITRQPLDAGF------------------------------------
59_3S_31    ----------------------------------CKGLRRLSMSGLLTDLVFKSIGAHADRLEMLS
59_3S_32    ----------------------------------CKGLRRLSMSGLLTDLVFKSIGAHADRLEMLS

Predicted   LAFAGDSDLGLNDILSGCKSLKKLEIRDCPFGDKALLANAAKLETMRSLWMNSCSLTVGGCRLLALK
F1_30_S_2   LAFAGDSDLGLNDILSGCKSLKKLEIRDCPFGDKALLANAAKLETMRSLWMNSCSLTVGGCRLLALK
58_3S_11    LAFAGDSDLGLNDILSGCKSLKKLEIRDCPFGDKALLANAAKLETMRSLWMNSCSLTVGGCRLLALK
58_3S_12    LAFAGDSDLGLNDILSGCKSLKKLEIRDCPFGDKALLANAAKLETMRSLWMNSCSLTVGGCRLLALK
59_3S_31    LAFAGDSDLGLNDILSGCKSLKKLEIRDCPFGDKALLANAAKLETMRSLWMNSCSLTVGGCRLLALK
59_3S_32    LAFAGDSDLGLNDILSGCKSLKKLEIRDCPFGDKALLANAAKLETMRSLWMNSCSLTVGGCRLLALK

Predicted   MPHLTVEIINDPGETCPVESLPFDSPVEKLYVYRTLAGPRSDTPDCVQIV
F1_30_S_2   MPHLTVEIINDPGETCPVESLPFDSPVEKLYVYRTLAGPRSDTPDCVQIV
58_3S_11    MPHLTVEIINDPGETCPVESLPFDSPVEKLYVYRTLAGPRSDTPDCVQIV
58_3S_12    MPHLTVEIINDPGETCPVESLPFDSPVEKLYVYRTLAGPRSDTPDCVQIV
59_3S_31    MPHLTVEIINDPGETCPVESLPFDSPVEKLYVYRTLAGPRSDTPDCVQIV
59_3S_32    MPHLTVEIINDPGETCPVESLPFDSPVEKLYVYRTLAGPRSDTPDCVQIV
```

Figure 3.12: Alignment of the predicted *Lolium* protein with the sequences from the different genotypes. In red is represented the only amino-acid different between the sequences.

3.5. General discussion

In order to identify the genes contained in the *S*-locus region, the aim was to construct the genomic sequence of the region at first, identify the genes within this region, and then to look at the expression of these genes in the tissues involved in SI: pollen and stigma.

Using the flanking markers as well as non-recombinants markers identified during the fine mapping of the *S*-locus, two different BAC libraries, LTS18 and NV#20F1-30, were screened in order to isolate BAC clones. A total of nine BAC clones were selected to be sequenced using pyrosequencing (454), with a coverage of 120 times per BAC clones. The coverage utilised was comparatively high (typically 40 fold) to maximise the chances of obtaining full length sequence recovery. However, from the screening of the BAC libraries and with the alignment of the BAC clones in order to confirm their order, it seems that a part of the region is potentially absent. Indeed, from the Figures 3.3 and 3.4 (section 3.4.1.1), no overlap could be identified between the BAC clones 520_7A and 520_10H and the BAC clone 573. The same observation was made for the BAC library NV#20F1-30 where the BAC clones 227 and 180 do not overlap with the BAC clone 202 and where the top flanking marker 05_02790 could not be used to identify BAC clone. However given the high number of repetitive elements in this region, the missing region could be non-coding and therefore not contain any genic locations.

The assembly of the reads from the nine BAC clones resulted in the creation of 70 scaffolds, with an average size of 10,232 bp and a N50 of 47,009 bp. These scaffolds were then annotated using the *Lolium* EST database in order to identify conserved genes as well as unidentified genes. From the annotations, only 16 scaffolds were annotated, with 27

annotations. Out of these annotations, six rice genes, with their *Brachypodium* syntenic genes, described in chapter 2 in the *S*-locus region were identified: Os05g0149900, Os05g0150000 (partially), Os05g0150300, Os05g0150400, Os05g0150500 and Os05g0150600. An additional gene was identified in *Brachypodium*, the gene Bradi2g35767.1. Four *Brachypodium* genes from the *S*-locus region were not annotated in the scaffolds: Bradi2g35807.1 (syntenic to the rice Os05g0149800), Bradi2g35790.1, Bradi2g35780.1 and Bradi2g35750.1. An additional rice gene was also not found in the *Lolium S*-locus region, Os05g0150100. The missing annotations could be due to a local breakdown in synteny between *Lolium* and the two model species used so far, *Brachypodium* and rice. However, the gene Os05g0149800/Bradi2g35807.1 missing from the *Lolium* sequence indicates that the *S*-locus region is not complete as the marker 05_02790 is mapped in this gene. Finally, another rice gene has been annotated in the *Lolium S*-locus, Os04g0142400, which is coding for an unknown protein. In addition to the annotations syntenic to the rice and Brachypodium *S*-locus region, five annotations have similarity with *Lolium perenne* retrotransposon *Camilla* and *Lotte* (Asp *et al*., 2011). Six annotations from the *Lolium S*-locus region have no synteny with any plants but have been identified with the *Lolium* EST database and therefore could potentially be unknown proteins. Another annotation has been identified as an unknown protein similar to one identified and uncharacterized in *Hordeum vulgare*. Finally, the last annotation has some similarity with the *Lolium* promoter of the cinnamoyl coA reductase.

In order to study the *Lolium S*-locus genes expression and to identify potential *S* candidate genes, the transcriptomes from pollen, stigma as well as from incompatible and compatible pollinations, were sequenced using NGS (Illumina). For pollen and stigma samples, as well as self-incompatible pollination and compatible pollination, three

biological replicates were used in order to compare them. In addition, three pollen and two stigma samples, with different genotypes for the *S*-locus, were used in order increase the allelic diversity and possibly help to associate a genic allele with the expected *S* allele.

In order to study the gene expression in the *S*-locus region, the transcriptomes were aligned to the *Lolium S*-locus scaffolds sequences. The expression level was assessed by assessing the proportion of raw reads for one sample that was aligned to one annotation. The proportion was then homogenized between the three sample replicates and compared with the other tissue or pollination.

From the expression level study between pollen and stigma samples, six genes syntenic between *Lolium*, rice and *Brachypodium* have been found to be over expressed in stigma samples: Os05g0149900/Bradi2g35800.1 by 3-fold, Bradi2g35767.1 by 3-fold, Os05g0150000/ Bradi2g35760.1 by 11.3-fold, Os05g0150300/Bradi2g35740.1 by 4.9-fold, Os05g0150400/Bradi2g35730.1 by 13-fold, and Os05g0150500/ Bradi2g35720.1 by 5.6-fold on average. For all these genes, the expression difference (fold) between pollen and stigma is important enough, with the P Value as a proof, to affirm that these genes are over-expressed in stigma tissues.

The *Lolium* gene syntenic to the rice gene Os05g0150300 is over represented and expressed among both tissues. These results are not surprising as the gene contains protein domains involved in a transcription factor mechanism, facilitating the chromatin unwinding for example. However, the over expression in stigma could be due to the involvement of the gene in the SI mechanism acting like an on/off switch to the pollen compatibility, by making the access to certain genes available or not. The *Lolium* gene syntenic to Os05g0150400/Bradi2g35730.1, also over expressed in stigma tissue, is also

involved in regulation of gene expression, and could be coupled with the first one, the physical distance between the two being small in *Lolium perenne* (less than 800bp according to the scaffold annotations).

As for the *Lolium* gene syntenic to the rice Os05g0149900 and over expressed in stigma, it has been described in the chapter 2 as a potential good candidate as it is coding a protein containing a tetratricopeptide-like (TPR-like) domain, involved in the protein-protein interaction. Therefore, this gene could be involved in the recognition of the self- or non-self-pollen by acting as a receptor on the stigma surface or in the stigma as from the expression study; it is expressed in stigma tissues. The gene mapping next to it in *Lolium*, homologous to the rice Os05g0150000, could also be an interesting candidate gene. The gene has been described in *Brachypodium* as an enzyme involved in the amino-acid biosynthesis, and therefore unlikely to be involved in a receptor-ligand interaction. However, the homology between the two genes is not complete and the *Lolium* gene therefore might code for another type of protein. Because of this partial homology and the fact that the gene is over expressed in stigma, the gene has to be considered as a candidate gene.

Two *Lolium* annotations (scaffold 42 from 1,590 to 2,067 bp and scaffold 56 from 225 to 1,043 bp), homologous to the same *Brachypodium* gene Bradi2g35767.1, are also showing an over-expression in stigmatic tissue. The *Brachypodium* gene Bradi2g35767.1, syntenic to rice Os07g0141700, is annotated as a protein containing an NB-ARC domain (Nucleotide-Binding – APAF-1, R protein and CED-4), involved in disease resistance in plants (Van Ooijen *et al*., 2008). Resistance mechanisms are based on the recognition of a pathogen and therefore, such a mechanism could function in SI, where the self-pollen

would be recognised and destroyed as a pathogen would be. This gene is a good candidate gene as the stigmatic S-gene.

As for the pollen specific expression, six annotations have been identified as over expressed in pollen. Out of these six annotations, only three have homology with other plant species: scaffold 1 from 45,508 bp, homologous to a predicted protein from *Hordeum vulgare*, scaffold 7 from 11,027 bp, homologous to the rice and *Brachypodium* gene Os05g0150600 and Bradi2g35707.1 respectively and the scaffold 8 from 11,000, homologous to a promoter from the *Lolium* cinnamoyl-CoA reductase gene (CCR1). However, the *Lolium* gene annotation homologous to the rice gene Os05g0150600, over expressed 4.1 times in pollen, cannot be a candidate gene as, as described in the previous chapter, the recombinant marker 05_02889 is included in its sequence.

As for the *Lolium* annotation on scaffold 1, homologous to a predicted *Hordeum vulgare*, its expression in pollen is 2.2 times higher than in stigma. This expression difference is not large and even if the P value (8.5E-04) is significant, the stigma samples are showing some expression, making this gene as an unlikely candidate gene. However, the protein function is unknown and no annotation should be removed from the candidate gene list so an evaluation of the allelic diversity of this annotation is necessary.

Finally, the *Lolium* annotation from scaffold 8, homologous to the promoter of the CCR1 gene is over expressed by 2-fold in pollen, so there is only a small difference in the expression between the two tissues. However, the differential expression is 5.3-fold when comparing the different genotypes. This annotation is similar to a promoter and therefore potentially involved in gene regulation and expression. Even so it cannot be part of a receptor/ligand interaction; the SI mechanism could be, as described previously for the

homologous of Os05g0150300 and Os05g0150400, involved in a regulation process that could control pollen germination.

The three annotations showing an over expression in pollen samples (with a 4 to 6.8-fold difference) are showing no homology with any plant species. Therefore, no assumption can be made on them as they could be novel genes, as it was the case for the poppy pollen gene *PrpS* (Wheeler *et al*., 2009).

The *S*-locus has been annotated with large sequence of retrotransposon elements. The expression analysis did not reveal that their expression was tissue specific. Retrotransposon elements are large repetitive motifs, very abundant in plant genomes as they represent over 50% of the maize and rice genome (SanMiguel and Bennetzen, 1998; Li *et al*., 2004). These repetitive motifs, unlike transposon, can replicate using RNA transcription and then reverse transcribe into DNA where they can re-integrate back into the genome (Boeke and Corces, 1989). By doing so, they can replicate within genes and therefore create mutations. However, only few retrotransposon elements are translated into protein, mainly in order to create the complex for their reverse transcription and therefore cannot be the prime *S* candidates, even if they could regulate *S*-genes expression.

The rice gene Os05g0150500, syntenic to the *Brachypodium* gene Bradi2g35720.1, has been a prime candidate as it is coding a transport inhibitor response gene, containing an F-box domain as well as a leucine rich region (LRR). This gene was found in the *Lolium S*-locus sequence and is over expressed in stigma samples 5.6-fold. Even if the F-box domain involved in the S-RNase GSI mechanism is the pollen gene, this gene remains a good candidate as a stigma-expressed gene. The analysis of the *Lolium* Os05g0150500-like gene using the transcripts from different genotypes did not show any major polymorphisms.

Indeed, only one amino-acid was changed from the sequence of the genotype F1-30, not thought to be significant enough to be involved in any *S*-specificity.

On the other hand, major polymorphisms were observed between the genotypes transcript sequences. However, the polymorphic part of the sequences was mapped after the stop codons, and therefore not translated to the protein sequence. But some studies have revealed that the translation of the mRNA into protein does not always stop at the stop codon (Namy and Rousset, 2010). Indeed, ribosome normally detach from the mRNA when a stop codon is reached, but in some cases the translation continues until the next stop codon. This phenomenon is called "stop codon readthrough". If this stop codon readthrough is applied to *Lolium*, the high variability observed for the CDS could be transferred to protein. The Figure 3.13 is illustrating the example of a shift of one stop codon in the translation of the protein. In this case, the amino acid sequence would be polymorphic: homozygous for the plant 58_3S (S_5/S_5) and the plant F1-30 (S_1/S_2) and heterozygous for the plant 59_3S (S_5/S_6). The heterozygosity is following the *S*-alleles except for the F1-30 plant which should be heterozygous but only have an additional amino-acid change compare to the sequence of what could be the S_6 allele (59_3S_31 in the Figure 3.13).

Figure 3.13: Alignment of the reads mapping to the end of the *Lolium* Os05g0150500-like gene, with one stop codon readthrough, increasing the length of the protein sequence. The red frame is indicating the extra protein sequence from the one stop codon readthrough. The coloured letters show the differences between sequences. The most variability between sequences occurs in this new region. In blue is represented the sequences associated with the allele S_5. The genotype 59_3 is heterozygous S_5S_6, with the allele S_6 sequence being 59_3S_31. Only one sequence for the genotype F1-30 is available, meaning the genotype is homozygous for the gene; however, F1-30 has been genotyped as S_1S_2.

Finally, another explanation as to why the polymorphism expected cannot be identified, if this gene is one of the *S*-genes, could be an alternative splicing of the coding sequence, generating different mRNA. Alternative splicing (AS) is a mechanism by which, a single coding sequence of a gene, with several introns and exons, can be transcribed into different mRNA, by not transcribing an exon (exon skipping) or by retaining an intron in the mRNA sequence (intron retaining) in most cases. In *Brachypodium distachyon*, Walters *et al*. (2013) have found 1219 AS mapped onto 941 genes, suggesting that

approximately 6.3% of the genes could exhibit AS, which is less than in rice (23.5%; Campbell *et al*., 2006). The effects of AS are wide: from generating different transcripts from the same coding sequence, causing a loss of function or regulating gene expression (Lareau *et al.*, 2007). Moreover, AS has been found to be tissue specific in some case and involve in protein-protein interaction mechanisms such as the plant defence response. In tobacco, the N gene encoding for a TIR receptor with a leucine rich region (LRR) has been found to have two splice variant leading to two proteins, one of which is missing its LRR and therefore its resistance to the tobacco mosaic virus. The expression of one variant over the other is activated by the pathogen itself, favouring the production of the truncated protein (Dinesh-Kumar and Baker, 2000).

Alternative splicing could be the source of the polymorphism for this gene, but it would have to be revealed by looking at the homology between the splicing sites of the gene genomic sequence. In the *Brachypodium* region syntenic to the *S*-locus region, the gene Bradi2g35730.1 (homologous to rice Os05g0150400) has been found to have an AS site, creating two splice variants of the gene (Gramene database; http://www.gramene.org/Brachypodium_distachyon/Info/Index).

When comparing the expression differences between pollination types (incompatible and compatible) in the *S*-locus region, only one annotation was found to be differentially expressed: scaffold 4 from 665 bp is over expressed in compatible pollination. However, the difference is small (2.1-fold) but and the P Value (0.02) is only just significant. Moreover, only few reads are from incompatible and compatible pollinations (13 and 23 reads on average respectively). It is therefore unlikely that this gene is involved in the incompatible cascade reaction. Among the other *S*-locus genes

expressed at the same level between both pollinations, the gene syntenic to Os05g0150300 is again representing the majority of the reads mapped to the region (75%).

Studying the *S*-locus gene expression in pollination samples was done in order to identify genes differentially expressed but also gene expression. However, because the samples have been collected approximately 2 hours after pollination, it is expected that, due to the fast acting SI mechanism, the expression of the *S*-locus genes involved in the pollen/stigma recognition would not be expressed anymore. No genes from this study have been found to be expressed in stigma and pollen tissues before pollination but not after pollination, even though the level of expression was varying. However, a better comparison of the samples has to be done, comparing before and after pollination, whatever the plant tissues (pollen or stigma). Moreover, it is possible that transcripts from the *S*-genes are present in a pollinated stigma, even at a low level and therefore no conclusion could be made from this comparison.

The Figure 3.14 summarizes the results from this detailed study of the S-locus region. However, some scaffolds with annotations could not be placed on this map because of a lack of synteny with rice and *Brachypodium* species. Therefore, most of the pollen-specific candidate genes, which have been annotated as unknown, are not represented but should not be left out of the results.

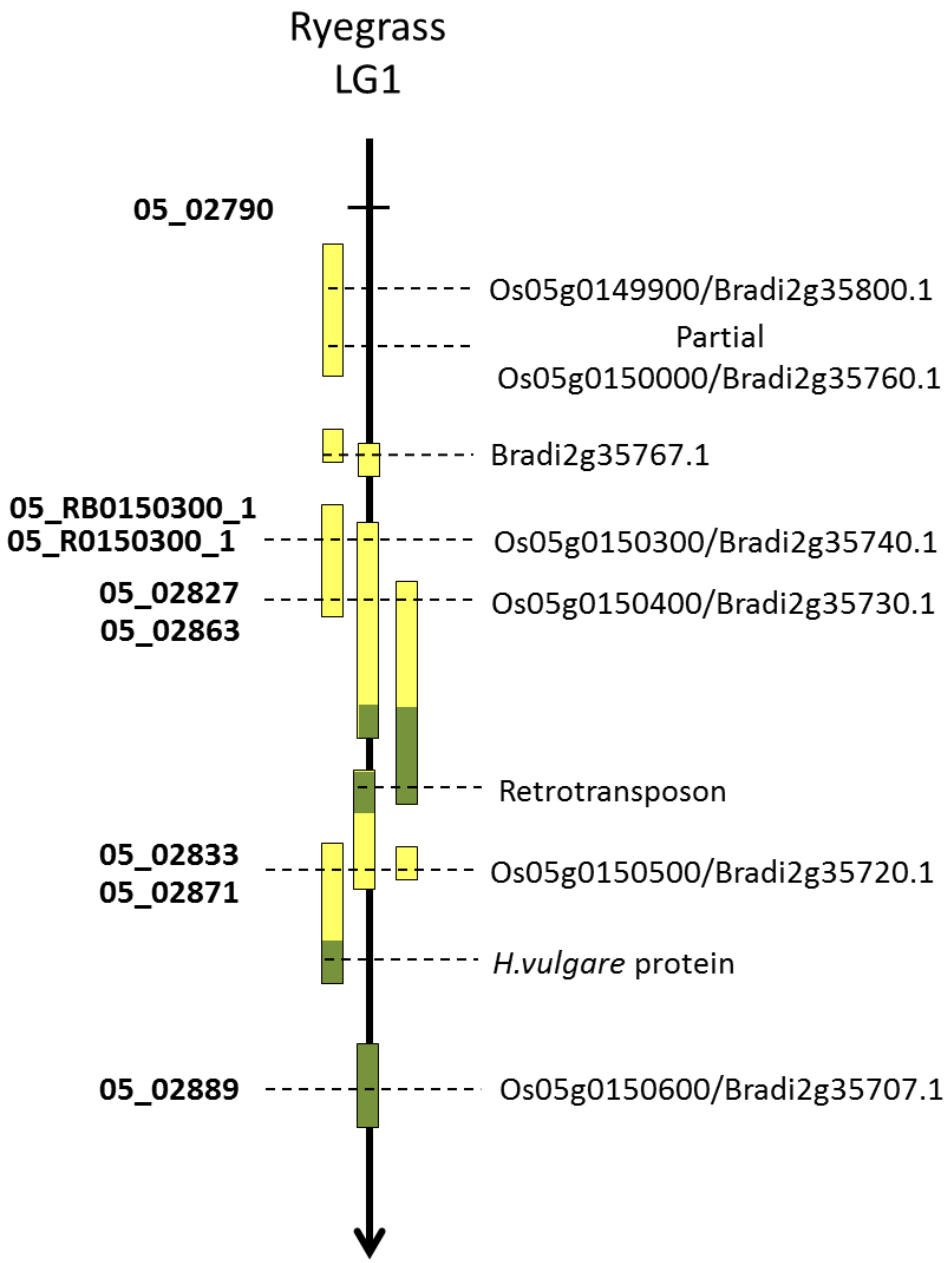


Figure 3.14: *Lolium perenne S*-locus detailed region, with gene annotations and gene expression. The scaffolds are represented in coloured rectangles, yellow when over expressed in stigma tissue and green when over expressed in pollen tissue. The annotations have been placed according to the homologies between *Lolium* scaffolds and the gene ordering in rice and *Brachypodium*. The *Lolium* molecular markers have been placed using their sequence synteny with the *Lolium* scaffolds.

Chapter 4:

Self-incompatibility allele genotyping

4.1. Introduction

4.1.1. Aims and strategy

Self-incompatibility is an advantage in evolution as it promotes outbreeding and therefore increases genetic diversity. In breeding programmes, SI can be used as a tool to produce largely intra-incompatible, inter-compatible parental populations, by reducing the number of SI alleles, and therefore be used in the creation of hybrid population. But SI can also be a constraint in breeding programmes as it prevents breeders from creating pure lines and can affect breeding programmes based on recurrent selection. Restricted numbers of incompatibility alleles may lead to population extinction because of the inability to set seed in a breeding population that has been subject to inbreeding (e.g. by half-sib family selection) or maybe a recently isolated population consisting of very few individuals.

In both instances, being able to predict the *S* and *Z* genotypes of the parents of a cross is of interest for breeders. By genotyping for *S* and *Z* a large number of plants quickly and cheaply, the breeder will be able to select plants on SI genotype according to his breeding strategy: a restricted number of *S* and *Z* alleles to produce parental population for hybrid population production or large allelic diversity in a breeding programme with recurrent selection that might otherwise be subject to inbreeding, incompatibility allele restriction and limitations for seed set.

Being able to routinely genotype natural plant populations would also enable empirical evaluation of the theories of SI population genetics and frequency dependent selection as originally proposed by Sewell Wright in 1939.

The aim of this study is to develop a fast and easy method to genotype both *S* and *Z* alleles without using time-consuming pollination tests as is the case currently. Closely

linked *S* and *Z* molecular markers have been developed in previous studies (chapter 2 and personal communication by Dr Bruno Studer). Using some of these polymorphic markers, combined with the HRM curve analysis, the assessment of *S* and *Z* allelic diversity was conducted in a breeding population of 55 plants, a thirteenth generation of half-sib crosses. The plants' SI loci genotypes were predicted using the genotypes of marker combinations that are assumed to be completely linked to the SI loci. Markers were screened using rapid HRM technology and more traditional genotyping methods separating out length polymorphisms using electrophoresis by capillarity.

4.1.2. Allelic diversity of self-incompatibility

Self-incompatibility has evolved as a system to prevent inbreeding which might otherwise lead to homozygosity and the exposure of lethal recessive genes. SI systems are governed by multi-allelic loci, and in most cases a multi-allelic single *S*-locus. The high allelic diversity is a key feature of SI as it increases the proportion of compatible crosses in a population, i.e. mate availability as referred to by Vekemans *et al*. (1998). However, in single *S*-locus systems, GSI and SSI differ in their mate availability. In SSI systems, the phenotype of a pollination will depend on the male parent and therefore will be the same for all the pollen: self-incompatible or 100% compatible pollination. In GSI however, the phenotype of the pollination varies as it is the genotype of the pollen that is recognized and not the genotype of the male parent. In the case of two plants with one *S*-allele in common, an additional phenotype for the pollination will occur: 50% compatibility, leading to higher mate availability in GSI systems compared to SSI (Vekemans *et al*., 1998).

The maintenance of high allelic diversity in populations is one key element for mate availability, as well as population size (Wright, 1939). New alleles derived either by mutation or through introgression from outside the population are at a selective advantage and increase in frequency over sexual generations (negative frequency dependent selection) and then, in theory, are subject to balancing selection such that allele frequencies become equal. In order to maintain compatibility in a population, at least 3 *S*-alleles are necessary in a single locus SI system. Many studies have been done over the years to evaluate and estimate *S*-allelic diversity of the single locus SI system (reviewed in Lawrence, 2000). The *S*-allelic diversity evaluated in those studies is large, with *S*-allele varying from 12 alleles in *Solanum carolinense* (Lu, 2006), 41 alleles in *Trifolium pratense* (Lawrence, 1996) and up to 46 S-alleles in *Prunus lannesiana* (Kato *et al*., 2007). However, these observations have been made on specific populations and the estimations of the allelic diversity are higher, with over 100 *S*-alleles estimated in *Trifolium* species (Lawrence, 2000). Again, like the mate availability, differences exist between SSI and GSI systems for the *S*-allele diversity. In finite populations and with no dominance interaction between alleles, SSI systems maintain more diversity (Busch *et al*., 2012), thought to be due to a stronger frequency dependent selection of rare alleles (due to a lower mate availability). However, because of dominance between *S*-alleles in many SSI systems, the allelic diversity is reduced due to the negative frequency dependent selection of recessive alleles (Billiard *et al*., 2007). Other factors influencing *S*-allele diversity are the population size, which as a positive effect on diversity and the recombination rate, creating more allelic diversity (Busch *et al*., 2012).

Because of the two-loci governing the GSI system of grasses, the mate availability as well as the allelic diversity is different. Unlike the single *S*-locus system, only two

alleles of each *S* and *Z* locus are necessary to maintain compatibility within a population. But as for the single locus system, the diversity within natural populations is much higher. Lundqvist (1962b, 1964 and 1969) identified 11 *S*-alleles and 12 *Z*-alleles in a *Festuca pratensis* population, and more recently, Fearon *et al*. (1994) have identified in *Lolium perenne* 17 *S*-alleles and 17 *Z*-alleles out of a population of 42 plants. However, the number of *S* and *Z* alleles in *Lolium perenne* is estimated to be higher, with more than 31 alleles at each locus, using the E_2 estimator described by O'Donnell and Lawrence (1984). The E_2 was used to estimate *Lolium S*-allelic diversity as it is adapted to take account of unequal allele frequency, as observed in the *Lolium* population where one *S* allele was present in 12 out of 38 plants and one *Z* allele was common between nine plants out of 39 (Fearon *et al*., 1994). Indeed, even if a rare allele, in any GSI system has an advantage of selection (frequency dependent selection), at equilibrium, it is thought that the allele frequency is equal (Lawrence, 2000).

Most allelic diversity has been estimated using crossing compatibilities. Currently, with the *S*-genes identified in the main SSI and GSI systems, the allelic diversity is evaluated by sequencing different alleles as for example in European pear (Sanzol and Robbins, 2008), in cherry (Kato and Mukai, 2004; Sonneveld *et al*., 2006) and in poppy (Wheeler *et al*., 2009). Another method developed to quickly identify the *S*-alleles in potatoes is PCR-based markers (Marcellán and Acevedo, 2012). However, in the grass two-loci SI system, none of the genes (*S* and *Z*) triggering the SI reaction is known, therefore cross-compatibility determination is the only option so far to determine the allelic diversity of a population. But due to the two loci system, the SI genotyping of the plants through crossing is laborious and time consuming. As described by Fearon *et al*. (1994), the method to determine *S* and *Z* alleles based on pollination tests requires several stages of

crosses to classify plants. Indeed, in order to genotype the SI alleles, pollination tests have to be made between plants having the same pair of *S* or *Z* alleles, leaving the other locus responsible for the phenotype of the pollination, which will be either incompatible, 50% compatible or 100% compatible (see Table 4.1). However, in the chapter 2 of this thesis, a number of close markers have been developed in perennial ryegrass, with a recombination frequency around 0.05%. A method using molecular markers such as these markers would facilitate the prediction of *S* and *Z* genotypes, enabling the study of *S* and *Z* allele diversity in natural and synthetic (breeding) populations of *Lolium perenne*.

Table 4.1: Phenotypes of pollination tests between plants with common *Z* alleles but unknown *S* alleles. The table classifies the phenotype of pollination according to the percentage of compatible pollen. By inter-crossing plants, it is possible to genotype the SI alleles with the condition that one locus has the same alleles, in this example, Z_{12}.

		Male parent genotype				
		$S_{aa}Z_{12}$	$S_{ab}Z_{12}$	$S_{cc}Z_{12}$	$S_{ac}Z_{12}$	$S_{cd}Z_{12}$
Female parent genotype	$S_{aa}Z_{12}$	0	50	100	50	100
	$S_{ab}Z_{12}$	50	0	100	50	100
	$S_{bb}Z_{12}$	100	50	100	100	100

4.2. Material and methods

4.2.1. Breeding population

In this study, fifty five inter-related plants of *Lolium perenne* L., which were the mother plants of the 13th generation of a breeding programme (F13), were genotyped using *S* and *Z* linked markers. The mother plants, F13, were selected from the breeding programme made at IBERS (Aberystwyth University, UK). The IBERS breeding programme was started by Dr Peter Wilkins in the late 1970's and is now continued by Dr Richard Hayes. The breeding method is based on half-sib recurrent selection of plants,

where selected plants from the previous generation are polycrossed, leading to the creation of half-sib populations. A half-sib family is a population harvested from one plant, the mother plant that has been pollinated by all of the other plants of the population in a pollen-proof isolation chamber. The IBERS breeding programme was started with seven genotypes: two S23 genotypes (IBERS, Aberystwyth University, UK), selected for ryegrass mosaic virus resistance, three ecotypes from the North Italian Alps, chosen for high digestibility, and two genotypes of the variety Vigor R.v.P (ILVO). Records of the breeding programme started only from the generation 5 (F5) in 1983, after five cycles of mass selection followed by seven cycles of half-sib family selection, presented in Table 4.2. Two more genotypes were added to the breeding programme to insert new gene diversity: one genotype Jumbo TC2 (cross between Jumbo and AberDart; IBERS, Aberystwyth University, UK) was introduced in generation F9 for its high spring yield and one genotype Twystar (KWS UK) in generation F10 for its overall performance.

The fifty-five mother plants F13 were selected at the 13th generation (current generation) out of the 400 genotypes initially selected from four F12 families. Those fifty-five plants were selected after a strong initial selection pressure for seed yield was applied and were polycrossed in 2011 in order to create the next generation. Seed yielding ability is critical to ensure the commercial success of any forage grass variety in addition to important agronomic traits such as grazing forage yield, silage forage yield, quality (digestibility, protein content, sugars concentration, fibres and lipids), persistence, cold tolerance, and resistance to biotic stresses such as *Drechslera siccans*, which are all selected for in extensive field evaluations (communication from Dr Richard Hayes).

Table 4.2: Design of IBERS breeding programme for *Lolium perenne*. The table is showing the different generation of half-sib crosses, with the number of plants polycrossed and the number of families obtained (one family correspond to the seeds of one genotype polycrossed). The breeding programme started with 7 genotypes and two more genotypes were introduced at a later stage: Jumbo TC2 at generation F9 and Twystar at F10. The number of family for generation F13 is not known at the breeding programme is still currently at the 13th generation stage.

Generation	Number of polycrossed genotypes	Number of families selected for the next generation
Parents	7	7
F1	Unknown	Mass selection
F2	Unknown	Mass selection
F3	Unknown	Mass selection
F4	Unknown	Mass selection
F5	90	2
F6	90	3
F7	170	4
F8	420	3
F9	540	2 + Jumbo TC2
F10	504	3 + Twystar
F11	600	3
F12	414	4
F13	400	?

4.2.2. Methods

4.2.2.1. DNA extraction

DNA was extracted from the fifty-five mother plants F13 following a CTAB extraction protocol. For each plant, one young leaf was selected and placed into a 2.2 ml screw-cap tube with a steel bead. Samples were frozen in liquid nitrogen before being ground twice for 1 minute at 20Hz using a Retsch-mill (Retsch, Haan, Germany). To each sample, 1 ml of preheated (65°C) 2X CTAB extraction buffer (2% CTAB, 0.1 M Tris pH 8.0, 20 mM EDTA pH 8.0, 1.4 M NaCl, 1% β-mercaptoethanol) was added and mix before

samples were incubated for 30 minutes at 65°C. A volume of 500 µl of chloroform:isoamyl alcohol (24:1) was added and mix by inversion. After 10 minutes incubation at room temperature, samples were centrifuged for 10 minutes at 8000 rpm and 900 µl of the supernatant was transferred to a new 1.5 ml Eppendorf tube. 500 µl of isopropanol was added and samples were incubated for 30 minutes at room temperature before been centrifuged for 20 minutes at 8000 rpm. The supernatant was discarded with care and the pellet was washed with 500 µl of 70% ethanol. After centrifugation for 5 minutes at 8000 rpm, the ethanol was removed and the pellets were left to dry for 15 minutes at room temperature, before been resuspended in 100 µl in R40 TE buffer (40 ng/ml RNase A, 1 X TE buffer). An additional RNase treatment was done by adding 0.5 µl RNase A (100 ng/ml) to each DNA sample with an incubation for 1 hour at 37°C.

4.2.2.2. Marker selection

In order to easily genotype plants for *S* and *Z* loci, STS markers using the HRM curve analysis for genotyping were selected. The selected markers were closely linked to the two SI loci, showing a complete linkage or a genetic distance of less than 1cM to *S* or *Z*. However, to confirm with more certainty the genotypes and the haplotypes of the plants, four STS markers were selected and screened for length polymorphism. Moreover, two new markers were designed using primer3 software (http://frodo.wi.mit.edu/primer3/, Rozen and Skaletsky, 2000), using the genomic sequence of the BAC clone contigs flanking the *S*-locus region, BAC_573_001 and BAC_764_022, with no repetitive sequences in order to have specific markers. The markers used in this experiment are presented in Table 4.3.

Table 4.3: *S* and *Z*-linked markers used in the genotyping of the mother plant F13 for *S* and *Z* loci. For the Z-locus linked markers, the primer sequences are not given as they are confidential and the markers were directly obtained from Dr Bruno Studer (ETH Zürich, Switzerland).

Oligoname	SI locus targeted	Forward sequence	Reverse sequence	PCR program
05_02720	*S*-locus	AGCCCAACAGCTATTCCAAG	TCTGGTTTTTGGTGAAGGTG	HiRes63
05_02790	*S*-locus	CATCGCCAGCATGCTTATAG	CCACTTGCTCTTCCTCTTCC	HiRes63
05_02827	*S*-locus	GCAGCAAGTGATGATGAAGG	AGCTGCTTCAACTTGCCATC	HiRes63
05_02833	*S*-locus	CTTGTTCCGTTTCGATTGGT	TTGTCTTTCCTGGCCTGTCT	HiRes63
05_02889	*S*-locus	GGACGTCAAGACAACCACTG	GGAGATTGTCTGTGCTTCGAG	HiRes65
05_02911	*S*-locus	AAGCGGGAGACGGTTGAG	ACCCGCTACGAGCCTGTG	HiRes63
05_02915	*S*-locus	CCGACAATTTCCCGTACACT	TTAGGAAAGATTGGGGTTGG	HiRes63
05_B35780.1	*S*-locus	ATTTCATGCCATTCCGAAAC	AGTCCTGGGCACCCATATC	HiRes63 / 65-55 touchdown
05_R0150300_1	*S*-locus	GGCTCTTGGACATCTTGGAA	CTGCCCTAGTCGAAAGCAAG	HiRes63
SBAC_175_r	*S*-locus	TGCAGCATACAACACCAACA	GAGCTTGGGATTGCAAAGAG	HiRes65 / 65-55 touchdown
BAC_573_001	*S*-locus	CAAATCGCGGAGGAGAGAC	AGCGGCGAACAGAAAGATT	65-55 touchdown
BAC_764_022	*S*-locus	GCAGTGATTGAAGGACACGA	TGCTTGTCAACACCTTCTGC	65-55 touchdown
LpUSP1	*Z*-locus	Confidential, markers provided by Dr Bruno Studer		65-55 touchdown
LpGK2	*Z*-locus	Confidential, markers provided by Dr Bruno Studer		HiRes63 / 65-55 touchdown
LpCadelp130	*Z*-locus	Confidential, markers provided by Dr Bruno Studer		HiRes63
OSB42350	*Z*-locus	Confidential, markers provided by Dr Bruno Studer		HiRes63
OSB39420	*Z*-locus	Confidential, markers provided by Dr Bruno Studer		HiRes63
BAC_HiBeg	*Z*-locus	Confidential, markers provided by Dr Bruno Studer		HiRes63
BAC_25-27	*Z*-locus	Confidential, markers provided by Dr Bruno Studer		HiRes63
5000	*Z*-locus	Confidential, markers provided by Dr Bruno Studer		HiRes63
6500	*Z*-locus	Confidential, markers provided by Dr Bruno Studer		HiRes63
10280	*Z*-locus	Confidential, markers provided by Dr Bruno Studer		HiRes63
12000	*Z*-locus	Confidential, markers provided by Dr Bruno Studer		HiRes63

4.2.2.3. Genotyping method using HRM curve analysis

The DNA of the mother plants F13 was amplified using PCR for each marker, following the protocol described in section 2.3.3 in chapter 2. The PCR program used, HiRes63 and HiRes65, are described in Table 2.5 (chapter 2, section 2.3.3). The amplified DNA was then analysed by HRM. The melting of the amplified DNA was done between 60 and 98°C using the 96-well LightScanner (Idaho Technology Inc., Salt Lake City, UT, USA) and the genotyping was done using the software provided, LightScanner® and Call-IT® (Idaho Technology Inc., Salt Lake City, UT, USA).

4.2.2.4. Genotyping method using electrophoresis by capillarity

For the markers genotyped for length polymorphism, the mother plants F13 DNA was amplified with the marker primers but the forward primer was fluorescently 5'-labelled with one of four fluorochrome moieties: 6-carboxyfluorescein FAM™ (blue; Metabion), VIC™ (green; Applied Biosystems, Foster City, CA, USA), NED™ (yellow; Applied Biosystems) and PET™ (red; Applied Biosystems). The amplification was done in a 7 µl volume reaction made out of 1 µl 10X PCR buffer (Applied Biosystems), 1 µl dNTPs (2 mM), 0.5 µl F primer labelled (10 µM), 0.5 µl R primer (10 µM), 0.05 µl AmpliTaq® Gold (5 U/µl, Applied Biosystems), 0.6 µl MgCl2 (25 mM) and 1.35 µl of sterilized deionized water with 2 µl of DNA. The amplification was then conducted using the 65-55 touchdown PCR program: 95°C for 10 minutes, followed by a touchdown (-1°C at each cycle, from 65 to 55°C) of 10 cycles (95°C for 1 minute, 65°C for 30 seconds and 72°C for 1 minute), an amplification of 25 cycles (94°C for 30 seconds, 55°C for 30 seconds and 72°C for 30 seconds) and a final amplification stage at 72°C for 10 minutes.

For the genotyping using electrophoresis by capillarity, 1 µl of the amplified products was mixed with 9.8 µl of formaldehyde and 0.2 µl of 500 LIZ (size standard; Applied Biosystems). The samples were denatured at 95°C for 5 minutes and loaded in the ABI3130 (Applied Biosystems, Foster City, CA, USA) to run the capillarity electrophoresis. The genotyping analysis of the ABI products was done with GeneMapper® (Applied Biosystems) software.

4.3. Results

4.3.1. *S*-allelic diversity

The results from the capillarity genotyping were obtained using the software GeneMapper® (Applied Biosystems). The allelic diversity, resulting from length polymorphism, of the four markers genotyped in the mother plants F13 is shown in Table 4.4.

Table 4.4: Allele sizes of the four *S*-linked markers genotyped using electrophoresis by capillarity. Each allele size has been called with a letter (a, b, c or d), that was used in the genotyping of the mother plants F13.

Marker name	Alleles size (bp)
05_B35780.1	183 (a) and 185 (b)
SBAC_175_r	243 (a) and 244 (b)
BAC_573_001	318 (a), 323 (b), 327 (c) and 337 (d)
BAC_764_022	314 (a) and 316 (b)

Only the marker 05_02827 appeared to be monomorphic in this population. The polymorphism detected for all the rest of the markers varies, from two groups for markers

05_02720, 05_02790, 05_02833, 05_02911 and 05_02915 (see Figure 4.1), to three groups (for markers 05_B35780.1 and 05_R0150300.1, see Figure 4.2) and up to five different melting profiles for the marker 05_02889 (see Figure 4.3).

The HRM curve analysis technique gave a melting profile for each plant and for each marker. However, a melting profile only corresponds to an overall genotype but is not able to discriminate allele combinations without *a priori* knowledge. For each marker, plants were grouped according to their melting profile and a letter was given as a "genotype" for each group. The Table 4.5 is showing the results of the genotyping of the mother plants F13 with *S*-linked markers, using both HRM method and electrophoresis by capillarity.

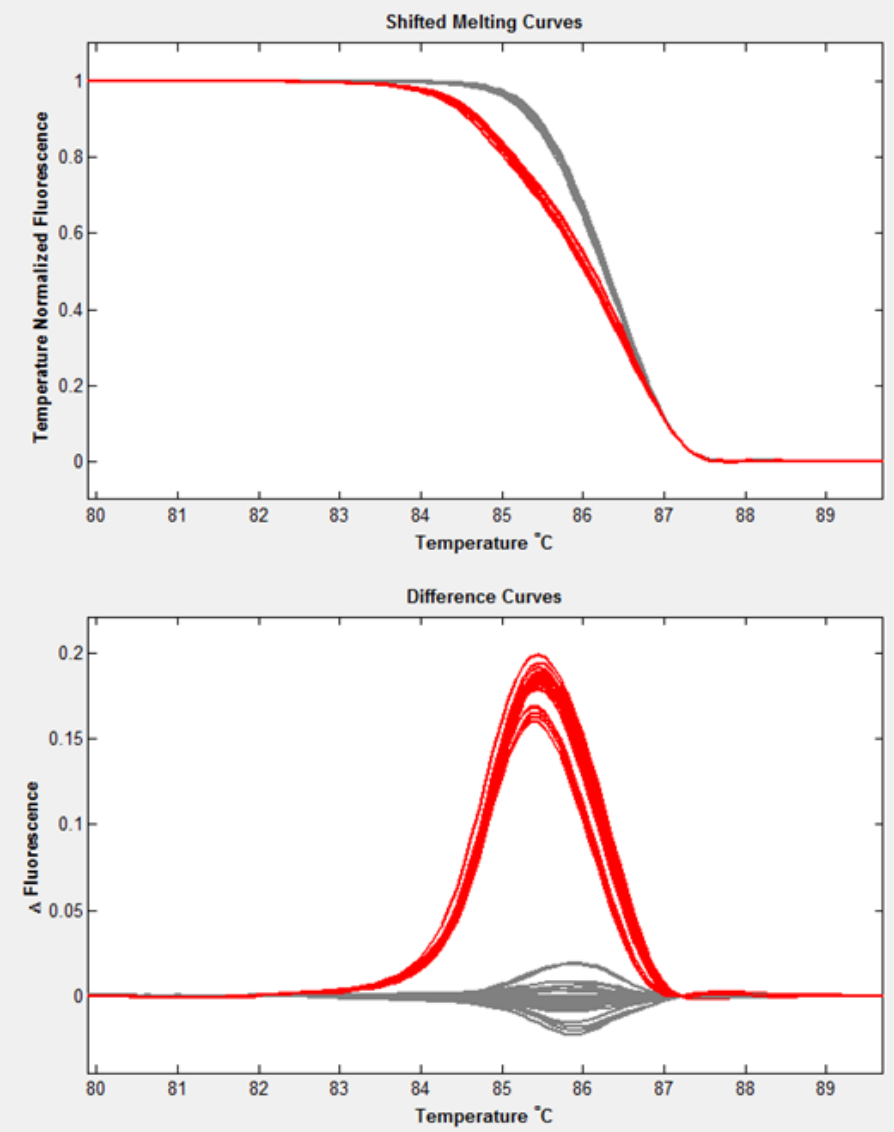


Figure 4.1: Melting profile of the mother plants F13 using the marker 05_02833 with the HRM method. This marker enables the classification of the plants into two groups, suggesting the plants with grey melting curves are homozygous and the plants with red melting curves are heterozygous.

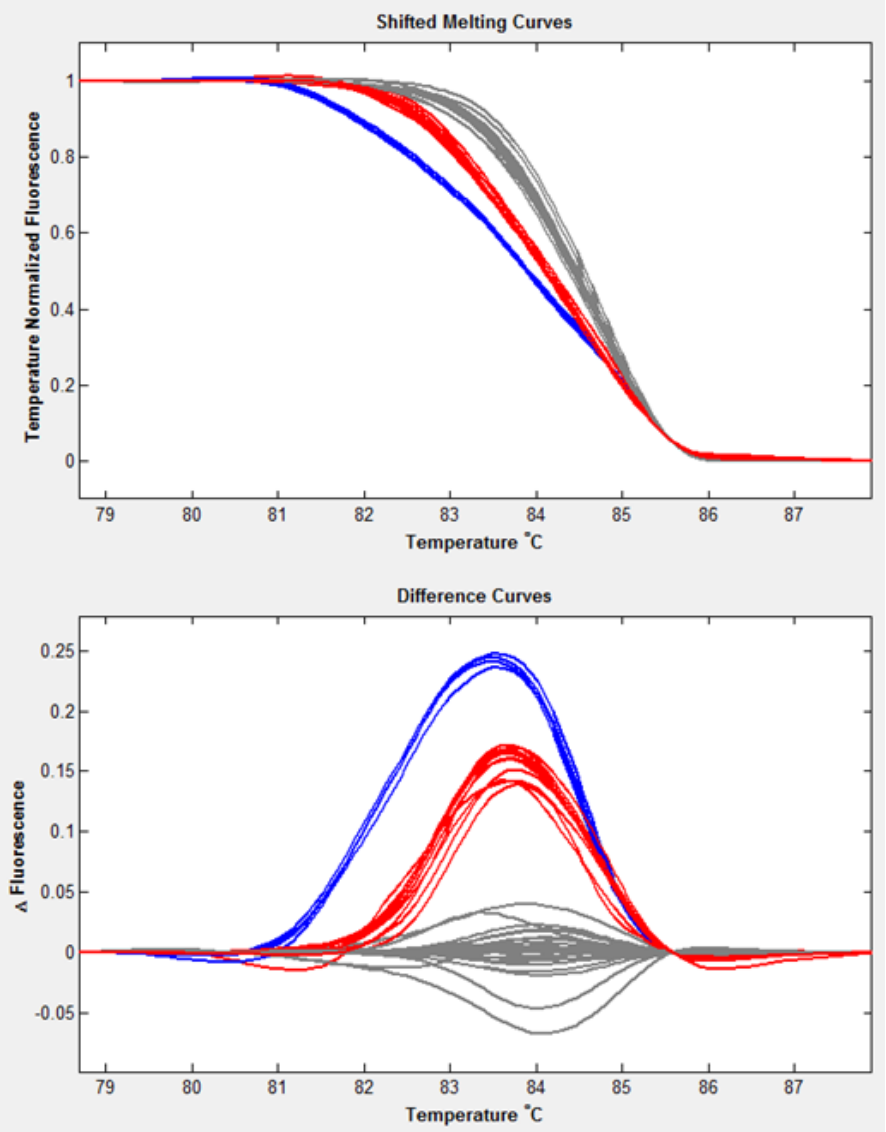


Figure 4.2: Melting profile of the mother plants F13 using the marker 05_R0150300 with the HRM method. This marker enables the classification of the plants into three groups, suggesting the plants with grey and red melting curves are homozygous but for different alleles and the plants with blue melting curves are heterozygous. However, the zygosity of the plants is uncertain as no control was used in this experiment, therefore, only the grouping can be used to classify the plants.

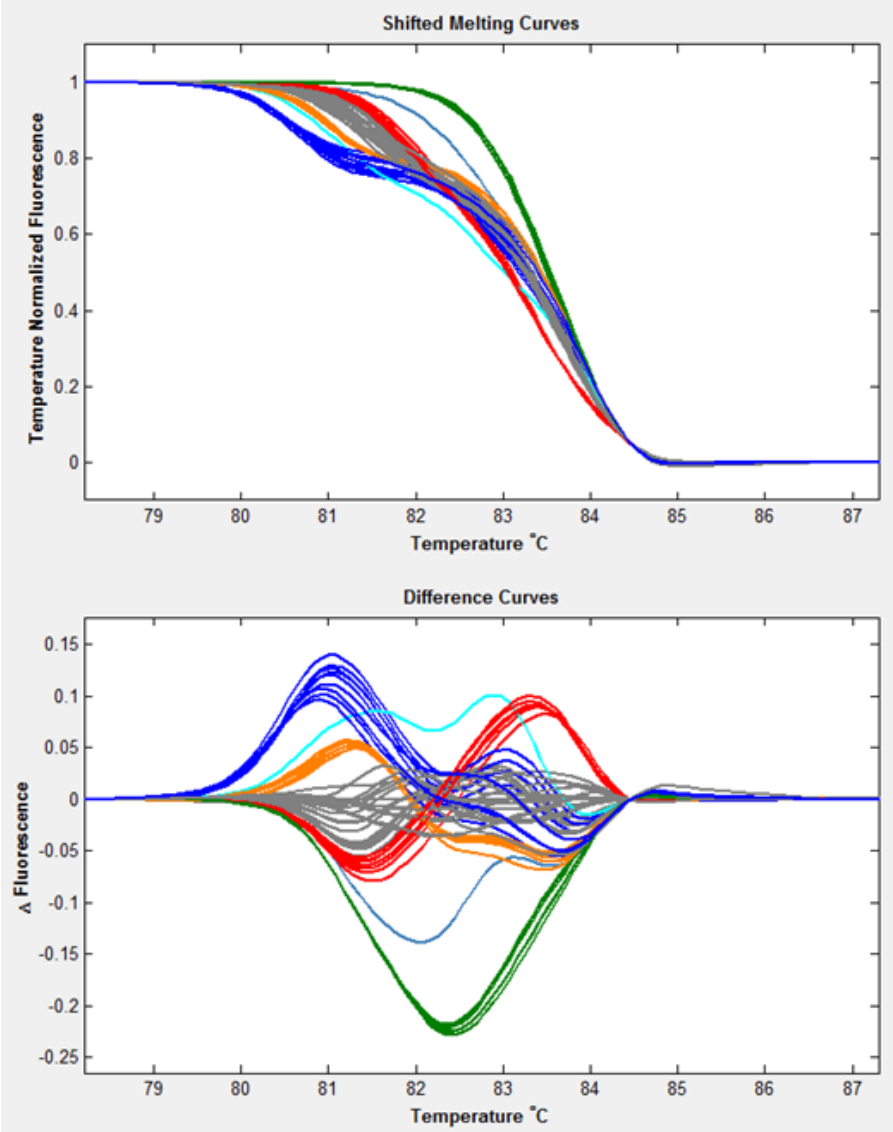


Figure 4.3: Melting profile of the mother plants F13 using the marker 05_02889 with the HRM method. This marker enables the classification of the plants into five groups: green, red, grey, orange and dark blue melting curves. Two additional curves were found with this marker, light blue and grey-blue, corresponding to either additional groups (made of one plant each) or misclassified plants.

Using two markers genotyped with electrophoresis by capillarity, 05_B35780.1_1 and BAC_573_001, putative *S*-alleles were given for each plant (see Table 4.5). Some data are missing in the table; however, the missing data for the marker BAC_573_001 seems to be allele specific so it has been ascribed as a null allele. Because the group created by these

two markers have sometimes been divided into several sub-groups, other alleles have been included.

In total, the fifty-five mother plants have been classified into 19 groups for the *S*-locus, of which eight groups are made out of a single plant, and eleven hypothetical *S*-alleles have been found grouped into 18 *S*-allele combinations.

Table 4.5: Results from the genotyping of the mother plants F13 for the *S*-locus, using STS markers scored with HRM and electrophoresis by capillarity (EC). The markers are placed in the table according to their position in the *S*-locus region. The classification of the plants into groups was made according to the haplotype combinations for all markers. The alleles are represented by letters and the melting curve profiles are represented by colors, white been missing data. The *S*-allele combinations were determined using the allele combination of two markers, BAC_573_001 and 05_B35780.1. (?) represents the missing allele for marker BAC_573_001 and (*) represents the *S*-allele combination were the profile of the plant genotype along all markers is different from the rest of the plants belonging to the same group.

Half-sib family	Plant number	Group	05_02720	05_02790	BAC_573_001 (EC)	05_02833	R0150300	BAC_764_022 (EC)	05_02889	SBAC_175_r (EC)	SBAC_175_r HRM	05_02911	05_02915	05_B35780.1 HRM	05_B35780.1 (EC)	Allele combination between markers BAC_573_001 and 05_B35780.1		Hypothetical *S*-allele combination
1	1	**1**			a					ab					c	a-c	?-c	2, 9
1	5	**1**			a					ab					c	a-c	?-c	2, 9
1	8	**1**			a					ab					c	a-c	?-c	2, 9
3	31	**1**			a			a		ab					c	a-c	?-c	2, 9*
4	46	**1**			a			a		ab					c	a-c	?-c	2, 9*
4	41	**2**			d					ab					bc	d-b	?-c	5, 9
4	52	**2**			d					ab					bc	d-b	?-c	5, 9
2	18	**3**			b			ab							bc	b-b	?-c	3, 9
2	12	**3**			b			ab							bc	b-b	?-c	3, 9
2	14	**3**			b			ab							bc	b-b	?-c	3, 9
2	15	**3**			b			b							bc	b-b	?-c	3, 9*
2	21	**3**			b			a							bc	b-b	?-c	3, 9*
1	4	**4**								ab					c	?-c	?-c	9, 9
1	11	**4**								ab					c	?-c	?-c	9, 9
3	36	**4**								ab					c	?-c	?-c	9, 9
2	16	**14**													c	?-c	?-c	9, 9*

Half-sib family	Plant number	Group	05_02720	05_02790	BAC_573_001 (EC)	05_02833	R0150300	BAC_764_022 (EC)	05_02889	SBAC_175_r (EC)	SBAC_175_r HRM	05_02911	05_02915	05_B35780.1 HRM	05_B35780.1 (EC)	Allele combination between markers BAC_573_001 and 05_B35780.1		Hypothetical *S*-allele combination
3	32	**5**			ab			b							b	a-b	b-b	1, 3
3	40	**5**			ab			b							b	a-b	b-b	1, 3
3	28	**5**			ab			b							b	a-b	b-b	1, 3
3	29	**5**			ab			b							b	a-b	b-b	1, 3
3	38	**5**																1, 3
3	25	**5**			ab			ab							b	a-b	b-b	1, 3*
3	37	**5**			ab			ab							b	a-b	b-b	1, 3*
1	6	**6**			c					ab					bc	c-b	?-c	7, 9
4	47	**6**			c					ab					bc	c-b	?-c	7, 9
4	51	**6**			c					ab					bc	c-b	?-c	7, 9
3	35	**6**			c					ab					bc	c-b	?-c	7, 9*
1	7	**6**			c			a		ab					bc	c-b	?-c	7, 9
4	42	**6**			c			a		ab					bc	c-b	?-c	7, 9
2	24	**16**			c			a							bc	c-b	?-c	7, 9*
1	2	**7**			a										bc	a-b	a-c	1, 2
3	30	**7**			a			b							bc	a-b	a-c	1, 2
3	34	**7**			a										bc	a-b	a-c	1, 2
4	44	**7**			a										bc	a-b	a-c	1, 2
4	48	**7**			a										bc	a-b	a-c	1, 2
4	50	**7**								ab					bc	a-b	?-c	1, 2*
4	54	**7**			ac										bc	a-c	c-b	2, 7

Half-sib family	Plant number	Group	05_02720	05_02790	BAC_573_001 (EC)	05_02833	R0150300	BAC_764_022 (EC)	05_02889	SBAC_175_r (EC)	SBAC_175_r HRM	05_02911	05_02915	05_B35780.1 HRM	05_B35780.1 (EC)	Allele combination between markers BAC_573_001 and 05_B35780.1		Hypothetical *S*-allele combination
2	17	**8**			b			ab		ab					bc	b-b	b-c	3, 4
2	20	**8**			b			a		ab					bc	b-b	b-c	3, 4
2	23	**8**			b			ab		ab					bc	b-b	b-c	3, 4
3	27	**9**			b			b		ab					bc	b-b	?-c	9, 8*
4	49	**9**			b			b		ab					bc	b-b	?-c	9, 8
4	43	**9**			b			b		ab					bc	b-b	?-c	9, 8
4	45	**9**			b			b		ab					bc	b-b	?-c	9, 8
3	39	**10**			a			a							c	a-c	?-c	9, 11
3	26	**10**																9, 11
4	55	**10**			a			a							c	a-c	?-c	9, 11
4	53	**11**			ac			a		ab					bc	a-c	c-b	2, 7*
2	22	**12**			bc			ab							bc	b-b	c-c	3, 6
1	9	**13**			ad					b					bc	a-c	d-b	2, 5
2	19	**15**			b			ab							b	b-b	b-b	3, 3
1	10	**17**			a					b					c	a-c	?-c	2, 9*
3	33	**18**			d					ab					c	d-c	?-c	9, 10
2	13	**19**			bd			a		b					bc	b-b	d-c	3, 10
1	3	**20**			a			a										1 or 2, ?

4.3.2. *Z*-allelic diversity

For this locus, the method of analysis was the same. Only two markers were showing some length polymorphism using the electrophoresis by capillarity, LpGK2 and LpUSB1. The allele diversity for these two markers in the mother plants F13 population is presented in Table 4.6.

Table 4.6: Allele sizes of the two *Z*-linked markers genotyped using electrophoresis by capillarity. Each allele size has been called with a letter (a, b, c, d, e or f), that was used in the genotyping of the mother plants F13.

Marker name	**Alleles size (bp)**
LpGK2	136 (a), 138 (b), 144 (c), 145 (d), 146 (e) and 147 (f)
LpUSP1	614 (a), 627 (b), 638 (c) and 651 (d)

Among the ten markers screened using the HRM method, the discrimination power varies, from two groups for markers OSB42350 and OSB39420 (see Figure 4.4), three groups for the marker LpCadelp130, four groups for the marker 5000, five groups for markers 6500, 12000 and BAC_HiBeg (see Figure 4.5), seven groups for markers LpGK2 and BAC_25-27 (see Figure 4.6) and up to 9 groups for the marker 10280 (see Figure 4.7).

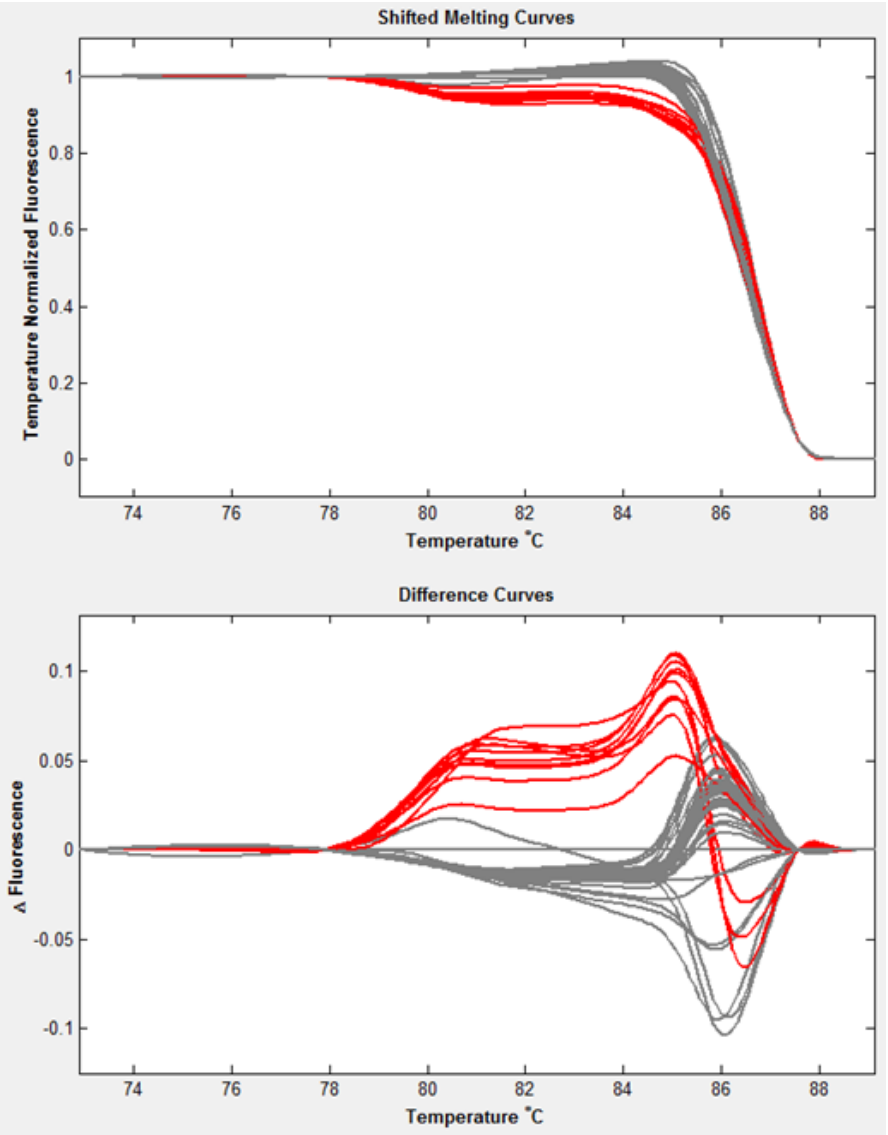


Figure 4.4: Melting profile of the mother plants F13 using the marker OSB42350 with the HRM method. This marker enables the classification of the plants into two groups, suggesting the plants with grey melting curves are homozygous and the plants with red melting curves are heterozygous.

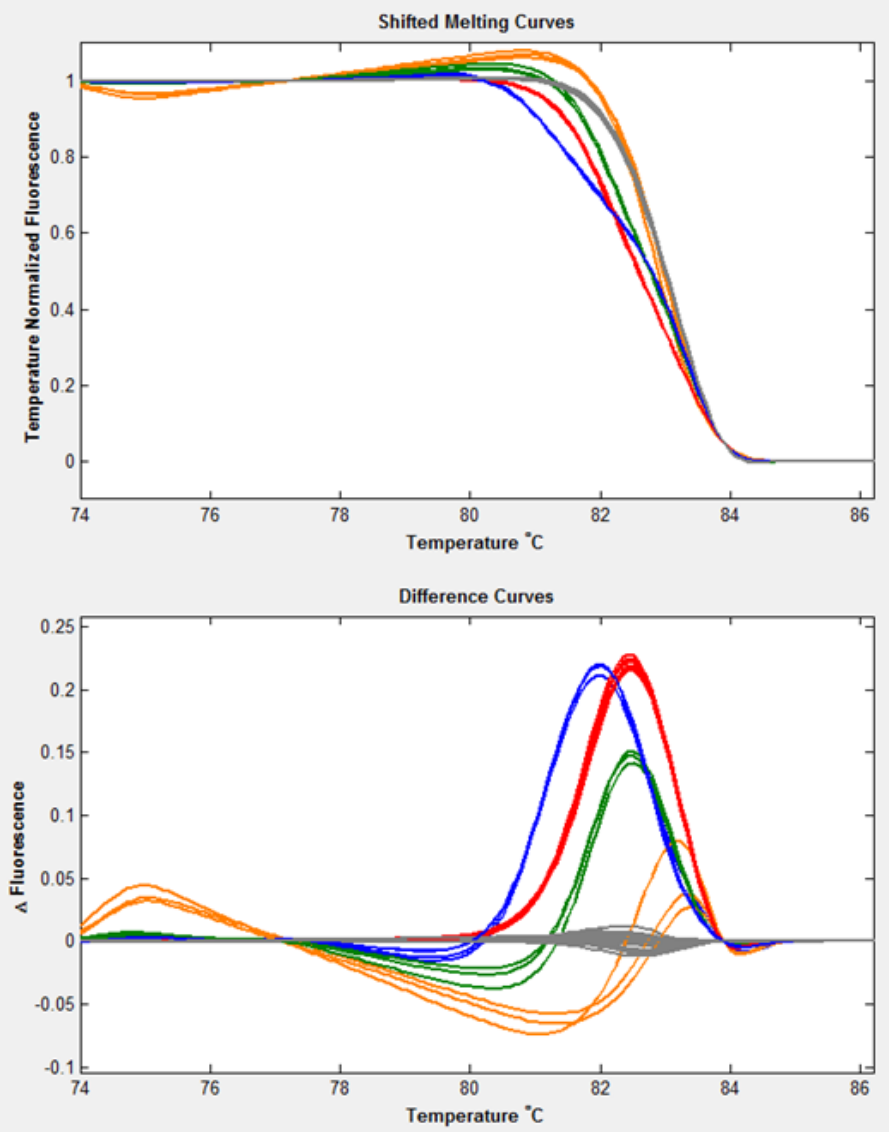


Figure 4.5: Melting profile of the mother plants F13 using the marker 12000 with the HRM method. This marker enables the classification of the plants into five groups.

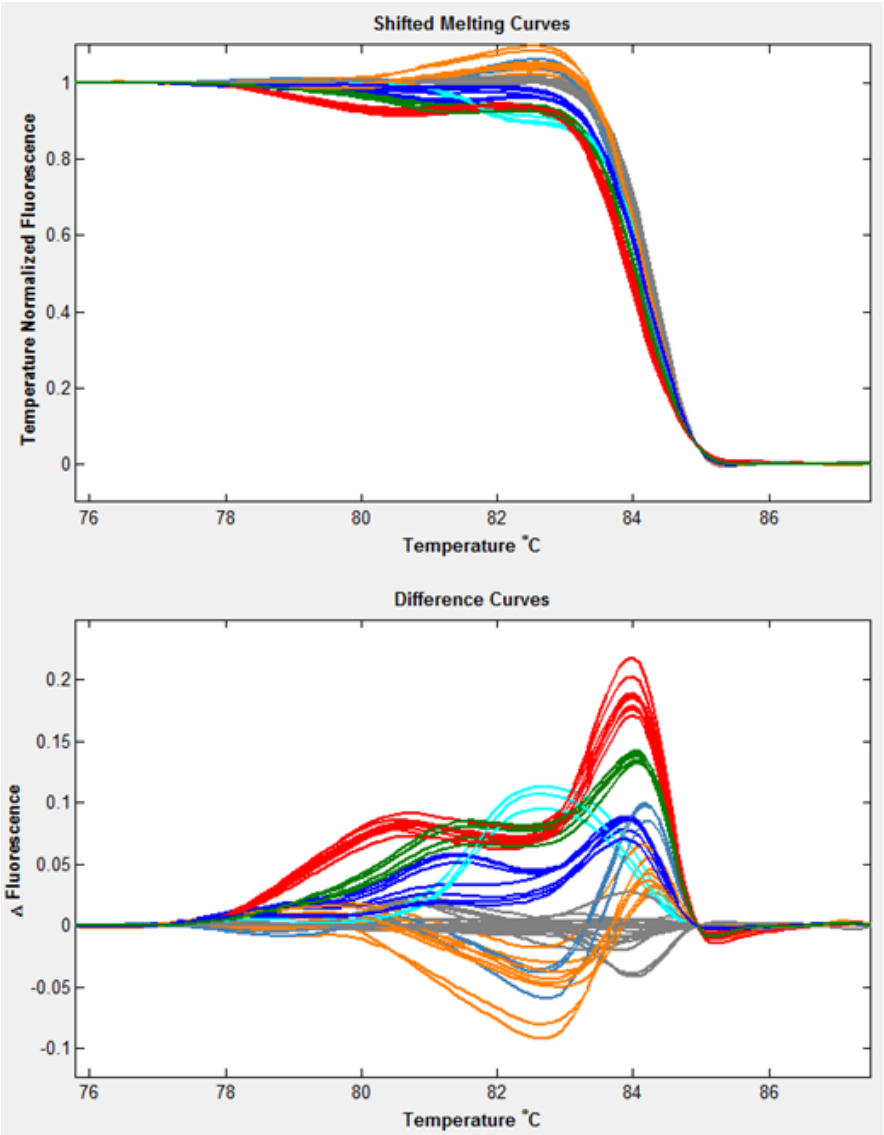


Figure 4.6: Melting profile of the mother plants F13 using the marker LpGK2 with the HRM method. This marker enables the classification of the plants into seven groups. These groups, according to the allele sizes with the electrophoresis by capillarity are a combination of 6 alleles.

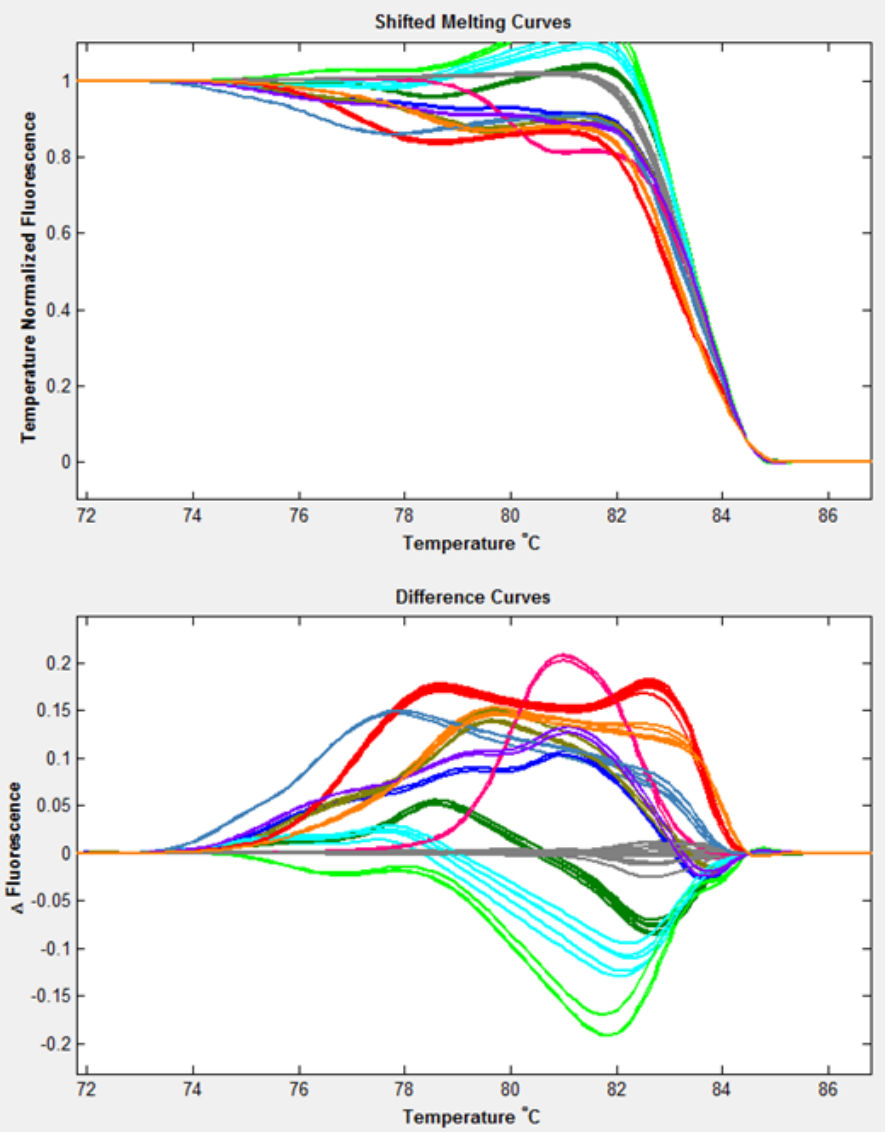


Figure 4.7: Melting profile of the mother plants F13 using the marker 10280 with the HRM method. This marker enables the classification of the plants into nine groups, easy to differentiate.

As for the genotyping with the *S*-linked markers, the mother plants F13 have been classified according to their genotypes and haplotypes and are presented in Table 4.7. Using the only genotypes available from the marker LpGK2 and LpUSP1, hypothetical *Z*-alleles have been given to the plants, with some variations according to the group formed with the rest of the markers, leading to 10 hypothetical *Z*-alleles, classified into 13 *Z*-allele combinations.

Table 4.7: Results from the genotyping of the mother plants F13 for the *Z*-locus, using STS markers scored with HRM and electrophoresis by capillarity (EC). The markers are placed in the table according to their position in the *Z*-locus region. The classification of the plants into group was made according to the haplotype combinations for all markers. The alleles are represented by letters and the melting curve profiles are represented by colors, white been missing data. The *Z*-allele combinations were determined using the allele combination of two markers, LpGK1 and LpUSP1. (*) represents the *Z*-allele combination were the profile of the plant genotype along all markers is different from the rest of the plants belonging to the same group.

Half-sib family	Plant number	Group	OSB42350	OSB39420	BAC_HiBeg	5000	6500	LpGK2 (EC)	LpGK2 HRM	10280	12000	LpUSP1 (EC)	BAC_25-27	LpCadelp130	Allele combination between markers LpGK2 and LpUSP1		Hypothetical *Z*-allele combination
3	25	**1**						a				d			**a-d**	**a-d**	1,1
3	27	**1**						a				d			**a-d**	**a-d**	1,1
3	28	**1**						a				d			**a-d**	**a-d**	1,1
3	37	**1**						a				d			**a-d**	**a-d**	1,1
4	49	**1**						a				d			**a-d**	**a-d**	1,1
4	52	**1**						a				d			**a-d**	**a-d**	1,1
1	11	**2**						ac				bd			**a-d**	**c-b**	1,2
3	31	**2**						ac				bd			**a-d**	**c-b**	1,2
4	47	**2**						ac				bd			**a-d**	**c-b**	1,2
3	29	**3**						ad				d			**a-d**	**d-d**	1,3
3	32	**3**						ad				d			**a-d**	**d-d**	1,3
3	35	**3**						ad				d			**a-d**	**d-d**	1,3
2	18	**4**						ae				ad			**a-d**	**e-a**	1,4
2	24	**4**						ae				ad			**a-d**	**e-a**	1,4
3	26	**4**						ae				ad			**a-d**	**e-a**	1,4
4	41	**4**						ae				ad			**a-d**	**e-a**	1,4
4	55	**4**						ae				ad			**a-d**	**e-a**	1,4

Half-sib family	Plant number	Group	OSB42350	OSB39420	BAC_HiBeg	5000	6500	LpGK2 (EC)	LpGK2 HRM	10280	12000	LpUSP1 (EC)	BAC_25-27	LpCadelp130	Allele combination between markers LpGK2 and LpUSP1		Hypothetical *Z*-allele combination
1	3	**5**						af				ad			**a-d**	**f-a**	1,5
1	5	**5**						af				ad			**a-d**	**f-a**	1,5
3	38	**5**						af							**-**	**-**	1,5
4	42	**5**						af				ad			**a-d**	**f-a**	1,5
4	44	**5**						af				ad			**a-d**	**f-a**	1,5
2	20	**6**						bc				b			**c-b**	**b-b**	2,6
2	23	**6**						bc				b			**c-b**	**b-b**	2,6
2	14	**7**						cd				b			**c-b**	**d-b**	2,7*
2	17	**7**						cd				b			**c-b**	**d-b**	2,7
4	48	**7**						cd				b			**c-b**	**d-b**	2,7
4	54	**7**						cd				b			**c-b**	**d-b**	2,7
2	13	**8**						ce				ab			**c-b**	**e-a**	2,4
2	15	**8**						ce							**-**	**-**	2,4
2	16	**8**						ce				ab			**c-b**	**e-a**	2,4
2	21	**8**						ce				ab			**c-b**	**e-a**	2,4*
2	22	**8**						ce				ab			**c-b**	**e-a**	2,4*
4	43	**8**						e				ab			**e?-b**	**e-a**	2,4*
4	45	**8**						ce				ab			**c-b**	**e-a**	2,4
4	50	**8**						ce				ab			**c-b**	**e-a**	2,4
4	51	**8**						ce				ab			**c-b**	**e-a**	2,4
2	19	**9**						cf				ab			**c-b**	**f-a**	2,5
4	46	**9**						cf				ab			**c-b**	**f-a**	2,5*
4	53	**9**						cf				ab			**c-b**	**f-a**	2,5
2	12	**9**						cf				ab			**c-b**	**f-a**	2,5*

Half-sib family	Plant number	Group	OSB42350	OSB39420	BAC_HiBeg	5000	6500	LpGK2 (EC)	LpGK2 HRM	10280	12000	LpUSP1 (EC)	BAC_25-27	LpCadelp130	Allele combination between markers LpGK2 and LpUSP1		Hypothetical *Z*-allele combination
3	30	**10**						de				a			**e-a**	**d-a**	4,8
3	36	**10**						de				a			**e-a**	**d-a**	4,8
3	33	**10**						de				a			**e-a**	**d-a**	4,8*
1	1	**11**						e				a			**e-a**	**e-a**	4,4
1	4	**11**						e				a			**e-a**	**e-a**	4,4*
1	7	**11**						e				a			**e-a**	**e-a**	4,4
1	8	**11**						e				a			**e-a**	**e-a**	4,4
1	10	**11**						e				a			**e-a**	**e-a**	4,4
1	9	**12**						e				a			**e-a**	**e-a**	4,10
3	34	**12**						e				a			**e-a**	**e-a**	4,10
3	39	**12**						e				a			**e-a**	**e-a**	4,10
3	40	**12**						e				a			**e-a**	**e-a**	4,10
1	2	**13**						e				ac			**e-a**	**e-c**	4,9
1	6	**13**						e				ac			**e-a**	**e-c**	4,9

4.4. Discussion

The possibility to genotype any plant for its self-incompatibility genes is important in a breeding programme, in order to estimate allelic diversity and pre-empt potential incompatibility problems. The SI genotype prediction is also important to design crossing schemes that maximise the occurrence of favourable heterotic combinations, for example in producing largely intra-incompatible populations but inter-compatible population combinations in F1 hybrid population schemes as proposed by England (1974). Genotyping would also open up opportunities to study the population genetics of SI in previously understudied grass species in more detail. In SI systems where the genes involved are known, genotyping is straightforward but in grasses, neither the *S* nor the *Z* genes are known. However, this does not mean that the genotyping is impossible and the prediction of compatibility or incompatibility can be done by using closely linked markers. From previous studies of the *S* and *Z* genes, *S* and *Z*-linked markers were selected in order to evaluate the diversity of their alleles as well as the potential of the HRM method for an easy and fast method to use.

The HRM method has been used in order to genotype a pair-cross population previously where genotypic variation is limited and predictable (see chapter 2) but not to genotype a more complex population where no *a priori* knowledge about the *S* or the *Z*-genes can be assumed. In order to assess the method, markers specifically developed for HRM (small product size and single polymorphism targeted) were selected but also tested for length polymorphism using electrophoresis by capillarity (EC), which, unlike HRM analysis enables allele calls. For the *S*-locus, a total of 11 markers were selected: 9 markers were screened using the HRM method, 2 markers were genotyped using EC and two markers were screened using both methods for a direct comparison of the results. As for

the *Z*-locus, 11 markers were also selected but only one marker (LpUSP1) was screened using EC and one marker (LpGK2) was screened with both methods.

From the comparison between the two methods, EC and HRM, the conclusions vary. For the *Z*-linked marker LpGK2, the comparison between EC and HRM is showing that the discrimination power of EC is higher than HRM, with 11 allele combinations identified using EC against 7 with HRM (see Table 4.8 **A**). However, the conclusions are reversed with the *S*-linked marker 05_B35780.1, were one allele combination (bc) from EC is divided into two groups with HRM (grey and red, see Table 4.8 **B**). The lower discrimination with HRM for the marker LpGK2 is probably due to the sensitivity applied when grouping the melting curves using the HRM software. A lower sensitivity using HRM might not discriminate every variation in the melting curves but it avoids the misclassification of the plants into separate groups and the overestimation of the diversity. Finally, in the case of the marker SBAC-175_r, the classification was identical between EC and HRM screening methods. However, a lot of the missing data from the EC screening did show good amplification as well as a clear melting curve using the HRM method, leading to the classification of a lot of the missing data into one single group. The reason for this could be due to the PCR conditions used for both methods, even though the annealing temperature used in the amplification previous to the EC was lower than the one for the HRM method (55°C against 65°C), so supposedly less specific.

Table 4.8: Evaluation of the discriminating power of HRM method using electrophoresis by capillarity (EC) to compare. **A** represents the comparison for the marker LpGK2 and **B** is the comparison for the marker 05_B35780.1.

A

Plant	EC	HRM
25	a	
27	a	
28	a	
37	a	
49	a	
52	a	
20	bc	
23	bc	
1	e	
4	e	
7	e	
8	e	
10	e	
9	e	
34	e	
39	e	
40	e	
2	e	
6	e	
29	ad	
32	ad	
35	ad	
14	cd	
17	cd	
48	cd	
54	cd	
30	de	
36	de	
33	de	

Plant	EC	HRM
18	ae	
24	ae	
26	ae	
41	ae	
55	ae	
3	af	
5	af	
38	af	
42	af	
44	af	
19	cf	
46	cf	
53	cf	
12	cf	
13	ce	
15	ce	
16	ce	
21	ce	
22	ce	
43	e	
45	ce	
50	ce	
51	ce	
11	ac	
31	ac	
47	ac	

B

Plant	EC	HRM
32	b	
28	b	
29	b	
19	b	
40	b	
25	b	
37	b	
1	c	
5	c	
8	c	
31	c	
46	c	
4	c	
11	c	
36	c	
55	c	
10	c	
33	c	
16	c	
39	c	

Plant	EC	HRM
52	bc	
6	bc	
35	bc	
7	bc	
42	bc	
24	bc	
2	bc	
30	bc	
44	bc	
54	bc	
53	bc	
9	bc	
18	bc	
12	bc	
21	bc	
17	bc	
20	bc	
23	bc	
27	bc	
43	bc	
45	bc	
22	bc	
13	bc	
41	bc	
14	bc	
15	bc	
47	bc	
51	bc	
34	bc	
48	bc	
50	bc	
49	bc	
38		
26		
3		

This study has also shown that only a small number of markers using HRM method for genotyping are needed in order to estimate the mate availability of a population. Indeed, in the case of the *Z*-locus, using only two markers genotyped with HRM, 10280 and 12000, the plants were classified into twelve groups, one less than the final classification. By adding the genotyping using the marker LpUSP1, the plants are classified into the final thirteen groups (see Table 4.9). For the *S*-locus, three HRM markers only are needed to classify the plants into the main ten groups (1-10): BAC_573_001, R0150300 and 05_02889.

Table 4.9: Classification of the mother plants F13 according to their *Z* haplotypes, using only two HRM markers, 10280 and 12000 and the marker LpUSP1. Plants are grouped in this table according to their *Z*-haplotypes, using only few markers. For the marker 10280 and 12000, the *Z*-haplotype scored using HRM has been given a capital letter, unlike for the LpUSP1, where the *Z*-genotype is written in small letters.

Plant	Group	10280	LpUSP1	12000
30	10	F	aa	E
33		F	aa	E
36		F	aa	E
1	11	I	aa	A
4		I	aa	A
7		I	aa	A
8		I	aa	A
10		I		A
9	12	A	aa	A
34		A	aa	A
39		A	aa	A
40		A	aa	A
13	8	G	ab	F
15		G		F
16		G	ab	F
21		G	ab	F
22		G	ab	F
43		G	ab	F
45		G	ab	F
50		G	ab	F
51		G	ab	F
12	9	H	ab	A
19		H	ab	A
46		H	ab	A
53		H	ab	A
2	13	A	ac	A
6		A	ac	A

Plant	Group	10280	LpUSP1	12000
18	4	C	ad	A
24		C	ad	A
26		C	ad	A
41		C	ad	A
55		C	ad	A
3	5	D	ad	B
5		D	ad	B
38		D		B
42		D	ad	B
44		D	ad	B
20	6	A	bb	A
23		A	bb	A
14	7	E	bb	D
17		E	bb	D
48		E	bb	D
54		E	bb	D
11	2	B	bd	A
31		B	bd	A
47		B	bd	A
25	1	A	dd	B
27		A	dd	B
28		A	dd	B
37		A	dd	B
49		A	dd	B
52		A	dd	B
29	3	A	dd	C
32		A	dd	C
35		A	dd	C

Using a large number of closely linked markers, the diversity for *S* and *Z* alleles has been estimated to be large in the mother plants F13. In this population, eleven *S*-alleles and ten *Z*-alleles have been estimated with potentially more due to some unique haplotype combinations. The unique haplotypes cannot be classified for sure into groupings as they

could be due to unique alleles or allele combination. From the eleven *S*-alleles, 18 *S*-allele combinations have been estimated, and from the ten *Z*-alleles, 13 *Z*-allele combinations have been estimated. The diversity at both *S* and *Z* is large even after twelve generations of half-sib families, generated from nine plants so with a maximum of 18 *S*-alleles and 18 *Z*-alleles. This high allelic diversity is in accordance to what has been observed and adopted as a general consent by many scientists (Wright, 1939; Lawrence, 2000): rare SI alleles are favoured in a population, as they do have a transmission advantage as they will more rarely lead to a SI pollination compare to a common alleles. Due to the advantage of rare alleles, a high allelic diversity is maintained in a large population, with an equal allelic frequency at equilibrium (Lawrence, 2000). However, the *S* and *Z*-alleles frequency is not homogeneous among the mother plants F13 population, with alleles such as S_9 and Z_4, present in 32 and 28 plants respectively. This over-representation of these two alleles could be due to the fact that they are present in the four half-sib family mothers; out of four mothers plants F12 used to produce the mother plants F13, four have the S_9 alleles and three have the Z_4 allele (see Table 4.10). The selection of the family to go to the next generation are made according to the genotype of the mother (half-sib populations) which explain why the alleles Z_9 and Z_4 are over-represented, Z_1 being also largely present in 22 plants out of 55. The rare alleles found in this population such as S_6, S_{10}, Z_6 and Z_9 would have been therefore transmitted through the pollen. Another factor that can change the distribution of *S* and *Z*-alleles is the selection applied to the breeding population. Indeed, plants are selected for a number of traits that could be associated, through linkage, with a specific *S* or *Z* allele. However, the uneven distribution is more likely due to the mother selection for the 13th generation. These findings are in accordance with many other studies, where the distribution of the alleles is even within a population, observation mainly done in SSI systems rather than GSI (Castric & Vekemans, 2004).

Table 4.10: *S* and *Z* genotypes of the four half-sib family mothers. The alleles have been identified as coming from the mother as either one of the two alleles for *S* and *Z* are present in the half-sib population.

Half-sib population (mother)	*S*-alleles	*Z*-alleles
1	S_2S_9	Z_1Z_4
2	S_3S_9	Z_2Z_4
3	S_1S_9	Z_1Z_4
4	S_2S_9	Z_1Z_2

By associating the *S* and *Z* genotypes given to the mother plants F13 (see Table 4.11), one plant (plant 4) were found to be homozygous for both *S* and *Z* loci. This suggests that this plant may have been produced by self-pollination. Another and more plausible explanation could be that the allele combinations are different to the one given, as annotated in the table with the "*".

Table 4.11: *S* and *Z* genotypes of the fifty-five mother plants F13. Plants are grouped according to their half-sib families. (*) is indicating that the allele combination is not certain, as some variation in the genotyping was observed compared to other plants with the sample alleles. Plant 4 and 10, from the half-sib family 1, are double-homozygous for *S* and *Z*, with could result from a self-pollination or miscall of the *S* or *Z* alleles.

Half-sib family	Plant number	Hypothetical *S*-allele combination	Hypothetical *Z*-allele combination
1	1	2, 9	4,4
1	4	9, 9	4,4*
1	6	7, 9	4,9
1	7	7, 9	4,4
1	8	2, 9	4,4
1	2	1, 2	4,9
1	9	2, 5	4,10
1	10	2, 9*	4,4
1	5	2, 9	1,5
1	11	9, 9	1,2
1	3	1 or 2, ?	1,5
2	15	3, 9*	2,4
2	16	9, 9*	2,4
2	18	3, 9	1,4
2	21	3, 9*	2,4*
2	24	7, 9*	1,4
2	13	3, 10	2,4
2	22	3, 6	2,4*
2	12	3, 9	2,5*
2	14	3, 9	2,7*
2	17	3, 4	2,7
2	19	3, 3	2,5
2	20	3, 4	2,6
2	23	3, 4	2,6
3	26	9, 11	1,4
3	33	9, 10	4,8*
3	36	9, 9	4,8
3	39	9, 11	4,10
3	30	1, 2	4,8
3	34	1, 2	4,10
3	40	1, 3	4,10
3	27	9, 8*	1,1
3	31	2, 9*	1,2
3	35	7, 9*	1,3
3	25	1, 3*	1,1
3	28	1, 3	1,1
3	29	1, 3	1,3
3	32	1, 3	1,3
3	37	1, 3*	1,1
3	38	1, 3	1,5
4	41	5, 9	1,4
4	43	9, 8	2,4*
4	45	9, 8	2,4
4	51	7, 9	2,4
4	55	9, 11	1,4
4	50	1, 2*	2,4
4	42	7, 9	1,5
4	46	2, 9*	2,5*
4	47	7, 9	1,2
4	49	9, 8	1,1
4	52	5, 9	1,1
4	44	1, 2	1,5
4	48	1, 2	2,7
4	53	2, 7 *	2,5
4	54	2, 7	2,7

The aim of this study was to genotype plants from a population for their S and Z loci, using molecular marker. Using HRM, a fast, cheap and simple method, molecular markers were scored for each locus, in 55 plants from a breeding population. This study has shown that it is possible to use HRM with some closely linked SI markers in order to estimate the allelic diversity and the mate availability of the IBERS breeding population. However, in order to fully assess the method, more work is needed. Indeed, in the classification of the plants, and especially regarding the *S*-locus, some plants could not be grouped due to variation in the haplotype pattern. An example is plant 38 for the *S*-locus; the plant was classified as belonging to the group 5, with alleles S_1S_3 even though the haplotype of the plant for the marker 05_02720 was different to the other plants within this group. This could be explained by a recombination event of this marker, mapped previously at approximately 1.9 cM. However, recombination events are rare for the markers used as they are closely linked so not all the haplotype combinations can be explain by it. Another explanation for some unique groupings could be the contamination of the population with some pollen. The ingression of some external germplasm could lead to some extra alleles in the population or some misclassification. However, if there was contamination with external germplasm, both *S* and *Z* alleles could be expected to be unique.

In order to fully assess the method used in this study, sequencing of the marker products is required. By comparing the nucleotide sequence of a few markers, within a group and between groups, the classifications can be confirmed and the unique groups could be reorganized into existing groups or confirmed as unique. Semi *in-vivo* pollinations have also to be made in order to confirm the group classifications of

genotypes. By scoring the compatibility phenotype of plants belonging to the same group and different groups, the *S* and *Z* haplotypes can be confirmed.

Chapter 5:

Study and mapping of self-compatibility

5.1. Introduction

5.1.1. Aims of the studies

Segregation distortions have been observed in different *Lolium* populations (Thorogood *et al*., 2002, 2005) and have led to the suggestion of two modifier loci involved in the self-incompatibility system. Starting from these studies, the main aim of these projects is to better understand the breakdown of the self-compatibility mechanism in *Lolium perenne*, by gaining better characterization of these loci using a mapping genetics approach.

In order to do so, the first part of the project is an initial mapping of the two loci, *F* and *T*, by using two specifically designed populations segregating for one of these loci only. Using the same mapping strategy as the one used for the *S*-locus in Chapter 2, the aim is to isolate a small genetic region for both loci. By using the results from Thorogood *et al*. (2002, 2005), markers on LG 3 and 5 will be selected and screened on the *F* and *T* populations respectively, to look for segregation distortion. Then, depending on the number of recombinants found, new markers will be selected or designed in order to reduce the region around the locus.

The second part of this project to understand if the segregation distortion observed on LG 3 could be due to the self-compatibility (SC) characterize in this population. The self-compatible character was analysed using self-pollination tests in order to 1) phenotype the SC character in the population; 2) study the segregation of SC in this population; and 3) observe the site of action of the SC. This study was done on using the *F*-locus and the F-population as nothing is known about the role of the *F*-locus in self-compatible reaction or the stage of activation at which the locus enables self-pollination.

By observing pollen-tube growth and determining proportions of compatible pollen through semi *in-vivo* pollination, it is possible to phenotype the self-compatibility status of individuals of an F2 population. Segregation patterns can be observed which can then be associated with marker segregations to determine linkage relationships. The ultimate outcome of a successful pollination is fertilisation of ovules and the formation of seed such that seed setting ability may well be associated with pollen compatibility. Thorogood *et al*. (2002), although proposing an additional locus (*F*) associated with the SI response, were still able to classify their pollinations based on a two locus (*S*/*Z*) model suggesting that the third locus did not affect the stigmatic inhibition associated with the S/Z system and that it may act at a different stage of the reproductive process. So, the observations of the incompatibility/compatibility may help to understand at which stage the *F*-locus influences the reproductive process. By looking at the pollen tube growth at different time after self-pollination, the site and mode of action of the *F*-locus might be better understood. Self-incompatibility in grasses is not known at a molecular level but the reaction has been observed under microscope for many years. Many experiments are reported in the literature to look at stigma-pollen interaction and pollen germination: stigmatic printing in *Papaveraceae* (Franklin-Tong *et al*., 1988), liquid media for pollen growth in *Nicotiana alata* (Gray *et al*., 1991). In this study, we are using a method developed by Lunqvist (1961) in which the pistils are placed on a solid media and the pollen is sprinkled on top of it.

5.1.2. Self-compatibility in *Lolium perenne*

Lolium perenne is a self-incompatible species, ensuring cross-pollination in order carry out successful fertilisation and seed development. However, this mechanism is imperfect and some plants have the ability to reproduce by self-pollination, making them self-compatible.

Self-compatibility (SC) has been reported in many SI species. The source of the SC can vary, depending on the SI mechanism, the species and even the population. The main source of SC however is from the mutation of a key gene in the SI system (de Nettancourt, 1977). Many mutations have been reported in the single *S*-locus SI system but the mutations are specific to populations and species. In *Prunus* species, a SI species governed by a gametophytic S-RNase system, Socias i Company *et al*. (2012) have reviewed different mutations affecting the pollen SFB or the pistil *S*-RNase. In sour cherry, several mutations in the coding sequence of *S*-RNase have been associated with a low RNase expression or activity, leading to a partial or complete SC (Tsukamoto *et al*., 2008). As for the pollen SFB gene, numerous mutations, artificially produced or naturally occurring, have been found in different *S*-alleles, leading to SC (Sonneveld *et al*., 2005; Ushijima *et al*., 2004). The mutations in the *S*-locus are important in evolution as exemplified by the self-compatibility of *Arabidopsis thaliana* resulting from a mutation in a gene, PUB8, encoding a U-box-containing protein linked to the *S*-locus and responsible for the regulation of SRK transcript levels (Liu *et al*., 2007).

In the grass SI system, even if the SI genes are unknown, using self-fertile populations, mutations in the *S* and *Z* genes were identified as linked to the self-fertility found in rye (Voylokov *et al*., 1993 and 1998). In *Pharasis coerulescens* also mutations on

the *S* and *Z* loci were found to be linked to the loss of self-incompatibility, either from pollen-only mutants or pollen-pistil mutants (Hayman and Richter, 1992). However, Hayman and Richter (1992) also found that sixteen of their mutations causing self-fertility were linked to LG 5. This locus, named *T*, is not one of the key genes in the recognition of self-pollen; however, when mutated, the SI is partially or fully inactivated, without any changes in the *S* or *Z* loci. The same observations were made in rye (Fuong *et al*., 1993; Voylokov *et al*., 1993), with self-fertility linked to a locus on LG 5, the *S5* locus, which is probably homologous to *T*.

In perennial ryegrass, a self-fertility locus on LG 5 has also been identified, probably homologous to the *T*-locus from rye and *Phalaris coerulescens*, because of the segregation distortion observed in a mapping family (Thorogood and Hayman, 1991; Thorogood *et al*., 2005). The segregation of this self-fertile locus is independent to both *S* and *Z* loci. Indeed, Thorogood *et al*. (2005) have observed a segregation ratio 1:1 leading to the conclusion that the *T*-locus is a single gene, with two alleles (*T* and *t*), one of which is not transmitted by the pollen (*t*).

Finally, an additional segregation distortion locus has been reported in *Lolium perenne*, named the *F*-locus (Thorogood *et al*., 2002). There is evidence of association between the *S*-locus and a locus on linkage group 3 linked to the isoenzyme glutamate oxalo-acetate transminase GOT/3 (Thorogood *et al.*, 2002). However, the *F*-locus has not been reported elsewhere.

5.2. Mapping of the *T* and *F* loci

5.2.1. Plant populations

5.2.1.1. *T*-locus

The family used in this mapping is an F3 population resulting from a self-pollination of a half-self-compatible F2 plant (code 02-05) from which marker information was demonstrated to be heterozygous at the *T*-locus (*Tt*), exhibiting a half-self-compatible pollination genotype on semi *in-vivo* pollination. The *t*-allele is not transmitted and the progeny would be expected to segregate into *TT* and *Tt* individuals with linked DNA markers showing subsequent distorted segregation favouring one homozygote over the other with the degree of distortion dependent on the recombination distance between marker and the *T*-locus (see figure 5.1).

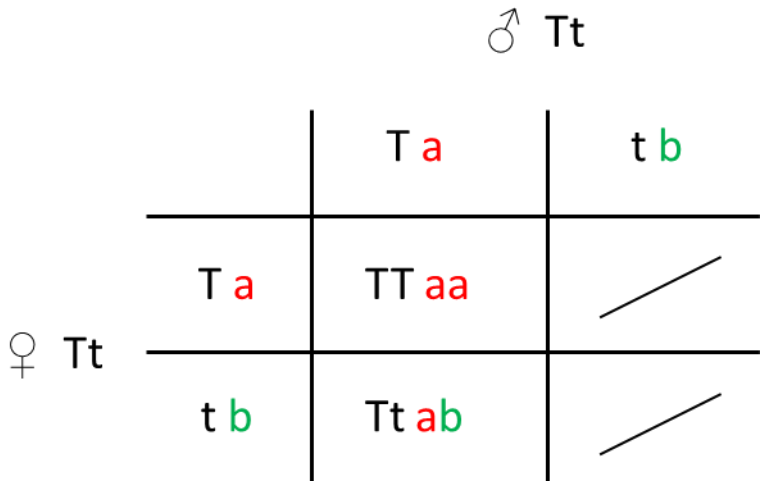


Figure 5.1: Segregation of the *T*-locus with a linked marker in the mapping population 0205. The population is resulting from a self-pollination of a heterozygous plant 0205. A closely linked marker M, with two alleles a and b is segregating with the T-locus.

The F2 population from which this mapping family is derived has been used extensively for mapping a range of agronomic traits, most notably, water soluble

carbohydrate content (Turner *et al*, 2006). In total, 167 plants from the *T*-populations have been screened in order to map the *T*-locus.

5.2.1.2. *F*-locus

The population used in this semi *in-vivo* assay is a *Lolium perenne* F2 population, the "F2 biomass" population described by Anhalt *et al.* (2008). This F2 biomass population results from a cross between two *Lolium* inbred lines, obtained from the cytoplasmic male sterility (CMS) programme in Teagasc (Oak Park) developed by Dr V. Connolly. Originally, several inter-specific cross was done between meadow fescue (*Festuca pratensis*) and perennial ryegrass. The hybrids were back-crossed to their ryegrass parent several times and selfed for nine or ten more generations, creating the inbred lines. The two inbred lines used to produce the F2 biomass population result from inter-specific crosses using two different ryegrass cultivars. The mother plant results from a cross between *Festuca* and the ryegrass cultivar S24 (IBERS, Aberystwyth University) whereas the father plant ryegrass parent is the cultivar Premo (Mommersteeg International BV). The F1 resulting from the cross between the two inbred lines was then selfed to obtain the F2 biomass population.

For this initial mapping of the *F*-locus, 96 plants from the F2 biomass population were screened and genotyped.

5.2.2. Methods of screening

5.2.2.1. DNA extraction

The plant DNA from both *F* and *T* populations was extracted using the DNA 96 Plant kit (Qiagen, Hilden, Germany). The extraction was done following the manufacturer instruction and described previously in the chapter 2 (section 2.3.1.1).

5.2.2.2. Primer selection and design

The initial markers selected for both *F* and *T*-locus study were SSRs markers from ViaLactia Biosciences (Auckland, New Zealand). The markers were selected at random to cover the linkage group 3 and 5 for the locus *F* and *T* respectively. The Table 5.1 shows the marker primer code used in the initial screening of the *T* and *F* populations.

Table 5.1: ViaLactia SSRs markers used in the initial screening of the *F* and *T*-locus. The sequences of the oligonucleotides are not shown in the table as they are under licence use. For each marker, the table describes the linkage group it is mapped onto, the annealing temperature (Ta), if the marker is specific and polymorphic and if so, the alleles size in the studied population.

Name	Linkage group	Ta	Polymorphism	Allele sizes (bp)
RV0154	LG3	60	No	134
RV0629	LG3	60	No	213
RV1063	LG3	60	No	135
RV1131	LG3	60	Yes	181, 201
RV1133	LG3	60	Yes	131, 154
RV1144	LG3	60	No amplification	/
RV0054	LG5	59	Yes	107, 115
RV0116	LG5	61	Yes	282, 299
RV0184	LG5	61	Yes	210, 224
RV0250	LG5	59	Yes	242, 250
RV0471	LG5	59	Yes	110, 152
RV0495	LG5	61	Yes	98, 105
RV0757	LG5	59	Yes	218, 224
RV0814	LG5	59	Yes	185, 205
RV1021	LG5	61	No	154
RV1030	LG5	61	Yes	152, 177
RV1159	LG5	59	No	212
RV1188	LG5	60	Yes	202, 208
RV1307	LG5	61	No amplification	/

Additional published SNP markers were selected for both loci. Regarding the *F*-locus, Thorogood *et al*. (2002) have found that the region around the markers CDO920/WG889 map on linkage group 3 in *Lolium perenne* showed linkage disequilibrium with markers in the region of the *S*-locus despite not being physically linked. Therefore, five markers close to the markers CDO920 (Van Deynze *et al*., 1998) were selected from the map developed by Studer *et al*. (2008) and used to screen the F2

biomass population (see Table 5.2). From this same published map (Studer *et al*., 2008), five markers were selected from LG 5 for the *T*-locus study and are listed in Table 5.2.

Finally, from the screening with the published markers, new markers were designed for both loci using the same method as described in chapter 2 for the *S*-locus. From the initial screening, a rice region was isolated (from chromosome 1 for the *F*-locus and from chromosome 12 for the *T*-locus) and *Lolium perenne* cDNA reads (Aarhus University, Denmark) were aligned to it using the CLC Genomics Workbench software (CLC bio, Aarhus, Denmark). Markers were designed using Primer3 software (http://frodo.wi.mit.edu/primer3/, Rozen and Skaletsky, 2000), targeting SNPs in the nucleotide sequence common between *Lolium perenne* and rice. Markers for the *T*-locus mapping are detailed in Table 5.3 and Table 5.4 lists the markers for the *F*-locus mapping.

Table 5.2: Published markers from Studer *et al*. (2008 & 2012) selected for the screening of the *T* and *F* loci. For each marker, the table contains the linkage group the marker is mapped onto, the accession number, both forward and reverse sequences, the PCR program used and if the marker is polymorphic.

Name	Linkage group	Accession number	Forward sequence	Reverse sequence	PCR program	Polymorphism
G03_089	LG 3	ES700225	TCACCAACACCACACTCCTC	GCTGCTCGTACTGCTCGTAG	HiRes63	Not good
G04_057	LG 3	ES699604	GTGCTCCTCCTTGTCTTTCG	ACATTGAGCCATCCAGCATT	HiRes63	No
G04_067	LG 3	ES699747	TGGCAGTACAAATAGCTGAACG	TTGCAGGTTTCTGTGCACTC	HiRes63	No
G07_071	LG 3	ES699543	CATCTATGCTCCTTGTCCACTCT	AACGAATCCGATCGAAACAG	HiRes63	Yes
LpRGA5	LG 3	/	GGGATAAGTTGGAGGCTGAA	ATCCCGACCGAAATTGATCT	HiRes63	Yes
G01_033	LG 5	ES699868	ATTAATCGCCCTCGTTCAAA	AAAATCCACGGGTCACACAT	HiRes63	Yes
G01_045	LG 5	ES699707	TGCTAAAACAGCAGCCAATG	TTGAACATGGGAATGCACTG	65-55 / 60	No
G01_095	LG 5	ES699091	CACTCTTCCTCCCGGATCA	GGATCCTCATGTCGATGTCC	65-55 / 60	Yes
G04_043	LG 5	ES699476	TCAAACCCACCACTTTGACA	CAGGCTCTCCTCCTCCTTCT	65-55 / 60	No
G05_044	LG 5	ES700454	GACCGATTGGAACCAACAAC	CGATGCTTTCAGCGGTTAAT	HiRes63	No
G05_065	LG 5	ES699367	ATGAAGATCGCCACTCACCT	TGCTGCTGCTACCGTTCTTA	65-55 / 60	Yes
G05_071	LG 5	EY458003	CTGGTCCGGTCAGTCAGTCT	TTTGCTGGACAAACTGCAAC	HiRes63	Yes
G06_096	LG 5	EY458028	GATCTTGCAGGCCGTCTAAC	GCCAGCGTCTTTATTTAGGC	HiRes63	Yes
G07_056	LG 5	ES699421	CAAAGAAGTCACGCACCAAA	GCTGGTGTAGCAGATGAGCA	HiRes63	No
PTA_394	LG 5	/	GCGACCGGAGATGTTCCTT	TTGTCGAGGCCGAGGAAG	HiRes63	No
PTA_845	LG 5	/	ACCAGAAGCTTTGCAATCGT	GCAAAAGGGATCTGAATGGA	HiRes63	No
PTA_1953	LG 5	/	TACAAGACGCTGAGGGACAG	CTTGTCGAAGAGGACGAAGG	HiRes63	Yes
PTA_2547	LG 5	/	CATGGCAGGCATGAGGTT	TCCTCCAACGAGGCACATA	HiRes63	Yes
gsa_001d	LG 5	/	ACGGACTGCGTTGAGACATT	ACCCTCATGGAGTGCTTCAG	HiRes63	No
rg1_012d	LG 5	/	GTTCCCAGAACCACAGCAAG	ATCTCACTAGAAACCTGGAG GTACA	HiRes63	No

Table 5.3: Markers for the *T*-locus mapping, designed on *Lolium perenne* cDNA reads aligned with the rice region 17 to 24 Mb of the chromosome 12. The name of the marker depends on its position on the rice physical map: 12 indicates the rice chromosome 12 and the following number indicates the position on the chromosome (e.g. 17432 meaning 17.432 Mb). For each marker, the table contains the forward and reverse sequence of the primer, the annealing temperature (Ta), the PCR program used and if the marker is specific and polymorphic.

Name	Forward sequence	Reverse sequence	Ta	PCR program	Polymorphism
12_17432	CAGGGTGGGCAGAGGAG	CGTCGCCTTCCAGTACCC	59	HiRes63	Not specific
12_18020	GTCACCGGCAGCGTCAG	GTCCGGTCGGGGTTCAG	58	HiRes63	No
12_18959	AACTTCGAAAAGGCCAGACA	CGACTCAGCAAGGAACATGA	61	HiRes63	No
12_19503	AGGAGGATCGTCGGAGATGT	CTGCGGGTAGGTCTTGGAG	60	HiRes63	Not specific
12_19557	CCCCAACAACTCAGATGTCC	CAATCCTTGATGGGCAGATT	59	HiRes63	Yes
12_19937	GGCAGGTGTAGATGAAGACGA	CATGAGCGGCATAGTCACC	60	HiRes63	Yes
12_20049	AGTGGTGCTGCACACATTG	CACCAAGCCAGATGATGATG	59	HiRes63	No
12_20543	TCCCATGACGACTAGGAACG	GCACGTTCCACCACTTCTC	59	HiRes63	No
12_20965	TAAGACGCTCGGCCTGATTT	GACGCCGTAGAGAATTGAGA	60	HiRes63	No
12_21248	AGGACGACTCACAGCAATCC	GCAAAGCTGTGCTTACCTCA	58	HiRes63	Yes
12_21463	GCAGTCTCCAAGAGATGCTCA	ATGCGATAATCCAGGCAAAG	60	HiRes63	No
12_21695	TTCCCGGTGCATCTTTTACT	TGTTTGCTTGTTTGCTTCCA	59	HiRes63	Yes
12_22211	TGCTTCGCTGAACTTTTGTG	CTCTTCACACAGCCAACGAG	59	HiRes63	Not specific
12_22465	ACGACGACGAGGCGGTA	GTCGGGGTGGACGAGAG	60	HiRes63	Yes
12_22744	TGGCAAGGTACCATGAGTG	GCCGCGATGAACAAGAAGT	59	HiRes63	Yes
12_23003	GACCCTTCTCAAGCATGTCA	CGAGCGTGGTGAAGGAGAT	59	HiRes63	No
12_23383	TCCGTCTTCAAGTCCCACAC	GAGCTCCTTCTGGTGCTTGT	60	HiRes63	Not specific
12_23502	AACCAGGAATCTGCTCATCC	GGAAACGGGTCACCAAGAA	61	HiRes63	No

Table 5.4: Markers for the *F*-locus mapping, designed on *Lolium perenne* cDNA reads aligned with the rice region 25 to 28Mb of the chromosome 1. The name of the marker depends on its position on the rice physical map: 01 indicates the rice chromosome 1 and the following number indicates the position on the chromosome (e.g. 25259 meaning 25.259 Mb). For each marker, the table contains the forward and reverse sequence of the primer, the annealing temperature (Ta), the PCR program used and if the marker is specific and polymorphic

Name	Forward sequence	Reverse sequence	Ta	PCR program	Polymorphism
01_25259	GTTGCTGTACGGCACGAC	CCGGCGAGCATAGTGTACTT	59	HiRes63	Not specific
01_25610	GGTCGTGTCCGACAGTTTCT	CCTCCACGGGGTTGTTG	60	HiRes63	No
01_26028	ATGGACTTCATGGGCTGGA	CGTCCCACCTCGACCTC	60	HiRes63	No
01_26395	ATCAGCCGCCGCTTCTC	GCACACGGGCCACTACA	60	HiRes63	No
01_26807	CTGAATGCTTCTCGCATCAA	AGCTCCTTCATTGCATCCTC	58	HiRes63	No
01_27035	TCGAGAAGACGGAGTTCGAT	CTTGGCCTCCTTCAGAGC	61	HiRes63	No
01_27295	ACCACACGGACGCCAAC	AAGGCCCTGGCCTCCTC	60	HiRes63	No
01_27495	GATGGAGAGGATGCTTGAGG	CCGAGGAGGAACTTTGTCTG	59	HiRes63	Yes
01_27732	GCTGCGAGAAGTGATAGGG	TGCTGCAAGAAAACCATACC	60	HiRes63	Yes
01_27917	ACTGGAGACCATCGCAAGTC	CAGGTTCAGCATTTGGAACA	58	HiRes63	No

5.2.2.3. Screening

Three methods were used to genotype the plants from both *T* and *F* populations: electrophoresis by capillarity, electrophoresis using acrylamide gels and HRM curve analysis. Electrophoresis was used to score SSRs markers as well as some SNPs (InDel) whereas HRM was used for the markers designed from *Lolium perenne* cDNA reads.

The initial screening for both *F* and *T*-locus mapping studies was done using markers from ViaLactia Biosciences, and then publically available markers (Studer *et al*., 2008 & 2012). Primers were tested for their polymorphism using few plants from the mapping populations as well as their parents: 0205 plant for the *T*-locus and the grand-parents as well as the parent of the F2 biomass population for the *F*-locus. The DNA was amplified in a 10µl PCR reaction containing 1µl of DNA (approximately 20ng), 0.25 unit DreamTaq DNA Polymerase (Fermentas), 1X DreamTaq Buffer, 200µM dNTPs, 0.2µM of each forward and reverse primer, with the forward primer fluorescently labelled with one of four fluorochrome moieties: 6-carboxyfluorescein FAM™ (blue; Metabion), VIC™ (green; Applied Biosystems, Foster City, CA, USA), NED™ (yellow; Applied Biosystems) and PET™ (red; Applied Biosystems). The amplification was then conducted in a thermocycler using a 65-55 touchdown PCR program: 95°C for 10 minutes, followed by a touchdown (-1°C at each cycle, from 65 to 55°C) of 10 cycles (95°C for 1 minute, 65°C for 30 seconds and 72°C for 1 minute), an amplification of 25 cycles (94°C for 30 seconds, Ta (see Table 5.1 and 5.2 for Ta) for 30 seconds and 72°C for 30 seconds) and a final amplification stage at 72°C for 10 minutes. The amplified DNA was genotyped using electrophoresis by capillarity: 1 µl of the amplified products was mixed with 9.8 µl of formaldehyde and 0.2 µl of 500 LIZ (size standard; Applied Biosystems). The samples were denatured at 95°C for 5 minutes and loaded in the ABI3130 (Applied Biosystems,

Foster City, CA, USA) to run the capillarity electrophoresis. The genotyping analysis of the ABI products was done with GeneMapper® (Applied Biosystems) software.

Once published markers were identified as polymorphic using electrophoresis by capillarity, the whole population (96 plants) was screened with it, using acrylamide gels to reduce the cost. The DNA was amplified using a standard PCR protocol: a 10µl PCR reaction containing 0.25 unit DreamTaq DNA Polymerase (Fermentas), 1X DreamTaq Buffer, 200µM dNTPs, 0.2µM of each forward and reverse primers and 1µl of DNA (approximately 20ng). The PCR program used was the same as the one used for the electrophoresis by capillarity procedure. The samples were then placed in a polyacrylamide gel previously cast. The gel was made with 67.56 g of urea 98% (Sigma-Aldrich®, St. Louis, MO, USA), 50 ml of distilled water, 30 ml of 5X TBE (Electran BDH, VWR International LLC, Radnor, PA, USA), 750 µl of Ammonium Persulfate (100mg/ml, Bio-Rad), 50 µl of TEMED (Electran BDH) and 16.8 ml of Acrylamide stock solution (Bio-Rad). The PCR products were mixed with the same volume of denaturing buffer and were placed on the top of the polyacrylamide gel. It was run for 2 hours approximately at 3000V/100W/400mA. The DNA was revealed by a series of washes: sodium carbonate solution (30g/L), silver nitrate solution (1g/L) (silver staining kit, Promega, Madison, WI, USA). The gel was scanned and the genotyping of the population was done by scoring the different alleles.

Finally, specifically designed markers for both loci were scored using HRM curve analysis (see chapter 2 section 2.3.3 for the detailed protocol). The DNA was amplified using the PCR program HiRes63 described in Table 2.5 (chapter 2 section 2.3.3). The melting profiles of the samples were obtained using the 96-well LightScanner (Idaho Technology Inc., Salt Lake City, UT, USA) and the genotyping was done using the

software provided, LightScanner® and Call-IT® (Idaho Technology Inc., Salt Lake City, UT, USA).

5.2.3. Results

5.2.3.1. *T*-locus

The initial screening of the parent 0205 as well as 10 progeny from the *T*-population was done using the ViaLactia markers. In total, thirteen markers have been tested by electrophoresis by capillarity. Out of the thirteen tested markers, ten markers are polymorphic in the *T*-population, two markers are monomorphic and one marker did not work (see Table 5.1). The polymorphic markers were then screened on 124 plants from the *T*-populations using polyacrylamide gels. The genotypes of the plants are shown in Table 5.5.

Once the initial screening with the ViaLactia markers was done, a subset of 16 plants was created out of the 167 plants, selecting the DNA of the parent 0205, the recombinants and few non recombinant plants. The markers from Studer *et al*. (2008 & 2012) were then tested to see if the recombination rate was lower and if the *T*-region could be reduced. The numbers of recombinants are detailed in the Table 5.6.

Table 5.5: Genotype distribution of the *T*-population plants for the ViaLactia markers. The expected genotype distribution for a marker closely linked to the *T*-locus is a 1:1 distribution, with one homozygous missing unless a recombination has occurred between the marker and the locus or the marker is too far away. The percentage of one genotype was calculated using the total number of genotypes (plant screened minus missing data).

Marker name	Number of plant screened	Number of Heterozygous plants	Number of Homozygous plants	Number of Homozygous recombinant	Missing data
RV0054	167	80 (48.8%)	82 (50.0%)	2 (1.2%)	3
RV0184	96	42 (45.7%)	43 (46.7%)	7 (7.6%)	4
RV0250	167	81 (49.4%)	81 (49.4%)	2 (1.2%)	3
RV0471	167	78 (48.8%)	81 (50.6%)	1 (0.6%)	7
RV0495	96	46 (48.4%)	38 (39.6%)	11 (11.6%)	1
RV0757	124	66 (54.5%)	51 (42.2%)	4 (3.3%)	3
RV0814	124	64 (52.4%)	54 (44.3%)	4 (3.3%)	2
RV1030	124	64 (52.4%)	51 (41.8%)	7 (5.8%)	2
RV1188	124	54 (47%)	49 (42.6%)	12 (10.4%)	9

Table 5.6: Number of recombinants from the *T*-population (out of 167 plants) for the publically available markers (Studer *et al.*, 2008 & 2012). The percentage of recombination is also indicated in the table.

Marker name	Number of homozygous recombinants
G01_033	2 (1.2%)
G01_095	5 (3.0%)
G05_065	1 (0.6%)
G05_071	2 (1.2%)
G06_096	1 (0.6%)
PTA_1953	3 (1.8%)
PTA_2547	1 (0.6%)

From the recombination frequency as well as the recombinant individuals, the markers were aligned around the *T*-locus. The marker sequences were used to query a BLAST search in order to locate the *T*-locus region on the rice genome. The sequence of

the rice region from 17 to 25 Mb of chromosome 12 was isolated and used to design *Lolium perenne* specific markers (see Table 5.3). Those markers were screened on the *T*-subset of 16 plants as well but genotyped using HRM curve analysis. A total of 18 markers have been screened on the *T*-subset; height markers were monomorphic and only six markers were polymorphic (see Table 5.3). For all of the six polymorphic markers, one recombinant plant remains, so no non recombinant marker have been found. The Figure 5.2 is representing the *T*-locus region with the markers screened during this mapping.

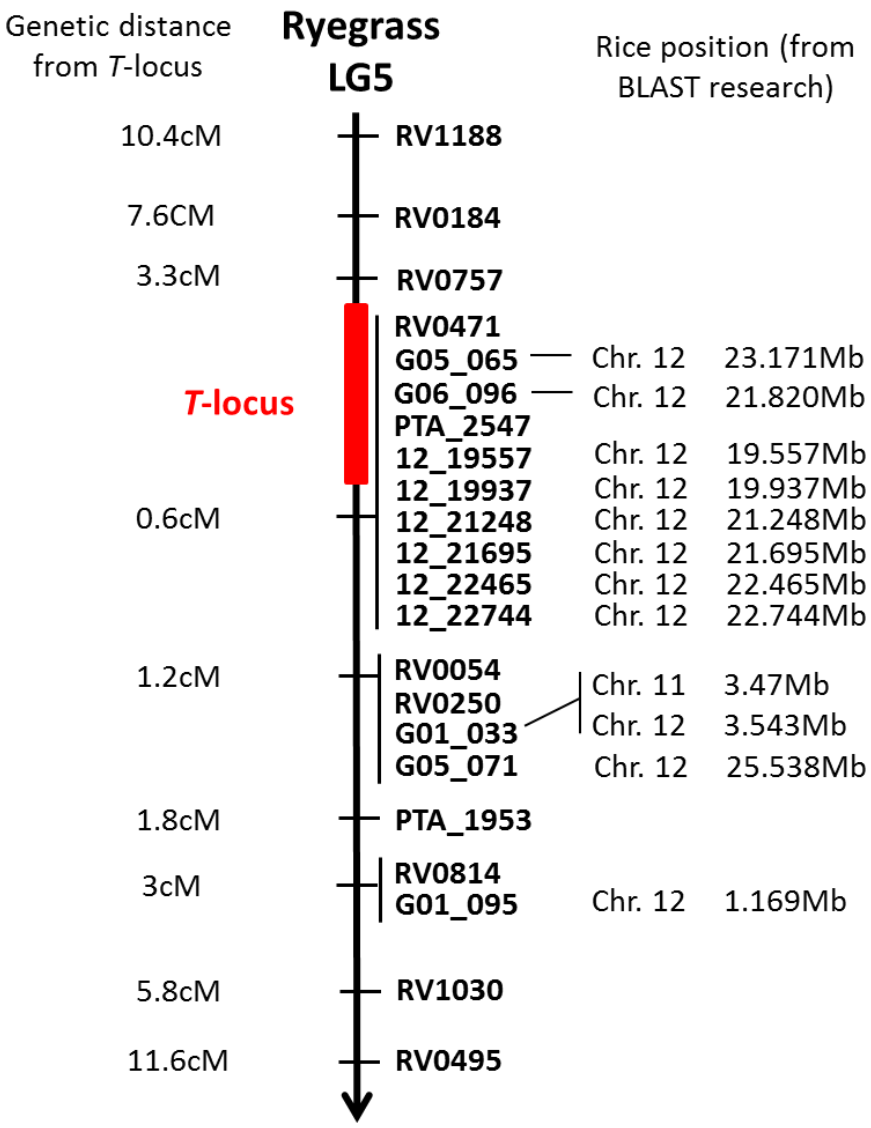


Figure 5.2: Map of the *T*-locus region in *Lolium perenne*. The markers are positioned onto the map according to their genetic distance, calculated as the percentage of recombination, therefore, the genetic distances are not exact as the whole *T*-population (167 plants) has not been screened by all markers. The red bar indicates the location of the *T*-locus, flanked by two groups of markers: RV0757 and the group of markers mapped at 0.6cM from *T*. The rice position of the G0X_0XX markers was obtained by querying a BLAST search of the marker sequence using NCBI.

5.2.3.2. *F*-locus

The mapping of the *F*-locus was conducted initially using SSRs markers from ViaLactia (see Table 5.1). To assess the marker polymorphism, the grand-parents (mother and father), the parent (F1) and five progeny of the F2 biomass population were genotyped, using electrophoresis by capillarity. The two polymorphic markers, RV1131 and RV1133, were then genotyped for the whole 96 F2 biomass plants. The genotype distribution of the F2 biomass population is detailed in Table 5.7. Because only two markers were polymorphic, more markers were needed for the mapping, leading to the selection of publically available *Lolium* markers (Studer *et al*., 2008 & 2012). Out of five markers selected, two markers were monomorphic and the three other markers did not work or were not easy to score (see Table 5.2). Therefore none of these markers were used to genotype the 96 F2 biomass plants.

Table 5.7: Genotype distribution of the *F*-population plants for the ViaLactia markers. The expected genotype distribution for a marker closely linked to the *F*-locus is a 1:1 distribution, with one homozygous group missing unless a recombination has occurred between the marker and the locus or the marker is too far away. The percentage of one genotype was calculated using the total number of genotypes (plant screened minus missing data).

Marker name	**Number of plant screened**	**Number of Heterozygous plants**	**Number of Homozygous plants**	**Number of Homozygous recombinant**	**Missing data**
RV1131	96	53 (55.8%)	29 (30.5%)	13 (13.7%)	1
RV1133	96	64 (66.7%)	26 (27.1%)	6 (6.2%)	0

Because not many markers tested were polymorphic, another strategy to isolate a region of interest was used. Thorogood *et al.* (2002) found that the region around the markers CDO920/WG889 map on linkage group 3 in *Lolium perenne* showed linkage

disequilibrium with markers in the region of the *S*-locus despite not being physically linked. The BLAST query of the marker CDO920 sequence (Van Deynze *et al*., 1998) has shown that the marker is syntenic to the rice region 27.37Mb on chromosome 1 (LOC_Os01g47820, www.gramene.org). Moreover, the sequences of markers around WG889 of the genetic map from Armstead *et al*. (2002) were used to query BLAST search on the rice genome (see Table 5.8). From the rice position of the F-locus *Lolium perenne* region, the rice region 25 to 28 Mb from chromosome 1 was selected to design the first *Lolium* specific markers.

Table 5.8: Results of the BLAST search of markers mapped around the marker WG889. The marker sequence was used to query a BLAST search of the marker in the rice genome. From the BLAST search, the rice gene and its position on rice chromosome 1 were obtained. The *Lolium* markers are arrange in the same order as on the genetic map from Armstead *et al*. (2002).

Marker name	**Rice gene**	**Rice position on chromosome 1**
R1613	LOC_Os01g07310	3.44 Mb
PSR394	LOC_Os01g41010	23.22 Mb
CDO328.1	LOC_Os01g14050	7.86 Mb
CDO920	LOC_Os01g38620	21.70 Mb
PSR370	LOC_Os01g36590	20.30 Mb
WG889	LOC_Os01g47820	27.37 Mb
GOT3(C250)	LOC_Os01g55540	32.00 Mb
BCD828	LOC_Os01g49190	28.27 Mb
CDO345	LOC_Os01g65310	37.88 Mb

By aligning *Lolium perenne* cDNA reads to this rice region, ten specific markers were designed to initially screen the *F*-population in order to better identify the *F*-locus. Out of the ten markers screened, only two markers (01_27495 and 01_27732) were polymorphic in the F2 biomass population. The two markers were screened on 21 plants of

the F2 biomass family as an initial screening. Both marker 01_027495 and 01_27732 were showing the same genotype distribution for the 21 plants and two recombinants (9.5%). The two recombinants are common with the marker RV1131 but the two markers are closer to the *F*-locus as the marker RV1131 has four recombinants out of those 21 plants.

5.2.4. Discussion

The method used in the mapping of both *F* and *T* locus was the same as the one used for the *S*-locus: numerous markers were screened and the mapping was done according to the recombination frequency. For these initial mappings, small populations were used (167 and 96 plants for the *T* and *F* mapping respectively) but large number of markers were tested (46 and 21 for the *T* and *F* locus respectively).

The mapping of the *T*-locus was initially conducted using SSR markers and publically available markers (Studer *et al*., 2008 and 2012). The results of this screening identified flanking markers: on one side, the marker RV0757 with 3.3% recombination frequency and on the other side, the marker RV0471, G05_065 and G06_096 with 0.6% recombination frequency. This mapping led to the identification of a syntenic rice region, mapped on chromosome 12 from 17 to 25 Mb. Using the rice region as a reference and aligning *Lolium* cDNA reads, new markers were developed in order to reduce the *T*-locus region. Out of the 18 markers designed, six markers were polymorphic in the *T*-population, all showing the same recombination rate (0.6%) and the same recombinant as the markers RV0471, G05_065 and G06_096. The addition of new markers did not help to reduce the *T*-locus region even though they are covering over 3 Mb of the rice genome, from 19.55 Mb to 22.74 Mb. Moreover, the *T*-locus region could not also be reduced on the other side,

with the flanking marker RV0757, which is mapped 3.3 cM away. This mapping strategy might not be as successful in this case as for the *S*-locus. Indeed, the synteny between the rice chromosome 5 and *Lolium* LG 5 in the *S*-locus region is good as shown with the Figure 5.3 from Sim *et al*. (2005). However, Sim *et al*. (2005) also demonstrated that the LG 5 from *Lolium* is syntenic with the rice chromosomes 9, 11 and 12, chromosome 12 being at the bottom of LG 5 (the last RFLP marker on the Figure 5.3). Because of this breakdown in the synteny of LG 5, an alternative strategy might be to design new markers for the mapping of the *T*-locus using the *Brachypodium* genome as a reference to assemble the *Lolium* cDNA reads or even the *Lolium* genome scaffold which should be available in 2013.

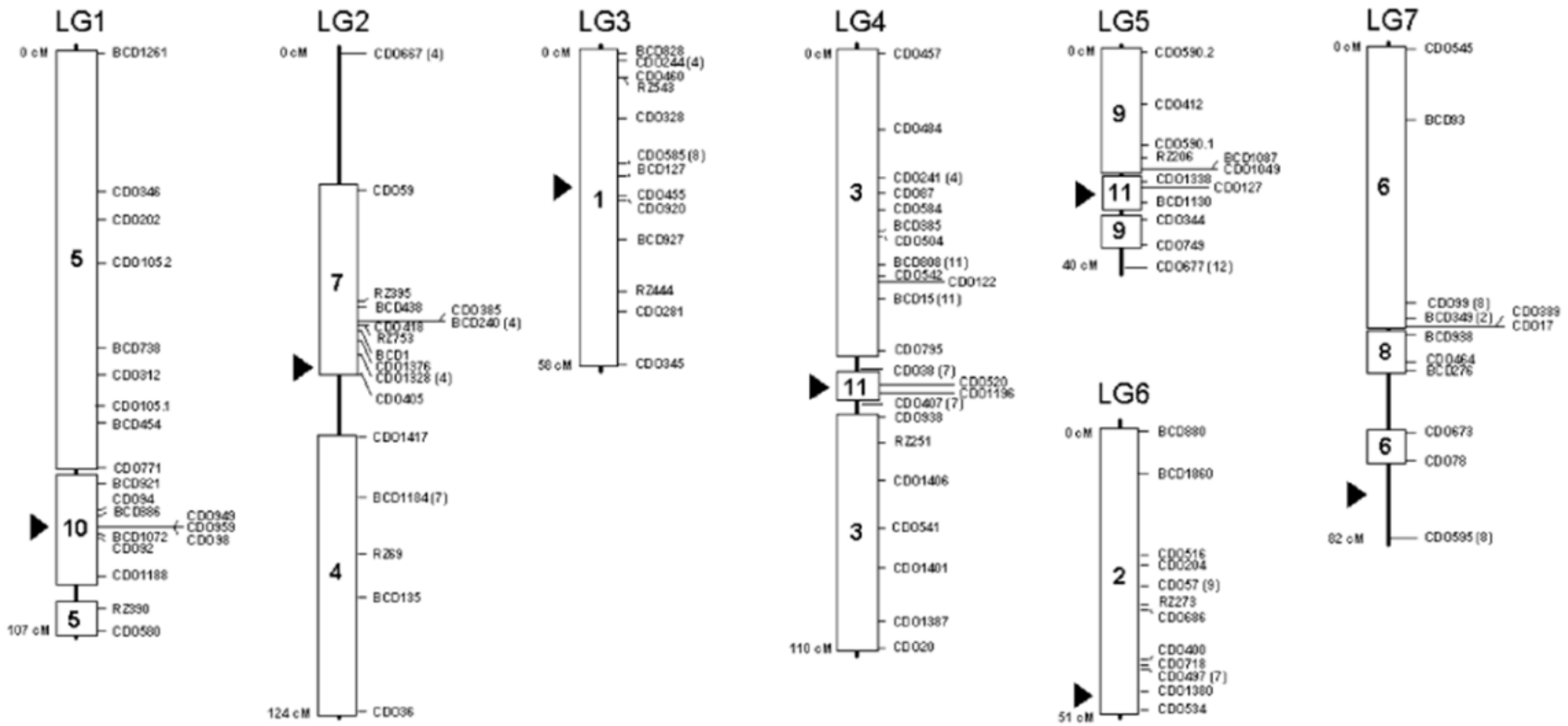


Figure 5.3: Comparative relationship between ryegrass and rice genomes. The bars represent the five LG from *Lolium*, with RFLP markers. The white boxes represent the rice chromosome syntenic to the LG, the number corresponding to the rice chromosome number. The *Lolium* centromeres are indicated with the black arrows. Figure from Sim *et al*. (2005)

Because the *T*-locus region could not be reduced to less than 3.9 cM, and the marker RV0757 could not be positioned on the rice or *Brachypodium* genome as its sequence was not available, the rice and *Brachypodium* region of interest could not be defined. Therefore, at this stage of the *T*-locus mapping, no candidate genes could be identified.

For the mapping of the *F*-locus, the screening of the F2 biomass population could only genotype two of the eleven markers screened. Because polymorphism appeared to be slow, the strategy adopted was to design new markers based on *Lolium* cDNA reads. From the results from Thorogood *et al*. (2002), the marker CDO920/WG889 was identified as the target region for the *T*-locus. Using the *Lolium* genetic map markers from Armstead *et al*. (2002) to search for syntenic region in rice, the rice region from 25 to 28 Mb from chromosome 1 was selected as the reference to assemble the *Lolium* cDNA reads in the *F*-locus region. Preliminary *T* mapping was initiated by designing ten markers of which two were found to be polymorphic in the F2 biomass populations (01_27495 and 01_27732), each with two recombinants out of 21 plants. These two markers mapped closer to the F-locus than the marker RV1131 but their genetic distance to the *F*-locus can not be estimated as only 21 plants have been screened.

The synteny between rice and *Lolium* on LG 3 is conserved as shown in Figure 5.3 by Sim *et al.* (2005). However, the microsynteny is not as good as the BLAST search of the markers from Armstead *et al*. (2002) genetic map showed that the marker order was not fully conserved (see Table 5.8). For this preliminary mapping strategy to pursued, more markers need to be designed, but the size of the rice region of interest also needs to be

increased and the *Brachypodium* genome should also be considered in order to exploit improve the synteny and comparative genomics knowledge more effectively.

5.3. Semi *in*-vivo self-pollination to study the *F*-locus

5.3.1. Materials and methods

The F2 biomass population was used in this study to look at pollen tube growth on self-pollinated stigma as the population was genotyped previously for two markers mapped on LG3 of *Lolium perenne*.

In this pollination test, 96 plants of the "F2 biomass" population have been used (the same plants as for the mapping of the *F*-locus). The plants were vernalized in a frost free glasshouse for 10 weeks and transferred into a controlled environmental unit, with long light days (16h) and temperature of 20°C. After approximately 6 weeks, inflorescence started to emerge and elongate. The flower maturity was reached after 9 weeks and at this stage, the humidity of the control environmental unit was reduced from 74% to 60% in order to facilitate the collection of viable pollen.

The pollination tests were made for each individual of the population. The plant was selected for the test when some pollen was released the day before. The plant was shaken to remove any pollen released the day before; all inflorescences were bagged early in the morning before anthesis for subsequent pollen collection. Around midday, the cellophane bag containing the freshly released pollen was removed. Ten pistils with ovaries were collected from a non-anthesed but mature floret. Pistils (still attached to ovaries) were placed on petri dishes onto an agar medium (100g/L sucrose, 20g/L agar-agar and 0.1g/L boric acid, sterilized before pouring into Petri dishes). The pollen bags

were shaken above the collected pistils in order to gently release the pollen onto the stigmas without overloading the stigmas with too much pollen as it would reduce the pollen germination rate.

In order to look at pollen tube growth in the style, observations were done at different time points after semi *in-vivo* pollination. In order to stop the pollen germination, the pistils were removed from the agar medium, the ovary removed with a dissecting needle and the stigma was placed on a microscope slide with a drop of aniline blue solution (0.1% w/v of aniline blue in 0.1M K_3PO_4; Martin, 1959) and a cover slip was placed on top. For each plant, at least two stigmas were taken at each time point. For most of the plants, stigmas were sampled at four time points after semi *in-vivo* pollination: 30 minutes, 60 minutes, 120 minutes and 4 hours; the rest of the plants had one time point for collection: approximately 4 hours after semi *in-vivo* pollination. This sampling method was applied to all the plants and self-pollination were repeated up to 4 times in some cases. Samples were observed under the microscope (x50) with UV light.

In parallel to the semi *in-vivo* pollination, the plants were also phenotyped *in vivo* for seed sets on self-pollination (SP) and open-pollination (OP) success. Once the semi *in-vivo* pollinations were completed, the plants were transferred to a normal glasshouse. Between five to ten inflorescences were bagged together in order to self-pollinate the plants, using cotton wool to separate the inflorescences and a cellophane bag to prevent pollen contamination. The rest of the inflorescences from the plant were left unbagged to enable open-pollination with the other plants from the F2 biomass family. After three months, the inflorescences from each plant were harvested, keeping the self-pollination ones separate to the open-pollinated. For each plant, a set of measurements was taken: weight of all the seeds, number of seeds (calculated by weighing 100 seeds and then using

the following formula: number of seeds = weight of 100 seeds * 100 / weight of all the seeds), total number of inflorescences in each unit, total number of spikelets and the number of florets per mid-spikelet. The total number of florets (i.e. seed setting sites) per unit could then be calculated. From the estimate, or in some cases actual counts, of seed numbers and the total number of florets, an estimation of the % seed set could be derived.

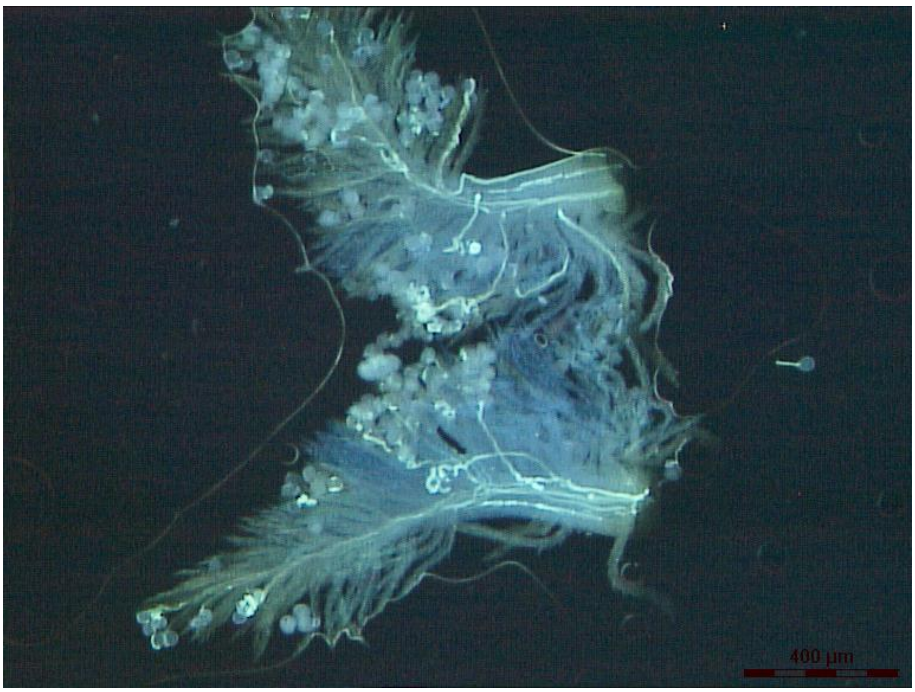
5.3.2. Results

Initially, in order to induce flowering, the plants from the F2 biomass family were vernalized and incubated at 20°C under long day light. Out of the 96 plants that underwent this treatment, only 80 plants flowered, enabling pollination tests. These 80 plants were phenotyped for their self-compatibility (SC) by using semi *in-vivo* self-pollination. Among the 80 plants phenotyped for their SC, three plants could not be phenotyped as they were partially sterile: not releasing much pollen or pollen not germinating.

The self-compatibility genotype was assessed under the microscope by looking at pollen tube growth and length as well as the occlusion (indicated by fluorescence of callose especially around the germinating pollen-tube tip) of the pollen grain. An incompatible pollen grain presents a coating of callose around it, appearing bright blue under UV light because of the aniline colouring, with usually a very short pollen tube that fails to penetrate the stigma surface. On the other hand, a compatible pollen grain will show a long pollen tube, penetrating and reaching the end of the transmitting tract after some time, appearing bright blue under UV light because of the callose deposit on the pollen tube wall. By assessing the proportion of compatible to incompatible pollen grains, the self-compatibility can be phenotyped as a percentage of compatible pollen grain. In the case of the "*F*-locus"

population, two classifications were found: fully self-compatible (100%) and half self-compatible (50%); no plants were found to be fully self-incompatible (0%). The Figure 5.4 and Figure 5.5 are representing the two classes of self-compatibility, fully self-compatible and half self-compatible respectively.

A



B

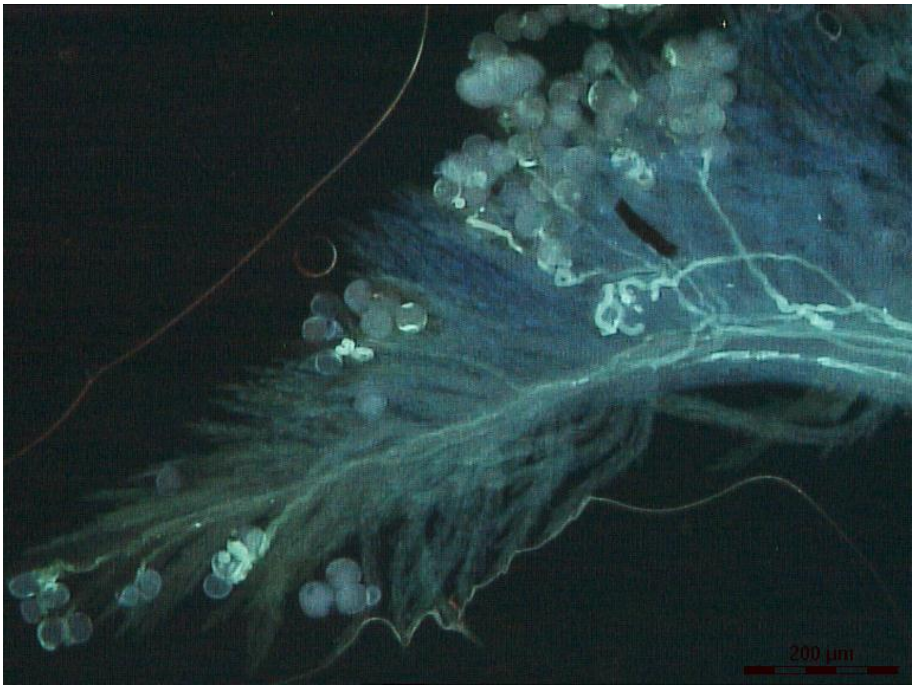
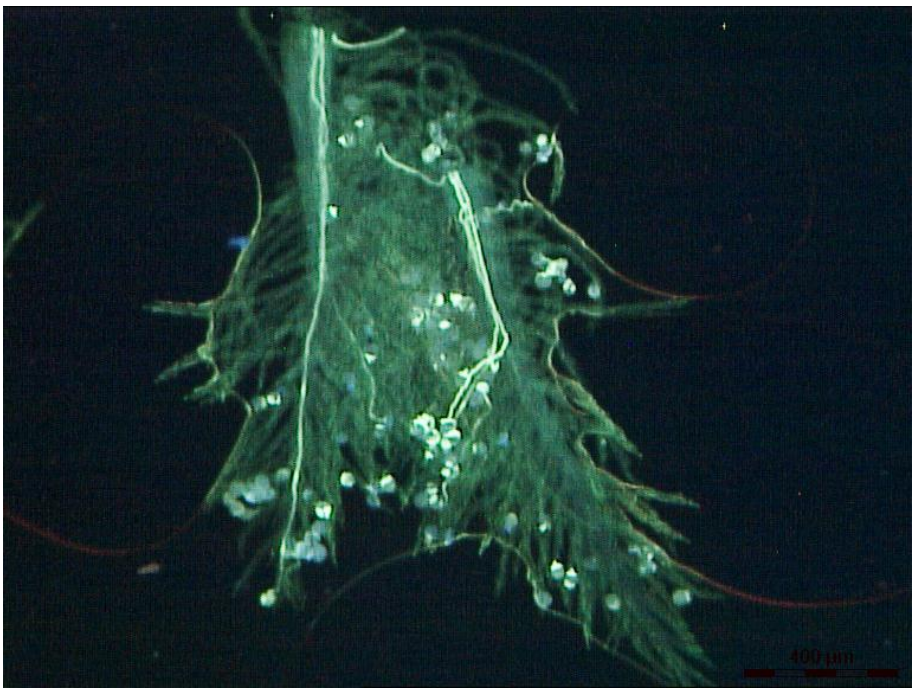


Figure 5.4: Photomicrograph of a fully compatible self-pollination. Aniline blue has been used as a dye and is revealed under UV light, zoom x50 (**A**) or x100 (**B**). Both pictures represent pollinated stigmas in which pollen tubes have grown into the transmitting tract of the style. Many pollen grains have not germinated and appear in transparent blue; their non-germination can be due to the short viability of mature pollen as well as some partial sterility.

A



B

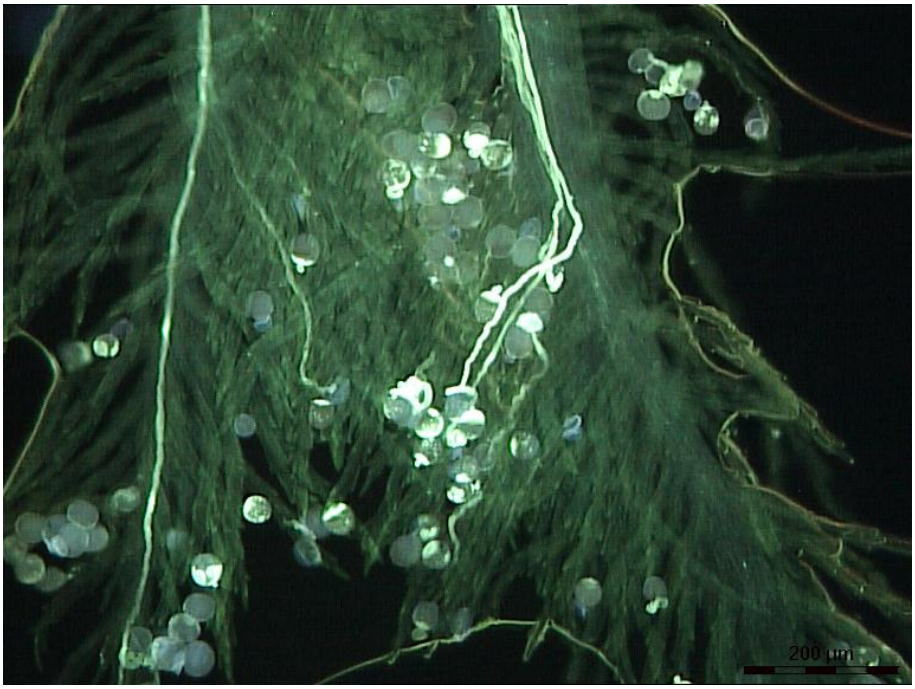


Figure 5.5: Photomicrograph of a half compatible self-pollination. Aniline blue has been used as a dye and is revealed under UV light, zoom x50 (**A**) or x100 (**B**). Both pictures represent pollinated stigmas where pollen tubes have grown into the transmitting tract of the style. Incompatible pollen grains differentiate from other pollen grains by a callose depot (in bright blue) partially covering the surface of the pollen, with sometimes a really short bright pollen tube, unlike non-germinated pollens which appear in transparent blue.

Pollen tube germination was studied using 27 plants from the F2 biomass family and sampling self-pollinated stigmas at different times: 30 minutes, 60 minutes, 120 minutes and 4 hours. Under microscope and with aniline blue staining, the pollen tube growth was assessed and no differences were observed between the different times. Indeed, 30 minutes after pollination, the pollen tube of compatible pollen was already reaching the bottom of the transmitting tract and incompatible pollen grains could be observed. Therefore, the pollinations as well as the incompatible reactions are fast-acting mechanisms.

The results of the semi *in-vivo* phenotyping of the F2 biomass plants for SC as well as the OP and SP seed setting data are presented in the Appendix 7.

Out of the 77 plants genotypes for SC using semi *in-vivo* self-pollination, two phenotypes were observed: half SC (50%) and fully SC (100%), with 42 and 35 plants per phenotype respectively. The distribution of the SC is in agreement with the 1:1 segregation ratio (test of goodness to fit: $\chi^2_{[1\ df]} = 0.31$, $p = 0.43$).

In order to study the association between phenotype and genotype, the genotypes of the markers RV1131 and RV1133 located on LG 3 were compared to the SC phenotypes (see Table 5.9). The genotypes of the markers are expected to follow the same pattern as the SC phenotype if the putative *F*-locus is involve in the phenotype. The segregation of both markers is distorted and do not follow the 1:2:1 ratio expected from a normal segregation (test of goodness to fit: $\chi^2_{[2\ df]} = 8.968$, $p = 0.0113$ for RV1131; $\chi^2_{[2\ df]} = 12.838$, $p = 0.0016$ for RV1133), with an under representation of one double homozygous (bb)

Out of the 42 plants that have been phenotyped as half self-compatible, 50% are heterozygous (ab) for the marker RV1131 and 64.3% for the marker RV1133; the rest of the plants are either homozygous aa or bb. For the 100% SC phenotype, the distribution is following the same tendency, with 60% of the plants heterozygous for the marker RV1131 and 71.5% for the marker RV1133.

Table 5.9: Distribution of the genotypes of two ViaLactia markers mapped on LG3 for the SC phenotype. The genotypes of each marker have been grouped according the SC phenotype, half SC (50% SC) or fully SC (100% SC). Within each group, the genotype distribution has been indicated by the number of individuals and by the percentage that the number represents within one phenotype.

	RV1131		**RV1133**	
	50% SC	**100% SC**	**50% SC**	**100% SC**
aa	17 (40.5%)	10 (28.5%)	11 (26.2%)	8 (22.8%)
ab	21 (50%)	21 (60%)	27 (64.3%)	25 (71.5%)
bb	4 (9.5%)	4 (11.5%)	4 (9.5%)	2 (5.7%)

The relation between semi *in-vivo* SC phenotypes and the seed sets data from OP and SP can be seen in the Appendix 7. The seed sets results are showing some variability. The Figure 5.6 is a graph representing the relation between the % seed set from OP inflorescences and the % seed set from SP inflorescences. The data have been separated according to the phenotype of the plants for SC, either 50% SC or 100% SC. For each relation, a linear regression has been applied. The coefficient of determination R^2 of the regression lines are between 0.45 and 0.47, therefore the R distribution coefficient is between 0.67 and 0.68, showing that the relationship between seed set in OP and seed set in SP is linear ($p < 0.001$, df = 76).

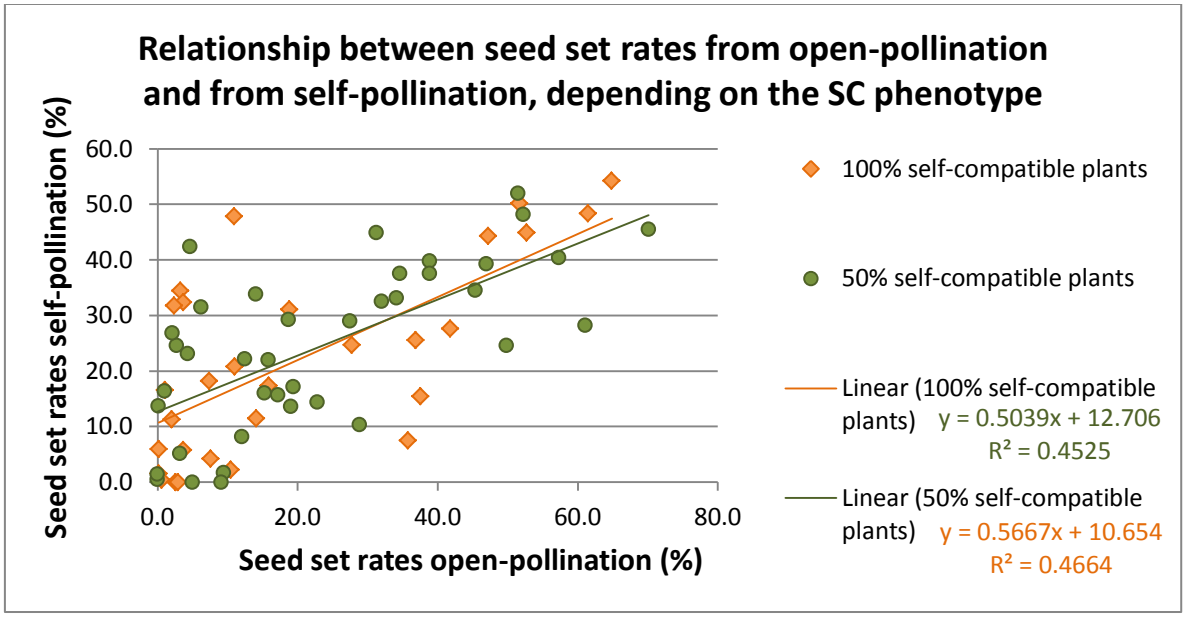


Figure 5.6: Representation of the relationship between the seed set rates from open-pollination and the seed sets rates of self-pollination, for each SC phenotype (50% and 100% self-compatible). Both seed set rates are represented as percentage.

The same observations can be made when comparing the phenotypes for SC and the seed setting rates in SP and OP. The Figure 5.7 is a graph representing the two SC phenotype classes (50% and 100%) according to the seed set rates of OP and SP. For the 50% self-compatible phenotype, the seed set rate average from OP does not statistically differ from the one from SP (23.5% against 24.5% respectively). A larger difference can be observed between the average seed set for the phenotype group 100% self-compatible, with the average seed set from OP being 19.3% and the average seed set from SP being 23%. However, the dispersion of the seed set for all four groups (50% SC - seed set OP, 100% SC - seed set OP, 50% SC – seed set SP and 100% SC – seed set SP) are important as the standard deviation are large (19.7%, 20.7%, 14.8% and 18.0% respectively).

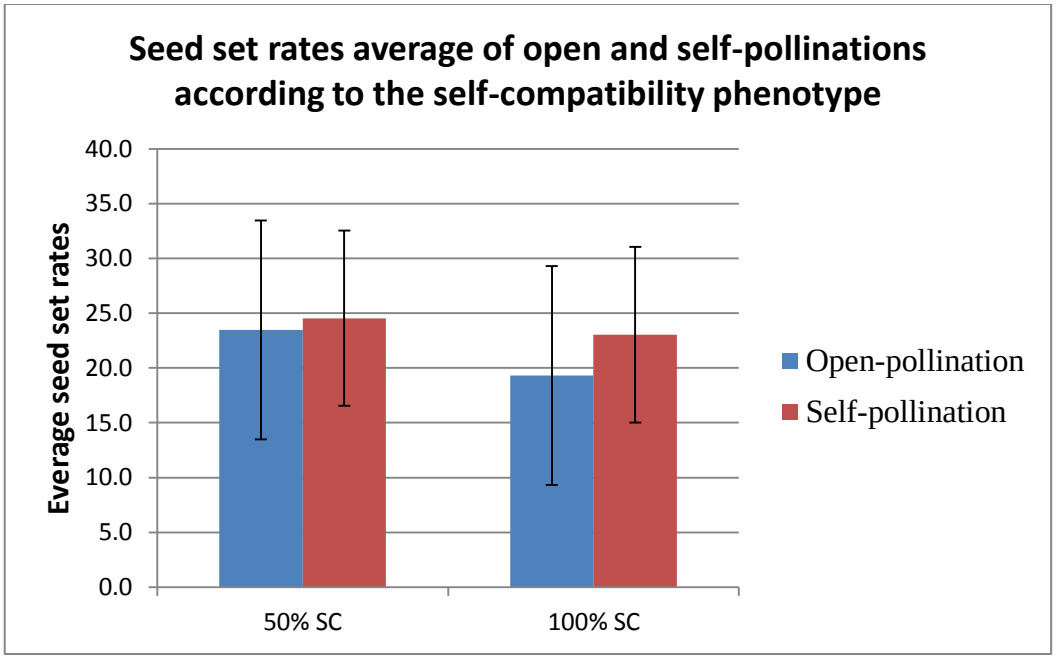
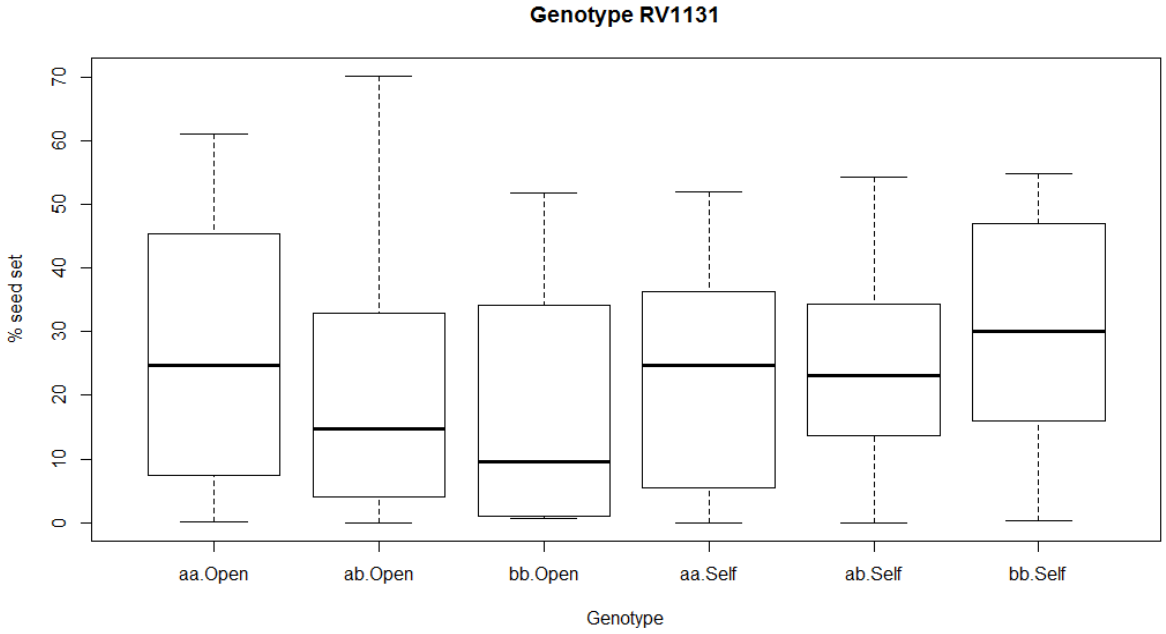


Figure 5.7: Representation of the seed set average of open-pollination and self-pollination for both SC phenotypes. The average seed set rates is represented as percentage and the error bars represent the segregation distortion of the seed set distribution.

Finally, in order to look at the significance of the differences, a comparison of all the data have been made using a two way ANOVA test with R (R Development Core Team, 2008) and the results are available in Appendix 8. The Figure 5.8 is a box plot representing the median and the data distribution; and shows that the differences between values are not significant for any of the measurement. Indeed, no linkages were found between the genotype of the markers with the seed sets, OP or SP, as well as the SC phenotype.

A)



B)

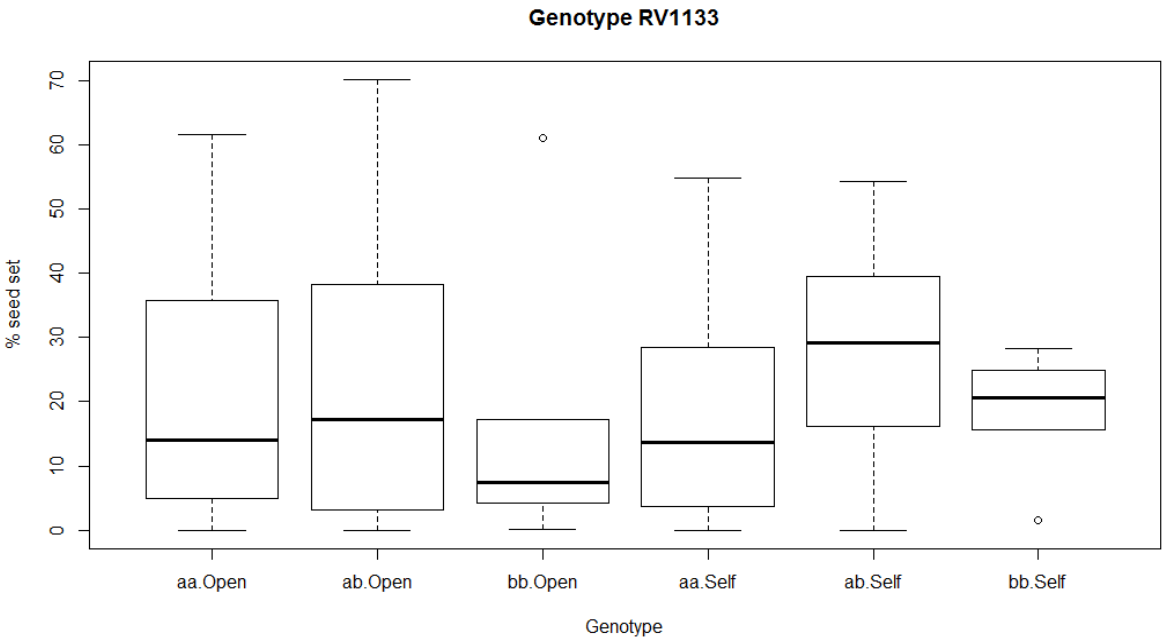


Figure 5.8: Box plot representations of the seed set from OP and SP for each genotype of the two markers linked to the *F*-locus, RV11331 (**A**) and RV1133 (**B**). Seed sets are represented for each allele combination, and for both pollination methods (OP and SP). For each box, the middle bar represents the median, the white box represents the 25 and 75% quantiles, the extreme lines represent the minimum and maximum values and the same circle represents the extreme outliers.

5.3.3. Discussion

This assay was designed to study the self-pollination reaction in a self-compatible population showing segregation distortion on LG 3 that was roughly similar to the one observed in the ILGI population by Thorogood *et al*. (2002). In order to do so, several aspects of a self-pollination were observed: self-pollination at a microscopic level, the segregation of the self-compatibility in the population and the success of seed setting in SP as well as in OP in order to determine the effectiveness of self-pollination to produce seed and relative efficacies of self- and open-pollinations in producing seed .

The self-pollinations were carried out using semi *in-vivo* pollination tests in order to determine the phenotype of the SC as well as to look at pollen tube growth and the incompatible reaction. On fixing pollinated stigmas at different times after self-pollination, it was observed that the pollen tube was fully grown (reaching the end of the transmitting tract) 30 minutes after self-pollination. The incompatible reaction was also fast as incompatible pollen grains were also observed after 30 minutes. The incompatible reaction in this SC population seems to be as fast as the incompatibility in other populations and suggests that the mechanism involved in the recognition of self-pollen in self-compatible plants is acting at an early stage in the pollen germination, at the stigmatic level. The self-compatibility in this population could be due to the inactivation (or activation) of a component of the SI cascade reaction.

From the phenotyping of 77 out of 96 plants from the F2 biomass family, two groups of phenotypes have been identified: half self-compatible (50%) and fully self-compatible (100%). The 1:1 segregation of the phenotype in the population is the same as the one observed by Thorogood *et al*. (2005) in another F2 population, and strongly

suggest that SC is controlled by a major unique gene. The 1:1 segregation ratio also supports the theory of gametic selection in which one of the gamete, and probably the pollen in this case, carrying the recessive allele is not transmitted to the progeny, leading to an absence of one homozygous genotype. However, the joint segregation of the two markers used in the *F*-mapping, RV1131 and RV1133 with SC phenotype does not suggest that the SC phenotype observed in this population is linked to the *F*-locus. Indeed, the segregation of both markers, although showing significant deviation from an expected 1:2:1 ratio and closer to a 1:1 ratio, is unlinked to the segregation of the SC phenotype and therefore, the SC observed in this population, even if there is distortion segregation on LG 3, must be linked to another *Lolium* LG or a mutation in the *S* or *Z* loci. Another locus, the *T*-locus, has been reported as a potential SC locus (Thorogood *et al.*, 2005) and could be responsible for the SC of the F2 biomass population. In order to test this hypothesis, markers from LG 5 have to be tested on the F2 biomass population. Other markers mapped on all LGs have to be tested as well to see if other segregation distortion regions could be involved in the SC.

The % seed set of the 77 plants from the F2 biomass family was assessed for open-pollination and self-pollination, as the ratio total number of seeds : total number of florets. When comparing the two seed sets from a same plant, the relationship between the two measurements could be associated to a linear distribution, showing that the two seed set distributions are strongly related ($p < 0.001$). However, the distribution tendency of the seed sets relationship seems to follow the same pattern between 50% SC and 100% SC with tendency line having nearly the same equation. The average seed set was calculated for each pollination (OP and SP) and separating the plant 50% SC to the ones 100% SC. The average differences between OP and SP within the same SC group are not important,

varying from 1% to 3.7%. However, the distribution of the seed set varies a lot around the average, with important standard deviation going from 14.8% and up to 20.7%. Because of this large dispersion of the seed set, no significant differences could be observed between OP and SP seed set for both 50% SC and 100% SC.

Another observation from the seed set data is that the average seed set for OP and SP are quite low, at 21.7% and 23.9% respectively. For the SP, the seed set varies from 0 to 54.8% which is a lower seed set than for OP, for which seed set varies from 0 to 70.1%. However, SI species rarely set anywhere near 100% seed. The low seed set in the OP could be due to the glasshouse environment the plants were in as no air circulation was applied, maybe preventing the pollen dispersion between plants. The low seed set for SP could be explained by the bag environment that is not favourable to pollen viability because of the increase of the temperature and humidity in the bags. Some sterility was observed in this population, where three plants were not genotyped for SC as no pollen was released. In this case, the sterility would be due to the pollen lack of fertility. However, the low seed set could also be due to a late acting mechanism preventing the fecundation in the ovule or the seed formation. These observations confirm the low seed set of self-compatible plant that has been observed by McCraw and Spoor (1983).

Finally, no relationship could be established between the seed set (from OP and SP) and the genotype of the two markers mapped on the LG 5. The small differences between the seed set could not be associated with a marker or even an allele combination leading to the conclusion that the *F*-locus is not playing a key role in the self-compatibility by interfering with the seed formation, and therefore is not responsible for the low seed set.

Chapter 6:

General discussion

6.1. Aims and background

The overall objective of this project was to study the genetic control of self-incompatibility in *Lolium perenne* as well as the breakdown of this mechanism, enabling self-compatibility. The genes involved in many SI systems have been discovered using different approaches: genomic mapping and transcriptome analysis (Stein *et al*., 1991; Suzuki *et al.*, 1999; Schopfer *et al.*, 1999), or proteomic studies (Hinata and Nishio, 1978; Anderson *et al*., 1986; Foote *et al.*, 1994). The grass SI mechanism is unique as it is governed by two loci, *S* and *Z.* The studies of grass SI have been conducted in different species using both genomic and transcriptomic approaches. Initially, a mapping approach was used in order to identify a broad region and candidate genes as done by Bian *et al*. (2004) who delimited the *S*-locus in *P. coerulescens* to a region of 0.26 cM on the short arm of chromosome 1. Kakeda *et al*. (2008, 2009) in *Hordeum bulbosum* and Yang *et al*. (2009) in *Lolium perenne* used a tissue-specific expression approach to identify pistil or pollen-specific genes within the *S* and *Z* loci.

Using the work of Yang *et al*. (2009) as a basis, this project focused on the study of the *S*-locus. The strategy used was: at first, to reduce the *S*-locus region in *Lolium perenne*, from the 0.93 cM region identified by Klaas *et al*. (2010), using a fine mapping approach; then identify and sequence the BAC clones covering the *S*-locus region in order to build up the *Lolium perenne S*-locus genomic sequence and annotate the genes included in it; and finally, study the expression of these genes in pollen and stigma in order to identify tissue-specifically-expressed genes within the delimited *S*-locus region.

Self-compatibility in *Lolium perenne*, as well as other grasses, has been identified in different populations and genetically linked to additional loci (*F* and *T* loci in perennial

ryegrass). In order to understand the SI mechanism and its breakdown, the mapping of these additional loci was conducted as well as a physiological study.

Identifying the genes and mechanism involved in the grass SI system opens up opportunities to use the SI in breeding programmes instead of being dependent on it. Being able to predict the genotypes of breeding material for SI genes will increase the potential of breeding programmes, by targeting a large number of alleles in recurrent selection programmes or a small number in order to increase the hybrids rate in a cross-pollination. Markers from the fine mapping were used in order to assess the *S* and *Z* allelic diversity in a breeding programme.

But the identification of the genes controlling the grass SI system would enable comparison with other well-known SI systems. This comparison would be interesting for the understanding of the evolution of sexual reproduction in angiosperms. SI is believed to have evolved independently at least three times (Charlesworth *et al*., 2005) and because the grass SI system is showing some characteristic from both GSI and SSI, solving this SI system might help understand the SI evolution. The gene identification, besides its potential application in breeding programmes, would increase the understanding of allelic diversity, induced by frequency dependent selection of SI.

6.2. Key findings

The fine mapping of *Lolium perenne* was done by genotyping 10,177 plants, from six different populations segregating for the *S*-locus. The *S*-locus region was reduced to 0.11cM, with the top flanking marker 05_02790 mapped at 0.09 cM from *S* and the bottom

flanking marker 05_02889 mapped at 0.016 cM. The recombination frequency in this region was lower than expected as the first fine mapping done on 1,393 plants identified the exact same region as the large-scale fine-mapping. The genetic distance of the region was reduced but no additional recombinants were found. However, the genetic distance is not representative of the physical distance in centromeric region as it is the case for the *S*-locus. By synteny with the *Lolium* molecular markers, the *S*-locus region in rice and *Brachypodium* was identified. The *S*-locus rice region, with a size of 60 Mb, contains eight genes and the *Brachypodium* region, a bigger size (138 Mb), contains eleven genes. However, only six of the rice genes and nine of the *Brachypodium* genes can be candidate genes.

The genetic study of the S-locus region was conducted by sequencing the BAC clones covering the region as well as studying the S-locus gene expression in pollen and pistil. The identification of the S-locus BAC clones revealed that the *Lolium S*-locus region is larger than in rice and *Brachypodium* for which the S-region is 60 and 138 Mb respectively. Indeed, the *S*-locus region could not be covered by a single BAC clone, which has an average size of 100 Mb. After sequencing and assembling, the *Lolium S*-locus genomic sequence is made out of 70 scaffolds (716,307 bases) but with some gaps in the sequence. A lot of retrotransposon elements were found in the *S*-region leading to difficulties in the sequence assembly process and the duplication of some scaffolds. Retrotransposon are highly expressed, with a tendency to be over-expressed in pollen tissue over stigma.

The annotation of the *Lolium S*-region indicates that six rice genes are homologous to the region, five of which were previously identified using the *Lolium*-rice synteny. As for the synteny with *Brachypodium*, six homologous genes were found in the *Lolium S*-region, all identified previously. But the annotation of the *Lolium S*-region indicates also other

genes homologous to different grass species: a *Lolium perenne* promoter of CCR1 gene and a *Hordeum vulgare* gene coding for a predicted protein. Several annotations were not homologous to any known plant genes.

The identified candidate genes mapped in the *Lolium S*-locus region are the following:

- Gene Os05g0149900-like, over-expressed in stigma tissue, contains a TPR-like domain, involved in protein-protein interactions.

- Gene Os05g0150000-like, over-expressed in stigma, is annotated as an enzyme involved in the amino-acid biosynthesis so is unlikely to be involved in the direct recognition of the pollen by the stigma.

- Gene Os05g0150300-like is the gene with the highest expression in the *S*-region, independently of the tissue or pollination type. However, it is over-expressed in stigma compared to pollen, and is annotated as a transcription factor-activator.

- Gene Os05g0150400-like, also over-expressed in stigma, is annotated as containing a ds-RNA binding domain and therefore is involved in expression regulation processes.

- Gene Bradi2g35767.1-like, annotated as involved in pathogen resistance mechanisms is over-expressed in stigma tissue.

- *Brachypodium* gene Bradi2g35750.1, homologous to rice Os05g0198000, and mapped next to the candidate gene HSP10 from Kakeda *et al*. (2008) is not syntenic to the *Lolium S*-region.

- Gene Os05g0150500-like, over-expressed in stigma, is annotated as a transport inhibitor protein containing an F-box domain, which makes it a prime candidate. However, the allelic diversity study of the Os05g0150500-like gene did not show any major polymorphism between *S*-alleles that could be linked to the specificity needed by the SI mechanism.

- Five annotations for unknown proteins, showing higher expression levels in pollen samples, of which only one is syntenic to a gene in another plant, *Hordeum vulgare*. The pollen *S*-gene is most likely unknown and coding for a novel protein or RNA.

A study of the SI allelic diversity was conducted in a breeding population. This study revealed that it is possible to determine the *S* and *Z* allelic diversity in a population by using closely few linked markers coupled with the HRM method. However, the allele calls using closely linked markers and the HRM method only is not easy in a population with large number of alleles. The technique needs to be proven by allele sequencing and confirmatory pollination tests. The results show that allelic diversity, at both *S* and *Z* loci, is maintained in breeding population even within recurrent selection breeding programmes. However, the allele frequency is uneven.

Finally, the self-compatibility study was done by mapping both *T* and *F* locus. The *T*-locus, supposed to be similar to the *Secale* S5 self-compatible gene (Fuong *et al*, 1993; Voylokov *et al*., 1998), is mapped to a region of 3.9 cM on LG 5. However, the lack of synteny between *Lolium* and rice for the LG 5 can be a problem in mapping approaches. The use of other species, closely related to *Lolium* is essential. The additional *F*-locus could not be closely mapped in *Lolium*, due to a breakdown in the micro-synteny between Lolium and rice. But, this initial mapping study demonstrates a segregation distortion region on LG 3. However, this segregation distortion is not linked to the self-compatibility phenotype in the studied population. Indeed, the segregation of the SC phenotype of this population follows a 1:1 ratio, with only two phenotypes: 50% self-compatible and 100% self-compatible, but no genotype could be associated with a molecular marker on LG 3. Another finding from this study is that the self-compatibility of a pollen grain is quickly

assessed as the SC/SI phenotype can be scored 30 minutes after pollination, suggesting that the SC is a mutation in a gene involved in the SI cascade response.

6.3. *S*-locus: hypothesis of the SI mechanism in *Lolium*

Few hypotheses can be derived from the *S*-locus sequence annotation and the gene expression study from pollen and stigma. However, because all of the pollen over-expressed identified genes are unknown or coding for unknown proteins, hypothesises are built on the stigma *S*-candidate.

The first hypothesis about the SI system would consider the gene Os05g0150500-like as the stigma-specific *S*-gene. This gene is annotated in rice and *Brachypodium* as a transporter inhibitor containing an F-box domain. F-box proteins are known to be involved in SI, as it is the case in the GSI S-RNase system. In this SI mechanism, the F-box is expressed by the pollen and is involved in the identification of S-RNase released by the stigma with the same *S*-allele, targeting them for their degradation in the proteasome. The first difference in *Lolium* is that the F-box containing protein is expressed in stigma and because this Os0g0150500-like gene is coding for a transporter inhibitor, it might be that the pollen ligand is secreted or anchored in the pollen exine, activating the stigmatic F-box protein. The second main difference between the S-RNase SI system and the *Lolium* SI is that the gene coding for the F-box protein in Solanaceae contains some hypervariable coding regions, responsible for the *S*-specificity of the SLF protein (Ushijima *et al*., 2004; Ikeda *et al*., 2004; Sonneveld *et al*., 2005). However, such hypervariable regions have not been identified in the cDNA sequences of the *Lolium* gene, and therefore, if this gene is *S*-

allele specific, the diversity is not due to allelic variation in the coding sequence but other mechanisms such as splicing variant or gene expression regulation.

However, the grass SI system is more complex than only one *S*-locus and the *Z*-locus has to be taken into consideration when hypotheses are made. F-box proteins are often part of SCF complex which have ubiquitination activity, targeting substrate for their degradation (Hershko and Ciechanover, 1998). Among the *Z*-candidate genes, one of them could be involved in such an ubiquitination/proteolysis complex. A putative ubiquitin-specific protease (UBP), LpTC116908, has been identified as co-segregating with the *Z*-locus (Shinozuka *et al*., 2010).

The homologous gene to Os0g0150500 in *Brachypodium*, Bradi2g35720.1 is well annotated and the transporter inhibitor it is coding for is involved in several biological processes: pollen maturation, stamen development but also in response to auxin stimulus, as Dharmasiri *et al.* (2005) demonstrated that TIR1 is an auxin receptor, part of the SCF^{TIR1} complex targeting for the degradation of the hormone. Auxin is well known to have an effect on SI. The Auxin Response Factor (ARF) is responsible for the expression of the Auxin Response gene by binding to their promoter but also recruiting Auxin/Indole-3-acetic acid (IAA) repressor (Guilfoyle and Hagen, 2007). By using transformed *Arabidopsis thaliana* exhibiting functional SI, Tantikanjana and Nasrallah (2012) observed the effect of ARF on SI and proved that the loss of the ARF function abolished SI and that the increase of ARF transcripts was enhancing SI, leading to the conclusion that ARF controls negatively the transcription level of the Auxin Response gene and activates the Aux/IAA repressor. Therefore, by reducing the auxin levels, SI is enhanced. No ARF genes have been found in the *S*-locus region but TIR1 could have the same effect, by directly targeting the auxin for degradation.

Another hypothesis for the self-incompatibility mechanism in grass could be that the stigmatic *S* component is a transcription factor and therefore a gene expression regulator element. This hypothesis is based on the Os05g0150300-like gene as the stigma *S*-gene. This gene, largely over-expressed in stigma, is annotated in rice and *Brachypodium* (Bradi2g35720) as a global transcription activator SNF2L1 (SWI/SNF related matrix associated actin dependent regulator of chromatin subfamily A member 1). SNF2 is a family of chromatin remodelling ATPases (Hargreaves and Crabtree, 2011) that can increase or decrease the accessibility of the genomic DNA by altering the histone-DNA interactions and therefore the DNA condensation (Clapier and Cairns, 2009). The interaction with the DNA is sequence specific and therefore could recognise a specific gene involved in the pollination and enhance or suppress its expression. In such a SI system, the S-Z system can be imagined as an on/off switch where both *S* and *Z* stigma components are transcription factors that will be secreted in the pollen and bind to the pollen *S* and *Z* specific genes which could be involved in the pollen tube germination. However, no such SI system has been discovered but transcription factors are known to be involved in many biological processes, such as stress resistance, cell growth, embryo development (Flaus *et al*., 2006).

The gene Os05g0150400-like is also over-expressed in stigma and closely mapped to the previously described gene (800 bp according to the scaffold annotations). This gene is also involved in expression regulation but binding to double-stranded (ds) RNA. Ds-RNA are recognised by the ds-RNA binding protein thanks to the sequence specificity of the protein ds-RNA Binding Motif (dsRBM). Such protein have several known functions such as pre-mRNA editing, RNA regulated kinase activity or RNA transport or gene-

specific silencing (Fierro-Monti and Mathews, 2000). As described with the gene Os05g0150300-like, this gene could also be involved in a SI mechanism involving gene expression regulation. Moreover, three *Z* candidate protein kinases have been isolated in perennial ryegrass by Van Daele *et al*. (2008). Protein kinases are involved in the regulation of many biological pathways and could play a part in this hypothetical SI mechanism.

Another hypothesis about the *Lolium* SI system can be made according to the stigma S-candidate gene. The Bradi2g35767-like gene is annotated as a protein containing an NB-ARC domain (Nucleotide-Binding – APAF-1, R protein and CED-4), involved in disease resistance in plants (Van Ooijen *et al*., 2008). Resistance mechanisms are based on the recognition of the "self" and the "non-self", which is the same recognition method as for "self" and "non-self" pollen.

Few hypotheses have been discussed here but no gene mapped within the *S*-locus region should be discarded as a potential candidate gene. The grass SI system is unique as it is governed by at least two loci, *S* and *Z* and therefore it is likely that the mechanism of recognition and repression of the "self" pollen will be unique too.

6.4. Future work

The identification of the *S* and *Z* genes is the key to understand the grass SI mechanism. The fine-mapping has delimited a region on linkage group 1 for the *S*-locus (and on linkage group 2 for the *Z*-locus, personal communication by Dr. Studer). From the sequencing, using 454, of BAC clones covering this region, 70 scaffolds covering the *S*-region have been built. However several scaffolds are showing the same annotations,

implying that the *S*-region assembly is not perfect. The *Lolium* genome sequencing project (Byrne *et al*., 2011) should be completed and available in 2013 (personal communication) and could therefore be used in order to check the assembly of the *S*-region as well as its entirety.

Once the *S*-locus region genomic sequence is completed, the annotation of the genes has to be analysed at the nucleotide level, by comparing homologous gene sequences from other species to the *Lolium* one, in order to look for variation in the coding sequence of specific domains such as the F-box domain, the ds-RNA binding domain, etc… but also in the whole coding sequence. By doing so, the gene annotation can be validated and the gene function assessed in *Lolium* as even if the *Lolium* genes are syntenic to other plant species, there is no certainty that the *S*-genes are present in rice and *Brachypodium*, both self-compatible species. The identification of some large sequence variation between the SC and SI grass species will give a better insight as some potential candidate genes. Moreover, the predicted protein sequence can be used in order to look at the predicted protein structure and localisation in the cell (e.g. transmembrane or not), by using a software program such as Protparam (Gasteiger *et al*., 2005).

The comparison work done and presented in this project for the gene Os05g0150500-like is also to be done for every candidate genes identified in the *S*-locus region. The polymorphism between transcript sequences, as well as protein predicted sequence of the different *S*-genotypes has to be assessed in order to identify allelic variations in these genes. The study has to be done separating the pollen and stigma transcripts as well in order to identify both expected components.

But the sequence analysis is not enough as alternative splicing (AS) of intron or exons is possible. The genomic sequences of every candidate gene should be processed using predicting splicing software in order to identify every splicing variant possible. The raw reads from the different cDNA libraries have to be aligned to the different splicing as the alignment can be affected by AS.

On the other hand, a "shotgun" proteomics approach can be done in order to identify the protein expressed in stigma and pollen. Shotgun proteomics is a novel method used to identify the protein and their sequences from a protein mixture using a tandem mass spectrometry (MS/MS) technique (Link *et al*., 1999). Proteins are extracted from a specific tissue (in this case, stigma and pollen separately), digested into peptides and separated according to their mass and isoelectric point. By comparing the results to some large database, it is possible to identify and characterize most of the proteins (Nesvizhskii, 2005). This technique has been used in many species, as a whole proteomic approach like in yeast (Washburn *et al*., 2001) or as a method to look at the protein present in a specific tissue such as pollen (Grobei *et al*.,2009). Like Shotgun genomic (NGS), Shotgun proteomics is becoming more and more popular and affordable. In the case of the SI study, the technique could be used in order to sequence the protein from stigma and pollen and associate them with the transcripts already available from this project. By doing so, it would be possible to identify the proteins that are actually expressed in both tissues and therefore any receptor that could be involve in the SI mechanism as well as novel proteins.

The ultimate method to identify a gene and its function (functional genomic) is to look at the phenotype of a plant with a modified gene. The most common methods are transformation using gene knockout, where a gene is mutated (by insertion of another nucleotide fragment) or the knockdown, using the RNA interference (RNAi) principle. RNAi is preferred nowadays but it has many drawbacks, especially in plants. A more preferred method would be to develop a perennial ryegrass TILLING (Targeting Induced Local Lesions in Genomes) population. TILLING has evolved as a key method for functional genomics in model plant species such as *Arabidopsis thaliana* (McCallum *et al*., 2000), as well as in major crop species such as rice (Till *et al*., 2007), wheat (Slade *et al*., 2005), maize (Till *et al*., 2004) and barley (Mejlhede *et al*., 2006). As a reverse genetics approach, it allows the identification of induced mutations in a specific genes of interest and to directly associate that mutation to its corresponding phenotype (Abe *et al*., 2012). TILLING populations are created by mutagenizing seeds from one plant, usually by using a chemical treatment such as ethyl methanesulphonate (EMS). The M1 seeds are then grown and self-pollinated in order to develop homozygous mutants (M2 plants). By screening M1 plants for candidate genes (in this project, SI genes), M2 plants from M1 plants with interesting mutations are then phenotyped. In the case of SI genes, the phenotype could not be observed directly as the original plant producing the M1 seeds has to be self-compatible. In the case of a self-compatibility of the parent due to either a mutation in *S* or *Z*, by crossing the M2 homozygous mutant with a normal self-incompatible plant twice, it is possible to reinstate the original *S* or *Z* gene and therefore if the phenotype is still self-compatible, it would be due to the mutation induced. Using the TILLING population, other loci involved in SI/SC such as *T* and *F* can be assessed, after further mapping studies.

The creation of such a TILLING population could benefit breeding programmes. Indeed, the creation of mutants for *S* or *Z* genes, the self-compatibility could be introduced in breeding populations in order to facilitate inbreeding to obtain pure lines, but the SC could also be removed by crossing plants back with the original parent in order to reinstall the SI genes, and therefore promote outbreeding and hybrid creation.

However, even without knowing the SI genes, it is possible to genotype or assess the allelic diversity of breeding populations. The method developed in this project, using closely linked markers and HRM method have shown good results. However, in order to validate and improve the method, the sequencing of the PCR products needs to be done in order to validate the sequence similarity between plants classified as having the same allelic combination. Moreover, pollination tests have to be realized, between plants within the same group for *S* and *Z* but also between groups, in order to validate the genotype prediction method.

The genetic control of self-incompatibility in grasses has been studied since its first discovery in 1956 (Lundqvist, 1956; Hayman, 1956). With this research project, the discovery of one of the SI genes, the *S*-locus, has been pushed forward. Even though more work needs to be done in order to affirm with certainty which genes are responsible for the control of SI, the answer to this mystery will be soon found with advances in genomic and proteomic sequencing technology. With the knowledge of the genetic control, the cascade reaction triggered by “self” pollen could be explored, leading to a better understanding of this unique SI system.

Appendix 1: Data from the tissue samples gene expression analysis for each *S*-locus annotation. For each annotation, the number of raw reads and the RPKM are given for the three repetitions of pollen and stigma samples, as well as the logFC, logCPM and the P Value.

- Before filtering, F1_30 samples

			Stigma						Pollen								
Scaffold name	Start position (bp)	Gene Length	F1_1 count	F1_1 RPKM	F1_2 count	F1_2 RPKM	F1_3 count	F1_3 RPKM	M1_1 count	M1_1 RPKM	M1_2 count	M1_2 RPKM	M1_3 count	M1_3 RPKM	logFC	logCPM	P Value
scaffold1	3,011	4486	353	52,144	509	58,769	444	63,093	95	24,527	56	17,064	44	24,754	-2.295	15.266	5.49E-14
scaffold1	45,508	2435	49	15,339	63	15,415	22	6,625	123	67,297	94	60,699	47	56,034	1.423	13.529	3.54E-05
scaffold2	905	4555	2,863	594,399	3,193	518,158	3,040	607,157	660	239,497	503	215,419	105	83,025	-2.395	18.024	9.51E-12
scaffold2	10,424	6899	319	5,587	673	9,213	349	5,880	45	1,378	28	1,012	25	1,668	-3.325	15.144	3.13E-20
scaffold3	31,205	896	465	23,075	642	24,901	377	17,997	1,104	95,753	881	90,181	644	121,710	1.272	16.958	2.46E-06
scaffold4	665	2609	10	2,919	14	3,194	6	1,685	97	49,482	79	47,563	9	10,004	3.059	12.615	1.43E-09
scaffold4	45,398	4555	95	2,559	86	1,811	51	1,321	90	4,237	119	6,612	71	7,283	0.718	13.925	0.016
scaffold6	905	5000	2,292	235,298	2,572	206,386	2,422	239,193	527	94,561	450	95,296	102	39,881	-2.308	17.721	1.35E-11
scaffold6	10,431	582	383	9,861	799	16,080	412	10,205	47	2,115	33	1,753	19	1,863	-3.559	15.368	4.00E-25
scaffold6	36,800	8521	174	24,822	245	27,319	143	19,625	730	182,020	786	231,301	534	290,134	2.313	16.397	7.97E-15
scaffold7	6,657	11000	12	132	6	52	4	42	40	768	35	793	25	1,046	2.616	11.878	9.55E-10
scaffold7	11,027	7000	73	1,764	95	1,794	46	1,069	404	17,064	254	12,662	132	12,149	2.330	14.900	6.97E-13
scaffold8	1,000	4652	167	2,883	207	2,793	85	1,411	280	8,448	286	10,184	142	9,335	1.072	15.099	4.54E-04
scaffold8	11,000	4693	76	9,183	114	10,767	60	6,974	192	40,549	148	36,889	92	42,338	1.235	14.323	2.73E-05
scaffold11	202	847	530	7,257	842	9,011	738	9,721	65	1,556	55	1,553	46	2,399	-3.219	15.841	1.04E-21
scaffold15	5,894	1177	149	3,870	196	3,979	125	3,123	62	2,815	66	3,536	23	2,275	-1.190	14.033	1.89E-04
scaffold17	2,518	414	123	1,683	121	1,294	107	1,408	9	215	12	339	6	313	-3.243	13.292	5.88E-16
scaffold17	6,053	477	32	849	53	1,099	57	1,455	24	1,113	10	547	8	808	-1.306	12.304	6.12E-04
scaffold23	190	386	2	28	3	33	0	0	6	149	3	88	1	54	1.396	8.739	0.148
scaffold42	1,590	818	10	2,533	35	6,930	30	7,311	14	6,199	8	4,180	4	3,859	-1.075	11.443	0.015
scaffold47	819	896	54	7,282	46	4,849	44	5,708	69	16,264	82	22,811	72	36,980	1.076	13.501	8.86E-04
scaffold56	225	2,609	23	1,065	28	1,014	27	1,203	14	1,133	3	287	6	1,058	-1.307	11.455	4.47E-03

- After filtering out the annotations with large number of reads to avoid distortion and misrepresentation of the results, F1_30 samples

			Stigma						Pollen								
Scaffold name	Start position (bp)	Gene Length	F1_1 count	F1_1 RPKM	F1_2 count	F1_2 RPKM	F1_3 count	F1_3 RPKM	M1_1 count	M1_1 RPKM	M1_2 count	M1_2 RPKM	M1_3 count	M1_3 RPKM	logFC	logCPM	P Value
scaffold1	3,011	4486	353	25,213	509	23,526	444	31,511	95	5,969	56	4,081	44	4,989	-2.594	16.226	4.62E-18
scaffold1	45,508	2435	49	6,448	63	5,364	22	2,876	123	14,237	94	12,620	47	9,818	1.125	14.31	8.55E-04
scaffold2	10,424	4555	319	22,439	673	30,634	349	24,393	45	2,784	28	2,010	25	2,792	-3.623	16.091	1.34E-30
scaffold3	31,205	6899	465	21,596	642	19,294	377	17,398	1,104	45,102	881	41,746	644	47,481	0.973	17.71	1.49E-04
scaffold4	665	896	10	3,576	14	3,240	6	2,132	97	30,513	79	28,823	9	5,109	2.762	13.392	1.25E-07
scaffold4	45,398	2609	95	11,667	86	6,835	51	6,223	90	9,723	119	14,911	71	13,842	0.419	14.709	0.21
scaffold6	10,431	4555	383	26,941	799	36,370	412	28,797	47	2,908	33	2,368	19	2,122	-3.858	16.321	1.58E-34
scaffold6	36,800	5000	174	11,150	245	10,160	143	9,105	730	41,150	786	51,389	534	54,323	2.014	17.083	1.77E-14
scaffold7	6,657	582	12	6,606	6	2,138	4	2,188	40	19,371	35	19,659	25	21,849	2.319	12.652	9.40E-08
scaffold7	11,027	8521	73	2,745	95	2,312	46	1,719	404	13,363	254	9,745	132	7,880	2.031	15.654	2.23E-09
scaffold8	1,000	11000	167	4,864	207	3,902	85	2,460	280	7,174	286	8,500	142	6,566	0.773	15.868	9.16E-03
scaffold8	11,000	7000	76	3,479	114	3,377	60	2,729	192	7,731	148	6,912	92	6,685	0.936	15.1	1.44E-03
scaffold11	202	4652	530	36,504	842	37,528	738	50,507	65	3,938	55	3,865	46	5,030	-3.518	16.822	5.32E-29
scaffold15	5,894	4693	149	10,173	196	8,659	125	8,480	62	3,724	66	4,597	23	2,493	-1.489	14.934	4.81E-07
scaffold17	2,518	847	123	46,529	121	29,620	107	40,219	9	2,995	12	4,631	6	3,603	-3.541	14.249	1.99E-21
scaffold17	6,053	1177	32	8,711	53	9,336	57	15,418	24	5,747	10	2,777	8	3,457	-1.604	13.211	2.47E-05
scaffold23	190	414	2	1,548	3	1,502	0	0	6	4,085	3	2,369	1	1,229	1.108	9.581	0.21
scaffold42	1,590	477	10	6,717	35	15,214	30	20,023	14	8,272	8	5,483	4	4,265	-1.372	12.319	1.99E-03
scaffold47	819	386	54	44,824	46	24,709	44	36,291	69	50,382	82	69,446	72	94,877	0.778	14.265	0.026
scaffold56	225	818	23	9,009	28	7,097	27	10,509	14	4,824	3	1,199	6	3,731	-1.604	12.347	2.11E-04

- Before filtering annotations, IBERS samples

			Stigma				Pollen						
Scaffold name	Start position (bp)	Gene Length	58_3_S count	58_3_S RPKM	59_3_S count	59_3_S RPKM	58_3_P count	58_3_P RPKM	59_3_P count	59_3_P RPKM	logFC	logCPM	P Value
scaffold1	3,011	4486	1,461	69,962	76	23,003	30	4,563	23	3,538	-3.958	14.382	5.31E-06
scaffold1	45,508	2435	331	33,590	65	41,692	92	29,652	59	19,231	-0.495	13.528	0.44
scaffold2	905	4555	3,407	229,306	400	170,162	998	213,334	494	106,794	-0.456	16.683	0.7
scaffold2	10,424	6899	1,046	5,939	135	4,845	74	1,334	68	1,240	-2.159	14.480	0.0021
scaffold3	31,205	896	3,429	55,161	887	90,188	1,194	61,004	3,164	163,486	0.910	17.800	0.22
scaffold4	665	2609	32	3,028	6	3,588	24	7,212	33	10,029	1.473	11.296	0.043
scaffold4	45,398	4555	4,460	38,944	207	11,425	208	5,768	131	3,674	-2.887	16.077	0.001
scaffold6	905	5000	2,776	92,387	417	87,717	850	89,845	432	46,180	-0.421	16.499	0.62
scaffold6	10,431	582	1,343	11,210	166	8,758	71	1,882	76	2,038	-2.463	14.763	0.00061
scaffold6	36,800	8521	1,552	71,775	598	174,801	2,739	402,310	1,463	217,322	1.862	17.486	0.019
scaffold7	6,657	11000	13	46	4	90	4	45	16	183	1.115	9.934	0.27
scaffold7	11,027	7000	177	1,387	37	1,832	211	5,250	153	3,850	1.661	13.933	0.011
scaffold8	1,000	4652	2,339	13,089	642	22,707	442	7,856	219	3,936	-1.277	16.281	0.022
scaffold8	11,000	4693	416	16,295	79	19,559	752	93,556	972	122,295	2.695	15.995	2.60E-05
scaffold11	202	847	1,438	6,383	114	3,198	31	437	23	328	-3.945	14.496	8.76E-07
scaffold15	5,894	1177	696	5,861	46	2,448	17	455	11	298	-3.824	13.407	5.58E-06
scaffold17	2,518	414	178	790	32	897	38	535	35	499	-0.627	12.542	0.38
scaffold17	6,053	477	46	396	9	489	14	382	16	442	0.021	10.925	1
scaffold23	190	386	15	69	2	58	8	117	3	44	0.263	9.306	0.77
scaffold42	1,590	818	5	411	3	1,557	3	782	18	4,748	2.243	9.747	0.061
scaffold47	819	896	341	14,908	94	25,974	201	27,909	461	64,734	1.501	14.899	0.038
scaffold56	225	2,609	8	120	10	949	12	572	14	675	1.410	10.415	0.52

- After filtering out the annotations with large number of reads to avoid distortion and misrepresentation of the results, IBERS samples

			Stigma				Pollen						
Scaffold name	Start position (bp)	Gene Length	58_3_S count	58_3_S RPKM	59_3_S count	59_3_S RPKM	58_3_P count	58_3_P RPKM	59_3_P count	59_3_P RPKM	logFC	logCPM	P Value
scaffold1	3,011	4486	1,461	16,834	76	5,258	30	1,080	23	730	-4.249	14.867	2.49E-05
scaffold1	45,508	2435	331	7,026	65	8,285	92	6,104	59	3,451	-0.786	13.907	0.42
scaffold2	10,424	4555	1,046	11,870	135	9,199	74	2,625	68	2,126	-2.450	14.893	3.90E-03
scaffold3	31,205	6899	3,429	25,692	887	39,904	1,194	27,959	3,164	65,311	0.618	18.019	0.44
scaffold4	665	896	32	1,846	6	2,078	24	4,327	33	5,245	1.182	11.573	0.096
scaffold4	45,398	2609	4,460	88,363	207	24,625	208	12,879	131	7,150	-3.178	16.559	2.30E-03
scaffold6	10,431	4555	1,343	15,240	166	11,311	71	2,518	76	2,376	-2.754	15.180	1.10E-03
scaffold6	36,800	5000	1,552	16,045	598	37,120	2,739	88,498	1,463	41,669	1.571	17.842	0.04
scaffold7	6,657	582	13	1,155	4	2,133	4	1,110	16	3,915	0.827	10.175	0.54
scaffold7	11,027	8521	177	1,074	37	1,348	211	4,000	153	2,557	1.369	14.267	0.033
scaffold8	1,000	11000	2,339	10,991	642	18,114	442	6,491	219	2,835	-1.569	16.664	0.043
scaffold8	11,000	7000	416	3,072	79	3,503	752	17,355	972	19,775	2.403	16.241	4.40E-04
scaffold11	202	4652	1,438	15,978	114	7,606	31	1,077	23	704	-4.236	14.961	7.11E-06
scaffold15	5,894	4693	696	7,666	46	3,042	17	585	11	334	-4.114	13.878	2.95E-05
scaffold17	2,518	847	178	10,863	32	11,726	38	7,248	35	5,885	-0.919	12.907	0.32
scaffold17	6,053	1177	46	2,020	9	2,373	14	1,922	16	1,936	-0.269	11.262	0.92
scaffold23	190	414	15	1,873	2	1,499	8	3,122	3	1,032	-0.024	9.697	0.9
scaffold42	1,590	477	5	542	3	1,952	3	1,016	18	5,374	1.956	9.928	0.13
scaffold47	819	386	341	45,664	94	75,581	201	84,124	461	170,080	1.209	15.107	0.11
scaffold56	225	818	8	506	10	3,794	12	2,370	14	2,437	1.121	10.698	0.63

Appendix 2: Data from the pollination samples gene expression analysis for each *S*-locus annotation. For each annotation, the number of raw reads and the RPKM are given for the three repetitions of self-pollination (SI) and two repetitions of the cross-pollination samples (SC), as well as the logFC, logCPM and the P Value (* indicates a significant change in the expression).

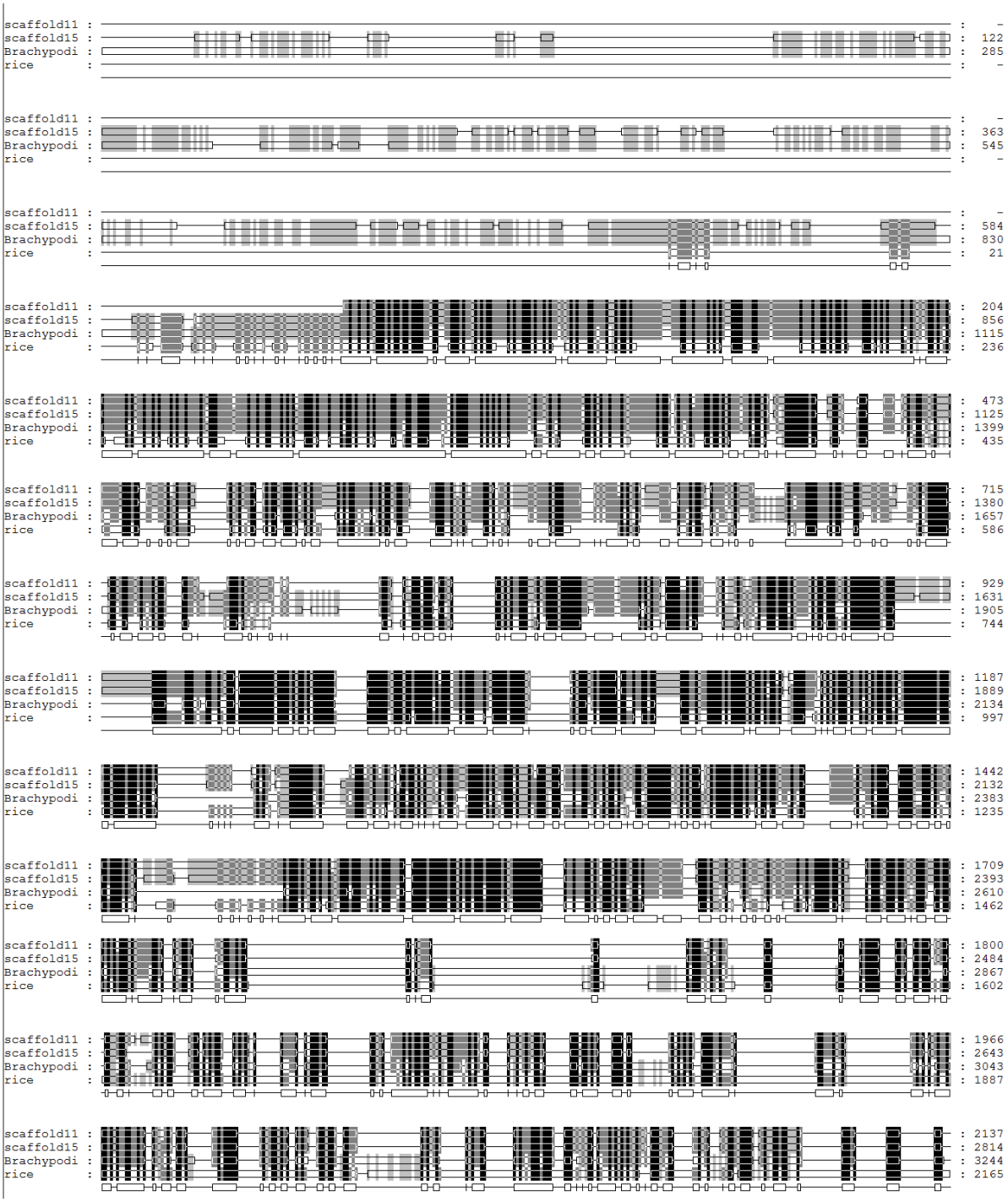
- Before filtering annotations

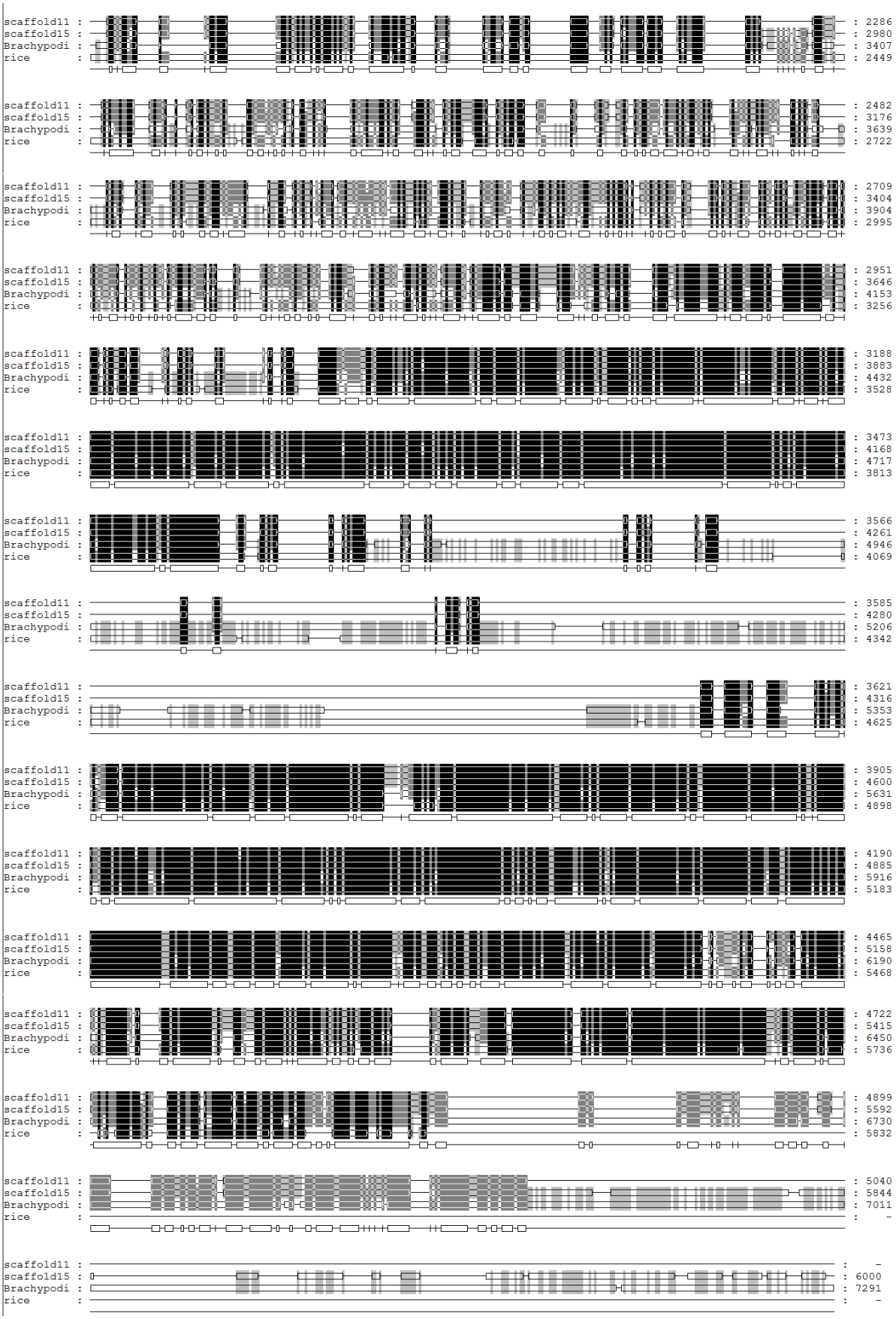
			Self-pollination (self-incompatible pollination)						Cross-pollination (compatible)						
Scaffold name	Start position (bp)	Gene Length	SI_1 count	SI_1 RPKM	SI_2 count	SI_2 RPKM	SI_3 count	SI_3 RPKM	SC_1 count	SC_1 RPKM	SC_2 count	SC_2 RPKM	logFC	logCPM	P Value
scaffold1	3,011	818	828	61,221	280	30,846	507	58,666	270	30,659	342	45,554	0.541	15.203	0.22
scaffold1	45,508	386	64	10,028	57	13,307	42	10,299	59	14,197	76	21,453	-0.586	12.361	0.15
scaffold2	905	582	4,925	511,806	3,034	469,772	2,549	414,550	3,206	511,666	2,394	448,181	0.049	18.065	0.92
scaffold2	10,424	6,899	1,143	10,020	992	12,957	639	8,767	955	12,858	757	11,955	-0.163	16.242	0.65
scaffold3	31,205	2,435	947	23,522	849	31,420	1,182	45,946	693	26,435	560	25,058	0.390	16.205	0.22
scaffold4	665	414	14	2,045	17	3,700	7	1,600	21	4,712	27	7,106	-1.189	10.573	0.018*
scaffold4	45,398	4,486	137	1,847	122	2,451	109	2,300	95	1,967	107	2,599	0.006	13.287	0.94
scaffold6	905	1,177	4,047	207,960	2,504	191,713	1,989	159,952	2,575	203,211	2,092	193,659	0.013	17.778	0.98
scaffold6	10,431	4,693	1,292	16,651	1,230	23,618	872	17,587	1,215	24,048	950	22,056	-0.211	16.567	0.62
scaffold6	36,800	847	405	28,920	399	42,451	659	73,643	265	29,061	273	35,118	0.584	15.148	0.11
scaffold7	6,657	11,000	19	104	24	197	11	95	11	93	26	258	-0.312	10.639	0.59
scaffold7	11,027	5,000	148	1,790	159	2,866	154	2,915	135	2,508	149	3,247	-0.160	13.700	0.82
scaffold8	1,000	7,000	274	2,367	346	4,454	241	3,259	217	2,879	311	4,841	-0.154	14.593	0.78
scaffold8	11,000	1,000	122	7,379	133	11,985	147	13,914	83	7,709	119	12,966	0.134	13.402	0.62
scaffold11	202	8,825	1,366	9,362	452	4,615	879	9,428	482	5,073	535	6,605	0.548	15.944	0.2
scaffold15	5,894	4,652	333	4,329	123	2,383	211	4,293	115	2,296	168	3,935	0.377	13.986	0.41
scaffold17	2,518	8,831	248	1,698	114	1,163	146	1,565	152	1,599	127	1,567	0.005	13.719	1
scaffold17	6,053	4,555	48	637	87	1,721	49	1,018	57	1,162	47	1,124	-0.036	12.317	0.94
scaffold23	190	8,521	0	0	10	106	3	33	3	33	8	102	-0.606	8.743	0.61
scaffold42	1,590	477	17	2,156	19	3,589	13	2,580	8	1,558	19	4,340	0.001	10.376	1
scaffold47	819	896	90	6,075	61	6,135	99	10,458	83	8,604	37	4,499	0.199	12.672	0.51
scaffold56	225	2,609	34	788	26	898	28	1,016	26	926	24	1,002	-0.043	11.235	1

- After filtering out the annotations with large number of reads to avoid distortion and misrepresentation of the results

			Self-pollination (self-incompatible pollination)						Cross-pollination (compatible)						
Scaffold name	Start position (bp)	Gene Length	SI_1 count	SI_1 RPKM	SI_2 count	SI_2 RPKM	SI_3 count	SI_3 RPKM	SC_1 count	SC_1 RPKM	SC_2 count	SC_2 RPKM	logFC	logCPM	P Value
scaffold1	3,011	4486	828	24,408	280	11,228	507	18,752	270	12,074	342	16,248	0.644	16.224	0.12
scaffold1	45,508	2435	64	3,476	57	4,211	42	2,862	59	4,861	76	6,652	-0.483	13.357	0.2
scaffold2	10,424	4555	1,143	33,183	992	39,177	639	23,276	955	42,058	757	35,420	-0.059	17.257	0.8
scaffold3	31,205	6899	947	18,152	849	22,137	1,182	28,427	693	20,150	560	17,300	0.493	17.181	0.16
scaffold4	665	896	14	2,066	17	3,413	7	1,296	21	4,702	27	6,422	-1.086	11.571	0.02 *
scaffold4	45,398	2609	137	6,944	122	8,412	109	6,932	95	7,304	107	8,741	0.110	14.283	0.77
scaffold6	10,431	4555	1,292	37,509	1,230	48,576	872	31,763	1,215	53,508	950	44,451	-0.107	17.573	0.71
scaffold6	36,800	5000	405	10,711	399	14,355	659	21,868	265	10,632	273	11,637	0.687	16.105	0.061
scaffold7	6,657	582	19	4,317	24	7,418	11	3,136	11	3,791	26	9,521	-0.209	11.635	0.72
scaffold7	11,027	8521	148	2,297	159	3,357	154	2,999	135	3,178	149	3,727	-0.057	14.680	0.89
scaffold8	1,000	11000	274	3,294	346	5,658	241	3,635	217	3,957	311	6,026	-0.050	15.573	0.89
scaffold8	11,000	7000	122	2,305	133	3,418	147	3,484	83	2,379	119	3,623	0.237	14.373	0.5
scaffold11	202	4652	1,366	38,831	452	17,478	879	31,351	482	20,785	535	24,511	0.651	16.964	0.12
scaffold15	5,894	4693	333	9,383	123	4,715	211	7,460	115	4,916	168	7,630	0.481	15.000	0.24
scaffold17	2,518	847	248	38,720	114	24,212	146	28,600	152	35,999	127	31,957	0.109	14.741	0.83
scaffold17	6,053	1177	48	5,393	87	13,297	49	6,907	57	9,715	47	8,511	0.067	13.306	0.9
scaffold23	190	414	0	0	10	4,345	3	1,202	3	1,454	8	4,118	-0.504	9.725	0.6
scaffold42	1,590	477	17	4,713	19	7,165	13	4,522	8	3,364	19	8,489	0.103	11.375	0.9
scaffold47	819	386	90	30,833	61	28,428	99	42,555	83	43,135	37	20,429	0.303	13.669	0.49
scaffold56	225	818	34	5,497	26	5,718	28	5,679	26	6,376	24	6,253	0.060	12.240	0.95

Appendix 3: Alignment of genomic sequences from Os05g0150500-like genes from rice, *Brachypodium* (Bradi2g35720.1) and scaffold 11 and 15 from *Lolium*. The sequences similarity are represented by shades of greys, the darker the more sequences are common, with black for the four sequences being identical. There are mainly three black blocks, which represent the coding region (exons).




scaffold11 : : 2286
scaffold15 : : 2980
Brachypodi : : 3407
rice : : 2449
scaffold11 : : 2482
scaffold15 : : 3176
Brachypodi : : 3639
rice : : 2722
scaffold11 : : 2709
scaffold15 : : 3404
Brachypodi : : 3904
rice : : 2995
scaffold11 : : 2951
scaffold15 : : 3646
Brachypodi : : 4153
rice : : 3256
scaffold11 : : 3188
scaffold15 : : 3883
Brachypodi : : 4432
rice : : 3528
scaffold11 : : 3473
scaffold15 : : 4168
Brachypodi : : 4717
rice : : 3813
scaffold11 : : 3566
scaffold15 : : 4261
Brachypodi : : 4946
rice : : 4069
scaffold11 : : 3585
scaffold15 : : 4280
Brachypodi : : 5206
rice : : 4342
scaffold11 : : 3621
scaffold15 : : 4316
Brachypodi : : 5353
rice : : 4625
scaffold11 : : 3905
scaffold15 : : 4600
Brachypodi : : 5631
rice : : 4898
scaffold11 : : 4190
scaffold15 : : 4885
Brachypodi : : 5916
rice : : 5183
scaffold11 : : 4465
scaffold15 : : 5158
Brachypodi : : 6190
rice : : 5468
scaffold11 : : 4722
scaffold15 : : 5415
Brachypodi : : 6450
rice : : 5736
scaffold11 : : 4899
scaffold15 : : 5592
Brachypodi : : 6730
rice : : 5832
scaffold11 : : 5040
scaffold15 : : 5844
Brachypodi : : 7011
rice : : -
scaffold11 : : -
scaffold15 : : 6000
Brachypodi : : 7291
rice : : -

Appendix 4: Sequences of the *Lolium perenne* Os05g0150500-like gene predicted with the software Augustus.

Coding sequence:

>CDS

atgggccgcgccggcggccccgccgcgcccccgtggcactcgctcccggacgaggtctgggagcacgccttctccttcctccccgccgac
gccgaccgcggcgccgccgcctgcgcctgccacggctggctccgcgccgagcgccgctcccgccgccgcctcgtcgtgcccaactgctac
gccaccgacccgcgcgacgccgtcgacaggttcccctccgtccgcgccgccgaggtcaagggcaagccgcacttcgccgacttcggcctc
ctgccgccctcctggggcgcctatgccgcgccctgggtcgccgccgcagccgacggctggccgctgctcgaggagctcagcttcaagcgc
atgttcgtcaccgacgagtgccttgagatgatcgcgtcatccttcaggaacttccaggtcctgcgactaaactcgtgcgagggcttcaccaccgc
cggcctcgccgccattaccgaaggttgcagaaatttaaaagagcttgacctgcaagagaactacattgatgattgttcaagtcattggctcagca
actttccagaatgctatacttctcttgaagctctgaatttttcatgcttacacggggaggtcaatttcactgtacttgagaggctagtaagcagatgcc
gcaacctcaagactctgaagctcaacaatgcaatccctcttgacaatgttgctagcctgcttcgttgtacaagccttagaaggctttctgggacttg
ggacgctgttccagattacctgccagcattctattgtgtatgtgaaggcctcacatctcttaatctgagttatgccaccgtgcaaggccctgagctc
atcaaattcattagcagatgcaagaatctgctgcagttatgggtgatggacctcattgaggaccatggtctatctgttgtggcatcaagttgcagta
aactgcaagagttgcgggtcttcccttccgatccttttggtcataatggcgggcaagttttcttgacagaaagaggtcttgttgatgtttctgccagtt
gtcccaaattggagtcagttctttacttctgcagccggatgacgaatgaggctcttgttatgattgcaaagaaccgtccaaacttcacttgcttccgc
ttagccctccttgagccccgttctccggattacatcacacggcagcctcttgatgctggtttcagtgccattgtggaatcatgcaaggggcttagg
cgcctctctatgtctggtcttctcacagatcttgtattcaaatcaatcggtgcacatgctgatcgtcttgagatgctatcactcgcatttgctggagac
agcgatctaggcctgaatgacatcctctctggctgcaagagcctgaagaagctagagatcagggactgcccgtttggggataaagcgttgctg
gcaaatgctgccaagctggagacaatgcgatccctttggatgaactcgtgctcgttgaccgtgggcgggtgccgactgcttgcactcaagatgc
ctcaccttactgtggagataataaacgatcctggagagacatgtccagtggagtcactcccgtttgatagccctgtcgagaaattgtatgtctacc
ggactcttgcaggtccaagatctgacacaccagactgtgtccagattgtttag

Protein sequence

>protein

MGRAGGPAAPPWHSLPDEVWEHAFSFLPADADRGAAACACHGWLRAERRSRRRLVVPNCYATDPRDAVDR
FPSVRAAEVKGKPHFADFGLLPPSWGAYAAPWVAAAADGWPLLEELSFKRMFVTDECLEMIASSFRNFQVLRL
NSCEGFTTAGLAAITEGCRNLKELDLQENYIDDCSSHWLSNFPECYTSLEALNFSCLHGEVNFTVLERLVSRCRNL
KTLKLNNAIPLDNVASLLRCTSLRRLSGTWDAVPDYLPAFYCVCEGLTSLNLSYATVQGPELIKFISRCKNLLQLWV
MDLIEDHGLSVVASSCSKLQELRVFPSDPFGHNGGQVFLTERGLVDVSASCPKLESVLYFCSRMTNEALVMIAK
NRPNFTCFRLALLEPRSPDYITRQPLDAGFSAIVESCKGLRRLSMSGLLTDLVFKSIGAHADRLEMLSLAFAGDSDL
GLNDILSGCKSLKKLEIRDCPFGDKALLANAAKLETMRSLWMNSCSLTVGGCRLLALKMPHLTVEIINDPGETCP
VESLPFDSPVEKLYVYRTLAGPRSDTPDCVQIV

Appendix 5: Sequences of the *Lolium perenne* Os05g0150500-like gene predicted with the software FGENESH.

Coding sequence:

>CDS

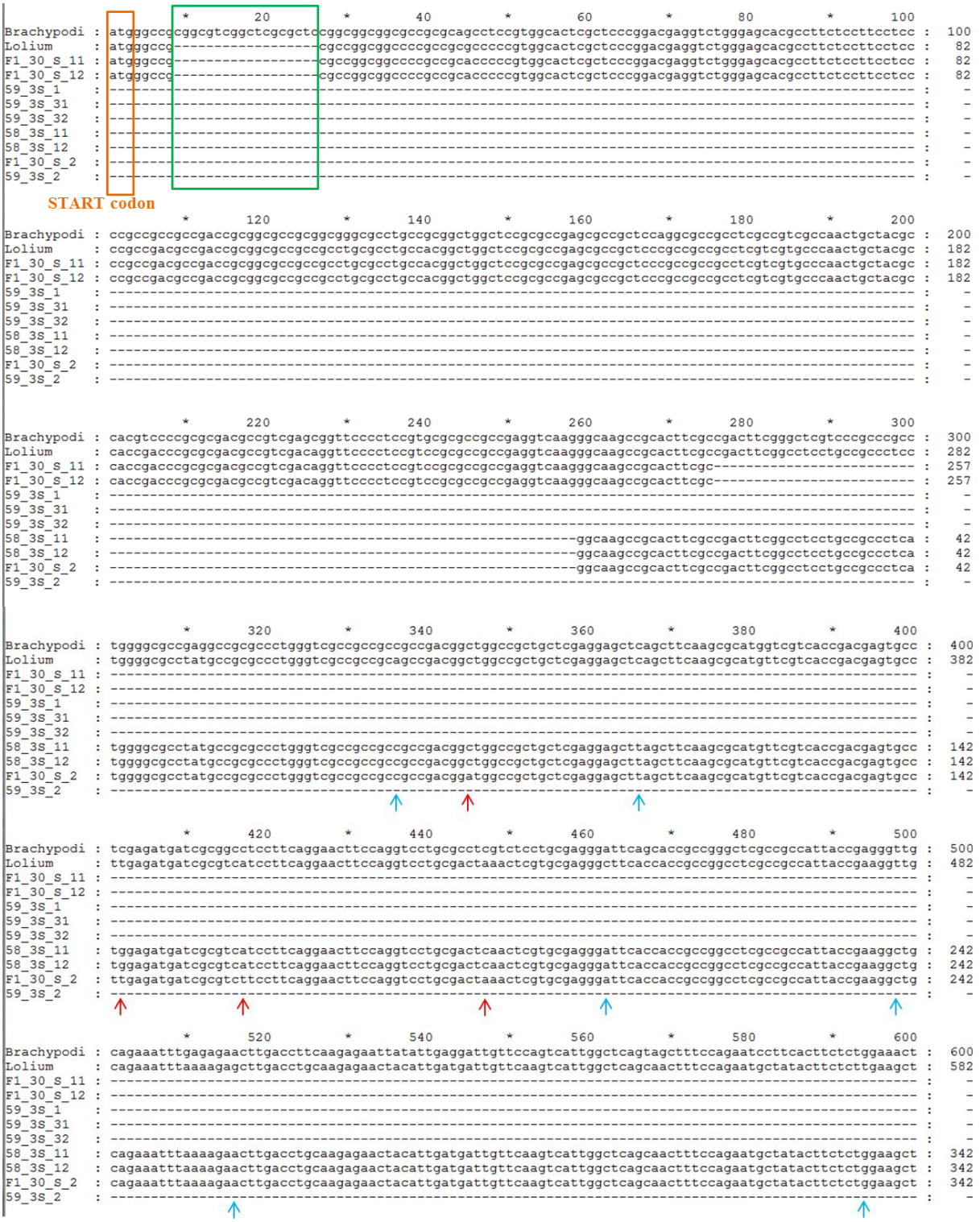
ATGGGCCGCGCCGGCGGCCCCGCCGCGCCCCCGTGGCACTCGCTCCCGGACGAGGTCTGGGAG
CACGCCTTCTCCTTCCTCCCCGCCGACGCCGACCGCGGCGCCGCCGCCTGCGCCTGCCACGGCT
GGCTCCGCGCCGAGCGCCGCTCCCGCCGCCGCCTCGTCGTGCCCAACTGCTACGCCACCGACCC
GCGCGACGCCGTCGACAGGTTCCCCTCCGTCCGCGCCGCCGAGGTCAAGGGCAAGCCGCACTTC
GCCGACTTCGGCCTCCTGCCGCCCTCCTGGGGCGCCTATGCCGCGCCCTGGGTCGCCGCCGCAG
CCGACGGCTGGCCGCTGCTCGAGGAGCTCAGCTTCAAGCGCATGTTCGTCACCGACGAGTGCCT
TGAGATGATCGCGTCATCCTTCAGGAACTTCCAGGTCCTGCGACTAAACTCGTGCGAGGGCTTC
ACCACCGCCGGCCTCGCCGCCATTACCGAAGGTTGCAGAAATTTAAAAGAGCTTGACCTGCAAG
AGAACTACATTGATGATTGTTCAAGTCATTGGCTCAGCAACTTTCCAGAATGCTATACTTCTCTT
GAAGCTCTGAATTTTTCATGCTTACACGGGGAGGTCAATTTCACTGTACTTGAGAGGCTAGTAA
GCAGATGCCGCAACCTCAAGACTCTGAAGCTCAACAATGCAATCCCTCTTGACAATGTTGCTAG
CCTGCTTCGTAAGGCTCCGCAAATAATAGAACTCGGAACTGGCAAATTTTCTGCTGACTATCAT
CCAGATCTTTTTGCGAAGGTTGAAGCGGCATTTGCAGGTTGTACAAGCCTTAGAAGGCTTTCTG
GGACTTGGGACGCTGTTCCAGATTACCTGCCAGCATTCTATTGTGTATGTGAAGGCCTCACATCT
CTTAATCTGAGTTATGCCACCGTGCAAGGCCCTGAGCTCATCAAATTCATTAGCAGATGCAAGA
ATCTGCTGCAGTTATGGGTGATGGACCTCATTGAGGACCATGGTCTATCTGTTGTGGCATCAAG
TTGCAGTAAACTGCAAGAGTTGCGGGTCTTCCCTTCCGATCCTTTTGGTCATAATGGCGGGCAA
GTTTTCTTGACAGAAAGAGGTCTTGTTGATGTTTCTGCCAGTTGTCCCAAATTGGAGTCAGTTCT
TTACTTCTGCAGCCGGATGACGAATGAGGCTCTTGTTATGATTGCAAAGAACCGTCCAAACTTC
ACTTGCTTCCGCTTAGCCCTCCTTGAGCCCCGTTCTCCGGATTACATCACACGGCAGCCTCTTGA
TGCTGGTTTCAGTGCCATTGTGGAATCATGCAAGGGGCTTAGGCGCCTCTCTATGTCTGGTCTTC
TCACAGATCTTGTATTCAAATCAATCGGTGCACATGCTGATCGTCTTGAGATGCTATCACTCGCA
TTTGCTGGAGACAGCGATCTAGGCCTGAATGACATCCTCTCTGGCTGCAAGAGCCTGAAGAAGC
TAGAGATCAGGGACTGCCCGTTTGGGGATAAAGCGTTGCTGGCAAATGCTGCCAAGCTGGAGA
CAATGCGATCCCTTTGGATGAACTCGTGCTCGTTGACCGTGGGCGGGTGCCGACTGCTTGCACT
CAAGATGCCTCACCTTACTGTGGAGATAATAAACGATCCTGGAGAGACATGTCCAGTGGAGTCA
CTCCCGTTTGATAGCCCTGTCGAGAAATTGTATGTCTACCGGACTCTTGCAGGTCCAAGATCTGA
CACACCAGACTGTGTCCAGATTGTTTAG

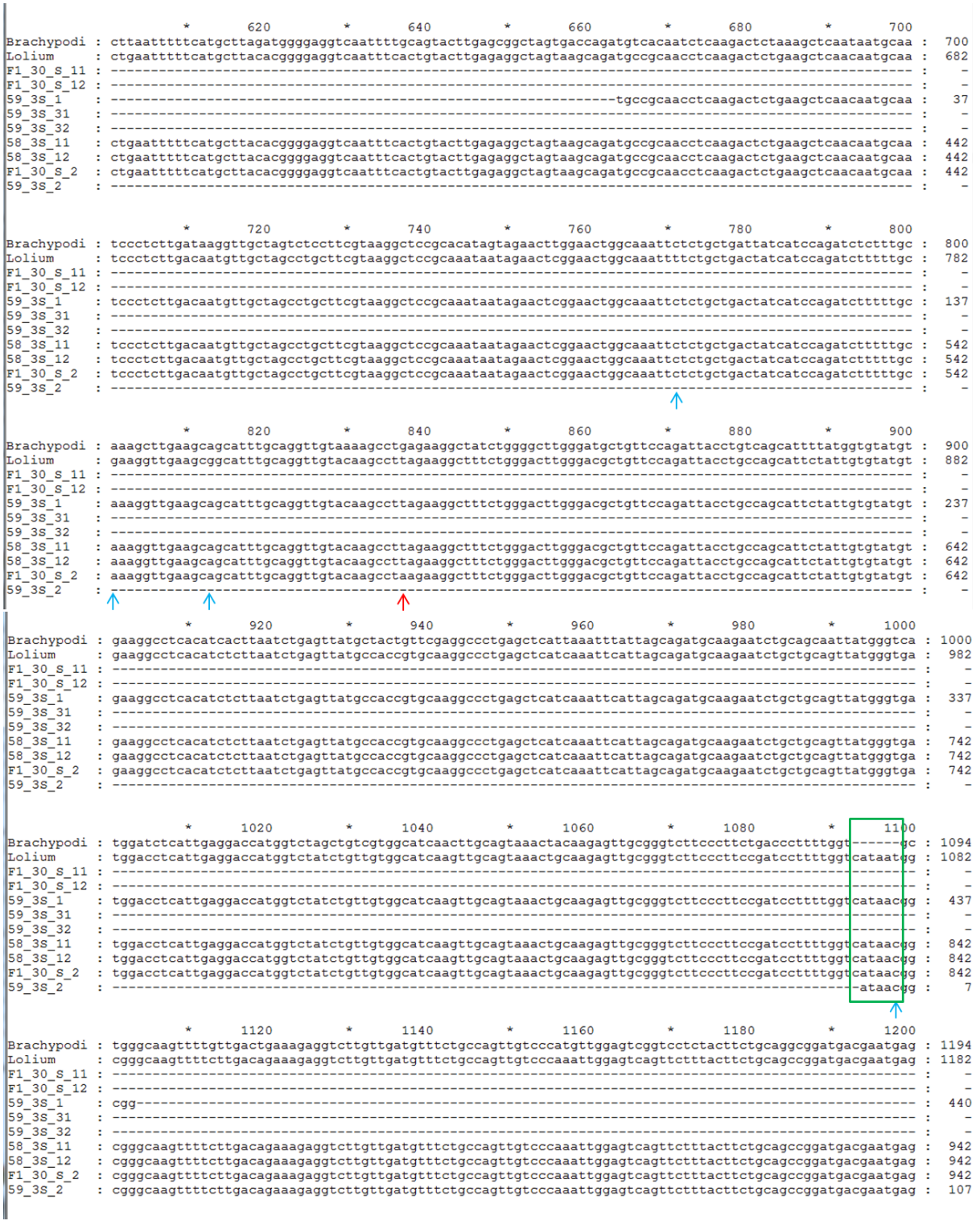
Protein sequence

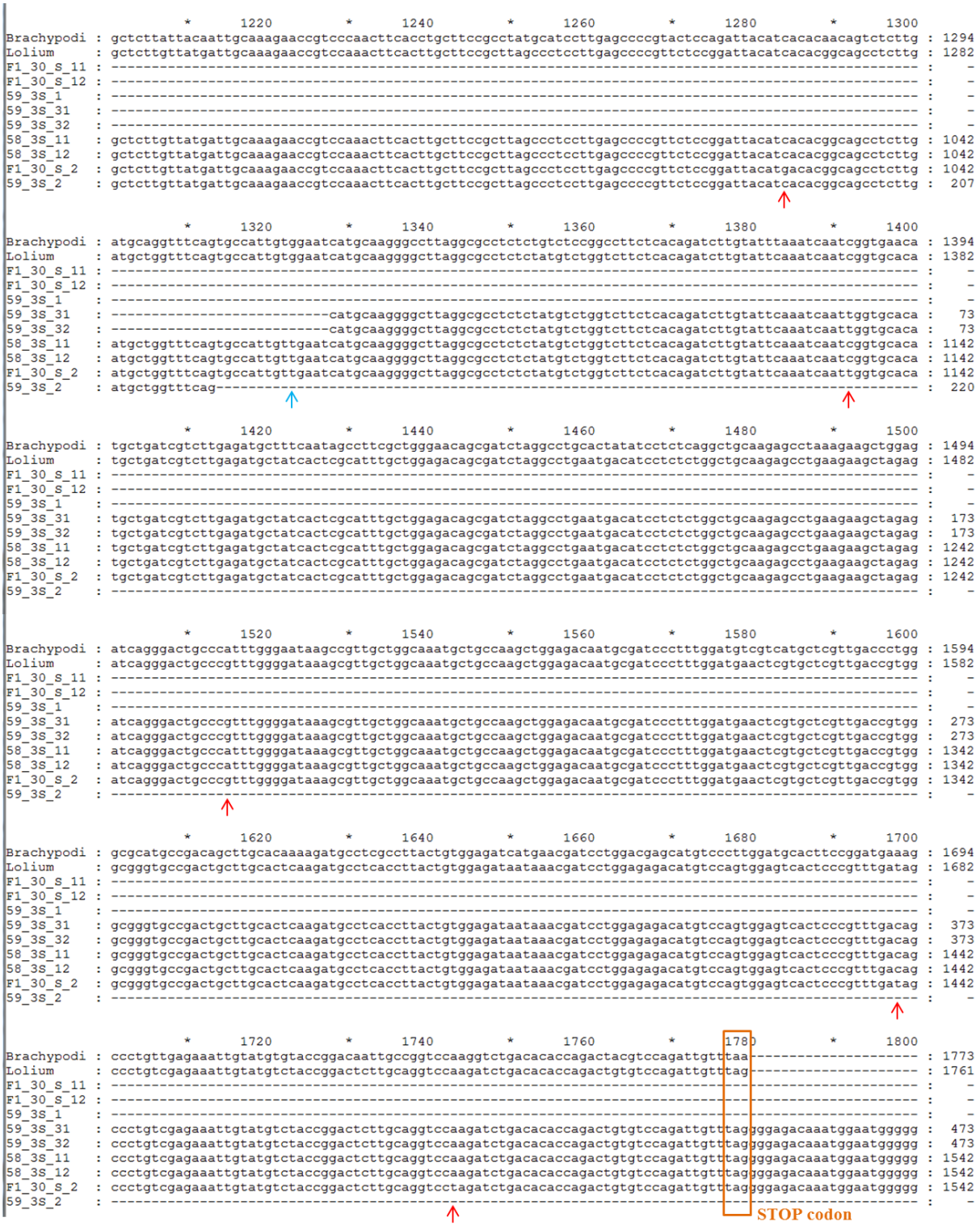
>protein

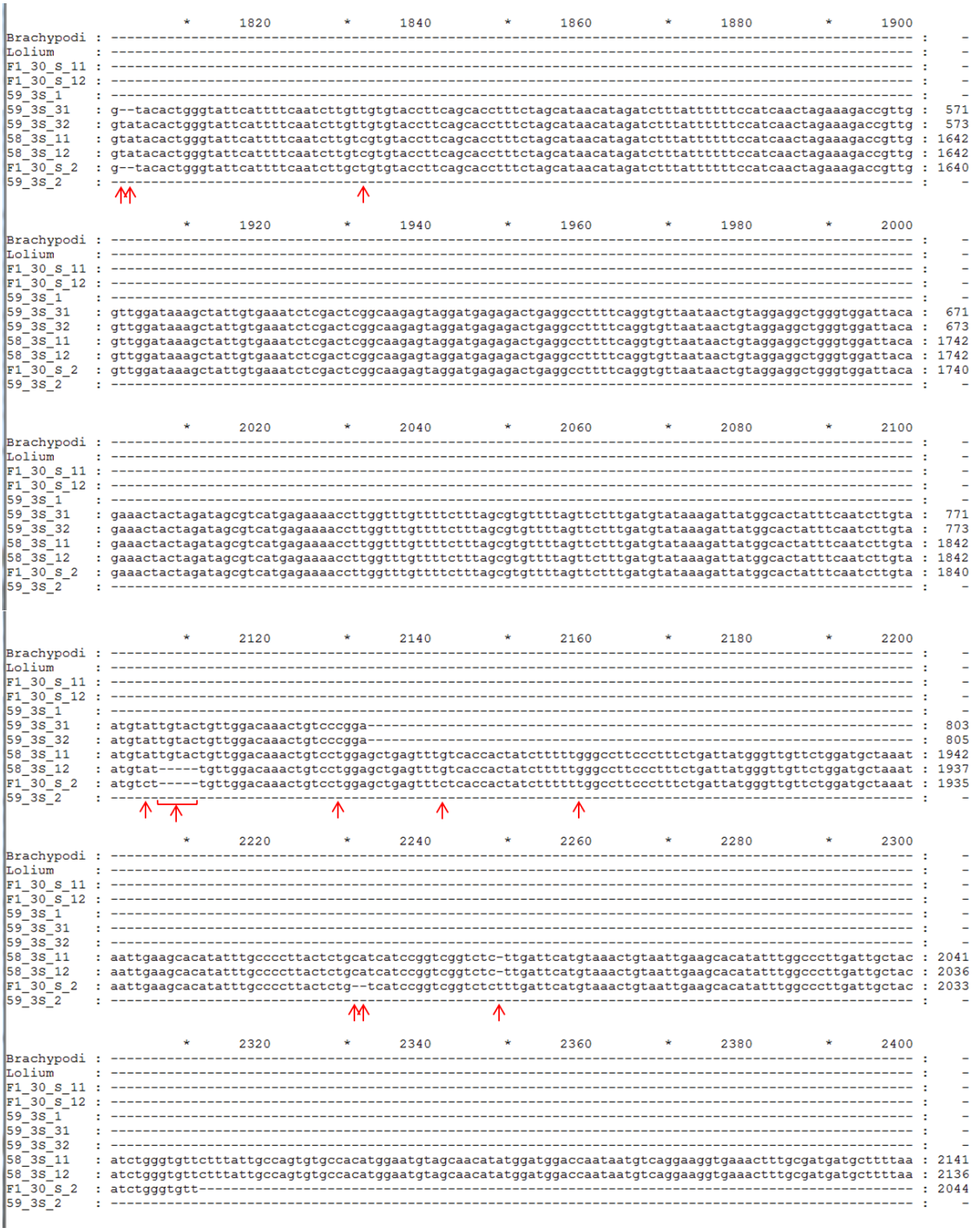
MGRAGGPAAPPWHSLPDEVWEHAFSFLPADADRGAAACACHGWLRAERRSRRR
LVVPNCYATDPRDAVDRFPSVRAAEVKGKPHFADFGLLPPSWGAYAAPWVAAA
ADGWPLLEELSFKRMFVTDECLEMIASSFRNFQVLRLNSCEGFTTAGLAAITEGCR
NLKELDLQENYIDDCSSHWLSNFPECYTSLEALNFSCLHGEVNFTVLERLVSRCRN
LKTLKLNNAIPLDNVASLLRKAPQIIELGTGKFSADYHPDLFAKVEAAFAGCTSLR
RLSGTWDAVPDYLPAFYCVCEGLTSLNLSYATVQGPELIKFISRCKNLLQLWVMD
LIEDHGLSVVASSCSKLQELRVFPSDPFGHNGGQVFLTERGLVDVSASCPKLESVL
YFCSRMTNEALVMIAKNRPNFTCFRLALLEPRSPDYITRQPLDAGFSAIVESCKGLR
RLSMSGLLTDLVFKSIGAHADRLEMLSLAFAGDSDLGLNDILSGCKSLKKLEIRDC
PFGDKALLANAAKLETMRSLWMNSCSLTVGGCRLLALKMPHLTVEIINDPGETCP
VESLPFDSPVEKLYVYRTLAGPRSDTPDCVQIV

Appendix 6: Alignment of the *Brachypodium* and *Lolium* predicted CDS for the Os05g0150500-like gene with *Lolium* transcripts from stigma samples with different genotypes. The red and blue arrows are indicating the polymorphism, between *Lolium* genotypes (transcripts) or between *Lolium* predicted CDS and *Lolium* transcripts, respectively. The green frames are representing the large insertion between *Brachypodium* and *Lolium* sequences. The start and stop codons are indicated by an orange frame.






STOP codon


* 1820 * 1840 * 1860 * 1880 * 1900
Brachypodi : ---------- : -
Lolium : ---------- : -
F1_30_S_11 : ---------- : -
F1_30_S_12 : ---------- : -
59_3S_1 : ---------- : -
59_3S_31 : g--tacactgggtattcattttcaatcttgttgtgtaccttcagcacctttctagcataacatagatctttatttttttccatcaactagaaagaccgttg : 571
59_3S_32 : gtatacactgggtattcattttcaatcttgttgtgtaccttcagcacctttctagcataacatagatctttatttttttccatcaactagaaagaccgttg : 573
58_3S_11 : gtatacactgggtattcattttcaatcttgtcgtgtaccttcagcacctttctagcataacatagatctttatttttttccatcaactagaaagaccgttg : 1642
58_3S_12 : gtatacactgggtattcattttcaatcttgtcgtgtaccttcagcacctttctagcataacatagatctttatttttttccatcaactagaaagaccgttg : 1642
F1_30_S_2 : g--tacactgggtattcattttcaatcttgctgtgtaccttcagcacctttctagcataacatagatctttatttttttccatcaactagaaagaccgttg : 1640
59_3S_2 : ---------- : -
* 1920 * 1940 * 1960 * 1980 * 2000
Brachypodi : ---------- : -
Lolium : ---------- : -
F1_30_S_11 : ---------- : -
F1_30_S_12 : ---------- : -
59_3S_1 : ---------- : -
59_3S_31 : gttggataaagctattgtgaaatctcgactcggcaagagtaggatgagagactgaggccttttcaggtgttaataactgtaggaggctgggtggattaca : 671
59_3S_32 : gttggataaagctattgtgaaatctcgactcggcaagagtaggatgagagactgaggccttttcaggtgttaataactgtaggaggctgggtggattaca : 673
58_3S_11 : gttggataaagctattgtgaaatctcgactcggcaagagtaggatgagagactgaggccttttcaggtgttaataactgtaggaggctgggtggattaca : 1742
58_3S_12 : gttggataaagctattgtgaaatctcgactcggcaagagtaggatgagagactgaggccttttcaggtgttaataactgtaggaggctgggtggattaca : 1742
F1_30_S_2 : gttggataaagctattgtgaaatctcgactcggcaagagtaggatgagagactgaggccttttcaggtgttaataactgtaggaggctgggtggattaca : 1740
59_3S_2 : ---------- : -
* 2020 * 2040 * 2060 * 2080 * 2100
Brachypodi : ---------- : -
Lolium : ---------- : -
F1_30_S_11 : ---------- : -
F1_30_S_12 : ---------- : -
59_3S_1 : ---------- : -
59_3S_31 : gaaactactagatagcgtcatgagaaaaccttggtttgttttctttagcgtgttttagttctttgatgtataaagattatggcactatttcaatcttgta : 771
59_3S_32 : gaaactactagatagcgtcatgagaaaaccttggtttgttttctttagcgtgttttagttctttgatgtataaagattatggcactatttcaatcttgta : 773
58_3S_11 : gaaactactagatagcgtcatgagaaaaccttggtttgttttctttagcgtgttttagttctttgatgtataaagattatggcactatttcaatcttgta : 1842
58_3S_12 : gaaactactagatagcgtcatgagaaaaccttggtttgttttctttagcgtgttttagttctttgatgtataaagattatggcactatttcaatcttgta : 1842
F1_30_S_2 : gaaactactagatagcgtcatgagaaaaccttggtttgttttctttagcgtgttttagttctttgatgtataaagattatggcactatttcaatcttgta : 1840
59_3S_2 : ---------- : -
* 2120 * 2140 * 2160 * 2180 * 2200
Brachypodi : ---------- : -
Lolium : ---------- : -
F1_30_S_11 : ---------- : -
F1_30_S_12 : ---------- : -
59_3S_1 : ---------- : -
59_3S_31 : atgtattgtactgttggacaaactgtcccgga---------- : 803
59_3S_32 : atgtattgtactgttggacaaactgtcccgga---------- : 805
58_3S_11 : atgtattgtactgttggacaaactgtcctggagctgagtttgtcaccactatctttttgggccttccctttctgattatgggttgttctggatgctaaat : 1942
58_3S_12 : atgtat-----tgttggacaaactgtcctggagctgagtttgtcaccactatctttttgggccttccctttctgattatgggttgttctggatgctaaat : 1937
F1_30_S_2 : atgtct-----tgttggacaaactgtcctggagctgagtttctcaccactatctttttgggccttccctttctgattatgggttgttctggatgctaaat : 1935
59_3S_2 : ---------- : -
* 2220 * 2240 * 2260 * 2280 * 2300
Brachypodi : ---------- : -
Lolium : ---------- : -
F1_30_S_11 : ---------- : -
F1_30_S_12 : ---------- : -
59_3S_1 : ---------- : -
59_3S_31 : ---------- : -
59_3S_32 : ---------- : -
58_3S_11 : aattgaagcacatatttgccccttactctgcatcatccggtcggtctc-ttgattcatgtaaactgtaattgaagcacatatttggcccttgattgctac : 2041
58_3S_12 : aattgaagcacatatttgccccttactctgcatcatccggtcggtctc-ttgattcatgtaaactgtaattgaagcacatatttggcccttgattgctac : 2036
F1_30_S_2 : aattgaagcacatatttgccccttactctg--tcatccggtcggtctctttgattcatgtaaactgtaattgaagcacatatttggcccttgattgctac : 2033
59_3S_2 : ---------- : -
* 2320 * 2340 * 2360 * 2380 * 2400
Brachypodi : ---------- : -
Lolium : ---------- : -
F1_30_S_11 : ---------- : -
F1_30_S_12 : ---------- : -
59_3S_1 : ---------- : -
59_3S_31 : ---------- : -
59_3S_32 : ---------- : -
58_3S_11 : atctgggtgttctttattgccagtgtgccacatggaatgtagcaacatatggatggaccaataatgtcaggaaggtgaaactttgcgatgatgcttttaa : 2141
58_3S_12 : atctgggtgttctttattgccagtgtgccacatggaatgtagcaacatatggatggaccaataatgtcaggaaggtgaaactttgcgatgatgcttttaa : 2136
F1_30_S_2 : atctgggtgtt---------- : 2044
59_3S_2 : ---------- : -

```
                    *      2420         *      2440         *      2460         *      2480         *      2500
Brachypodi : ---------------------------------------------------------------------------------------------------- :     -
Lolium     : ---------------------------------------------------------------------------------------------------- :     -
F1_30_S_11 : ---------------------------------------------------------------------------------------------------- :     -
F1_30_S_12 : ---------------------------------------------------------------------------------------------------- :     -
59_3S_1    : ---------------------------------------------------------------------------------------------------- :     -
59_3S_31   : ---------------------------------------------------------------------------------------------------- :     -
59_3S_32   : ---------------------------------------------------------------------------------------------------- :     -
58_3S_11   : cttttgatagaaactcaatagatgctaaataaactggtgttgatctacatatagggtggatgactcatttgctctgatcttaagaaattgcatgggtgtt : 2241
58_3S_12   : cttttgatagaaactcaatagatgctaaataaactggtgttgatctacatatagggtggatgactcatttgctctgatcttaagaaattgcatgggtgtt : 2236
F1_30_S_2  : ---------------------------------------------------------------------------------------------------- :     -
59_3S_2    : ---------------------------------------------------------------------------------------------------- :     -


                    *      2520         *      2540         *      2560         *      2580         *      2600
Brachypodi : ---------------------------------------------------------------------------------------------------- :     -
Lolium     : ---------------------------------------------------------------------------------------------------- :     -
F1_30_S_11 : ---------------------------------------------------------------------------------------------------- :     -
F1_30_S_12 : ---------------------------------------------------------------------------------------------------- :     -
59_3S_1    : ---------------------------------------------------------------------------------------------------- :     -
59_3S_31   : ---------------------------------------------------------------------------------------------------- :     -
59_3S_32   : ---------------------------------------------------------------------------------------------------- :     -
58_3S_11   : catccctactttttttttttttttttgcaattacatggaagctgagttagtcacccctttctttggcattccctttctgcttatggcatgtatgatgtgctat : 2341
58_3S_12   : catccctactttttttttttttttttgcaattacatggaagctgagttagtcacccctttctttggcattccctttctgcttatggcatgtatgatgtgctat : 2336
F1_30_S_2  : ---------------------------------------------------------------------------------------------------- :     -
59_3S_2    : ---------------------------------------------------------------------------------------------------- :     -


                    *      2620         *      2640         *      2660         *      2680         *      2700
Brachypodi : ---------------------------------------------------------------------------------------------------- :     -
Lolium     : ---------------------------------------------------------------------------------------------------- :     -
F1_30_S_11 : ---------------------------------------------------------------------------------------------------- :     -
F1_30_S_12 : ---------------------------------------------------------------------------------------------------- :     -
59_3S_1    : ---------------------------------------------------------------------------------------------------- :     -
59_3S_31   : ---------------------------------------------------------------------------------------------------- :     -
59_3S_32   : ---------------------------------------------------------------------------------------------------- :     -
58_3S_11   : gggtcatctggtgttccctttagttcatggaaaccaactgactgctcctccgctttatactgatgatatgccagtgatctacctctacatggtatgttat : 2441
58_3S_12   : gggtcatctggtgttccctttagttcatggaaaccaactgactgctcctccgctttatactgatgatatgccagtgatctacctctacatggtatgttat : 2436
F1_30_S_2  : ---------------------------------------------------------------------------------------------------- :     -
59_3S_2    : ---------------------------------------------------------------------------------------------------- :     -
```

```
                    *      2720         *      2740         *      2760         *      2780         *      2800
Brachypodi : ---------------------------------------------------------------------------------------------------- :     -
Lolium     : ---------------------------------------------------------------------------------------------------- :     -
F1_30_S_11 : ---------------------------------------------------------------------------------------------------- :     -
F1_30_S_12 : ---------------------------------------------------------------------------------------------------- :     -
59_3S_1    : ---------------------------------------------------------------------------------------------------- :     -
59_3S_31   : ---------------------------------------------------------------------------------------------------- :     -
59_3S_32   : ---------------------------------------------------------------------------------------------------- :     -
58_3S_11   : gaactgaagctgtcagcatggaagcttgattgccacatctgagtgttctgggtcaactgatgaacccgtatgtgagaaacagaaactttcaagactttga : 2541
58_3S_12   : gaactgaagctgtcagcatggaagcttgattgccacatctgagtgttctgggtcaactgatgaacccgtatgtgagaaacagaaactttcaagactttga : 2536
F1_30_S_2  : ---------------------------------------------------------------------------------------------------- :     -
59_3S_2    : ---------------------------------------------------------------------------------------------------- :     -


Brachypodi : - :    -
Lolium     : - :    -
F1_30_S_11 : - :    -
F1_30_S_12 : - :    -
59_3S_1    : - :    -
59_3S_31   : - :    -
59_3S_32   : - :    -
58_3S_11   : t : 2542
58_3S_12   : t : 2537
F1_30_S_2  : - :    -
59_3S_2    : - :    -
```

Appendix 7: Results from the phenotyping of a self-compatible population using semi in-vivo pollination, open-pollination and self-pollination tests and from the genotyping of two genetic markers, RV1131 and RV1133, mapped on Lolium LG 3.

				Open-pollination					Self-pollination				
Plant number	**SC phenotype (%)**	**Genotype RV1131**	**Genotype RV1133**	Total seed weight (g)	Seed number	Total number of spikelet	Total number of floret	% seed set OP	Total seed weight (g)	Seed number	Total number of spikelet	Total number of floret	% seed set SP
2	50	ab	ab	0.33	220	144	704	31.3	0.75	395	166	879.8	44.9
3	100	bb	ab	0.40	364	148	703	51.7	0.56	431	179	859.2	50.1
4	100	ab	bb		2	173	865	0.2		12	164	803.6	1.5
5	100	ab	ab	0.64	400	158	758	52.7	0.63	371	165	825.0	44.9
6	50	ab	ab		1	101	626	0.2	0.11	127	211	928.4	13.7
7	100	aa	aa		2	163	1032	0.2	0.07	55	185	925.0	5.9
8	100	ab	aa	0.03	23	163	619	3.7	0.03	35	141	606.3	5.8
9	50	aa	bb	*	0	103	567		0.37	218	187	878.9	24.8
10	100	bb	ab	*	*	*	*		0.59	347	158	790.0	43.9
11	50	aa	bb	0.62	443	145	725	61.1	0.40	235	170	833.0	28.2
12	50	ab	ab	0.23	128	150	683	18.7	0.42	280	168	957.6	29.2
14	50	aa	ab	0.54	386	154	821	47.0	0.29	290	154	739.2	39.2
15	50	ab	ab	0.09	131	187	1085	12.1	0.05	86	181	1049.8	8.2
16	100	ab	ab	0.74	493	121	761	64.9	0.81	506	173	934.2	54.2
17	50	ab	aa		8	29	160	5.0	0.00	0	45	165.0	0.0
18	50	bb	ab	0.32	267	142	781	34.1	0.29	290	175	875.0	33.1
19	50	aa	ab	0.07	19	155	592	3.2	0.03	25	132	488.4	5.1
20	50	aa	ab	0.52	325	164	836	38.9	0.37	247	163	619.4	39.8
21	50	aa	aa	0.23	110	125	578	19.0	0.09	70	143	514.8	13.6

				Open-pollination					Self-pollination				
Plant number	**SC phenotype (%)**	**Genotype RV1131**	**Genotype RV1133**	Total seed weight (g)	Seed number	Total number of spikelet	Total number of floret	% seed set OP	Total seed weight (g)	Seed number	Total number of spikelet	Total number of floret	% seed set SP
22	50	aa	ab	0.50	313	164	689	45.4	0.16	105	112	304.0	34.5
23	50	ab	ab		0	69	466	0.0		3	163	652.0	0.5
25	100	ab	aa	0.15	167	81	466	35.8	0.05	53	187	710.6	7.5
28	50	ab	ab	0.14	103	99	673	15.3	0.19	158	190	988.0	16.0
29	100	ab	ab	0.06	98	141	893	11.0	0.56	431	162	900.0	47.9
30	50	ab	aa		0	174	1201	0.0		13	161	901.6	1.4
31	50	aa	ab	0.38	211	188	730	28.9	*	*	*	*	
32	100	ab	ab	0.19	124	143	658	18.9	0.56	329	188	1059.6	31.1
33	100	ab	aa	0.08	91	161	644	14.1	0.05	88	192	768.0	11.5
34	50	aa	ab	0.51	364	134	697	52.3	0.46	418	167	868.4	48.2
35	50	bb	bb	0.17	96	124	558	17.2	0.21	122	162	777.6	15.7
36	50	aa	aa	0.21	233	78	468	49.9	0.11	138	140	560.0	24.6
37	100	ab	ab	0.25	179	89	378	47.2	0.69	383	156	865.1	44.3
41	50	aa	ab	0.10	92	87	319	28.8	0.06	47	126	456.8	10.3
42	50	ab	aa	0.16	133	96	688	19.4	0.27	245	196	1430.8	17.2
43	100	ab	aa	0.76	447	169	727	61.5	0.58	387	160	800.0	48.3
46	50	ab	ab	0.55	500	128	713	70.1	0.52	400	187	878.9	45.5
47	50	aa	ab	0.89	556	180	1080	51.5	0.81	476	173	916.9	52.0
48	100	bb	aa		7	181	1104	0.6		4	193	984.3	0.4
49	50	aa	ab	0.47	261	143	815	32.0	0.43	253	134	777.2	32.5
50	100	ab	ab	0.13	130	156	819	15.9	0.12	146	194	840.7	17.4
51	100	aa	ab	*	*	*	*		0.70	467	197	1221.4	38.2

				Open-pollination					Self-pollination				
Plant number	**SC phenotype (%)**	**Genotype RV1131**	**Genotype RV1133**	Total seed weight (g)	Seed number	Total number of spikelet	Total number of floret	% seed set OP	Total seed weight (g)	Seed number	Total number of spikelet	Total number of floret	% seed set SP
52	50	ab	aa	0.11	98	166	697	14.1	0.45	265	174	783.0	33.8
53	50	ab	ab	0.19	173	146	628	27.5	0.27	225	152	775.2	29.0
54	50	ab	ab	0.52	306	171	787	38.9	0.60	353	174	939.6	37.6
55	100	ab	ab	0.15	96	161	869	11.0	0.29	161	158	774.2	20.8
56	100	aa	ab		7	72	270	2.6	0.00	0	*	*	0.0
57	100	aa	ab	0.48	400	355	1620	24.7			0	*	
58	100	aa	ab	0.11	92	179	877	10.5	0.02	23	181	1049.8	2.2
60	50	ab	ab										
62	50	aa	aa	0.54	338	133	589	57.3	0.51	364	184	901.6	40.4
63	100	ab	ab	0.06	32	119	422	7.6	0.08	38	150	900.0	4.2
64	50	ab	aa	0.06	62	172	994	6.2	0.37	308	188	977.6	31.5
65	100	aa	aa	0.14	200	140	543	36.9	0.30	188	171	735.3	25.5
66	100	ab	ab	0.27	193	142	694	27.8	0.34	227	170	918.0	24.7
67	100	ab	ab		5	115	460	1.1	0.14	126	186	762.6	16.5
68	100	aa	ab	0.01	17	159	583	2.9	0.00	0	*	*	0.0
69	50	aa	ab	0.14	57	114	456	12.5	0.08	62	115	279.3	22.2
70	50	ab	ab	0.14	20	164	738	2.7	0.21	210	200	854.5	24.6
71	100	aa	ab	0.33	2	147	545	0.4			0	*	
72	50	bb	ab	0.27	8	160	800	1.0	0.16	130	152	793.8	16.4
74	50	bb	ab	0.39	15	137	700	2.1	0.21	210	170	782.0	26.9
75	50	ab	ab	0.41	273	148	789	34.6	0.55	367	155	976.5	37.5

				Open-pollination					Self-pollination				
Plant number	**SC phenotype (%)**	**Genotype RV1131**	**Genotype RV1133**	Total seed weight (g)	Seed number	Total number of spikelet	Total number of floret	% seed set OP	Total seed weight (g)	Seed number	Total number of spikelet	Total number of floret	% seed set SP
76	100	ab	ab	0.33	330	172	879	37.5	0.13	130	179	841.3	15.5
77	100	ab	ab	*	*	*	*		0.22	147	180	882.0	16.6
79	50	ab	ab	0.24	160	188	1008	15.9	0.24	171	166	780.2	22.0
80	50	aa	aa	0.06	58	129	613	9.5		10	173	597.6	1.7
84	100	ab	ab	0.35	292	143	697	41.8	0.22	200	181	724.0	27.6
85	50	ab	aa	0.08	65	61	285	22.8	0.08	115	190	798.0	14.4
86	100	ab	ab	0.34	23	122	697	3.3	0.22	244	169	709.8	34.4
87	50	ab	bb	0.26	17	70	392	4.3	0.14	156	153	673.2	23.1
88	100	ab	ab	0.12	18	157	879	2.0	0.09	116	183	1024.8	11.3
90	100	bb	aa	0.82	22	182	1001		0.64	492	176	897.6	54.8
92	50	aa	aa		14	44	154	9.1		0	131	524.0	0.0
93	50	ab	ab	0.41	21	80	448	4.7	0.40	364	195	858.0	42.4
94	100	aa	bb	0.06	13	44	176	7.4	0.12	96	135	526.5	18.2
95	100	aa	ab	0.34	18	79	487	3.7	0.45	225	133	694.6	32.4
96	100	ab	ab	0.64	21	160	889	2.4	0.34	340	191	1069.6	31.8

Appendix 8: Two way ANOVA tables from R software program.

- Test marker RV1133 against Pollination method (OP or SP)

	Degree of freedom	Sum of square	Mean Square	F value	Pr(>F)
RV1133	2	1264	631.9	1.970	0.143
Pollination	1	243	243.4	0.759	0.385
RV1133:Pollination	2	365	182.6	0.569	0.567
Residuals	138	44262	320.7		

- Test marker RV1131 agains Pollination method (OP or SP)

	Degree of freedom	Sum of square	Mean Square	F value	Pr(>F)
RV1131	2	153	76.7	0.235	0.791
Pollination	1	217	217.1	0.666	0.416
RV1131:Pollination	2	747	373.3	1.144	0.321
Residuals	138	45017	326.2		

References

Abe, A., Kosugi, S., Yoshida, K., Natsume, S., Takagi, H., Kanzaki, H., Matsumura, H., Yoshida, K., Mitsuoka, C., Tamiru, M., Innans, H., Cano, L., Kamoun, S. and Terauchi, R. (2012). Genome sequencing reveals agronomically important loci in rice using MutMap. *Nature Biotechnoloy* **30**: 174-179.

Ai, Y.J., Singh, A., Coleman, C.E., Ioerger, T.R., Kheyrpour, A. and Kao, T.H. (1990). Self-incompatibility in *Petunia inflate* - isolation and characterization of cDNAs encoding 3 *S*-allele-associated proteins. *Sexual Plant Reproduction* **3**: 130–138.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990). Basic Local Alignment Search Tool. *Journal of Molecular Biology* **215**: 403-410.

Anderson, M.A., Cornish, E.C., Mau, S.L., Williams, E.G., Hoggart, R., Atkinson, A., Bonig, I., Grego, B., Simpson, R., Roche, P.J., Haley, J.D., Penschow, J.D., Niall, H.D., Tregear, G.W., Coghlan, J.P., Crawford, R.J. and Clarke, A.E. (1986). Cloning of cDNA for a stylar glycoprotein associated with expression of self-incompatibility in *Nicotiana alata*. *Nature* **321:** 38-44.

Anderson, M.A., McFadden, G.I., Bernatzky, R., Atkinson, A., Orpin, T., Dedman, H., Tregear, G., Fernley, R., Clarke, A.E. (1989). Sequence variability of three alleles of the self-incompatibility gene of *Nicotiana alata*. *Plant Cell* **1:** 483–491.

Anhalt, U.C.M., Heslop-Harrison, P.J.S., Byrne, S., Guillard, A. and Barth, S. (2008). Segregation distortion in *Lolium*: evidence for genetic effects. *Theoretical and applied genetics* **117**: 297-306.

Armstead, I.P., Turner, L.B., King, I.P., Cairns, A.J. and Humphreys, M.O. (2002). Comparison and integration of genetic maps generated from F-2 and BC1-type mapping populations in perennial ryegrass. *Plant breeding* **121**: 501-507.

Asp,T., Byrne,S., Gundlach,H., Bruggmann,R., Mayer,K.F., Andersen,J.R., Xu,M., Greve,M., Lenk,I. and Lubberstedt,T. (2011). Comparative sequence analysis of VRN1 alleles of *Lolium perenne* with the co-linear regions in barley, wheat, and rice. *Molecular Genetics and Genomics* **286**: 433-447.

Bai, C., Sen, P., Hofmann, K., Ma, L., Goebl, M., Harper, J.W. and Elledge, S.J. (1996). SKP1 connects cell cycle regulators to the ubiquitin proteolysis machineray through a novel motif, the F-box. *Cell* **86**: 263-274.

Barreiro, L.B., Laval, G., Quach, H., Patin, E. and Quintana-Murci, L. (2008). Natural selection has driven population differentiation in modern humans. *Nature Genetics* **40**: 340-345.

Bateman, A., Coggill, P. and Finn, R.D. (2010). DUFs: families in search of function. *Acta Crystallographia. Section F, Structural Biology and Crystallization Communications* **66**: 1148–1152.

Bateson, W., Saunders, E.R. and Punnett, R.C. (1905) Experimental studies in the physiology of heredity. *Reports to the Evolution Committee of the Royal Society* **2**, 1–55, 80–99.

Baumann, U., Juttner, J., Bian, X.Y. and Langridge, P. (2000). Self-incompatibility in the grasses. *Annals of Botany* **85**: 203-209.

Bennett, M.D., Smith, J.B. (1991). Nuclear DNA amounts in angiosperms. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **334**: 309-345.

Bentley, D.R., Balasubramanian, S., Swedlow, H.P., Smith, G.P., Milton, J., Brown, C.G., Hall, K.P., Evers, D.J., Barnes, C.L., Bignell, H.R. *et al.* (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**: 53-59.

Bian, X.Y., Friedrich, A., Bai, J.R., Baumann, U., Hayman, D.L., Barker, S.J. and Langridge, P. (2004). High-resolution mapping of the *S* and *Z* loci of *Phalaris coerulescens*. *Genome* **47:** 918-930.

Billiard, S., Castric, V. and Vekemans, X. (2007). A general model to explore complex dominance patterns in plant sporophytic self-incompatibility systems. *Genetics* **175**: 1351-1369.

Boeke, J.D. and Corces, V.G. (1989). Transcription and reverse transcription of retrotransposons. *Annual Review of Microbiology* **43**: 403-434.

Boetzer, M., Henkel, C.V., Jansen, H.J., Butler, D. and Pirovano, W. (2011). Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**: 578-579.

Bower M. S., Matias D. D., Fernandes-Carvalho E., Mazzurco M., Gu T., Rothstein S. J. and Goring D. R. (1996). Two members of the thioredoxin-h family interact with the kinase domain of a *Brassica* S locus receptor kinase. *Plan Cell* **8**: 1641-1650.

Bradford, M.M. (1976). A rapid and sensitive for the quantitation of microgram quantitites of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry* **72***:* 248-254.

Brewbaker, J.L. and Natarajan, A.T. (1960). Centric fragments and pollen part mutation of incompatibility alleles in *Petunia*. *Genetics* **45**: 699-704.

Burd, C.G. and Dreyfuss, G. (1994). Conserved structures and diversity of functions of RNA-binding proteins. *Science* **265**: 615-621.

Busch, J.W., Witthuhn, T. and Joseph, M. (2012). Fewer *S*-alleles are maintained in plant populations with sporophytic as opposed to gametophytic self-incompatibility. *Plant Species Biology*. doi: 10.1111/j.1442-1984.2012.00386.x

Byrne, S., Hedegaard, J., Panitz, F., Bendixen, C., Studer, B., Farrell, J.D., Swain, S., Armstead, I., Caccamo, M. and Asp, T. (2011). *De novo* genome sequencing of perennial ryegrass (*Lolium perenne* L.). **In** Barth, S. (ed) *: EUCARPIA - 29th Fodder Crops and Amenity Grass Section Meeting*. Dublin, Ireland.

Cabrillac, D., Cock, M., Dumas, C., and Gaude, T. (2001). The *S*-locus receptor kinase is inhibited by thioredoxins and activated by pollen coat proteins. *Nature* **410**: 220–223.

Campbell, M.A., Haas, B.J., Hamilton, J.P., Mount, S.M., Buell, C.R. (2006). Comprehensive analysis of alternative splicing in rice and comparative analyses with *Arabidopsis*. *BMC Genomics* **7**: 327.

Castric, V. and Vekemans, X. (2004). Plant self-incompatibility in natural populations: a critical assessment of recent theoretical and empirical advances. *Molecular Ecology* **13**: 2873-2889.

Chambers, G.K. and MacAvoy, E.S. (2000). Microsatellites: consensus and controversy. *Comparative biochemistry and physiology (part B)* **126**: 455-476.

Chagné, D., Gasic, K., Crowhurst, R.N., Han, Y., Bassett, H.C., Bowatte, D.R., Lawrence, T.J., Rikkerink, E.H.A., Gardiner, S.E. and Korban, S.S. (2008) Development of a set of SNP markers present in expressed genes of the apple. *Genomics* **92**:353–358.

Charlesworth, D., Vekemans, X., Castric, V. and Glémin, S. (2005). Plant self-incompatibility systems: a molecular evolutionary perspective. *New phytologist* **168**: 61-69.

Cheng, J.H., Han, Z.H., Xu, X.F. and Li, T.Z. (2006). Isolation and identification of the pollen-expressed polymorphic F-box genes linked to the *S*-locus in apple (*Malus* x *domestica*). *Sexual Plant Reproduction* **19 :** 175-183.

Chevreux, B., Pfisterer, T., Drescher, B., Driesel, A.J., Muller, W.E.G., Wetter, T., Suhai, S. (2004). Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome Research* **14**: 1147-1159.

Ching, A., Caldwell, K.S., Jung, M., Dolan, M., Smith, O.S., Tingey, S., Morgante, M. and Rafalski, A.J. (2002). SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. *BMC Genetics* **3** (19): 3-19.

Clapier, C.R. and Cairns, B.R. (2009). The biology of chromatin remodelling complexes. *Annual Review of Biochemistry* **78**: 273-304.

Clark, K.R., Okuley, J.J., Collins, P.D., Sims, T.L. (1990). Sequence variability and developmental expression of *S*-alleles in self-incompatible and pseudo-self-compatible petunia. *Plant Cell* **2**: 815–826.

Connor, H.E. (1979). Breeding systems in the grasses: a survey. *New Zealand Journal of Botany* **17**: 547-574.

Cornish, M.A., Hayward, M.D. and Lawrence, M.J. (1979). Self-incompatibility in ryegrass. 2. Joint segregation of *S* and *Z* in *Lolium perenne* L. *Heredity* **43**: 129-136.

Cornish MA, Hayward MD, Lawrence MJ. (1980). Self-incompatibility in ryegrass. 3. The joint segregation of S and PGI-2 in *Lolium perenne* L. *Heredity* **44**: 55–62.

Cornish E. C., Pettitt J. M., Bonig I. and Clarke A. E. (1987). Developmentally controlled expression of a gene associated with self-incompatibility in *Nicotiana alata*. *Nature* **326**: 99-102.

Craig, K.L and Tyers, M. (1999). The F-box: a new motif for ubiquitin dependent proteolysis in cell cycle regulation and signal transduction. *Progress in biophysics and molecular biology* **72**: 299-328.

Croxford, A.E., Rogers, T., Caligari, P.D.S. and Wilkinson, M.J. (2008) High-resolution melt analysis to identify and map sequence-tagged site anchor points onto linkage maps: a white lupin (*Lupinus albus*) map as an exemplar. *The New Phytologist* **180**: 594–607.

D'Andrea, L.D. and Regan, L. (2003). TPR proteins: the versatile helix. *Trends in Biochemical Sciences* **28**: 655-662.

Darwin Charles (1862). On the various contrivances by which British and foreign orchids are fertilized by insects, and on the good effects of intercrossing. *London, John Murray.*

Darwin Charles (1876). The effects of cross-fertilisation in the vegetable kingdom. *London, UK: John Murray*.

de Bruijn, N.G. (1946). A Combinatorial Problem. *Koninklijke Nederlandse Akademie v. Wetenschappen* **49**: 758–764.

de Graaf, B.H., Vatovec, S., Juárez-Díaz, J.A., Chai, L., Kooblall, K., Wilkins, K.A., Zou, H., Forbes, T., Franklin, F.C. and Franklin-Tong, V.E. (2012). The Papaver self-incompatibility pollen *S*-determinant, *PrpS*, functions in *Arabidopsis thaliana*. *Current Biology* **22**: 154-159.

de Nettancourt, D. (1977). *Incompatibility in angiosperms*. New York, NY, USA: Springer Verlag.

de Nettancourt, D. (2001). *Incompatibility and incongruity in wild and cultivated plants*. Berlin: Springer-Verlag.

Deshaies, R.J. (1999). SCF and Cullin/Ring H2-based ubiquitin ligases. *Annual review of cell and developmental biology* **15**: 435-467.

Dharmasiri, N., Dharmasiri, S. and Estelle, M. (2005). The F-box protein TIR1 is an auxin receptor. *Nature* **435**: 441-445.

Dinesh-Kumar, S.P. and Baker, B.J. (2000). Alternatively spliced N resistance gene transcripts: their possible role in tobacco mosaic virus resistance. *Proceedings of the National Academy of Sciences of the United States of America* **97**:1908-1913.

Donis-Keller, H., Green, P., Helms, S., Cartinhour, S., Weiffenbach, B., Stephens, K., Keith, T.P., Bowden, D.W., Smith, D., Lander, E.S., Botstein, D., Akots, G., Rediker, K.S., Gravius, T., Brown, V.A., Rising, M.B., Parker, C., Powers, J.A., Watt, D.E., Kauffman, E.R., Bricker, A., Phipps, P., Muller-Kahle, H., Fulton, T.R., Ng, S., Schumm, J.W., Braman, J.C., Knowlton, R.G., Barker, D.F., Crooks, S.M., Lincoln, S.E., Daly, M.J. and Abrahamsont, J. (1997) A genetic linkage map of the human genome. *Cell* **51**: 319-337.

Draper, J., Mur, L. A. J., Jenkins, G., Ghosh-Biswas, G. C., Bablak, P., Hasterok, R. and Routledge, A.P.M. (2001). *Brachypodium distachyon*: a new model system for functional genomics in grasses. *Plant Physiology* **127**: 1539–1555.

East E. M., Mangelsdorf A. J. (1925). A new interpretation of the hereditary behaviour of self-sterile plants. *Proceedings of the National Academy of Sciences of the United States of America* **11**: 166-171.

Ebert P.R., Anderson M. A., Bernatzky R., Altschuler M. and Clarke A. E. (1989). Genetic polymorphism of self-incompatibility in flowering plants. *Plant Cell* **56**: 255-262

Eid, J., Fehr, A., Gray, J., Luong, K., Lyte, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., Bibill, A., Bjornson, K., Chaudhuri, B., Christians, F. *et al*. (2009). Real-Time DNA Sequencing from Single Polymerase Molecules. *Science* **323**: 133-138.

Eisen, J.A. (1999). Mechanistic basis for microsatellite instability. **In**: Goldstein D.B. and Schlotterer, C. (eds) Microsatellites: evolution and applications. *Oxford University Press*, Oxford: 34-48

England, F J W. (1974). The use of incompatibility for the production of F_1hybrids in forage grasses. *Heredity* 32: 183-188.

Entani, T., Takayama, S., Iwano, M., Shiba, H., Che, F.S. and Isogai, A. (1999). Relationship between polyploidy and pollen self-incompatibility phenotype in *Petunia hybrida* Vilm. *Bioscience, Biotechnology and Biochemistry* **63**: 1882-1888.

Entani, T., Zwano, M., Shiba, H., Che, F.S., Isogai, A. and Takayama, S. (2003). Comparative analysis of the self-incompatibility (S-) locus region of *Prunus mume*: identification of a pollen-expressed F-box gene with allelic diversity. *Genes to Cells* **8**: 203-213.

Farrar, K., Asp, T., Lübberstedt, T., Xu, M., Thomas, A.M., Christiansen, C., Humphreys, M.O., Donnison, I.S. (2007). Construction of two Lolium perenne BAC libraries and identification of BACs containing candidate genes for disease resistance and forage quality. *Molecular Breeding* **19**: 15-23.

Fearon, C.H., Hayward, M.D. and Lawrence, M.J. (1984). Self-incompatibility in ryegrass. VII. The determination of incompatibility genotypes in autotetraploid families of *Lolium perenne* L. *Heredity* **53**: 403-413.

Fearon, C.H., Cornish, M.A.., Hayward, M.D and Lawrence M.J. (1994). Self-incompatibility in ryegrass. X. Number and frequency of alleles in a natural population of *Lolium perenne* L. *Heredity* **73**: 254-361.

Fierro-Monti, I. and Mathews, M.B. (2000). Proteins binding to duplexed RNA: one motif, multiple functions. *Trends in Biochemical Sciences* 25: 241-246.

Flaus, A., Martin, D.M.A., Barton, G.J. and Owen-Hughes, T. (2006). Identification of multiple distinct Snf2 subfamilies with conserved structural motifs. *Nucleic Acids Research* **34**: 2887-2905.

Foote, H.C.C., Ride, J.P., Franklin-Tong, V.E., Waler, E.A., Lawrence, M.J., Franklin, F.C.H. (1994). Cloning and expression of a distinctive class of self-incompatibility (*S*) gene from *Papaver rhoeas* L. *Proceedings of the National Academy of Sciences of the United States of America* **91**: 2265-2269.

Franklin-Tong, V.E., Lawrence, M.J. and Franklin, F.C.H. (1988). An *in vitro* bioassay for the stigmatic product of the self-incompatibility gene in *Papaver rhoeas* L. *New Phytologist* **110**: 109-118.

Franklin-Tong, V.E. and Franklin, F.C.H. (1992) Gametophytic self-incompatibility in *Papaver rhoeas* L. *Sexual Plant Reproduction* **5**: 1-7.

Franklin-Tong, V.E., Ride, J.P., Read, N.D., Trewawas, A.J. and Franklin, F.C.H. (1993). The self-incompatibility response in *Papaver rhoeas* is mediated by cytosolic free calcium. *Plant Journal* **4**: 163–177.

Franklin-Tong, V.E., Hackett, G. and Hepler, P.K. (1997). Ratio-imaging of [Ca2+]i in the self-incompatibility response in pollen tubes of *Papaver rhoeas*. *The Plant Journal* **12**: 1375-1386.

Franklin-Tong, V.E. (1999). Signalling in pollination. *Current Opinion in Plant Biology* **2**: 490–495.

Fuong, F.T., Voylokov, A.V. and Smirnov, V.G. (1993). Genetic studies of self-fertility in rye (*Secale cereale* L.). 2. The search for isozyme marker genes linked to self-incompatibility loci. *Theoretical and Applied Genetics* **87**: 619-623.

Ganders, F.R. (1979). Heterostyly in *Lithospermum cobrense* (Boraginaceae). *American Journal of Botany* **66**: 746-748.

Garg, R., Patel, R.K., Jhanwar, S., Priya, P., Bhattacharjee, A., Yadav, G., Bhatia, S., Chattopadhyay, D., Tyagi, A.K., Jain, M. (2011). Gene discovery and tissue-specific transcriptome analysis in chickpea with massively parallel pyrosequencing and web resource development. *Plant Physiology* **156**: 1661-1678.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. **In**: Walker, J.M. (ed): The Proteomics Protocols Handbook, *Humana Press*: 571-607.
Ge, L.L., Tian, H.Q. and Russell, S.D. (2007). Calcium function and distribution during fertilization in angiosperms. *American Journal of Botany* **94**: 1046–1060.

Geitmann, A., Snowman, B.N., Emons, A.C. and Franklin-Tong, V.E. (2000). Alterations in the actin cytoskeleton of pollen tubes are induced by the self-incompatibility reaction in *Papaver rhoeas*. *The Plant Cell* **12**: 1239-1252.

Gentleman, R.C., Carey, V.J., Bates, D.M, Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik. K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, S., Li, C., Maechler, M., Rossini, A.J., Sawitzki, G., Smith, C., Smyth, G., Tierney, L., Yang, J.Y.H. and Zhang, J. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology* **5**: R80.

Gertz, A. and Wricke, G. (1989). Linkage between the incompatibility locus-*Z* and a betaglucosidase locus in rye. *Plant Breeding* **102:** 255-259.

Giranton, J-L., Dumas, C., Cock, J.M. and Gaude, T. (2000). The integral membrane S-locus receptor kinase of Brassica has serine/threonine kinase activity in a membranous environment and spontaneously forms oligomers in planta. *PNAS* **97** (7): 3759-3764.

Goldraij, A., Kondo, K., Lee, C.B., Hancock, C.N., Sivaguru, M., Vazquez-Santana, S., Kim, S., Phillips, T.E., Cruz-Garcia, F. and McClure, B. (2006). Compartmentalization of S-RNase and HT-B degradation in self-incompatible *Nicotiana*. *Nature* **439**: 805-810.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N. and Regev, A. (2011). Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nature Biotechnology* **29**: 644-52.

Gray, J.E., McClure, B.A., Bönig, I., Anderson, M.A. and Clarke, A.E. (1991). Action of the style products of the self-incompatibility gene of *Nicotiana alata* (S-RNase) on *in vitro*-grown pollen tubes. *The Pant Cell* **3**: 271-283.

Grobei, M.A., Qeli, E., Brunner, E., Rehrauer, H., Zhang, R., Roschtzki, B., Basler, K., Ahrens, C.H. and Grossniklaus, U. (2009). Deterministic protein inference for shotgun proteomics data provides new insights into *Arabidopsis* pollen development and function. *Genome Research* 19: 1786-1800.

Gu T., Mazzurco M., Sulaman W, Matias D. D. and Goring D. R. (1998). Binding of an arm repeat protein to the kinase domain of the *S* locus receptor kinase. *Proceedings of the National Academy of Sciences of the United States of America* **95**: 382-387.

Guilfoyle, T.J. and Hagen, G. (2007). Auxin response factors. *Current Opinion in Plant Biology* **10** (5): 453-460.

Gupta, P.K., Varshney, R.K., Sharma, P.C., Sing, R., Kumar, N. and Balyan, H.S. (1999). Molecular markers and their applications in wheat breeding. *Plant breeding* **118**: 369-390.

Hackauf, B. and Wehling, P. (2005). Approaching the self-incompatibility locus *Z* in rye (*Secale cereale* L.) via comparative genetics. *Theoretical and Applied Genetics* **110**: 832-845.

Hackett, C.A. and Broadfoot, L.B. (2003). Effects of genotyping errors, missing values and segregation distortion in molecular markers data on the construction of linkage maps. *Heredity* **90**: 33-38.

Haldane, J.B.S. (1919) The combination of linkage values, and the calculation of distance between the loci of linked factors. Journal genetics: 299-309.

Hanson, L., Leitch, I.J., Bennett, M.D. (2002). Unpublished data from the Jodrell Laboratory, Royal Botanic Gardens, Kew

Harbord, R.M., Napoli, C.A. and Robbins, T.P. (2000). Segregation distortion of T-DNA markers linked to the self-incompatibility (*S*) locus in *Petunia hybrid*. *Genetics* **154**: 1323-1333.

Hargreaves, D.C. and Crabtree, G.R. (2011). ATP-dependent chromatin remodelling: Genetics, genomics and mechanisms. *Cell Research* **21**: 396-420.

Harushima, Y., Kurata, N., Yano, M., Nagamura, Y., Sasaki, T. and Nakagahra, M. (1996). Detection of segregation distortions in an *indica-japonica* rice cross using a high-resolution molecular map. *Theoretical Applied Genetics* **92**: 145–150.

Harushima, Y., Yano, M., Shomura, A., Sato, M., Shimano, T., Kuboki, Y., Yamamoto, T., Lin, S.Y., Antonio, B.A., Parco, A., Kajiya, H., Huang, N., Yamamoto, K., Nagamura, Y., Kurata, N., Khush, G.S. and Sasaki, T. (1998). A high-density rice genetic linkage map wit 2275 markers using a single F2 population. *Genetics* **148**: 479-494.

Hauck, N.R., Yamane, H., Tao, R and Iezzoni, A.F. (2006). Accumulation of non-functional S-haplotypes results in the breakdown of gametophytic self-incompatibility in tetraploid *Prunus*. *Genetics* **172**: 1191-1198.

Hayashi, H. (1995). Pyridoxal enzymes: mechanistic diversity and uniformity. *Journal of Biochemistry* **118**: 463-473.

Hayman, D.L. (1956). The genetic control of incompatibility in *Phalaris coerulescens* Desf.
Australian Journal of Biological Sciences **9**: 321-331.

Haymann, D.L. and Richter, J. (1992). Mutations affecting self-incompatibility in *Phalaris coerulescens* Desf (Poaceae). *Heredity* **68**: 495-503.

Hayward, M.D., Forster, J.W., Jones, J.G., Dolstra, O., Evans, C., McAdam, N.J., Hossain, K.G., Stammers, M., Will, J., Humphreys, M.O. and Evans, G.M. (1998) Genetic analysis of *Lolium*. I. Identification of linkage groups and the establishment of a genetic map. *Plant Breeding* **117**: 451–455.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H. and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular Cell* **38**: 576-589.

Heslop-Harrison, J. (1982). Pollen-stigma interaction and cross-incompatibility in the grasses. *Science* **215**: 1358-1364.

Hershko, A. and Ciechanover, A. (1998). The Ubiquitin System. *Annual review of biochemistry* **67**: 425-479.

Hinata K. and Nishio T. (1978). *S*-allele specificity of stigma proteins in *Brassica oleracea* and *B. campestris*. *Heredity* **41**: 93-100.

Hinata, K., Okazaki, K., Nishio, T. (1983). Gene analysis of self-incompatibility in *Brassica campestris* var. yellow sarson (a case of recessive epistatic modifier). *Proceedings of the 6th International Rapeseed Conference, Paris, France, May 17–19, 1983* (Groupe Consultatif International de Recherche sur le Colza, Paris) **Vol 1**: 354-359.

Hiscock, S.J. (2000). Genetic control of self-incompatibility in *Senecio squalidus* L. (Asteraceae): a successful colonizing species. *Heredity* **85**: 10-19.

Hiscock, S.J., McInnis, S.M., Tabah, D.A., Henderson, C.A. and Brennan, A.C. (2003a). Sporophytic self-incompatibility in *Senecio squalidus* L. (Asteraceae) - the search for *S*. *Journal of Experimental Botany* **54**: 169-174.

Hiscock, S.J. and Tabah, D.A. (2003b). The different mechanisms of Sporophytic self-incompatibility. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **358**: 1037-1045.

Hiscock, S.J. and McInnis, S.M. (2003c). Pollen recognition and rejectionduring the sporophytic self-incompatibility response: *Brassica* and beyond. *TRENDS in Plant Sciences* **8** (12): 606-613.

Holdaway-Clarke, T.L. and Hepler, P.K. (2003). Control of pollen tube growth: role of ion gradients and fluxes. *New Phytologist* **159**: 539–563.

Holsinger, K.E. (2000). Reproductive systems and evolution in vascular plants. *Proceedings of the National Academy of Sciences* **97** (13): 7037-7042.

Holt, C. and Yandell, M. (2011). MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *Bioinformatics* **12**: 491.

Hsu, S.C., TerBush, D., Abraham, M. and Guo, W. (2004). The exocyst complex in polarized exocytosis. *International Review of Cytology* **233**:243–265.

Hua, Z. and Kao, T.H. (2006). Identification and characterization of components of a putative *Petunia* S-locus F-box-containing E3 ligase complex involved in S-RNase-based self-incompatibility. *The Plant Cell* **18**: 2531-2553.

Hua, Z., Meng, X.Y. 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**: 3593-3609.

Hua, Z.H., Fields, A. and Kao, T.H. (2008). Biochemical models for S-RNas-based self-incompatibility. *Molecular Plant* **1**: 575-585.

Huang, J., Zhao, L., Yang, Q. and Xue, Y. (2006). AhSSK1, a novel SKP1-like protein that interacts with the S-locus F-box protein SLF. *The Plant Journal* **46**: 780-793.

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

Humphreys, M.W., Yadav, R.S., Cairns, A.J., Turner, L.B., Humphreys, J. and Skøt L. (2006). A changing climate for grassland research. *New Phytologist* **169**: 9-26.

Huo, N., Garvin, D.F., You, F.M., McMahon, S., Luo, M-C., Gu, Y.Q., Lazo, G.R. and Vogel, J.P (2011). Comparison of a high-density genetic linkage map to genome features in the model grass *Brachypodium distachyon*. *Theoretical and Applied Genetics* **123**: 455-464.

Igic, B. and Kohn, J.R. (2001). Evolutionary relationships among self-incompatibility RNases. *Proceedings of the National Academy of Sciences of the United States of America* **98**: 13167-13171.

Ikeda, K., Igic, B., Ushijima, K., Yamane, H., Hauck, N.R., Nakano, R., Sassa, H., Iezzoni, A.F., Kohn, J.R., and Tao, R. (2004). Primary structural features of the *S* haplotype-specific F-box protein, SFB, in *Prunus*.*Sex. Plant Reproduction* **16**: 235-243.

Ioerger, T.R., Clark, A.G., Kao, T.H. (1990). Polymorphism at the self-incompatibility locus in Solanaceae predates speciation. *Proceedings of the National Academy of Sciences of the United States of America* **87**: 9732–9735.

Ioerger, T.R., Gohlke, J.R., Xu, B., and Kao, T.H. (1991). Primary structural features of the self-incompatibility protein in Solanaceae. *Sexual Plant Reproduction* **4**: 81–87.

Iwano M, Shiba H, Matoba K, Miwa, T., Funato, M., Entani, T., Nakayama, P., Shimosato, H. Takaoka, A., Isogai, A. and Takayama, S. (2007). Actin dynamics in papilla cells of *Brassica rapa* during self- and cross-pollination. *Plant Physiology* **144**: 72–81.

Jensen, L.B., Andersen, J.R., Frei, U., Xing, Y., Taylor, C., Holm, P.B. and Lübberstedt, T. (2005a). QTL mapping of vernalization response in perennial ryegrass (*Lolium perenne* L.) reveals co-location with an orthologue of wheat *VRN1*. *Theoretical and applied genetics* **10** (3): 527-536.

Jensen, L.B., Muylle, H., Arens, P., Andersen, C.H., Holm, P.B., Ghesquiere, M., Julier, B., Lübberstedt, T., Nielsen, K.K., De Riek, J., Roldan-Ruiz, I., Roulund, N., Taylor, C., Vosman, B. and Barre, P. (2005b). Development and mapping of a public reference set of SSR markers in *Lolium perenne* L. *Molecular Ecology Notes* **5** (4)**:** 951-957.

Jones, E.S., Mahoney, N.L., Hayward, M.D., Armstead, I.P., Jones, J.G., Humphreys, M.O., King, I.P., Kishida, T., Yamada, T., Balfourier, F., Charmet, G. and Forster, J.W. (2002). An enhanced molecular marker based genetic map of perennial ryegrass (*Lolium perenne*) reveals comparative relationships with other Poaceae genomes. *Genome* **45**: 282–295.

Jordan, N.D., Ride, J.P., Rudd, J.J., Davies, E.M., Franklin-Tong, V.E. Franklin, F.C.H. (2000a). Inhibition of self-incompatible pollen in *Papaver rhoeas* involves a complex series of cellular events. *Annals of Botany* **85** (Supplement A): 197-202.

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

Jungreis, I., Lin, M.F., Spokony, R., Chan, C.S., Negre, N., Victorsen, A., White, K.P. and Kellis, M. (2011). Evidence of abundant stop codon readthrough in *Drosophila* and other metazoan. *Genome Research* **21**: 2096-2113.

Kachroo, A., Schopfer, C.R., Nasrallah, M.E., and Nasrallah, J.B. (2001). Allele-specific receptor ligand interactions in *Brassica* self-incompatibility. *Science* **293:** 1824–1826.

Kakeda, K., Jordan, N.D., Conner, A., Ride, J.P., Franklin-Tong, V.E. and Franklin, F.C. (1998). Identification of residues in a hydrophilic loop of the *Papaver rhoeas S* protein that play a crucial role in recognition of incompatible pollen. *The Plant Cell* **10**: 1723-1732.

Kakeda, K., Ibuki, T., Suzuki, J., Tadano, H., Kurita, Y., Hanai, Y., Kowyama, Y. (2008). Molecular and genetic characterization of the S locus in Hordeum bulbosum L., a wild self-incompatible species related to cultivated barley. *Molecular Genetics and Genomics* **280**: 509–519.

Kakeda, K. (2009). S locus-linked F-box genes expressed in anthers of Hordeum bulbosum. *Plant Cell Reports* **28**: 1453-1460.

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*. *Plant Cell* **19**:3961–3973.

Kantety, R.V., Rota, M.L., Matthews, D.E. and Sorrells, M.E. (2002). Data mining for simple-sequence repeats in expressed sequence tags from barley, maize, rice, sorghum, and wheat. *Plant Molecular Biology* **48**: 501-510.

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 of the United States of America* **93**: 12059-12065.

Kao, T.H. and Tsukamoto, T. (2004). The molecular and genetic bases of S-RNase-based self-incompatibiliy. *The Plant Cell* **16**: 72-83.

Kato, S. and Mukai, Y. (2004). Allelic diversity of S-RNase at the self-incompatibility locus in natural flowering cherry populations (*Prunus lannesiana* var. speciosa). *Heredity* **92**: 249-256.

Kato, S., Iwata, H., Tsumura, Y. and Mukai, Y. (2007). Distribution of *S*-alleles in island populations of flowering cherry, *Prunus lannesiana* var. speciosa. *Genes and Genetic Systems* **82**: 65-75.

Kearsey, M.J. and Pooni, H.S. (1996) The genetical analysis of quantitative traits. 1st edition. Chapman and Hall, London.

Khan, M. and Stace, C. (1999). Breeding relationships in the genus Brachypodium (Poaceae: Pooideae). *Nordic Journal of Botany* **19** (3):257–269.

Kheyr-Pour, A., Bintrim, S.B., Ioerger, T.R., Remy, R., Hammond, S.A. and Kao, T.H. (1990). Sequence diversity of pistil *S*-proteins associated with gametophytic self-incompatibility in *Nicotiana alata*. *Sexual Plant Reproduction* **3**: 88–97.

Kipreos, E.T. and Pagano, M. (2000). The F-box protein family. *Genome Biology* **1** (5): review 3002.

Kircher, M. and Kelso, J. (2010). High-throughput DNA sequencing - concepts and limitations. *Bioessays* **32**: 524-536.

Kitashiba, H., Liu, P., Nishio, T., Nasrallah, J.B., Nasrallah, M.E. (2006). Functional test of *Brassica* self-incompatibility modifiers in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America* **108** (44): 18173-18178.

Klaas, M., Yang, N., Bosch, M., Thorogood, D., Manzanares, C., Armstead, I.P., Franklin, F.C.H. and Barth, S. (2011). Progress towards elucidating the mechanisms of self-incompatibility in the grasses: further insights from studies in *Lolium*. *Annals of Botany* **108**: 677-685.

Knox, R.B. and Heslop-Harrison, J. (1971). Pollen-wall proteins - fate of intined-held antigens on stigma in compatible and incompatible pollinations of *Phalaris tuberosa* L. *Journal of Cell Science* **9:** 239-251.

Konishi, T., Yano, Y. and Abe, K. (1992). Geographic distribution of alleles at the *Ga2* locus for segregation distortion in barley. *Theoretical Applied Genetics* **85**: 419–422.

Kopecký, D., Havránková, M., Loureiro, J., Castro, S., Lukaszewski, A.J., Bartoš, J., Kopecká, J. and Doležel, J. (2010). Physical distribution of homoeologous recombination in individual chromosomes of *Festuca pratensis* in *Lolium multiflorum*. *Cytogenetic Genome Research* **129**: 162-172.

Kosambi, D.D. (1944). The estimation of map distance from recombination.

Kota, R., Varshney, R.K., Thiel, T., Dehmer, K.J. and Graner, A. (2001). Generation and comparison of EST-derived SSRs and SNPs in barley (*Hordeum vulgare* L.). *Hereditas* **135**: 145-151.

Kowyama, Y., Tsuchiya, T. and Kakeda, K. (2000). Sporophytic self-incompatibility in *Ipomoea trifida*, a close relative of sweet potato. *Annals of Botany* **85**: 191-196.

Kurata, N., Nagamura, Y., Yamamoto, K., Harushima, Y., Sue, N., Wu, J., Antonio, B.A., Shomura, A., Shimizu, T., Lin, S-Y., Inoue, T., Fukuda, A., Shimano, T., Kuboki, Y., Toyama, T., Miyamoto, Y., Kirihara, T., Hayasaka, K., Miyao, A., Monna, L., Zhong, H.S., Tamura, Y., Wang, Z-X., Momma, T., Umehara, Y., Yano, M., Sasaki, T. and Minobe, Y. (1994) A 300 kilobase interval genetic map of rice including 883 expressed sequences. *Nature Genetics* **8**: 365–372.

Kurata, N., Umehara, Y., Tanoue, H. and Sasaki, T. (1997). Physical mapping of the rice genome with YAC clones. *Plant Molecular Biology* **35**: 101-113.

Kurian, V. and Richards, A.J. (1997). A new recombinant in the heteromorphy '*S*' supergene in *Primula*. *Heredity* **78:** 383-390.

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*. *Plant Cell* **13:** 627-643.

Lai, Z., Ma, W., Han, B., Liang, L., Zhang, Y. Hong, G. and Xue, Y. (2002). An F-box gene linked to the self-incompatibility (S) locus of *Antirrhinum* is expressed specifically in pollen and tapetum. *Plant Molecular Biology* **50**: 29-42.

Lam, E. and del Pozo, O. (2000). Caspase-like proteases involvement in the control of plant cell death. *Plant Molecular Biology* **44**: 417-428.

Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* **10**: R25.

Lareau, L.F., Brooks, A.N., Soergel, D.A.W., Meng, Q. and Brenner, S.E. (2007). The coupling of alternative splicing and nonsense mediated mRNA decay. **In**: Blencowe, B.J. and Graveley, B.R. (eds.), Alternative Splicing in the Postgenomic Era. *Landes Biosciences*, Austin, TX: 191-212.

Lawrence, M.J. (1975). The genetics of self-incompatibility in *Papaver rhoeas*. *Proceedings of the Royal Society Lond. B*: **188**: 275-285.

Lawrence, M.J. (1996). Number of incompatibility alleles in clover and other species. *Heredity* **76**: 610-615.

Lawrence, M. J. (2000). Population genetics of the homomorphic self-incompatibility polymorphisms in flowering plants. *Annals of Botany* **85**: 221-226.

Lee, H.S., Huang, S. and Kao, T. (1994). S-proteins control rejection of incompatible pollen in *Petunia inflata*. *Nature* **367**: 560-563.

Levinson, G. and Gutman, G.A. (1987). Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Molecular Biology and Evolution* **4**: 203-221.

Lewis, S.E., Searle, S.M.J., Harris, N., Gibson, M., Iyer, V., Ritcher, J., Wiel, C., Bayraktaroglu, L., Birney, E., Crosby, M.A., Kaminker, J.S., Matthews, B.B., Prochnik, S.E., Smith, C.D., Tupy, J.L., Rubin, G.M., Misram S., Mungall, C.J. and Clamp, M.E. (2002). Apollo: a sequence annotation editor. *Genome Biology* **3** (12): R82.

Li, X.M., Nleld, J., Hayman, D. and Langridge, P. (1994). Cloning a putative selfincompatibility gene from the pollen of the grass *Phalaris coerulescens*. *Plant Cell* **6:** 1923-1932.

Li, X.M., 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 Molecular Biology* **34**: 223-232.

Li, H., Ruan, J. and Durbin, R. (2008). Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Research* **18**: 1851-1858.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and 1000 Genome Project Data Processing subgroup (2009a). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078-2079.

Li, R., Yu, C., Li, Y., Lam, T.W., Yiu, S.M., Kristiansen, K. and Wang, J. (2009b). SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* **25**: 1966-1967.

Li, J., Wang, X., Dong, R., Yang, Y., Zhou, J., Yu, C., Cheng, Y., Yan, C. and Chen, J. (2011) Evaluation of high-resolution melting for gene mapping in rice. *Plant Molecular Biology Reporter* **29** (4): 979-985.

Liu, P., Sherman-Broyles, S., Nasrallah, M.E and Nasrallah, J.B. (2007). A cryptic modifier causing transient self-incompatibility in *Arabidopsis thaliana*. *Current Biology* **17**: 734-740.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L. and Law, M. (2012). Comparison of Next-Generation Sequencing Systems. *Journal of Biomedicine and Biotechnology* **2012**.

Link, A.J., Eng, J., Schieltz, D.M., Carmack, E., Mize, G.J., Morris, D.R., Garvik, B.M. and Yates, J.R. (1999). Direct analysis of protein complexes using mass spectrometry. *Nature Biotechnology* **17**: 676-682.

Lowry, O.H., Rosebrough, N.J., Farr, A.L. and Randall, R.J. (1951). Protein measurement with the Folin-Phenol reagents. *Journal of Biological Chemistry* **193**: 265-275.

Lu, Y. (2006). Historical events and allelic polymorphism at the gametophytic self-incompatibility locus in Solanaceae. *Heredity* **96**: 22-28.

Lundqvist, A. (1954). Studies on self-sterility in rye, *Secale cereale* L. *Hereditas* **40**: 278-294.

Lundqvist, A. (1957). Self-incompatibility in rye. II. Genetic control in the autotetraploid. *Hereditas* **43**: 467-511.

Lunqvist, A. (1961). A rapid method for the analysis of incompatibility in grasses. *Hereditas* **47:** 705-707.

Lundqvist, A. (1962). Self-incompatibility in diploid *Hordeum bulbosum* L. *Hereditas* **48**: 138-152.

Lundqvist, A. (1962b). The nature of the two-loci incompatibility system in grasses. II. Number of alleles at the incompatibility loci in *Festuca pratensis* Huds. *Hereditas* **48**: 169-181.

Lundqvist, A. (1964). The nature of the two-loci incompatibility system in grasses. III. Frequency of specific alleles in a population of *Festuca pratensis* Huds. *Hereditas* **52**: 189-195.

Lundqvist, A. (1968). The mode of origin of self-fertility in grasses. *Hereditas* **59**: 413-426.

Lundqvist, A. (1969). The identification of the self-incompatibility alleles in a grass population. *Hereditas* **61**: 345-352.

Luu, D.T., Qin, X., Morse, D. and Cappadocia, M. (2000). S-RNase uptake by compatible pollen tubes in gametophytic self-incompatibility. *Nature* **407**: 649-651.

Lyttle T.W. (1991). Segregation distortion. *Annual Review of Genetics* **25**: 511-557.

Maize Genome Sequencing Project, Schnable, P.S. *et al*. (2009). The B73 Maize genome: complexity, diversity and dynamics. *Science* **326**: 1112-1115.

Malho, R., Read, N.D., Pais, M.S. and Trewavas, A.J. (1994). Role of cytosolic free calcium in the reorientation of pollen- tube growth. *The Plant Journal* **5**: 331-341.

Marcellán, O. and Acevedo, A. (2012). Development of PCR-based markers for the identification of the S-RNase alleles in wild potato species. *Euphytica* **187**: 19-29.

Margulies ,M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z., Dewell, S.B., Du, L., et al. (2005). Genome sequencing in microfabricated high-density picoliter reactors. *Nature* **437**: 376-380.

Martin, F.W. (1959). Staining and observing pollen tubes in the style by means of fluorescence. *Stain Technology* **34**: 125-128.

Matton, D.P., Maes, O., Laublin, G., Qin, X., Bertrand, C., Morse, D. and Cappadocia, M. (1997). Hypervariable domains of self-incompatibility RNases mediate allele-specific pollen recognition. *Plant Cell* **9**: 1757-1766.

Maynard Smith, J. (1971). What use is sex? *Journal of Theoretical Biology* **30**: 319-335.

McCallum, C.M., Comai, L., Greene, E.A., Henikoff, S. (2000). Targeting Induced Local Lesions IN Genomes (TILLING) for plant functional genomics. *Plant Physiology* **123**: 439-442.

McClure, B.A., Haring,V., Ebert, P.R., Anderson, M.A., Simpson, R.J., Sakiyama, F. and Clarke, A.E. (1989). Style self-incompatibility gene products of *Nicotiana alata* are ribonucleases. *Nature* **342**: 955-957.

McClure, B.A., Gray, J.E., Anderson, M.A., Clarke, A.E. (1990). Self-incompatibility in *Nicotiana alata* involves degradation of pollen rRNA. *Nature* **347:** 757-760.

McClure, B.A. and Franklin-Tong, V. (2006). Gametophytic self-incompatibility: understanding the cellular mechanisms involved in "self" pollen tube inhibition. *Planta* **224**: 233-245.

McCraw, J.M. and Spoor, W. (1983). Self-incompatibility in Lolium species. 1. *Lolium rigidum* Gaud and *Lolium multiflorum* L. *Heredity* **50**: 21-27.

Mejlhede, N., Kyjovska, Z., Backes, G., Burhenne, K., Rasmussen, S.K. and Jahoor, A. (2006). EcoTILLING for the identification of allelic variation in the powdery mildew resistance genes *mlo* and *Mla* of barley. *Plant Breeding* **125**: 461-467.

Mendel, J.G. (1866) *Versuche über Pflanzenhybriden*Verhandlungen des naturforschenden Vereines in Brünn, Bd. IV für das Jahr, **1865** *Abhandlungen*: 3–47. Translated by **Druery, C.T. and Bateson, W. (1901)** "Experiments in plant hybridization". *Journal of the Royal Horticultural Society* **26**: 1–32.

Meng, X. Sun, P. and Kao, T.H. (2011). S-RNase-based self-incompatibility in *Petunia inflata*. *Annals of Botany* **108**: 637-646.

Meng, X.Y., Hua, Z.H., Wang, N., Fields, A.M., Dowd, P.E. and Kao, T.H. (2009). Ectopic expression of S-RNase of *Petunia inflata* in pollen results in its sequestration and non-cytotoxic function. *Sexual Plant Reproduction* **22**: 263-275.

Morgan, T.H. (1910) Sex-limited inheritance in *Drosophila*. *Science* **132**: 120–122.

Morgan, T.H. (1911) Random segregation versus coupling in Mendelian Inheritance. *Science* **34:** 384.

Morgante, M. and Olivieri, A.M. (1993). PCR-amplified microsatellites as markers in plant genetics. *The Plant Journal* **3** (1): 175-182.

Morozova, O. and Marra, M.A. (2008). Applications of next-generation sequencing technologies in functional genomics. *Genomics* **92**: 255-264.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L. and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* **5**: 621-628.

Munson, M., Novick, P. (2006). The exocyst defrocked, a framework of rods revealed. *Nature Structural and Molecular Biology* **13**: 577–581.

Murase, K., Shiba, H., Iwano, M., Che, F.S., Watanabe, M., Isogai, A. and Takayama, S. (2004). A membrane-anchored protein kinase involved in *Brassica* self-incompatibility signaling. *Science* **303**: 1516–1519.

Murfett, J., Atherton, T.L., Mou, B., Gasser, C.S. and McClure, B.A. (1994). S-RNase expressed in transgenic *Nicotiana* causes S-allele-specific pollen rejection. *Nature* **367**: 563-566.

Namy O. and Rousset, J.P. (2010). Specification of standard amino acids by stop codons. **In**: Atkins, J.F., Gesteland, R.F. Expansion of decoding rules enriches gene expression. *Springer*, New York: 79-100.

Nasrallah, J.B. (2002). Recognition and rejection of self in plant reproduction. *Science* **296**: 305-308.

Nesvizhskii, A.I. (2007). Protein identification by tandem mass spectrometry and sequence database searching. *Methods in Molecular Biology* **367**: 87-119.

O'Donnell, S. and Lawrence, M. J. (1984). The population genetics of the self-incompatibility polymorphism in *Papaver rhoeas*. IV. Estimation of the number of alleles in a population. *Heredity* **53**: 495-507.

O'Donoughue, L.S., Wang, Z., Röder, M., Kneen, B., Leggett, M., Sorrells, M.E. and Tanksley, S.D. (1992). An RFLP-based linkage map of oats based on a cross between two diploid taxa (*Avena atlantic* × *A. hirtula*). *Genome* **35**: 765-771.

Olsen, M., Hood, L., Cantor, C. and Botstein, D. (1989). A common language for physical mapping of the human genome. *Science* **245**: 1434-1435.

Palanivelu, R. and Preuss, D. (2000). Pollen tube targeting and axon guidance: parallels in tip growth mechanisms. *Trends in Cell Biology* **10**: 517-524.

Pandey, K.K. (1977). Origin of complementary incompatibility systems in flowering plants. *Theoretical and Applied Genetics* **49**: 101–109.

Parry, S., Newbigin, E.D., Craik, D., Nakamura, K.T., Bacic, A. and Oxley, D. (1998). Structural analysis and molecular model of a self-incompatibility RNase from wild tomato. *Plant Physiology* **116**: 463-469.

Paterson, A.H., Damon, S., Hewitt, J.D., Zamir, D., Rabinowitch, H.D., Lincoln, S.E., Lander, E.S. and Tanksley, S.D. (1991). Mendelian factors underlying quantitative traits in tomato: Comparative across species, generations, and environments. *Genetics* **127**: 181–197.

Paterson. A.H., Bowers, J.E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., Haberer, G., Hellsten, U., Mitros, T., Poliakov, A., Schmutz, J., Spannagl, M., Tang, H., Wang, X., Wicker, T., Bharti, A.K., Chapman, J., Feltus, F.A., Gowik, U., Grigoriev, I.V., Lyons, E., Maher, C.A., Martis, M., Narechania, A., Otillar, R.P., Penning, B.W., Salamov, A.A., Wang, Y., Zhang, L., Carpita, N.C., Freeling, M., Gingle, A.R., Hash, C.T., Keller, B., Klein, P., Kresovich, S., McCann, M.C., Ming, R., Peterson, D.G., Mehboob-ur-Rahman, Ware, D., Westhoff, P., Mayer, K.F.X., Messing, J. and Rokhsar, D.S. (2009). The *Sorghum bicolor* genome and the diversification of grasses. *Nature* **457**: 551-556.

Pedersen, C., Giese, H. and Linde-Laursen, I. (1995). Towards an integration of the physical and the genetic chromosome maps of barley by in situ hybridization. *Hereditas* **123**: 77-88.

Pierson, E.S., Miller, D.D., Callaham, D.A., van Aken, J., Hackett, G. and Hepler, P.K. (1996). Tip-localized calcium entry fluctuates during pollen tube growth. *Developmental Biology* **174**: 160-173.

Poulter, N.S. (2009). Investigating the role of the cytoskeleton and signalling in the self-incompatibility response of *Papaver rhoeas*. PhD thesis, University of Birmingham.

Qiao, H., Wang, F., Zhao, L., Zhou, J., Lai, Z., Zhang, Y., Robbins, T.P. and Xue, Y. (2004). The F-box protein AhSLF-S2 controls the pollen function of S-RNase-based self-incompatibility. *The Plant Cell* **16**: 2307-2322.

R Development Core Team (2008). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria.

Richard, G.F., Kerrest, A. and Dujon B. (2008). Comparative genomics and molecular dynamics of DNA repeats in Eukaryotes. *Microbiology and Molecular Biology Reviews* **72** (4): 686-727.

Ride, J.P., Davies, E.M., Franklin, F.C. and Marshall, D.F. (1999). Analysis of *Arabidopsis* genome sequence reveals a large new gene family in plants. *Plant Molecular Biology* **39**: 927-932.

Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G. and Mesirov, J.P. (2011). Integrative genomics viewer. *Nature Biotechnology* **29**: 24-26.

Robinson, M.D., McCarthy, D.J. and Smyth, G.K. (2010) edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139.

Robbins, T.P., Harbord, R.M., Sonneveld, T. and Clarke, K. (2000). The molecular genetics of self-incompatibility in *Petunia hybrid*. *Annals of Botany* **85**: 105-112.

Ronaghi, M. (2000). Improved performance of pyrosequencing using single-stranded DNA-binding protein. *Analytical Biochemistry* **286**: 282-288.

Rothberg, J.M., Hinz, W., Rearick, T.M., Schltz, J., Mileski, W., Davey, M., Leamon, J.H., Johnson, K., Milgrew, M.J., Edwards, M., Hoon, J., Simons, J.F. et al. (2011). An integrated semiconductor device enabling non-optical genome sequencing. *Nature* **475**: 348-352.

Royo, J., Kunz, C., Kowyama, Y., Anderson, M.A., Clarke, A.E. and Newbigin, E. (1994). Loss of a histidine residue at the active site of S-locus ribonuclease is associated with self-compatibility in *Lycopersicon peruvianum*. *Proceedings of the National Academy of Sciences of the United States of America* **91**: 6511-6514.

Rozen, S. and Skaletsky, H.J. (2000). Primer3 on the WWW for general users and for biologist programmers. **In**: Krawetz, S. and Misener, S. (eds) Bioinformatics methods and protocols: Methods in molecular biology. *Humana Press,* Totowa, NJ, USA.

Rudd, J.J., Osman, K., Franklin, F.C. and Franklin-Tong, V.E. (2003). Activation of a putative MAP kinase in pollen is stimulated by the self-incompatibility (SI) response. *FEBS Letters* **547**: 223-227.

Sachidanandam, R., Weissman, D., Schmidt, S.C., Kakol, J.M., Stein, L.D., Marth, G., Sherry., S. Mullikin, J.C., Mortimore, B.J., Willey, D.L., Hunt, S.E., Cole, C.G., Coggill, P.C., Rice, C.M., Ning, Z., Rogers, J., Bentley, D.R., Kwok, P.Y., Mardis, E.R., Yeh, R.T., Schultz, B., Cook, L., Davenport, R., Dante, M., Fulton L., Hillier, L., Waterson, R.H., McPherson, J.D., Gilman, B., Schaffner, S., Van Etten, W.J., Reich, d., Higgins, J., Daly, M.J., Blumenstiel, B., Baldwin, J., Stange-Thomann, N., Zody, M.C., Linton, L., Lander, E.S. and Altshuler, D. (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**: 928-933.

Samuel, M.A., Chong, Y.T., Haasen, K.E., Aldea-Brydges, M.G., Stone, S.L. and Goring, D.R. (2009). Cellular pathways regulating responses to compatible and self-incompatible pollen in *Brassica* and *Arabidopsis* stigmas intersect at Exo70A1, a putative component of the exocyst complex. *Plant Cell* **21**:2655–2671.

Sandler, L., Hiraizumi, Y. and Sandler, I. (1959). Meiotic drive in natural populations of Drosophila melanogaster. I. The cytogenetic basis of segregation distortion. *Genetics* **44**: 233-250.

SanMiguel, P. and Bennetzen, J.L. (1998). Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. *Annals of Botany* **82** (Suppl A): 37-44.

Sanzol, J., Sutherland, B.G. and Robbins, T.P. (2006). Identification and characterization of genomic DNA sequences of the *S*-ribonuclease gene associated with self-incompatibility alleles S_1 to S_5 in European pear. *Plant Breeding* **125**: 513-518.

Sanzol, J. and Robbins, T.P. (2008). Combined analysis of *S*-alleles in European pear by pollinations and PCR-based *S*-genotyping; correlation between *S*-phenotypes and S-RNase genotypes. *Journal of the American Society for Horticultural Science* **133**: 213-224.

Sassa, **H., Hirano, H.** and **Ikehashi, H. (1992).** Self-incompatibility-related RNases in styles of Japanese pear (*Pyrus serotina* Rehd.). *Plant and Cell Physiology* **33**: 811-814.

Schaller, A. (2004). A cut above the rest: the regulatory function of plant proteases. *Planta* **220**: 183-197.

Schierup, M.H., Mable, B.K., Awadalla, P. and Charlesworth, D. (2001). Identification and characterization of a polymorphic receptor kinase gene linked to the self-incompatibility locus of *Arabidopsis lyrata*. *Genetics* **158**: 387-399.

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

Semagn, K., Bjørnstad, Å and Ndjiondjop, M.N. (2006). An overview of molecular marker methods for plants. *African Journal of Biotechnology* **5** (25): 2540-2568.

Shinozuka, H., Cogan, N.O.I., Smith, K.F., Spangenberg, G.C, 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**: 343-355.

Shivanna, K.R., Heslop-Harrison, Y. and Heslop-Harrison, J. (1982). The pollen-stigma
interaction in the grasses. 3. Features of the self-incompatibility response. *Acta Botanica Neerlandica* **31:** 307-319.

Sijacic, P., Wang, X., Skirpan, A.L., Wang, Y., Dowd, P.E., McCubbin, A.G., Huang, S. and Kao, T.H. (2004). Identification of the pollen determinant of S-RNase-mediated self-incompatibility. *Nature 429*: 302-305.

Silva N.F. and Goring D.R. (2001). Mechanisms of self-incompatibility in flowering plants. *Cellular and Molecular Life Sciences* **58**: 1988-2007

Sim, S., Chang, T., Curley, J., Warnke, S.E., Barker, R.E., Jung, G. (2005). Chromosomal rearrangements differentiating the ryegrass genome from the Triticeae, oat, and rice genomes using common heterologous RFLP probes. *Theoretical and Applied Genetics* **110**: 1011-1019.

Sims, T.L. and Robbins, T.P. (2009). Gametophytic self-incompatibility in Petunia. **In**: Gerats, T and Strommer, J. (eds) Petunia: evolutionary, developmental and physiological genetics. *Springer-Verlag*: 85-106.

Skowyra D., Craig, K.L., Tyers, M., Elledge, S.J. and Harper, J.W. (1997). F-box proteins are receptors that recruit phosphorylated substrates to the SCF ubiquitin-ligase complex. *Cell* **91** (2): 209-219.

Slade, A.J., Fuerstenberg, S.I., Loeffler, D., Steine, M.N. and Facciotti, D. (2005). A reverse genetic, nontransgenic approach to wheat crop improvement by TILLING. *Nature Biotechnology* **23**:75-81.

Snowman, B.N., Kovar, D.R., Shevchenko, G., Franklin-Tong, V.E. and Staiger, C.J. (2002). Signal-mediated depolymerization of actin in pollen during the self-incompatibility response. *The Plant Cell* **14**: 2613-2626.

Socia i Company, R., Fernandez I Marti, A, Kodad, O. and Alonso, J.M. (2012). Self-compatibility in *Prunus* species: diversity of mutations. *19th Eucarpia General Congress*, Budapest, Hungary.

Song, X.L., Sun, X.Z. and Zhang, T.Z. (2006). Segregation distortion and its effect on genetic mapping in plants. *Journal of Agricultural Biotechnology* **14**(2): 286–292.

Sonneveld, T., Kenneth, R., Tobutt, K.R., Simon, P., 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**: 37-51.

Sonneveld, T., Robbins, T.P. and Tobutt, K.R. (2006). Improved discrimination of self-incompatibility S-RNase alleles in cherry and high throughput genotyping by automated sizing of first intron polymerase chain reaction products. *Plant Breeding* **125**: 305-307.

Stanke, M., Steinkamp, R., Waack, S. and Morgenstern, B. (2004). AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Research* **32** (suppl 2): W309-W312.

Stein J. C., Howlett B., Boyes D. C., Nasrallah M. E. and Nasrallah J. B. (1991). Molecular cloning of a putative receptor protein kinase gene encoded at the self-incompatibility locus of *Brassica oleracea*. *Proceedings of the National Academy of Sciences of the United States of America* **88**: 8816-8820.

Stirling J. (1932). Studies of flowering in heterostyled and allied species. Part I. The Primulaceae. *Publication of the Hartley Botanical Laboratory* **8**: 3–42

Studer, B., Asp, T., Frei, U., Hentrup, S., Meally, H., Guillard, A., Barth, S., Muylle, H., Roldan-Ruiz, I., Barre, P., Koning-Boucoiran, C., Uenk-Stunnenberg, G., Dolstra, O., Skøt, L., Skøt, K.P., Turner, L.B., Humphreys, M.O., Kölliker, R., Roulund, N., Nielsen, K.K. and Lübberstedt, T. (2008). Expressed sequence tag-derived microsatellite markers of perennial ryegrass (*Lolium perenne* L.). *Molecular Breeding* **21** (4): 533-548.

Studer, B., Jensen, L., Fiil, A. and Asp, T. (2009) "Blind" mapping of genic DNA sequence polymorphisms in *Lolium perenne* L. by high resolution melting curve analysis. *Molecular Breeding* **24**: 191–199.

Studer, B., Byrne, S., Nielsen, R.O., Panitz, F., Bendixen, C., Islam, S., Pfeifer, M., Lübberstedt, T. and Asp, T. (2012). A transcriptome map of perennial ryegrass (*Lolium perenne* L.). *BMC Genomics* **13**: 140.

Sutton I. (1918). Report on tests of self-sterility in plums, cherries and apples at the John Innes Horticultural Institution. *Journal of Genetics* **7**: 281-300.

Suzuki, G., Kai, N., Hirose, T., Fukui, K., Nishio, T., Takayama, S., Isogai, A., Watanabe, M. and Hinata, K. (1999). Genomic organization of the *S*-locus: Identification and characterization of genes in *SLG*/*SRK* region of *S*9 haplotype of *Brassica campestris* (syn.
rapa). *Genetics* **153**: 391-400.

Suzuki, T., Kusaba, M., Matsushita, M., Okazaki, K. and Nishio, T. (2000). Characterization of *Brassica S*-haplotypes lacking *S*-locus glycoprotein. *FEBS Letters* **482**: 102-108.

Syvänen, A.C. (2001). Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nature reviews genetics* **2**: 930-942.

Takasaki, T., Hatakeyama, K., Suzuki, G., Watanabe, M., Isogai, A. and Hinata, K. (2000). The *S* receptor kinase determines self-incompatibility in *Brassica* stigma. *Nature* **403**: 913-916.

Takayama, S., Shiba, H., Iwano, M., Shimosato, H., Che, F.S., Kai, N., Watanabe, M., Suzuki, G., Hinata, K., and Isogai, A. (2000). The pollen determinant of self-incompatibility in *Brassica campestris*. *Proceedings of the National Academy of Sciences of the United States of America* **97**: 1920–1925.

Takayama, S., Shimosato, H., Shiba, H., Funato, M., Che, F.S., Watanabe, M., Iwano, M., and Isogai, A. (2001). Direct ligand receptor complex interaction controls *Brassica* self-incompatibility. *Nature* **413**: 534–538.

Talbert, L.E., Blake, N.K., Chee, P.W., Blake, T.K. and Magyar, G.M. (1994). Evaluation of sequence-tagged-site-facilitated PCR products as molecular markers in wheat. *Theoretical and Applied Genetics* **87**: 789-794.

Tan, L. W. and Jackson, J. F. (1988). Stigma proteins of the two loci self-incompatible grass Phalaris coerulescens. *Sexual Plant Reproduction* **1**(1): 25-27

Tantikanjana, T. and Nasrallah, J.B. (2012). Non-cell-autonomous regulation of crucifer self-incompatibility by Auxin Response Factor ARF3. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 19468-19473.

Tao, R., Yamane, H., Sassa, H., Mori, H., Gradziel, T.M., Dandekar, A.M. and Sugiura, A. (1997). Identification of stylar RNases associated with gametophytic self-incompatibility in almond (*Prunus dulcis*). Plant and Cell Physiology **38**: 304-311.

Tao, R., Yamane, H., Sugiura, A., Murayama, H., Sassa, H. and Mori, H. (1999). Molecular typing of *S*-alleles through identification, characterization and cDNA cloning for S-RNases in sweet cherry. *Journal of American Society for Horticultural Science* **124**: 224-233.

Tautz, D. and Renz, M. (1984). Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Research* **12**: 4127-4138.

The International Brachypodium Initiative (2010). Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* **463**: 763-768.

The international HapMap consortium (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**: 851-861.

The Rice Genome Sequencing Project (2005). The map-based sequence of rice genome. *Nature* **436**: 793-800.

Thomas, S.G. and Franklin-Tong, V.E. (2004). Self-incompatibility triggers programmed cell death in *Papaver* pollen. *Nature* **429**: 305-309.

Thorogood, D. and Hayward, M.D. (1991). The genetic control of self-incompatibility in an inbred line of *Lolium perenne* L. *Heredity* **67**: 175-181.

Thorogood, D., Kaiser, W.J., Jones, J.G. and Armstead, I. (2002). Self-incompatibility in ryegrass 12. Genotyping and mapping the *S* and *Z* loci of *Lolium perenne* L. *Heredity* **88**: 385-390.

Thorogood, D., Armstead, I.P., Turner, L.B., Humphreys, M.O. and Hayward, M.D. (2005). Identification and mode of action of self-incompatibility loci in *Lolium perenne* L. *Heredity* **94**: 356-363.

Till, B.J., Reynolds, S.H., Weil, C., Springer, N., Burtner, C., Young, K., Bowers, E., Codomo, C.A., Enns, L.C., Odden, A.R., Greene, E.A., Comai, L. and Henikoff, S. (2004). Discovery of induced point mutations in maize genes by TILLING. *BMC Plant Biology* **4**: 12.

Till, B.J., Cooper, J., Tai, T.H., Colowit, P., Greene, E.A., Henikoff, S. and Comai, L. (2007). Discovery of chemically induced mutations in rice by TILLING. *BMC Plant Biology* **7**: 19.

Tomaszewski, C., Byrne, S.L., Foito, A., Kildea, S., Kopecký, D., Doležel, J., Heslop-Harrison, J.S., Stewart, D. and Barth, S. (2012). Genetic linkage mapping in a F2 perennial ryegrass population using DArT markers. *Plant Breeding* **131**: 345-349.

Tsai, D.S., Lee, H.S., Post, L.C., Kreiling, K.M., and Kao, T.H. (1992). Sequence of an *S*-protein of *Lycopersicon peruvianum* and comparison with other solanaceous S-proteins. *Sexual Plant Reproduction* **5**: 256–263.

Tsukamoto, T., Tao, R. and Iezzoni, A.F. (2008). PCR markers for mutated *S*-haplotypes enable discrimination between self-incompatible and self-compatible sour cherry selections. *Molecular Breeding* **21**: 67-80.

Turner, L.B., Cairns, A.J., Armstead, I.P., Ashton, J., Skot, K., Whittaker, D. and Humphreys, M.O. (2006). Dissecting the regulation of fructan metabolism in perennial ryegrass (*Lolium perenne*) with quantitative trait locus mapping. *New Phytologist* **169**: 45–57.

Ushijima, K., Sassa, H., Dandekar, A.M., Gradziel, T.M., Tao, R. and Hirano, H. (2003). Structural and transcriptional analysis of the self-incompatibility locus in almond: identification of a pollen-expressed F-box gene with haplotype-specific polymorphism. *The Plant Cell* **15**: 771-781.

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-compatible haplotypes of *Prunus avium* and *P. mume*. *The Plant Journal* **39**: 573-586.

Van Bakel, H., Stout, J.M., Cote, A.G., Tallon, C.M., Sharpe, A.G., Hughes, T.R. and Page, J.E. (2011). The draft genome and transcriptome of *Cannabis sativa*. *Genome Biology* **12**: R102.

Van Daele, I., Van Bockstaele, E., Martens, C. and Roldán-Ruiz, I. (2008). Identification of transcribed derived fragments involved in self-incompatibility in perennial ryegrass (*Lolium perenne* L.) using cDNA-AFLP. *Euphytica* **163**: 67-80.

Van Deynze, A.E., Sorrells, M.E., Park, W.D., Ayres, N.M., Fu, H., Cartinhour, S.W., Paul, E. and McCouch, S.R. (1998). Anchor probes for comparative mapping of grass genera. *Theoretical and Applied Genetics* **97**: 356-369.

Vekemans, X., Schierup, M. H. and Christiansen, F.B. (1998). Mate availability and fecundity selection in multi-allelic self-incompatibility systems in plants. *Evolution* **52**: 19-29.

Vergnaud, G. (1989). Polymers of random short oligonucleotides detect polymorphic loci in the human genome. *Nucleic Acids Research* **17**: 7623-7630.

Voylokov, A.V., Fuong, F.T. and Smirnov, V.G. (1993). Genetic studies of self-fertility in rye (*Secale cereale* L.). 1. The identification of genotypes of self-fertile lines for the *Sf* alleles of self-incompatibility genes. *Theoretical and Applied Genetics* **87:** 616-618.

Voylokov, A.V., Korzun, V. and Borner, A. (1998). Mapping of three self-fertility mutations in rye (*Secale cereale* L.) using RFLP, isozyme and morphological markers. *Theoretical and Applied Genetics* **97:** 147-153.

Walker, E.A., Ride, J.P., Kurup, S., Franklin-Tong, V.E., Lawrence, M.J. and Franklin, F.C. (1996). Molecular analysis of two functional homologues of the *S3* allele of the *Papaver rhoeas* self-incompatibility gene isolated from different populations. *Plant Molecular Biology* **30**: 983-994.

Wallace, R.B., Shaffer, J., Murphy, R.F., Bonner, J., Hirose, T. and Itakura, K. (1979). Hybridization of synthetic oligodeoxyribonucleotides to phi 174 DNA: the effect of single base pair mismatch. *Nucleic Acids Research* **6** (11): 3543-3557.

Wang, Y., Wang, X., Skirpan, A.L. and Kao, T.H. (2003). S-RNase-mediated self-incompatibility. *Journal of Experimental Botany* **54**: 115-122.

Wang, Y., Tsukamoto, T., Yi, K.W, Wang, X., Huang, S., McCubbin, A.G. and Kao, T.H. (2004). Chromosome walking in the *Petunia inflata* self-incompatibility (S-) locus

and gene identification in an 881-kb contig containing S2-RNase. *Plant Molecular Biology* **54**: 727-742.

Wang, Z., Gerstein, M. and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* **10**: 57-63.

Wanlong Li, W., Zhang, P., Fellers, J.P., Friebe, B. and Gill, B.S. (2004). Sequence composition, organization, and evolution of the core Triticeae genome. *The Plant Journal* **40**: 500-511.

Washburn, M.P., Wolters, D. and Yates, J.R. (2001). Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnology* **19**: 242-247.

Watson, J.D. and Crick, F.H.C. (1953). A structure for Deoxyribose Nucleic Acid. *Nature* **171**: 737-738.

Wehling, P., Hackauf, B. and Wricke, G. (1994). Phosphorylation of pollen proteins in relation to self-incompatibility in rye (*Secale cereale* L.). *Sexual Plant Reproduction* **7:** 67-75.

Wheeler, M.J., de Graaf, B.H.J., Hadjiosif, N., Perry, R.M., Poulter, N.S., Osman, K., Vatovec, S., Harper, A., Franklin, F.C.H. and Franklin-Tong, V.E. (2009). Identification of the pollen self-incompatibility determinant in *Papaver rhoeas*. *Nature* **459**: 992-995.

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. *Journal of Experimental Botany* **61**: 2015-2025.

Wittwer, C.T., Reed, G.H., Gundry, C.N., Vandersteen, J.G. and Pryor, R.J. (2003) High-resolution genotyping by amplicon melting analysis using LCGreen. *Clinical chemistery* **49**: 853-860.

Wilkins, P.W. and Humphreys, M.O. (2003). Progress in breeding perennial forage grasses for temperate agriculture. *Journal of Agricultural Science* **140**: 129-150.

Woltering, E.J., van der Bent, A. and Hoeberichts, F.A. (2002). Do plant caspases exist? *Plant Physiology* **130**: 1764-1769.

Wright, S. (1939). The distribution of self-sterility alleles in populations. *Genetics* **24**: 538-552.

Wu, J., Wang, S., Gu, Y., Zhang, S., Publicover, S.J. and Franklin-Tong, V.E. (2011). Self-incompatibility in *Papaver rhoeas* activates nonspecific cation conductance permeable to Ca^{2+} and $K^{+[W]}$. *Plant physiology* **155**: 963-973.

Wu, S.B., Tavassolian, I., Rabiei, G., Hunt, P., Wirthensohn, M., Gibson, J.P., Ford, C.M. and Sedgley, M. (2009) Mapping SNP-anchored genes using high-resolution melting analysis in almond. *Molecular genetics and genomics* **282** (3): 273-281.

Xu, B., Grun, P., Kheyr-Pour, A. and Kao, T.H. (1990). Identification of pistil-specific proteins associated with three selfincompatibility alleles in *Solanum chacoense*. *Sexual Plant Reproduction* **3**: 54–60.

Xu, Y., Zhu, L., Xiao, J., Huang, N and McCouch, S.R. (1997). Chromosomal regions associated with segregation distortion of molecular markers in F_2, backcross, double haploid and recombinant inbred population in rice (*Oryza sativa* L.). *Molecular and General Genetics* **253**: 535-545.

Xue, Y., Carpenter, R., Dickinson, H.G. and Coen, E.S. (1996). Origin of allelic diversity in *Antirrhinum S* locus RNases. *Plant Cell* **8**: 805-814.

Yamane, H., Ikeda, K., Ushijima, K., Sassa, H. and Tao, R. (2003). A pollen-expressed gene for a novel protein with an F-box motif that is very tightly linked to a gene for S-RNase in two species of cherry, *Prunus cerasus* and *P. avium*. *Plant and Cell Physiology* **44**: 764-769.

Yang, B., Thorogood, D., Armstead, I. P., Franklin, F.C.H. and Bart, S. (2009). Identification of genes expressed during the self-incompatibility response in perennial ryegrass (*Lolium perenne* L.). *Plant Molecular Biology* **70**: 709-723.

Yang, B., Thorogood, D., Armstead, I. P. and Bart, S. (2008). How far are we from unravelling self-incompatibility in grasses? *New Phytologist* **178**: 740-753.

Zamir, D. and Tadmor, Y. (1986). Unequal segregation of nuclear genes in plants. *Botanical Gazette* **147** (3): 355-358.

Zhao, L., Huang, J., Zhao, Z., Li, Q., Sim, T. and Xue, Y. (2010). The Skp1-like protein SSK1 is required for cross-pollen compatibility in S-RNase-based self incompatibility. *The Plant Journal* **62**: 52–63.

Zöllner, S., Wen, X., Hanchard, N.A., Herbert, M.A., Ober, C. and Pritchard, J.K. (2004). Evidence for extensive transmission distortion in the human genome. *American Journal of Human Genetics* **74** (1): 62-72.

Zurek, D.M., Mou, B., Beecher, B. and McClure, B. (1997). Exchanging sequence domains between S-RNases from *Nicotiana alata* disrupts pollen recognition. *The Plant Journal* **11**: 797-808.

Supplemental files

Method for the 454 library preparation:

1-1-1.1. DNA samples quantification

Before the preparation of the library, the DNA extracted from each BAC clone had to be quantified accurately. In order to do so, the Quant-iT™ PicoGreen® dsDNA assay kit (Invitrogen, Grand Island, NY, US) was used, following the manufacturer instructions. A DNA standard curve was prepared using different dilutions of a lambda solution and samples were diluted 50 times using 1X TrisEDTA (TE). An equal volume of the PicoGreen solution (previously diluted to 2X) was added to each samples and standards. The fluorescence was read three times using a Real-Time PCR machine and using the standard curve, the concentration of each sample was calculated.

1-1-1.2. DNA nebulization

The first step of the library preparation is the nebulization of the DNA. For each sample, a volume corresponding to 500 ng of DNA was placed into a tube and TE buffer was added to a final volume of 100 µl. For each sample, 100 µl was placed into a Nebulizer and topped up with 500 µl of Nebulization Buffer. A pressure of 30psi (2.1 bar) of nitrogen was applied for 1 minute under a fume cabinet before adding 2.5 ml of PBI Buffer. The nebulized DNA samples were purified using a MinElute PCR Purification kit (Qiagen, Hilden, Germany). The column was loaded with 750 µl of the sample, centrifuged for 15 sec at 13,000 rpm and the flow-through was discarded. Those steps were repeated with the rest of the sample before the column was eluted in a new tube with 16 µl of TE buffer.

1-1-1.3. Fragment-ends repair

The second step of the library preparation was the fragment end repair. To the 16 µl from each sample was added 9 µl of End Repair mix consisting of 2.5 µl RL 10X PNK Buffer, 2.5 µl RL ATP, 1 µl RL dNTP, 1 µl RL T4 polymerase, 1 µl RL PNK and 1 µl RL Taq polymerase. After a quick vortex and spin, samples were placed in a thermocycler and the End Repair program was run for 20 min at 25°C followed by 20 min at 72°C. After the program finished 1 µl of RL MID adaptor was added to each sample as well as 1 µl of RL ligase. Samples were quickly vortex and spun before being incubated at 25°C for 10

minutes in order to ligate the adaptor. The MID adaptors are specific to each sample and will enable pooling them for the sequencing (see Table 1 for the corresponding MID to the samples).

Table 1: MID adaptors sequences for each BAC clones

MID	Sequence	BAC clone
1	ACACGACGACT	202
2	ACACGTAGTAT	175
3	ACACTACTCGT	180
4	ACGACACGTAT	227
6	ACGCGTCTAGT	714
7	ACGTACACACT	764
8	ACGTACTGTGT	520-10H
9	ACGTAGATCGT	573
10	ACTACGTCTCT	520-7A

1-1-1.4. Small fragments removal

The third step of the protocol was to remove the small fragment. AMPure beads were prepared for each sample as followed: 125 µl of AMPure beads were placed into a tube and the tube was placed onto a magnetic particle concentrator (MPC). The supernatant was removed when the beads were pelleted on the side of the tube and 73 µl of TE buffer was added. The tube was quickly vortex before 500 µl of Sizing Solution was added. After a quick vortex and spin, each DNA sample was added to a tube of AMPure beads and incubated at room temperature for 5 minutes. The tubes containing the samples were placed onto the MPC and once the beads had fully pelleted on the wall of the tube, the supernatant was discarded. The beads were then washed twice using 100 µl of TE buffer and 500 µl of sizing solution with a quick vortex and spin and an incubating step of 5 minutes at room temperature. The supernatant was discarded each time using the MPC.

Another two washes were applied, this time using 1 ml of 70% ethanol. The supernatant was removed and the pellet was air dried at room temperature for 2 minutes. To elute the DNA of the beads, 53 µl of TE buffer was added to each sample, the tubes were placed back onto the MPC and 50 µl of the supernatant was transferred to a new tube.

1-1-1.5. Quality assessment and quantification of the libraries

The final step of the library preparation is the library quantification and the library quality assessment. To quantify the DNA library, a standard curve was made using a series of 8 dilutions (2.5×10^9, 1.67×10^9, 1.11×10^9, 7.41×10^8, 4.94×10^8, 3.29×10^8, 2.19×10^8 and 1.46×10^8) from the initial RL standard (5×10^9 molecule/µl). Each dilution was then placed into a cuvette and the fluroscence was read using a TBS 380 Mini-Fluorometer (Turner BioSystems, Sunnyvale, California, US). Using the standard curve, the DNA concentration was calculated for each sample and the samples were diluted down to 1×10^7 molecule/µl.

For the quality assessment, 1 µl of the DNA library (before dilution) was loaded on an Experion® High Sensitivity DNA chip (BioRad Laboratories Inc., Hercules, CA, US) and the chip was placed onto the Experion® Automated Electrophoresis Station (BioRad Laboratories Inc., Hercules, CA, US) to be analysed. The quality assessment was then done by using the report of the electropherogram.

Method for the Illumina library preparation:

1. Purification and fragmentation of the mRNA

The purification stage will remove the poly-A tail of the mRNA as it can interfere with the sequencing, as well as any other type of RNA (e.g. ribosomal). During this purification, the RNA is also fragmented and random adapters are added for the cDNA synthesis step following.

The RNA samples were diluted to 50 µl using nuclease-free water and 50 µl of the RNA purification beads was added (the beads are coated with oligo dT that will bind to the poly-A RNA0. The samples are mixed gently by pipetting up and down 6 times and incubated for 5 minutes at 65°C before been placed on ice to cool down. Samples were left incubating 5 minutes at room temperature before been placed on a magnetic stand in order to separate the beads to the rest of the solution containing rRNA and other non-messenger RNA. The supernatant was discarded and the beads were washed with 200 µl and mixed by pipetting up and down 6 times. The tubes were placed again on the magnetic stand and after 5 minutes, the supernatant was removed and 50 µl of the Elution Buffer was added and mixed by pipetting up and down 6 times. The samples were incubated for 2 minutes at 80°C first and then for 2 minutes at 25°C before 50 µl of Bead Binding Buffer was added and mixed by pipetting up and down 6 times. The tubes were placed on the magnetic stand for 5 minutes and the supernatant was discarded. The beads were washed with 200 µl of Bead Washing Buffer, mixed by pipetting up and down 6 times and the supernatant was discarded after the tubes stood 5 minutes on the magnetic stand. 19.5 µl of Elute, Prime, Fragment Mix was added to the beads and mixed by pipetting up and down 6 times. The samples were incubated for 8 minutes at 94°C and then placed on ice to cool down before been briefly centrifuged.

1. First strand cDNA synthesis

The first strand cDNA synthesis is the reverse transcription of the RNA, using the random primer attached in the previous step as primers and reverse transcriptase.

The samples were placed on the magnetic stand for minutes and 17 µl of the supernatant was transferred into new PCR tubes. The SuperScript II (SSII) was mixed to

the First Stand Master Mix (FSM), 1 µl of SSII for each 9 µl of FSM. After mixing it, 8 µl of this SSII/FSM mix was added to each sample and mixed by pipetting up and down 6 times. The samples were placed in a thermocycler with the following program: lid heated to 100°C, 10 minutes at 25°C, 50 minutes at 42°C, 15 minutes at 70°C and hold at 4°C. The samples were processed to the next step of the protocol immediately.

2. Second strand cDNA synthesis

This step is to complete the cDNA synthesis and have the cDNA double-stranded (ds). At the end, all RNA will be removed from the samples.

To each samples, 25 µl of the Second Strand Master Mix was added and mixed by pipetting up and down 6 times. The samples were incubated for one hour at 16°C. The AMPure XP beads were vortexed and 90 µl of it was added to each sample and mixed by pipetting up and down 10 times. The samples were incubated for 15 minutes at room temperature before been placed on the magnetic stand for 5 minutes and 135 µl of the supernatant discarded. With the tubes still on the magnetic stand, the beads were washed by adding 200 µl of freshly made 80% ethanol, without disturbing the beads. After 30 seconds incubation at room temperature, the ethanol was removed and the washing step was repeated two more time. The samples were left for 15 minutes at room temperature to dry before been removed from the magnetic stand. 52.5 µl of Resuspension Buffer was added to each sample and mixed by pipetting up and down 10 times. The samples were incubated for 2 minutes at room temperature before been placed on the magnetic stand and 50 µl of the ds cDNA was transferred to clean tubes.

3. Repair ends

After the trimming of the RNA in the first step of this library protocol, the ends of the fragment need to be repaired and made into blunt ends.

To each samples, 40 µl of the End Repair Mix was added and mixed by pipetting up and down 10 times. The samples were incubated for 30 minutes at 30°C before 160 µl of well-mixed AMPure XP beads were added and mixed by pipetting up and down 10 times. After 15 minutes incubation at room temperature, the samples were place on the magnetic stand and incubated for a further 5 minutes. 250 µl of the supernatant was discarded in two

times (2*127.5 µl) before the beads were washed three time with 200 µl of 80% ethanol, the ethanol been discarded after 30 seconds incubation at room temperature between washes. Samples were left to dry 15 minutes at room temperature and removed from the magnetic bead. The pellet was resuspended with 17.5 µl of Resuspension Buffer, mixed by pipetting up and down 10 times and incubated for 2 minutes at room temperature. Samples were placed on the magnetic stand for 5 minutes and 15 µl of the clear supernatant was transferred into new tubes.

4. Adenylate 3' Ends

This step of the protocol will add a single nucleotide "A" to the blunt 3' end of the fragments, preventing the ligation between them during the adaptor ligation, adapter that have a "T" nucleotide on the 3' end, providing a complementary overhang.

12.5 µl of A-Tailing Mix and 2.5 µl of Resuspension Buffer were added to each sample and mixed by pipetting up and down 10 times. The samples were incubated for 30 minutes at 37°C and then cool down before the next stage.

5. Adapters ligation

This process ligates an index to each fragment, index specific to each library that will enable multiplexing samples.

To each sample, 2.5 µl of Resuspension Buffer, 2.5 µl of Ligation Buffer and 2.5 µl of a specific RNA Adapter Index were added and mixed by pipetting gently 10 times. The Table 1 give the specific indexes that were used for each samples. The samples were incubated for 10 minutes at 30°C before 5 µl of Stop Ligation Buffer was added and mixed by pipetting 10 times.

The samples are then cleaned-up by adding 42 µl of AMPure XL beads and mixing by pipetting up and down 10 times. After 15 minutes incubation at room temperature, the tubes were placed on the magnetic stand and left for 5 minutes before 79.5 µl of the supernatant was discarded. While remaining on the magnetic stand, the samples were washed three times with 200 µl of 80 % ethanol, discarding the supernatant between each wash after 30 seconds incubation. After the washes, the samples were left to dry for 15

minutes at room temperature and removed from the magnetic stand to rediscolve the pellet using 52.5 µl of Resuspension Buffer. After a gentle mix by pipetting and incubation for 2 minutes at room temperature, the tubes were placed back on the magnetic stand and 50 µl of the supernatant was pipetted out after 5 minutes and transferred to a new tube. Another clean-up was done by adding 50 µl of AMPure XL beads to each sample. After 15 minutes incubation at room temperature, the samples were placed on the magnetic stand and left for 5 minutes before 95 µl of the supernatant was removed. With the tubes remaining on the magnetic stand, the samples were washed three times with 200 µl of 80 % ethanol, discarding the supernatant between each wash after 30 seconds incubation. After 15 minutes at room temperature to dry, the tubes were removed from the magnetic stand and the pellets were resuspended with 22.5 µl of Resuspension Buffer. The samples were mixed by pipetting up and down 10 times, incubated 2 minutes at room temperature and a further 5 minutes on the magnetic stand before 20 µl of the clear supernatant was transferred to a new tube for each sample.

Table 1: Indexed Adapter used for the library preparation for each samples.

Sample	Indexed Adapter	Sequence
F1_30_1	AR018	GTCCGC
F1_30_2	AR019	GTGAAA
F1_30_3	AR020	GTGGCC
Male_1	AR021	GTTTCG
Male_2	AR022	CGTACG
Male_3	AR023	GAGTGG
SI_1	AR008	ACTTGA
SI_2	AR009	GATCAG
SI_3	AR010	TAGCTT
SC_1	AR011	GGCTAC
SC_2	AR013	AGTCAA
SC_3	AR015	ATGTCA
58/3 S	AR006	GCCAAT
58/3 P	AR007	CAGATC
59/3 S	AR012	CTTGTA
59/3 P	AR004	TGACCA
59/21 P	AR005	ACAGTG

6. PCR amplification

This step of the preparation is to amplify specifically the fragments that have properly ligated to the adapters on both ends. The number of cycles in the amplification needs to minimize to avoid skewing the representation of different fragments within a library.

To each sample, 5 µl of PCR Primer Cocktail and 25 µl of PCR Master Mix were added and mixed by pipetting up and down 10 times. The samples were amplified using a thermocycler with the following program: heated lid at 100°C, 30 seconds at 98°C, 15 cycles of 10 seconds at 98°C, 30 seconds at 60°C and 30 seconds at 72°C followed by one hold for 5 minutes at 72°C. The PCR products were then cleaned-up by adding 50 µl of well-mixed AMPure XL beads and mixing by pipetting up and down 10 times. The samples were incubated 15 minutes at room temperature and placed on the magnetic stand. After 5 minutes, 95 µl of the supernatant was discarded and the beads were washed three times using 200 µl of 80% ethanol, discarded between each wash after 30 seconds incubation. The beads were dried for 15 minutes at room temperature and the tubes were removed from the magnetic stand. The pellets were resuspended with 32.5 µl of Resuspension Buffer and gently mixed by pipetting up and down 10 times. After 2 minutes incubation at room temperature, the tubes were placed back on the magnetic stand and leaved for 5 minutes before 30 µl of the clear supernatant was transferred to a new tube.

At the stage, the libraries can be kept at -20°C for up to a week and are ready for sequencing but before doing so, the quality of them as well as the quantity needs to be checked.

7. Quality check

The quality of the libraries was assessed by loading 1 µl of each cDNA library onto an Experion® High Sensitivity DNA chip (BioRad Laboratories Inc., Hercules, CA, US). From the results, the quality and the concentration of the libraries were assessed.