Hybrid breeding in perennial grasses based on self-incompatibility and self-fertility

Andrea Sofía Arias Aguirre

Iowa State University

Follow this and additional works at: http://lib.dr.iastate.edu/etd

Part of the Agricultural Science Commons, Agriculture Commons, Agronomy and Crop Sciences Commons, Genetics Commons, and the Molecular Biology Commons

Recommended Citation

Arias Aguirre, Andrea Sofía, "Hybrid breeding in perennial grasses based on self-incompatibility and self-fertility" (2013). Graduate Theses and Dissertations. 13008.

http://lib.dr.iastate.edu/etd/13008

This Dissertation is brought to you for free and open access by the Graduate College at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
Hybrid breeding in perennial grasses based on self-incompatibility and self-fertility

by

Andrea Arias Aguirre

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Plant Breeding

Program of Study Committee:
Thomas Lübberstedt, Major Professor
Shuizhang Fei
Michael Lee
Reid Palmer
Erik Vollbrecht

Iowa State University

Ames, Iowa

2013

Copyright © Andrea Arias Aguirre, 2013. All rights reserved
THESIS DEDICATION

I would like to dedicate this thesis to the most important people in my life, my parents, sisters, nice and nephew and my in-laws. They have been the light and inspiration in my life.

I want to extend my gratitude to Thomas Lübberstedt and all the members of his group, in special to Uschi and Elizabeth for their support throughout this research. All of them along with the members of my committee and other professors here at ISU gave me great advice and were great teachers for my career and life. I thank the PSI for their support during my graduate education. Finally and most importantly, I want to give a special dedication of this thesis to my loving husband, for all his support and patience and for being the best partner in this incredible journey.
Table of Contents

List of Figures ........................................................................................................................................ vi
List of Tables ........................................................................................................................................ viii
General Introduction .......................................................................................................................... 1
  Organization of the Thesis ............................................................................................................... 3
  References ........................................................................................................................................... 3
Prospects for hybrid breeding in bioenergy grasses ......................................................................... 6
  ABSTRACT ......................................................................................................................................... 6
  INTRODUCTION ............................................................................................................................. 6
  Current population and synthetic breeding schemes in switchgrass and perennial ryegrass .......... 9
  Prospects for hybrid breeding in perennial grasses ....................................................................... 11
  Identification of heterotic patterns and development of heterotic groups .................................. 11
  Biological mechanisms for effective hybrid seed production ....................................................... 13
Hybrid breeding strategies for perennial grasses ............................................................................. 16
  Population hybrids ....................................................................................................................... 16
  Chance hybrids ............................................................................................................................ 17
  Single cross hybrids ..................................................................................................................... 20
CONCLUSIONS .............................................................................................................................. 24
REFERENCES ........................................................................................................................................ 26
PollenCALC: Software for estimations of pollen compatibility of self-incompatible allo- and autotetraploid species .............................................................................................................. 31
  ABSTRACT ......................................................................................................................................... 31
  Background ....................................................................................................................................... 31
  Results ............................................................................................................................................... 32
  Conclusion ......................................................................................................................................... 32
  Keywords .......................................................................................................................................... 32
  BACKGROUND ............................................................................................................................... 32
  Implementation ............................................................................................................................... 37
  Calculator parameters for meiosis ............................................................................................... 37
  Calculator frequencies estimation for segregation of S and Z ....................................................... 37
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calculator input and settings</td>
<td>38</td>
</tr>
<tr>
<td>Calculator output</td>
<td>40</td>
</tr>
<tr>
<td>RESULTS AND DISCUSSION</td>
<td>44</td>
</tr>
<tr>
<td>Difference between preferential and non-preferential pairing</td>
<td>44</td>
</tr>
<tr>
<td>Model testing</td>
<td>47</td>
</tr>
<tr>
<td>Allele composition</td>
<td>49</td>
</tr>
<tr>
<td>CONCLUSION</td>
<td>52</td>
</tr>
<tr>
<td>Availability and requirements</td>
<td>52</td>
</tr>
<tr>
<td>Abbreviations</td>
<td>53</td>
</tr>
<tr>
<td>Competing interests</td>
<td>53</td>
</tr>
<tr>
<td>Authors’ contributions</td>
<td>53</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>53</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>53</td>
</tr>
<tr>
<td>Mapping a new source of self-fertility in perennial ryegrass</td>
<td>55</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>55</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>56</td>
</tr>
<tr>
<td>MATERIALS AND METHODS</td>
<td>57</td>
</tr>
<tr>
<td>Plant materials</td>
<td>57</td>
</tr>
<tr>
<td>Pollination tests</td>
<td>58</td>
</tr>
<tr>
<td>Phenotypic segregation analysis</td>
<td>58</td>
</tr>
<tr>
<td>Marker data</td>
<td>59</td>
</tr>
<tr>
<td>Linkage map construction</td>
<td>60</td>
</tr>
<tr>
<td>Phenotypic and marker segregation analysis</td>
<td>60</td>
</tr>
<tr>
<td>Hypothesis testing</td>
<td>61</td>
</tr>
<tr>
<td>RESULTS</td>
<td>61</td>
</tr>
<tr>
<td>Phenotypic data</td>
<td>61</td>
</tr>
<tr>
<td>Linkage mapping</td>
<td>63</td>
</tr>
<tr>
<td>Marker segregation analysis for S and Z markers</td>
<td>64</td>
</tr>
<tr>
<td>Hypothesis testing</td>
<td>67</td>
</tr>
<tr>
<td>DISCUSSION</td>
<td>69</td>
</tr>
<tr>
<td>Population size, genetic map and segregation distortion</td>
<td>69</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure 2.1 Breeding schemes using self-incompatibility to produce population hybrids ........17
Figure 2.2 A breeding scheme using self-incompatibility to produce population hybrid using two populations grown in a 1:1 ratio ........................................................................................................20
Figure 2.3 Flow diagram of phases and decision making processes for single cross hybrid production................................................................................................................................................21
Figure 3.1 Scheme of model differences for self-incompatible species with a S-Z system.....35
Figure 3.2 Input box for meiotic parameters for S-Z and genotype of two individuals .......40
Figure 3.3 First output with input parameter summaries .................................................................41
Figure 3.4 Screenshot of pollen haplotype table .............................................................................43
a) First 19 (out of 100) rows of pollen compatibility calculation for one of the two reciprocal crosses, underlined rows are examples of compatible pollen .........................43
Figure 3.5 Summary output of different runs (changes in genotypes). ........................................44
Figure 3.6 Program output 3 ...........................................................................................................48
Figure 3.7 Hypothesis testing of SI models ....................................................................................49
Figure 4.1. *In vitro* pollination tests for pollen-stigma interaction.............................................62
Figure 4.2. Genetic linkage maps for markers on linkage groups 1, 2, 3, and 5 in the F2 population ..................................................................................................................................................64
Table 4.4. Deviation from expected segregation ratios under four single gene models.
Significant differences are shown at p =0.05 level .....................................................................65
Figure 4.3. Segregation of polymorphic markers on linkage group 3. ........................................66
Figure 4.4. Segregation of polymorphic markers on linkage group 5. ........................................67
Figure 4.5. Proposed gene action model ..........................................................................................71
Figure 5.1 Scheme of autotetraploid population development using colchicine doubling.
Alleles from S are assigned as letters and alleles for Z are assigned numbers .....................79
Figure 5.2 Shifted melting curves for primer ZB1 shows a clear segregation of the 43 genotypes into two categories ..............................................................................................................85
Figure 5.3 Shifted melting curves for primer ZB2 that similarly to ZB1 show a segregation of the population into two groups.

Figure 5.4 Difference curves for primer ZB3 showed up to three different curves for the 43 genotypes in the population.
LIST OF TABLES

Table 3.1 Formulas for pollen haplotype calculations ..........................................................41
Table 3.2 Pollen haplotypes and their respective frequencies for a male abcd1234 with
disomic inheritance ................................................................................................................44
Table 3.3 Pollen haplotypes and their respective frequencies for a male ABCD1234 with
tetrasomic inheritance ...........................................................................................................45
Table 3.4 Differences in reciprocal crosses between genotypes that have alleles identical
by descent (IBD) .........................................................................................................................50
Table 4.1. Markers used for mapping of chromosome regions affecting self-
compatibility. Markers are organized by name, linkage group, and reference for
sequence information ..............................................................................................................59
Table 4.2. Observed and expected counts used in chi-square test for goodness of fit for
different one gene models. P >0.05 means there is no significant difference between the
observed and expected segregation .......................................................................................63
Table 4.3. Deviation from expected segregation ratios under the four single gene
models. Significance is shown at the 0.05 level ....................................................................64
Table 4.4. Deviation from expected segregation ratios under four single gene models.
Significant differences are shown at p =0.05 level ..............................................................65
Table 4.5. Test for independence between marker segregation and phenotypic
segregation for markers in all linkage groups .......................................................................68
Table 5.1 Possible pollen genotypes from Z locus produced by male AABB1122 and
compatibility to the AABB 2233 under each model ............................................................79
Table 5.2 Expected genotypic frequencies for Z locus for all female gametes and
compatible male gametes for the doubled F1-30 and F1-39 parents under Model I and
Model II ..................................................................................................................................80
Table 5.3 Expected genotypic classes and frequencies for the Z locus tetraploid
population progeny under Model I using weighed frequencies ........................................81
Table 5.4 Expected genotypic classes and frequencies for the Z locus tetraploid
population progeny under Model II using weighed frequencies ........................................81
Table 5.5 Melting curves calls for markers ZB1, ZB2 and ZB3 and the respective haplotype designation .............................................................86
Table 5.6 Chi Square test for hypothesis testing of Model I and Model II. ........................................87
CHAPTER 1: GENERAL INTRODUCTION

For a long time, petroleum-based fuels have been used to provide energy, but limited supplies along with environmental effects of fossil fuels created the need of alternative and clean energy sources. Biofuels represent renewable sources that can substantially reduce carbon dioxide emissions. Efforts were undertaken by the Department of Energy (DOE) since the 1970’s to identify renewable energy sources. The DOE created the Bioenergy Feedstock Development Program (BFDP) and the Oak Ridge National Laboratory (ORNL) with the purpose of identifying and evaluating plant species with promising biofuel potential (McLaughlin and Walsh, 1998; McLaughlin and Kszos, 2005; Shepherd, 2000; Walsh et al., 2003). An initial screening of woody and herbaceous species identified four woody species and switchgrass as potential bioenergy crops and as a model species for bioenergy research. However, grasses in general have a low lignin content compared to woody species which is desirable for ethanol production (Heaton et al., 2008; Tilman et al., 2006). Furthermore, other features such as: low water and nutrient requirements, and a high rate of carbon fixation that make grasses desirable for biomass production (Lewandowski et al., 2003). Perennial grasses have also environmental advantages such as reduced soil organic matter and tillage requirements compared with annual species (Lewandowski et al., 2003). There are grass species currently grown for biofuel production, but the urgent need to maximize yield for specific eco-climates requires the development of successful breeding programs.

Perennial ryegrass has been increasingly used in Europe as forage grass. In 2004, it represented 40% of the grass seed production area in the European Union (Kolliker et al., 2005). This cross pollinated species has two main ploidy series: diploids (2n=2x=14) and tetraploids (Thorogood and Hayward, 1991). Ryegrass is forced to outcross due self-incompatibility (SI). Self-
incompatibility can be defined in a broad sense as the inability of a plant to produce seed after self-pollination (Allard, 1999). SI can be either gametophytic, in which the genetic control of SI at the paternal side is determined by the haploid pollen genome: or sporophytic, in which the genetic control depends on the diploid genome of the pollen producing paternal plant (Langridge and Baumann, 2008; Yang et al., 2008). SI in grasses is controlled by a two-locus systems S-Z which has also been described for rye (Hackauf and Wehling, 2005a; Lundqvist, 1954) and switchgrass (Martinez-Reyna and Vogel, 2002). For some SI systems, the gene(s) products have been isolated, including RNases, involved in RNA and consequently pollen degradation, or proteins that can trigger calcium-dependent cascades that also cause pollen degradation (Baumann et al., 2000; Klaas et al., 2011b; Langridge and Baumann, 2008; Yang et al., 2008). In grasses, the gene products of S and Z are not known yet. However, it is known that pollen in an S-Z system is rejected, when both the pollen S and Z allele match the pistil S and Z alleles (Yang et al., 2008). In diploid ryegrass, pollen grains that have at least one different allele from the S or Z alleles in the stigmata are compatible.

For allogamous grasses, selfing rates are considerably low due to SI and the progeny developed by selfing display a strong inbreeding depression (ID). This suggests that grasses are excellent candidates for hybrid breeding. To ensure 100% hybrid seed production, controlled pollinations are required. For ryegrass, male and female organs coexist in the same floral structure and emasculation can be laborious. Conveniently, SI is a built-in pollination control system that can be exploited in hybrid seed production. One imperative requirement for breeding of two-way hybrids is the development of inbred lines. For SI grasses there are two alternatives to develop inbred lines: self-fertility (SF) and pseudo-compatibility (PC). In species with single-locus gametophytic incompatibility, self-fertility arose from S locus mutations (de Nettancourt, 1977). In contrast, a study in rye suggests that self-fertility is caused by a locus different from S and Z
(Lundqvist, 1968). This was later confirmed for ryegrass (Thorogood and Hayward, 1991; Thorogood et al., 2005). In addition, SI systems can be overcome by high temperatures during pollination (30 C) in rye (Gertz and Wricke, 1991) and in perennial ryegrass, causing pseudo-compatibility (Wilkins and Thorogood, 1992). Finally, once the line development and pollination control systems are clearly identified, an additional effort needs to be in place to develop heterotic pools to have a successful breeding program.

Ryegrass will be used in this study to answer questions pertaining to biological and molecular mechanisms that could allow it to become a hybrid crop. In the longer run, ryegrass can serve as a model species to identify candidate genes or regions by exploiting the reported synteny among grasses (Gaut, 2002). The objectives of this research are to; i) discuss the perspectives for hybrid breeding in bioenergy grasses; ii) develop a tool to calculate pollen compatibility for self-incompatible allo-and autotetraploid species; iii) map a new source of self-fertility in ryegrass; and iv) determine the inheritance of polymorphic markers close to the Z gene to validate the self-incompatibility model in tetraploid ryegrass and its effectiveness at a higher ploidy level.

**Organization of the Thesis**

This thesis contains a general and brief introduction (Chapter 1) since most topics are discussed in the review (Chapter 2) and within each manuscript in the other chapters; a published review article (Chapter 2), a published manuscript (Chapters 3), a manuscript submitted to a journal (Chapter 4) and a manuscript in preparation for submission (Chapter 5). A general conclusion of all studies performed is summarized in the last chapter (Chapter 6).

**References**

CHAPTER 2: PROSPECTS FOR HYBRID BREEDING IN BIOENERGY GRASSES

A paper published in BioEnergy Research

Andrea Arias¹, Bruno Studer², Ursula Frei¹, Thomas Lübberstedt¹

ABSTRACT

Biofuels obtained from biomass have the potential to replace a substantial fraction of petroleum-based hydrocarbons that contribute to carbon emissions and are limited in supply. With the ultimate goal to maximize biomass yield for biofuel production, this review aims to evaluate prospects of different hybrid breeding schemes to optimally exploit heterosis for biomass yield in perennial ryegrass (Lolium perenne L.) and switchgrass (Panicum virgatum), two perennial model grass species for bioenergy production. Starting with a careful evaluation of current population and synthetic breeding methods, we address crucial topics to implement hybrid breeding, such as the availability and development of heterotic groups, as well as biological mechanisms for hybridization control such as self-incompatibility (SI) and male sterility (MS). Finally, we present potential hybrid breeding schemes based on SI and MS for the two bioenergy grass species, and discuss how molecular tools and synteny can be used to transfer relevant information for genes controlling these biological mechanisms across grass species.

INTRODUCTION

Biofuels have the potential to replace a substantial fraction of petroleum-based hydrocarbons that contribute to carbon emissions and are limited in supply. Maize, sugarcane, and soybean are currently used to produce biofuels (Coyle, 2006). However, additional crops that are not directly compromising feed and food production are required to implement more sustainable bioenergy
production systems. As part of the efforts to tap renewable energy sources, the Department of Energy (DOE) created the Bioenergy Feedstock Development Program (BFDP) to identify and evaluate plant species with promising biofuel potential (McLaughlin and Kszos, 2005; Shepherd, 2000; Somers et al., 2003; Walsh et al., 2003). The screening for energy crops included 100 species of short-rotation trees and 35 herbaceous species (Kszos et al., 2000), and identified four woody species and switchgrass (*Panicum virgatum*) as model species for bioenergy research (McLaughlin and Walsh, 1998).

Grasses, especially perennial grasses grown for forage production, are advantageous for bioenergy production because due to their perennial nature, farming practices such as tillage are not required as often as for annual crops (Lewandowski et al., 2003). Moreover, they have low water and nutrient requirements, a high rate of carbon fixation, and are consequently effective in biomass production (Searchinger, 2008). In addition, perennial grasses have a low lignin/cellulose ratio, convenient for biomass conversion.

Both of our target species are characterized by a highly effective self-incompatibility (SI) system which promotes cross pollination, thereby maintaining a high level of heterozygosity (de Nettancourt, 1997). Such allogamous species are usually improved as population or synthetic varieties, which exploit heterosis only partially. To increase productivity and sustainability of grasslands for bioenergy production, novel tools and innovative breeding strategies are needed. Hybrid varieties have the potential to outperform populations and synthetic cultivars. Indeed, heterosis for traits related to biomass yield such as plant height and dry matter yield have been described for both, perennial ryegrass and switchgrass hybrids. For switchgrass, hybrids have been assessed for yield on single plant and sward level in field trials: in a spaced plant study, the hybrids of selected genotypes from the cultivars ‘Kanlow’ and ‘Summer’ yielded between 0.18 (15.4%) and 0.35 (30%) kg per plant more biomass than the average of both parents (Martinez-
Reyna and Vogel, 2008). When the corresponding cultivars were evaluated in swards, high parent heterosis ranged from 30 to 38% (Vogel and Mitchell, 2008). Similar studies to determine heterosis for dry matter yield (ADMY) in swards were performed for ryegrass (Boller B, 2003; Posselt, 2010a; Posselt, 2003). The genetically most distant crosses yielded 15.5 t/ha ADMY, which represented panmictic mid-parent heterosis of 13% (Posselt, 2010a). For spaced plants, heterosis was as high as 31% in population hybrids of perennial ryegrass (Esparza Martínez and Foster, 1998; Foster, 1971).

Self-incompatibility (SI) and male sterility (MS) are important biological mechanisms which can be exploited to control pollination in hybrid breeding schemes. Despite intense research efforts during the last decades, little is known about the genetic control of SI and MS in perennial grasses such as ryegrass and switchgrass. For the gametophytic SI system which is assumed to be conserved in grass species (Yang et al., 2008), molecular genetic characterization is most advanced in perennial ryegrass (Klaas et al., 2011a), enabling the opportunity to use ryegrass as a model species for the genetically more complex switchgrass. Similarly, cytoplasmic MS has been characterized and commercially used in rye (Geiger and Miedaner, 2009) and other annual grass species such as maize (Laughnan, 1983).

Conserved synteny between grass genomes provide the opportunity to transfer genetic knowledge and sequence resources between grass species (Bennetzen and Freeling, 1993). Approximately 30 large chromosomal blocks are present in most grass genomes with different rearrangements (Devos, 1997; Devos, 2000; Devos, 2005; Feuillet and Keller, 2002; Gale, 1998). The level of synteny among grass genomes of species such as rice, barley, maize, sorghum, Brachypodium and others is between 40 and 73% (Gaut, 2002). More recent studies proposed that grass genomes have evolved from a common ancestor genome with a minimal size of 33.6 Mb structured in five
proto-chromosomes (Salse, 2008) from which grass species evolved by whole genome or segmental duplications, diploidization, small-scale rearrangements (translocations, gene conversions), and gene copying events (Bolot, 2009). Microsynteny can be used for identification of orthologous genes or genome regions in related species. Information available on complete and ongoing grass genome sequencing projects such as for rice, maize, sorghum, Brachypodium and possibly soon for bioenergy grass species such as ryegrass, switchgrass and *Miscanthus sinensis* will benefit marker development and gene identification in non-model bioenergy grass species (Devos, 1997; Donnison, 2009).

This review will focus on two perennial grass species for bioenergy production, namely switchgrass (*Panicum virgatum*) and perennial ryegrass (*Lolium perenne* L.). As perennial ryegrass is one of the few diploid perennial grass species for which substantial genomic tools are available (Farrar et al., 2007; Studer B et al., 2010), the potential to use synteny for the transfer of genetic knowledge and sequence resources between grass species is elucidated. In particular, this review aims to 1) evaluate current population and synthetic breeding schemes, 2) to review heterotic patterns that suggest possibilities to exploit heterosis, and 3) to address limitations and opportunities provided by the reproductive biology of these grasses.

**Current population and synthetic breeding schemes in switchgrass and perennial ryegrass**

Ryegrass and switchgrass are mainly developed as population or synthetic cultivars due to SI systems, which promote outcrossing in both species (Cornish et al., 1979; Fearon et al., 1984a; Martinez-Reyna and Vogel, 2002). Whereas population breeding is characterized by a continuous improvement through recurrent selection, synthetic breeding refers to crosses among a limited number of selected parents followed by multiplication through repeated open pollinations. Both,
population and synthetic cultivars, represent panmictic populations since they are based on open pollination ensuring random mating (Posselt, 2010). These population-based breeding schemes have been continuously improved. For example, Casler and Brummer (2008) determined by simulation studies that among and within family selection yielded higher genetic gains compared to recurrent selection for half sib and full sib families in most cases. For yield related traits, population improvement based on phenotypic or genotypic selection has been applied for several years resulting in low genetic gains. Humphreys (1997) reported an average yield gain of 4% per decade for forage yield which is approximately four times lower than the gains achieved for grain crops. This limited progress is likely due to negative correlations between the different target traits forage yield, seed yield, and forage quality (Casler, 2001; Casler and Brummer, 2008) and long breeding cycles in allogamous perennial grass species (Humphreys, 1997). Another reason might be incomplete exploitation of heterosis in population and synthetic breeding schemes (Casler and Brummer, 2008).

On the positive side, open pollinated cultivars are heterozygous and heterogeneous. Heterozygosity can result in stable yields as reported for sorghum (Haussmann, 2000; Reich and Atkins, 1970). Other studies in faba bean (Vicia faba) (Stelling, 1994) and barley (Einfeldt, 1999) have demonstrated that genetic diversity within a population can provide a buffer in response to biotic or abiotic stress. For yield improvement, narrow based synthetics might be advantageous due to higher selection intensity compared to broad based synthetics or traditional population varieties obtained by recurrent selection schemes. However, narrow based synthetics are susceptible to inbreeding depression (Posselt, 2010b). Therefore, the optimal number of founders in a synthetic cultivar might be a trade-off between the genetic diversity available and the selection intensity required for efficient polycross breeding. Choosing a diverse set of parents aided by molecular markers is a possible strategy to reduce the impact of inbreeding depression,
in particular with low-cost markers becoming increasingly available (Posselt, 2010b). In perennial ryegrass, progeny from wide crosses are consistently higher yielding. Molecular markers can be efficiently used to assess diversity among parental genotypes to determine the optimum number of genotypes of a polycross (Kolliker et al., 2005). A major advantage of population or synthetic compared to hybrid varieties is, that seed production is possible at substantially lower costs. Thus, from a farmer’s perspective, lower yields might be compensated for by lower seed costs.

**Prospects for hybrid breeding in perennial grasses**

Hybrid performance is mainly determined by the degree of heterosis, which is defined as superiority of the heterozygous hybrid over its homozygous parents (Lamkey and Edwards, 1999). Heterosis displayed by F_1 progeny of two random mating populations is known as panmictic-midparent heterosis (Posselt, 2010b). In allogamous grass species, the contribution of heterosis to hybrid performance is difficult to estimate due to heterozygosity of parents and confounding polyploidy effects (Breese, 1981; Dewey, 1980; Zeven, 1980). Biologically, there is a potential to more efficiently exploit heterosis in hybrid breeding schemes in perennial bioenergy grasses, because these cross-pollinated species are (i) wind pollinated, (ii) produce a large amount of seed per plant, and (iii) exhibit a strong self-incompatibility mechanism which can be used for cost-effective hybrid seed production.

**Identification of heterotic patterns and development of heterotic groups**

An important requirement to efficiently exploit heterosis in plant breeding is the identification or development of heterotic groups as available in maize (Lamkey and Edwards, 1999). Heterotic
groups are complementary gene pools for creation of high-performing hybrids. Most likely, breeding efforts based on recurrent reciprocal selection (RRS) will be required to develop heterotic groups, as it has been the case for maize (Hallauer et al., 2010). For identification and further improvement of potential heterotic groups, molecular markers applied to characterize genetic diversity within elite populations will play a crucial role. Different marker technologies were consistently able to discriminate heterotic groups in maize (Lubberstedt et al., 2000). It is in addition possible to use markers to identify genetically divergent populations. In ryegrass, various marker technologies including diversity arrays technology (DArT), simple sequence repeats (SSR), and single nucleotide polymorphisms (SNP) have been used to study genetic diversity (Kopecky et al., 2011; Kopecký, 2009; Kubik et al., 2001). When studying genetic variation within and among various ryegrass varieties and accessions, most of the genetic variation was found within rather than between ryegrass varieties or accessions (Kölliker, 1999) and among population variation only explained around 29% of the genetic variation (Bolaric et al., 2005). Analyses of molecular variance (AMOVA) showed that Northern and Southern German ecotypes of ryegrass can be classified as distinct germplasm pools (Bolaric et al., 2005). A more recent comprehensive analysis of population structure in European ryegrass varieties identified two groups representing maritime and continental varieties, respectively (Brazauskas, 2011). Both studies show that geographically distinct groups of accessions might form different gene pools and, therefore, represent an excellent starting point for RRS programs to establish continuously improved heterotic pools.

Martinez-Reyna and Vogel (2008) assessed different accessions of switchgrass for potential heterotic patterns based on geographic separation. Tetraploid upland and lowland cultivars as well as western and eastern populations and combinations of separate subspecies of switchgrass were evaluated for combining ability. In a spaced plant study, the tetraploid lowland cultivar
“Kanlow” and upland cultivar “Summer” represented two distinct heterotic pools (Martinez-Reyna and Vogel, 2008). Hultquist, S et al. (1996) used restriction fragment length polymorphisms (RFLP) to discriminate upland and lowland ecotypes, which have been demonstrated to be potential heterotic groups. Thus, geographically separated populations are a starting point to assess potential heterotic groups and this can be further aided by the use of new marker technologies.

**Biological mechanisms for effective hybrid seed production**

The production of hybrid seed requires a directional cross between a pollen donor and a pollen recipient. In current production schemes, self-pollination of the recipient line or genotype has to be prevented. This can be achieved either by manual emasculation or genetic mechanisms, among which SI and cytoplasmic MS are the most important non-transgenic systems. In perennial grasses like ryegrass and switchgrass with small hermaphrodite flowers, manual emasculation to control pollination is laborious and time consuming in terms of commercial seed production. Therefore, use of SI or MS is crucial for effective hybrid breeding.

SI is a genetic mechanism that allows flowering plants to prevent self-fertilization, thereby maintaining a high degree of heterozygosity (de Nettancourt 1977). The gametophytic SI mechanism in grasses is controlled by two loci, S and Z (Lundqvist 1962). Involvement of additional loci independent of S and Z has been revealed by studies on the breakdown of SI in perennial ryegrass (Thorogood et al., 2005). Recently, genes differentially expressed during a SI response have been identified using cDNA-AFLP (Van Daele, 2008a) and suppression subtractive hybridization (Yang et al., 2009). Still, none of the SI determinants in grasses has been cloned to date. Approaches to find SI loci in ryegrass included genetic mapping and exploitation of synteny between rye and rice (Hackauf and Wehling, 2005b; Yang et al., 2009). Subsequent fine-scale comparative genetic and physical mapping found a high degree of micro-
synteny for the Z locus and suggested that combined synteny and map-based cloning strategies are promising for isolating genes involved in SI response in grasses (Shinozuka et al., 2009). As the map-based isolation of SI loci is most advanced in perennial ryegrass, information on genetic mechanisms and markers for genes controlling SI as well as SI-based hybrid breeding strategies might soon become available and transferrable between perennial grass species (Klaas et al., 2011a). For several reasons, SI is a promising tool for hybrid seed production in perennial grasses. SI can be used for combining any pair of genotypes without the necessity to develop male sterile and maintainer genotypes as in the case of cytoplasmic male sterility (CMS). Moreover, it benefits from the ability to vegetatively propagate perennial grasses to multiply parental components for hybrid seed production. More importantly, SI can either be overcome by self-fertility genes as described for perennial ryegrass (Thorogood et al., 2005), or by environmental conditions such as high temperature, modifying the expression of incompatibility genes and finally enabling self-fertility by pseudo-compatibility (Wilkins and Thorogood, 1992).

For switchgrass, a two-locus SI system has been described (Martinez-Reyna and Vogel, 2002) as in ryegrass. Switchgrass is a species with reported ploidy series ranging from $2n = 2x = 18$ to $2n = 12x = 108$ (Nielsen, 1944). During meiosis, polyploids may show preferential pairing of duplicated genomes. Occasionally, chromosomes of unlike genomes pair to form quadrivalents or pairs of bivalents. Initial studies in tetraploid genotypes suggest preferential pairing of chromosomes (Missaoui, 2005). Furthermore, a recent genetic mapping experiment confirmed a great amount of preferential pairing for almost all chromosomes (Okada, 2010). High ploidy levels make it difficult to determine the genetics of SI, among others due to the expected large numbers of alleles per locus. Therefore, it is not clear yet, if there are dominance interactions between alleles at the loci involved in SI, which are absent in diploid gametophytic SI (Yang et al., 2008). Even though the SI process is not clearly elucidated in switchgrass, progress in fine
mapping of S and Z in ryegrass provides an opportunity to identify SI determinants in various perennial grasses with conserved SI system based on S and Z (Klaas et al., 2011b).

MS has been described for ryegrass (Wit, 1974) but not for switchgrass so far (Vogel and Lamb, 2007). Male sterility can be structural or functional (Fehr, 1993). Structural MS can cause abnormal dehiscence of anthers, preventing pollination, whereas functional MS is caused by interruption in either micro- or macrogametogenesis driven by cytoplasmic or nuclear genetic factors (Palmer et al., 1992). Male sterility can either be genic or cytoplasmic, which affects the appropriate breeding strategy (Fehr, 1993). In perennial ryegrass, there are two reported sources of MS, one was induced by intergeneric crosses to Festuca pratensis and the other by interspecific crosses between L. perenne and L. multiflorum (Wit, 1974), the latter causing CMS. In addition, nuclear restorers were identified for L. perenne (Kiang et al., 1993; Kiang and Kavanagh, 1996). Recently, molecular studies have shown that the presence of an element named LpCMSi in the mitochondrial genome of ryegrass affects CMS (McDermott, 2008). This information can be used to obtain male sterile lines (commonly called A lines, genotypes or populations) by crossing elite genotypes lacking restorer genes onto genotypes carrying CMS or MS genes. CMS systems are generally species-specific due to a high diversity of mitochondrial genes (Moore, 1997; Ruge et al., 2002). In addition, male sterility and restoration take place in different stages of pollen development for each plant species (Chase, 2007; Choi, 2007). Restorer genes from Oryza sativa, Petunia hybrid, Raphanus sativus, and Sorghum bicolor are members of the pentatricopeptide-repeat (PPR) protein family and are all involved in decreasing the accumulation of CMS locus transcripts (Chase, 2007; Choi, 2007). This initially suggested homology between restorer genes, but the protein structure and mechanisms for fertility restoration differ between and within species. In maize, different restorer genes have been cloned and characterized such as Rf1, Rf2, Rf3, and Rf9 which have a high allelic diversity (Gabay-
Laughnan et al., 2009). In contrast to other restorers described previously, Rf2 in maize is an aldehyde dehydrogenase (Cui, 1996). This indicates fertility restoration is not conserved within or across species. Therefore, it is not likely that genetic information for genes involved in ryegrass (C)MS can be transferred across a broader range of grass species based on synteny.

Hybrid breeding strategies for perennial grasses

The different hybrid breeding strategies applicable in perennial grasses are varying in the levels of inbreeding and genetic heterogeneity of hybrid seed, as well as the biological mechanism to be used for hybrid seed production.

Population hybrids

Crosses between pairs of populations which are normally a set of subpopulations or cultivars have been used for the development of population hybrids in switchgrass (Martinez-Reyna and Vogel, 2008). As a result, progeny will be a mixture of inter- and intrapopulation crosses. With the goal to maximize the amount of hybrid seed (derived from interpopulation crosses), different strategies can be applied: populations can be grown side by side but at different ratios. As a consequence, one population will contribute more pollen to the pollen cloud compared to the other, which can be used as pollen recipient population for seed production. For example, two populations can be grown in a 3:1 ratio of pollen donor to pollen recipient (Figure 1.1a). Therefore, the seed harvested from the pollen recipient would consist of at least 75% hybrid seed. Moreover, populations used as pollen donors could be improved by conventional breeding schemes for high quantities of pollen production and grown side by side, with the difference that one population will contribute for example three times more pollen to the pollen cloud compared to the recipient population (Figure 1.1b).
Figure 2.1 Breeding schemes using self-incompatibility to produce population hybrids

a) Pollen donor (A) and pollen receptor (B) population are grown in a 3:1 ratio
b) Pollen donor (A) is a genotype or a set of genotypes that shed more pollen than B genotypes. Therefore A contributes, e.g., three times more pollen to the pollen cloud than B genotypes

Chance hybrids

The percentage of hybrid seed can be increased if three or more populations are used for intercrossing, as suggested by Brummer (1999). If more populations are added, the amount of non-hybrid seed will decrease even more. The concept of producing a mixture of inter- and intrapopulation hybrids resulting from a cross between two populations has been described as chance hybrids (Burton, 1948), also known as semihybrids (Brummer, 1999). In contrast to population hybrids, a basic semihybrid design as the one proposed by Brummer et al. (1999), requires the identification of at least two populations with good combining ability. In addition, use of at least Syn2 seed within each population would increase the frequency of desired alleles
and the inbreeding coefficient within each population which will increase the chance of obtaining hybrid seed. However, as the number of heterotic combinations between groups is limited, the use of more than a pair of populations will result in little or no genetic gain.

The chances of capturing heterosis in a semihybrid design are greater to what is currently obtained with synthetic breeding for different reasons. For instance, in broad-based synthetics, selection causes increase of frequencies desired alleles from Syn1 to Syn3 (Brummer, 1999), resulting in an increased inbreeding coefficient. In contrast, in a semihybrid breeding scheme, different alleles will be enriched in the two parental populations prior to hybrid seed production, thereby maximizing the chances to capture heterosis. Alternatively, F1 hybrids can be advanced for hybrid seed production resulting in a similar or higher heterozygosity level as the one described above. (Charles Brummer Personal Communication). In this case SI can be used to avoid self-pollination of plants in hybrid seed production fields.

In polyploid switchgrass, preferential pairing reduces the number of possible gametes. This increases the chance to capture heterosis, if genotypes with different allele frequencies are crossed.

The challenge is to select true hybrid progeny in a semihybrid design. Herbicide resistance has been suggested as a possibility to select hybrid progeny in a semihybrid (Brummer, 1999). To have an efficient selection method, the two populations will have to differ in their ability to resist to two different herbicides. The source of resistance can occur naturally in a population or can be introduced via genetic transformation. The progeny of the cross between these two populations need to be sprayed with both herbicides. As a result, most of the surviving progeny should be hybrids between two populations (Brummer, 1999). However, it may be difficult to keep herbicide tolerances restricted to particular parental populations. Moreover, fixation of respective
herbicide tolerance genes might be challenging in polyploid species, which leads to a substantial fraction of susceptible hybrids in hybrid seed production.

An alternative strategy is to limit the number of S-Z haplotypes within each of the parent populations (Syn2 population or population hybrids). Thereby, the amount of intrapopulation offspring can be significantly decreased (Figure 1.2) and the amount of hybrid seed increased. This strategy will require molecular markers for S and Z. Linked markers are already available for S and Z in ryegrass (Shinozuka et al., 2009; Van Daele, 2008a; Van Daele, 2008b; Zeven, 1980), which can be used to predict S and Z haplotypes. In the near future, functional markers for S and Z may become available, which would increase accuracy of SI haplotype prediction, in particular because. Since S and Z are not linked loci, the possible number of haplotype combinations is large. Selection of a low number of S and Z haplotypes would decrease the chance of intrapopulation crosses. SI underlies frequency dependent selection (FDS), i.e., rare alleles are at a selective advantage. Thus, the number of initial SI haplotypes in a semihybrid design could be rather low and FDS would still ensure sufficient inter-population compatibility and fertility.

S and Z markers could be used in breeding programs to select for unique S-Z haplotypes within a population and different between populations. In a practical example, a single pair cross between two diploid genotypes completely heterozygous for S and Z would result in progeny with sixteen possible genotypes. Around 1092 individuals need to be screen to obtain at least 50 individuals with a unique S-Z genotype. The same has to be done in the other heterotic group, thus approximately 4400 data points are needed to identify parents for each population that accomplish two basic requirements: have genotypes with identical S-Z haplotypes within populations but different S-Z haplotypes among populations. It is noteworthy that the effectiveness of this design relies on the possibility of keeping populations with good combining
ability apart and to increase the number of favorable alleles only within groups. With such a scheme, testing thousands of plants for S and Z markers would be a time efficient and non-expensive procedure.

Figure 2.2 A breeding scheme using self-incompatibility to produce population hybrid using two populations grown in a 1:1 ratio
Dashed arrows represent limited compatibility among individuals within the same population and full arrow represents complete pollen-compatibility. $S_i Z_j$ represents a limited group of haplotypes within the population that differ from the $S_k Z_l$ set of haplotypes from the other population.

**Single cross hybrids**

In a strict sense, the parents of a single cross hybrid are two inbred lines. However, inbred line development in perennial grasses is impaired due to SI and inbreeding depression. Single cross hybrids from heterozygous parent genotypes are a more likely scenario for perennial grass species, and result in segregating F1 populations, comparable to double cross hybrids in maize.

The use of inbred lines will require initial inbreeding and an additional step to restore self-incompatibility. Moreover, inbred line based hybrids are genetically homogenous. This is in contrast to the situation where heterozygous parents are used and the resulting hybrid seed is
heterogeneous (Figure 1.3). If S-Z haplotypes are different between hybrid parents, exclusively hybrid seed will be produced. In the short term, obtaining inbred lines from switchgrass may not be feasible due to SI and the fact that SF genes have not been identified and described yet. The fact that heterotic patterns were reported for biomass yield in spaced plant evaluations in switchgrass (Martinez-Reyna and Vogel, 2008) and ryegrass (Bolaric et al., 2005; Posselt, 2010a) suggest, that it may be possible to identify heterozygous genotypes with superior specific combining ability.

Repeated testing and maintenance of genotypes that display good combining ability is essential for hybrid seed production based on heterozygous parents. Perennial grasses are easy to propagate vegetatively. As a result, selected ryegrass and switchgrass clones can be grown side by side in production fields and a high degree of hybridization will be ensured by SI. Preliminary data show that SI in switchgrass is not fully effective (Todd et al., 2011) which may require an initial screening of selected genotypes to avoid self-fertilization. Hybrid seed can be harvested from both parents, maximizing seed yield. The main limitation of this approach is the time required to obtain a sufficient number of clones and the time that it will take to establish these clones in production fields. Under greenhouse conditions at Iowa State University, a well-established switchgrass or ryegrass plant can produce up to twelve clones every 6 weeks and it can take up...
to 5 months before resulting plants can be transplanted to the field. However, in vitro culture techniques can aid to produce large numbers of plantlets in a short period of time. For example, nodal culture of a single switchgrass plant can produce 500 plants ready for planting within 12 weeks (McLaughlin and Kszos, 2005).

In the longer run, it is preferable to develop inbred lines mainly because seed in large amounts is more desirable and cost effective for establishing hybrid seed production fields compared to generating plantlets by vegetative propagation. This is doable in ryegrass, where self-fertility loci have been reported (Thorogood and Hayward, 1991). However, there are reports of strong inbreeding depression (Thorogood and Hayward, 1992; Thorogood et al., 2005). This poses a challenge for line development in terms of fitness and seed production. Reduced fitness of inbred lines resembles the initial stages of maize hybrid breeding. However, this bottleneck has been overcome by using RRS in maize (Hallauer et al., 2010).

Another challenge is production of inbred lines using self-fertility mechanisms, while maintaining the possibility to use SI for controlled crosses in hybrid seed production fields. Self-fertility in perennial ryegrass is monogenically inherited and dominant. Two independent loci affecting SF have been described (Thorogood and Hayward, 1991). Self-fertility loci would be fixed during inbred line development. Therefore, a mechanism to restore SI is required, before controlled crosses can be performed effectively. Up to date, there are no reports of mechanisms to interrupt or break down SF, once SF genes are introduced to a genotype which makes this approach not practical for hybrid seed production. Alternatively, pseudo-compatibility induced by high temperatures can be used to overcome SI temporarily (Wilkins and Thorogood, 1992; Wricke, 1978). In a study on pseudo-compatibility, seed set increased from an average of 2.3%, obtained under natural conditions, to an average of 31% after temperature treatment (Wilkins and Thorogood, 1992). However, control of temperature at a large scale might not be achievable in terms of costs and logistics. Instead, it is possible to use environments with optimal temperatures
to achieve pseudo-compatibility for inbred seed production. Still, the risk of using temperature-based mechanisms is, that selfing can be triggered by high temperatures in hybrid seed production fields. Therefore, a mechanism for pseudo-compatibility that can be better manipulated (such as application of a sprayable compound) would be preferable.

In ryegrass, the main limitation is the low amount of seed produced by self-pollination due to a limited number of reproductive tillers. This limitation affects seed yield and also the vegetative propagation coefficient discussed above. Heritability for tiller number is rather low (Boelt and Studer, 2009). Therefore, several cycles of selection might be required to obtain inbred lines with sufficient seed yield. In contrast, switchgrass may require a large-scale phenotypic screening of various genotypes or a TILLING population, as a first step towards identification of SF mechanisms as observed in ryegrass.

CMS is an alternative to produce single cross hybrids. The production scheme requires a male sterile A line, and a B line that can be used to maintain male sterility. When using CMS to produce F1 hybrids based on inbred parents, it will be required to introgress self-fertility into elite germplasm or to induce pseudo-compatibility for inbred line production. Male sterility would be introduced to inbred lines within one heterotic group to be pollinated by lines from the other heterotic group. Therefore, this approach is limited to the use of inbred parents because A and B lines have to be isogenic except for their CMS element. In contrast to, e.g., maize CMS breeding schemes (Duvick, 1959), restorers are not required since the target trait is biomass yield and not seed production. However, additional efforts are required to maintain nurseries for A and B lines. Another limitation of using CMS in ryegrass is, that it might be leaky under certain environmental conditions, so that viable pollen is produced by “male sterile” genotypes (Connolly and Wright-Turner, 1984), resulting in a certain fraction of self-fertilization (Boller B, 2003; Posselt, 2003). Currently, there are no cultivars of ryegrass on the market, developed by
using CMS. This is not due to the leaky systems but because the variety candidates developed did not meet the distinctiveness requirements for variety registration (Posselt, personal communication). New sources of CMS have been reported and seem to be more promising for hybrid seed production due to reduced susceptibility to temperatures (Ruge et al., 2002). In summary, SI or CMS are promising biological mechanisms to enable hybrid seed production in perennial grasses. However, CMS based system can be harvested only in females in contrast to a SI-based system in which hybrid seed can be harvested on both. Anyway, a major effort to move towards hybrid breeding is to develop heterotic groups to ensure sufficient levels of heterosis and hybrid performance for economic seed production. Finally it is noteworthy that there are no reports of male sterility in switchgrass. Therefore, using self-incompatibility for hybrid breeding in switchgrass may be the method of choice, as proposed by Martinez-Reyna and Vogel (2008). This scheme might be further improved by using the information available from ryegrass, e.g., sequences for S-Z determinants. This sequence information can be used to identify candidate genes in switchgrass that can be used as markers for parent selection in breeding programs.

**CONCLUSIONS**

Reportedly up to 30% of increase yield performance has been achieved by traditional breeding (McLaughlin, 2002). Hybrid breeding may represent an increase in seed production cost, therefore the yield gains obtained by hybrid breeding should be significantly higher than 30%. For instance, Perrin, et al (2008) report a $16.39 ha\(^{-1}\) seed cost for an annualized yield of 5.0 Mg ha\(^{-1}\) (Perrin, 2008), if seed production increases seed cost to $24.59 ha\(^{-1}\) (50% increase), hybrid cultivars obtained from any breeding strategy should yield at least 7.5 Mg ha\(^{-1}\) for farmers to
break even. Significant heterosis effects for biomass yield found in initial studies for switchgrass and ryegrass suggest a potential to exploit heterosis to obtain high yielding hybrids. To efficiently capture heterosis, efforts to identify and develop heterotic groups exhibiting good combining abilities are required. Molecular markers, and novel genotyping approaches using second generation sequencing will constitute an efficient tool to achieve these goals. Moreover, collaborative efforts of public breeders to identify these genetic patterns will aid to avoid mixing different gene pools. For our model species, both SI and MS are prospective biological mechanisms to control pollination for efficient hybrid seed production. SI may be the method of choice for population hybrids, if molecular markers allow haplotype identification at S and Z. Allelic markers for SI loci will be very useful for parent selection in population hybrids designs and help establishing a very resource efficient breeding system. CMS is also promising, particularly because no restoration of fertility is required. This is especially important, if inbred lines are used for hybrid seed production. However, it still requires the development of isogenic or nearly isogenic male sterile and maintainer lines, which may increase the time and effort needed to complete a breeding cycle. For species or genotypes, in which SF genes or CMS systems are not available, large-scale phenotypic of various genotypes constitute a first step towards the identification of similar mechanisms in the species of interest. Synteny between grass genomes will allow to use perennial ryegrass as a model for SI and SF, and enable a candidate gene approach to identify orthologues in switchgrass and other bioenergy grasses.

Acknowledgements

We are grateful to Dr. U. K. Posselt (Universität Hohenheim) for his insight regarding some points in this review. We also want to thank the R.F. Baker Center for Plant Breeding (Iowa State University) that supports Andrea Arias Aguirre.
REFERENCES

CHAPTER 3: POLLENCALC: SOFTWARE FOR ESTIMATIONS OF POLLEN COMPATIBILITY OF SELF-INCOMPATIBLE ALLO- AND AUTOTETRAPLOID SPECIES

Manuscript published in BMC Bioinformatics

Andrea Arias Aguirre¹*
* Corresponding author
Email: andreasa@iastate.edu

Bernd Wollenweber²*
* Corresponding author
Email: Bernd.Wollenweber@agrsci.dk

Ursula K Frei¹
Email: ufrei@iastate.edu

Thomas Lübberstedt¹
Email: thomasl@iastate.edu

¹ Department of Agronomy, Iowa State University, Agronomy Hall, Ames, IA, USA

² Department of Agroecology, Faculty of Science and Technology, Aarhus University, Aarhus, Denmark

ABSTRACT

Background

Self-incompatibility (SI) is a biological mechanism to avoid inbreeding in allogamous plants. In grasses, this mechanism is controlled by a two-locus system (S-Z). Calculation of male and female gamete frequencies is complex for tetraploid species. We are not aware of any software available for predicting pollen haplotype frequencies and pollen compatibility in tetraploid species.
Results

PollenCALC is a software tool written in C++ programming language that can predict pollen compatibility percentages for polyploid species with a two-locus (S, Z) self-incompatibility system. The program predicts pollen genotypes and frequencies based on defined meiotic parameters for allo- or autotetraploid species with a gametophytic S-Z SI system. These predictions can be used to obtain expected values for diploid and for (allo- or autotetraploid SI grasses.

Conclusion

The information provided by this calculator can be used to predict compatibility of pair-crosses in plant breeding applications, to analyze segregation distortion for S and Z genes, as well as linked markers in mapping populations, hypothesis testing of the number of S and Z alleles in a pair cross, and the underlying genetic model.

Keywords

Disomic inheritance, Polyploidy, Self-incompatibility, Software, Tetrasomic inheritance

BACKGROUND

Self-incompatibility (SI) is a biological mechanism to avoid inbreeding in allogamous plants that can be triggered before, during, or after pollination [1]. There are two SI systems differing mainly in their genetic control. Pollen compatibility determined by the diploid genotype of the pollen parent is referred to as sporophytic SI (SSI) [2]. In contrast, SI determined by the haploid pollen genotype is referred to as gametophytic SI (GSI) [3]. In the latter system, the male gamete does not display any dominance in diploid species because it is haploid. Thus, every allele is
functional. SI can be controlled by either a single locus or few loci depending on the species. For example, in *Brassicaceae* and *Solanaceae*, SI is controlled by a single S gene, whereas for self-incompatible grasses, SI is controlled by a two-locus (S-Z) system. These two independent loci have among others been described for rye [4], perennial ryegrass [5], and Italian ryegrass [6]. The S-Z SI system results in low selfing rates and is functional even in polyploid grasses such as switchgrass, where tetraploid plants showed 0.35% self-compatibility, whereas octoploids reached 1.39% [7].

The proposed mechanism of SI in diploid grasses suggests that, if at least one allele at S or Z in the pollen differs from the recipient S-Z alleles, the pollen will be compatible with the stigma [2]. Fifty percent of the pollen produced by a genotype is compatible to pollen receptor genotype. If the pollen donor has one different allele at each S and Z, then the 75% of the pollen is compatible with the pollen receptor genotype [2]. If the pollen donor has one different allele at each S and Z, then the 75% of the pollen is compatible with the pollen receptor genotype. Finally, if the pollen donor has two different alleles at one of the two loci, then pollen is 100% compatible. The GSI system in diploid species with an S-Z system might lead to differences in reciprocal crosses [2].

In polyploid species the interaction between pollen and stigma alleles is more complicated due to the presence of more than two alleles at a locus. Throughout the manuscript, letters are used to identify alleles at S locus and numbers are used to identify alleles at Z. Here we propose two possible modes of interaction, named Model I and Model II. If any allele at either S or Z in the pollen differs from the alleles for S or Z on the pistil, pollen is compatible (Model I) [8]. If at least one allele at S and one allele at Z locus present in the pollen grain matches the pistil, then pollen is incompatible (Model II) [9,10]. For example pollen grain with genotype AB15 will be compatible under Model I because it has one different allele compared to the ABCD1234 female.
In contrast, if Model II is true, pollen AB15 is incompatible with the female ABCD1234 because it shares one allele at S (A or B) and another at Z (1) (Figure 2.1). In addition to pollen-stigma interactions, the estimation of pollen compatibility percentages can be even more complex in polyploids due to inheritance patterns that can result in a high number of gametic haplotypes. Polyploidization is a common mechanism that occurred during evolution and it is the result of the fusion between unreduced gametes of different species (allopolyplloyd) or of the same species (autotetraploidy). Genome analysis demonstrated that many plant genomes such as maize and soybean are ancient polyploids that went through diploidization events [11]. The nature of gamete fusion determines the behavior of chromosomes during meiosis. Allopolyploids show mostly preferential pairing of chromosomes in structures called bivalents (disomic), whereas autotetraploids pair with their respective homologous or quadrivalents (tetrasomic). In disomic inheritance there is no pairing between homeologues and, therefore, homologous chromosomes do not end up together in a gamete. Thus, for a given locus with genotype ABCD and preferential pairing between AB/CD, only four gametes (AC, BD, BC, and AD) are possible. In contrast, tetrasomic inheritance results in all six possible combinations for pairs of alleles (AB, AC, AD, BC, BD, and CD). Normal chromosome segregation implies that the two sister chromatids are distributed to different gametes. However, during quadrivalent formation in autotetraploids, crossing over between non-sister chromatids can occur. The probability that two sister chromatids end up in the same gamete is known as $\alpha$. This gamete carries the alleles which are identical by descent [12] due to chromatid segregation and double reduction. For example, a plant with genotype ABCD for the S locus can produce AB, AC, AD, BC, BD, CD gametes under tetrasomic inheritance and AA, BB, CC and DD gametes if chromatid segregation and double reduction occur.
**Model I: Allelic combination of S-Z**

Model I: any different allele at either S or Z in the pollen different from S or Z on the pistil makes pollen compatible.

**Model II. One different allele system**

Model II: any S-Z allele combination present in the pollen grain that matches the pistil makes pollen incompatible.

Consequently, it is possible in a tetraploid species to find gametes with alleles that are identical by descent (IBD), which results in a total of 10 possible different allele combinations in a gamete. When considering both SI loci, S and Z, respectively, 100 different gametes are possible. Thirty six gametes have the genotypes expected under meiosis with no crossing over, 48 are the result of
crossing over between two non-sister chromatids and double reduction at either locus, and 16 gametes result from a double reduction at both loci. Chromatid segregation and double reduction increase the number of possible pollen genotypes and, therefore, increase the probability of cross-pollination between individuals with similar S and Z genotypes.

A software for numerical estimation and prediction of allelic and genotypic frequencies calculation for sporophytic SI systems has been published [13]. However, this model does not apply to gametophytic SI systems that do not display dominance. Segregation models for disomic and tetrasomic inheritance [14] as well as models for estimation of allele frequencies in polyploids [15] also have been published. However, these models lack computerization of calculation. In addition, these models are generic for any locus, but do not deal with self-incompatibility or calculate pollen compatibility.

The program that we present here has the objective to support hypothesis testing under the different scenarios described above, especially for uncharacterized species. It can serve four different purposes: i) determine the number of progeny needed to differentiate between allo- and autotetraploidy; ii) obtain expected values of pollen compatibility for a pair-cross to maximize seed yield; iii) determine the SI mechanism (differentiate between Model I and II) in tetraploids; and iv) display information regarding S and Z distortion that can provide useful knowledge about segregation distortion of markers at these loci or linked markers for linkage mapping and population genetic studies.
Implementation

The program was written in the C++ programming language using the freely available Qt compiler (Nokia, Germany) and compiled for the Windows™ operating system. A version for Mac OSX and Linux is under development.

Calculator parameters for meiosis

The calculator is based on estimable meiotic parameters that influence chromosome segregation during meiosis of autotetraploids. For instance, parameter $\beta$ determines, whether or not there is preferential pairing (disomic inheritance) between chromosomes. $\beta$ is restricted to have values of 0 or 1. In case of tetrasomic inheritance or no preferential pairing $\beta = 1$. With four alleles per locus (i.e. ABCD) six gamete combinations are possible (AB, AC, AD, BC, BD, CD) under random pairing. In case of preferential pairing of homologues, $\beta = 0$. Furthermore, preferential pairing between genomes results in only four gamete combinations. To illustrate this further we named each genome $jklm$ and alleles at a locus correspond to each genome in that order. For example, a genotype ABCD with preferential pairing between $jk/lm$ has gametes with haplotypes AC, AD, BC, and BD. In case of preferential pairing, only one out of the three possible pairing scenarios $jk/lm$, $jl/km$, or $jm/kl$ can be true. This has been implemented in our calculations by defining a $\Delta$ value to test all three possible scenarios sequentially. Technically, if the $\Delta$ value is equal to 1 for one scenario, it is 0 for the other. Usually, the values for $\beta$ and $\Delta$ do not differ for S and Z.

Calculator frequencies estimation for segregation of S and Z

S and Z are independent loci. Therefore, normal chromosome segregation of a genotype with four different alleles for S or Z and with no preferential pairing between chromosomes $jk$, $jl$, $jm$, $kl$,  

$km$, and $lm$ will produce six possible pollen haplotype combinations for each locus. Thus, the probability of a haplotype is $1/6$. It is possible that recombination during multivalent formation in meiosis followed by double reduction can cause sister chromatids to segregate together during meiosis. Therefore, there is a chance of two identical alleles to end up in the same gamete by double reduction [12]. If we consider a single chromatid, only 1 out of the seven remaining is its corresponding sister chromatid. Therefore, the probability ($\alpha$) of two sister chromatids to end up in the same gamete ranges from 0 to $1/7$. The value of $\alpha$ is species-specific, and depends on the distance between the locus and the centromere [12,16]. For the $jklm$ genotype, there are four possible gamete combinations that have two identical by descent alleles (IBD) i.e., $jj$, $kk$, $ll$, $mm$. Therefore, the frequency of each of these pollen genotypes is $\alpha/4$. The closer $S$ and $Z$ are linked with the centromere, the smaller is the $\alpha$ value [16].

The input form of the program has a cell to enter $\alpha$ values to calculate gamete frequencies that can be different for each locus. For example, *Phalaris coerulescens* is a diploid species for which $S$ and $Z$ have been mapped. The $S$ locus is located in a sub-centromeric region of chromosome 1, and the $Z$ locus is located on the long arm at the end of chromosome 2 [17]. In *P. coerulescens*, the $\alpha$ value for $S$ will, therefore, be close to 0, whereas the one for $Z$ should be larger. If the $\alpha$ value is unknown, a default value such as $1/14$ (average between the extremes of 0 and $1/7$) could be used.

**Calculator input and settings**

The input window (Figure 3.2) has boxes for meiotic parameters for both loci ($S$ and $Z$) as described above and for a given pair of genotypes. The default setting is disomic inheritance and preferential pairing between $jl/km$ for both loci, which can be used as a null hypothesis. The parameters settings have some restrictions. For example, if $\beta = 1$, all $\Delta$ values are set to 0. If an $\alpha$
value is given, then $\beta = 1$. It is possible to enter the genotype for a pair of individuals. To our knowledge there is no information about allelic diversity for S and Z loci since the determinants have not been identified yet. However, we developed the software assuming that the maximum difference between two genotypes is eight different alleles at each locus. Therefore, the program accepts values for S defined by letters from A-H and numbers from 1-8 for Z. Settings will be summarized in words once input parameters are set. All the input settings will be summarized in words once input parameters are set.
Figure 3.2 Input box for meiotic parameters for S-Z and genotype of two individuals

**Calculator output**

The initial output (Figure 3.3) obtained is a summary table of the input values. After the analysis is performed, the pollen haplotypes for each genotype and each locus are displayed. Pollen haplotypes are calculated based on the formulas in Table 3.1. Stift (2008) published formulas to calculate pollen haplotypes, which differ from ours in nomenclature. The second output contains tables for pollen genotypes and frequencies as well as pollen compatibility in reciprocal crosses between genotypes 1 and 2 (Figure 3.4a). For example for a genotype ABCD1234, the matrix for pollen compatibility has the 10X10 possible pollen haplotypes. Pollen haplotypes are arranged in column 1. In the column header of columns 2-5 and 6-9, the four female alleles for both, S and Z are displayed. Under each female allele, the numbers of identical alleles in the pollen are counted. Since pollen genotypes have only two alleles for each locus, the number of alleles it can possibly share with a female genotype at the S or Z locus, respectively, has to be 0, 1, or 2 alleles respectively. \( \alpha \) is the probability of double reduction and ranges from 0-0.14. \( \beta \) is a parameter for chromosome pairing during meiosis and can have values of 0 (disomic) or 1 (tetrasomic). \( \Delta \) is a parameter that defines which chromosomes show preferential pairing.
Figure 3.3 First output with input parameter summaries

Table 3.1 Formulas for pollen haplotype calculations

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>p(jj)</td>
<td>1/4α</td>
</tr>
<tr>
<td>p(kk)</td>
<td>1/4α</td>
</tr>
<tr>
<td>p(ll)</td>
<td>1/4α</td>
</tr>
<tr>
<td>p(mm)</td>
<td>1/4α</td>
</tr>
<tr>
<td>p(jk)</td>
<td>1/6β-1/6α + (1-β)(1/4Δ2 + 1/4Δ3)</td>
</tr>
<tr>
<td>p(jl)</td>
<td>1/6β-1/6α + (1-β)(1/4Δ1 + 1/4Δ3)</td>
</tr>
<tr>
<td>p(jm)</td>
<td>1/6β-1/6α + (1-β)(1/4Δ1 + 1/4Δ2)</td>
</tr>
<tr>
<td>p(kl)</td>
<td>1/6β-1/6α + (1-β)(1/4Δ1 + 1/4Δ2)</td>
</tr>
<tr>
<td>p(km)</td>
<td>1/6β-1/6α + (1-β)(1/4Δ1 + 1/4Δ3)</td>
</tr>
<tr>
<td>p(lm)</td>
<td>1/6β-1/6α + (1-β)(1/4Δ2 + 1/4Δ3)</td>
</tr>
</tbody>
</table>

Genomes are named generically jklm. Therefore, alleles for S (i.e. ABCD) correspond to genomes jklm respectively; and alleles for Z (i.e. 1234) also correspond to genome jklm.
For each possible pollen-stigma combination the number of shared alleles at the S and Z locus are given and summarized independently for S and Z. ΣS and ΣZ reflect the sum of shared alleles at S and Z, respectively (columns 10 and 11). In autotetraploids, for a given allele, a locus can be monoallelic (AAAA), biallelic (AABB) or triallelic (AABC) [12]. Therefore, the program is designed to avoid double counting of alleles.

Pollen compatibility (C) is calculated for each pollen-stigma combination for the two models described above. Mathematically, the Model I algorithm states that if ΣS or ΣZ < 2, the pollen is compatible, therefore C = 1; otherwise C = 0. For Model II, the algorithm states that if ΣS or ΣZ = 0 then pollen is compatible and C = 1; otherwise C = 0. Thereafter, each C value is multiplied with the respective pollen haplotype frequency. The percentage of compatible pollen for a particular pollen-stigma genotype combination is the sum of the products between C values and pollen frequencies (Figure 3.4b).
**Figure 3.4 Screenshot of pollen haplotype table**

a) First 19 (out of 100) rows of pollen compatibility calculation for one of the two reciprocal crosses, underlined rows are examples of compatible pollen.

b) Last 16 (out of 100) pollen haplotypes compatibility and final compatibility percentage under two proposed models

Our program allows the user to continue the analysis with similar meiotic parameters. Thus, the final output (Figure 3.5) is a summary table containing meiotic parameters used, genotype of individuals and the compatibility percentages under both models.
Figure 3.5 Summary output of different runs (changes in genotypes).
First cross, genotypes have one different allele at Z and a pollen compatibility of 50% under Model I and 0% under Model II. In the following case, genotypes differ in one allele at S and one at Z and pollen compatibility increase 25% under Model I and remains 0% under Model II. Meiotic parameters are kept constant for S and Z.

RESULTS AND DISCUSSION

Difference between preferential and non-preferential pairing

A male with genotype ABCD1234 will produce gametes with different haplotypes depending on the mode of pairing during meiosis. Tables 2.2 & 2.3 show possible pollen haplotypes, where shaded pollen haplotypes were produced under the assigned pairing and meiotic parameters. If such a male is crossed onto a female with different alleles at S and Z (i.e., EFGH5678), it is possible to determine the meiotic behavior based on the progeny. Based on output information, it is also possible to determine the number of progeny needed, to screen to find a certain haplotype.

Table 3.2 Pollen haplotypes and their respective frequencies for a male abcd1234 with disomic inheritance

<table>
<thead>
<tr>
<th>Disomic inheritance: $\beta = 0 \Delta 1 = 1$</th>
<th>0</th>
<th>0.25</th>
<th>0.25</th>
<th>0.25</th>
<th>0.25</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>AB</td>
<td>AC</td>
<td>AD</td>
<td>BC</td>
<td>BD</td>
<td>CD</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>12</td>
<td>AB12</td>
<td>AC12</td>
<td>AD12</td>
<td>BC12</td>
<td>BD12</td>
</tr>
<tr>
<td>0.25</td>
<td>13</td>
<td>AB13</td>
<td>AC13</td>
<td>AD13</td>
<td>BC13</td>
<td>BD13</td>
</tr>
<tr>
<td>0.25</td>
<td>14</td>
<td>AB14</td>
<td>AC14</td>
<td>AD14</td>
<td>BC14</td>
<td>BD14</td>
</tr>
<tr>
<td>0.25</td>
<td>23</td>
<td>AB23</td>
<td>AC23</td>
<td>AD23</td>
<td>BC23</td>
<td>BD23</td>
</tr>
<tr>
<td>0.25</td>
<td>24</td>
<td>AB24</td>
<td>AC24</td>
<td>AD24</td>
<td>BC24</td>
<td>BD24</td>
</tr>
<tr>
<td>0</td>
<td>34</td>
<td>AB34</td>
<td>AC34</td>
<td>AD34</td>
<td>BC34</td>
<td>BD34</td>
</tr>
</tbody>
</table>

Disomic inheritance: $\beta = 0 \Delta 2 = 1$

<table>
<thead>
<tr>
<th>0.25</th>
<th>0</th>
<th>0.25</th>
<th>0.25</th>
<th>0</th>
<th>0.25</th>
</tr>
</thead>
<tbody>
<tr>
<td>AB</td>
<td>AC</td>
<td>AD</td>
<td>BC</td>
<td>BD</td>
<td>CD</td>
</tr>
<tr>
<td>0.25</td>
<td>12</td>
<td>AB12</td>
<td>AC12</td>
<td>AD12</td>
<td>BC12</td>
</tr>
</tbody>
</table>
Disomic inheritance: \( \beta = 0 \) \( \Delta_3 = 1 \)

<table>
<thead>
<tr>
<th>0</th>
<th>13</th>
<th>AB13</th>
<th>AC13</th>
<th>AD13</th>
<th>BC13</th>
<th>BD13</th>
<th>CD13</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>14</td>
<td>AB14</td>
<td>AC14</td>
<td>AD14</td>
<td>BC14</td>
<td>BD14</td>
<td>CD14</td>
</tr>
<tr>
<td>0.25</td>
<td>23</td>
<td>AB23</td>
<td>AC23</td>
<td>AD23</td>
<td>BC23</td>
<td>BD23</td>
<td>CD23</td>
</tr>
<tr>
<td>0</td>
<td>24</td>
<td>AB24</td>
<td>AC24</td>
<td>AD24</td>
<td>BC24</td>
<td>BD24</td>
<td>CD24</td>
</tr>
<tr>
<td>0.25</td>
<td>34</td>
<td>AB34</td>
<td>AC34</td>
<td>AD34</td>
<td>BC34</td>
<td>BD34</td>
<td>CD34</td>
</tr>
</tbody>
</table>

Possible pollen haplotypes produced under the designed preferential pairing are in bold (\( \Delta_1 - 3 \))

**Table 3.3** Pollen haplotypes and their respective frequencies for a male ABCD1234 with tetrasomic inheritance

Tetrasomic inheritance: \( \beta = 1 \) all \( \Delta = 0 \) \( \alpha = 0 \)

<table>
<thead>
<tr>
<th>0.17</th>
<th>0.17</th>
<th>0.17</th>
<th>0.17</th>
<th>0.17</th>
<th>0.17</th>
</tr>
</thead>
<tbody>
<tr>
<td>AB</td>
<td>AC</td>
<td>AD</td>
<td>BC</td>
<td>BD</td>
<td>CD</td>
</tr>
<tr>
<td>0.17</td>
<td>12</td>
<td>AB12</td>
<td>AC12</td>
<td>AD12</td>
<td>BC12</td>
</tr>
<tr>
<td>0.17</td>
<td>13</td>
<td>AB13</td>
<td>AC13</td>
<td>AD13</td>
<td>BC13</td>
</tr>
<tr>
<td>0.17</td>
<td>14</td>
<td>AB14</td>
<td>AC14</td>
<td>AD14</td>
<td>BC14</td>
</tr>
<tr>
<td>0.17</td>
<td>23</td>
<td>AB23</td>
<td>AC23</td>
<td>AD23</td>
<td>BC23</td>
</tr>
<tr>
<td>0.17</td>
<td>24</td>
<td>AB24</td>
<td>AC24</td>
<td>AD24</td>
<td>BC24</td>
</tr>
<tr>
<td>0.17</td>
<td>34</td>
<td>AB34</td>
<td>AC34</td>
<td>AD34</td>
<td>BC34</td>
</tr>
</tbody>
</table>

Tetrasomic Inheritance + Chromatid Segregation + Double Reduction

<table>
<thead>
<tr>
<th>0.14</th>
<th>0.14</th>
<th>0.14</th>
<th>0.14</th>
<th>0.14</th>
<th>0.14</th>
<th>0.04</th>
<th>0.04</th>
<th>0.04</th>
<th>0.04</th>
</tr>
</thead>
<tbody>
<tr>
<td>AB</td>
<td>AC</td>
<td>AD</td>
<td>BC</td>
<td>BD</td>
<td>CD</td>
<td>AA</td>
<td>BB</td>
<td>CC</td>
<td>DD</td>
</tr>
<tr>
<td>0.17</td>
<td>12</td>
<td>AB12</td>
<td>AC12</td>
<td>AD12</td>
<td>BC12</td>
<td>BD12</td>
<td>CD12</td>
<td>AA12</td>
<td>BB12</td>
</tr>
<tr>
<td>0.17</td>
<td>13</td>
<td>AB13</td>
<td>AC13</td>
<td>AD13</td>
<td>BC13</td>
<td>BD13</td>
<td>CD13</td>
<td>AA13</td>
<td>BB13</td>
</tr>
<tr>
<td>0.17</td>
<td>14</td>
<td>AB14</td>
<td>AC14</td>
<td>AD14</td>
<td>BC14</td>
<td>BD14</td>
<td>CD14</td>
<td>AA14</td>
<td>BB14</td>
</tr>
<tr>
<td>0.17</td>
<td>23</td>
<td>AB23</td>
<td>AC23</td>
<td>AD23</td>
<td>BC23</td>
<td>BD23</td>
<td>CD23</td>
<td>AA23</td>
<td>BB23</td>
</tr>
<tr>
<td>0.17</td>
<td>24</td>
<td>AB24</td>
<td>AC24</td>
<td>AD24</td>
<td>BC24</td>
<td>BD24</td>
<td>CD24</td>
<td>AA24</td>
<td>BB24</td>
</tr>
<tr>
<td>0.17</td>
<td>34</td>
<td>AB34</td>
<td>AC34</td>
<td>AD34</td>
<td>BC34</td>
<td>BD34</td>
<td>CD34</td>
<td>AA34</td>
<td>BB34</td>
</tr>
<tr>
<td>0.04</td>
<td>11</td>
<td>AB11</td>
<td>AC11</td>
<td>AD11</td>
<td>BC11</td>
<td>BD11</td>
<td>CD11</td>
<td>AA11</td>
<td>BB11</td>
</tr>
<tr>
<td>0.04</td>
<td>22</td>
<td>AB22</td>
<td>AC22</td>
<td>AD22</td>
<td>BC22</td>
<td>BD22</td>
<td>CD22</td>
<td>AA22</td>
<td>BB22</td>
</tr>
<tr>
<td>0.04</td>
<td>33</td>
<td>AB33</td>
<td>AC33</td>
<td>AD33</td>
<td>BC33</td>
<td>BD33</td>
<td>CD33</td>
<td>AA33</td>
<td>BB33</td>
</tr>
<tr>
<td>0.04</td>
<td>44</td>
<td>AB44</td>
<td>AC44</td>
<td>AD44</td>
<td>BC44</td>
<td>BD44</td>
<td>CD44</td>
<td>AA44</td>
<td>BB44</td>
</tr>
</tbody>
</table>

Pure tetrasomic inheritance produces all 36 possible genotypes are shown in regular format. Haplotypes in bold are the ones formed by chromatid segregation and haplotypes in italic are formed by double reduction. The probabilities of each \( S \) or \( Z \) allele combination are given; the gamete frequency can be calculated as joined probability of the frequency at each locus.
Assuming that markers are available to distinguish alleles for S and Z, the stepwise procedure to determine meiotic behavior for S and Z loci would be as follows: first, distinguish between disomic and tetrasomic inheritance. If disomic inheritance occurs, under any type of preferential pairing only 16 pollen haplotypes are possible. In contrast, if tetrasomic inheritance occurs, 36 pollen haplotypes are possible. To distinguish between both, it is necessary to evaluate the presence or absence of any of the 20 (36-16) haplotypes, that are indicative of tetrasomic inheritance.

If disomic inheritance is true, it can be determined, which genomes are pairing during meiosis. One haplotype that distinguishes between \( \text{jklm}, \text{jlkm} \) or \( \text{jmkl} \) needs to be identified. In the example presented in Table 3.2, AB12 is a haplotype that can differentiate between \( \text{jklm} \) and \( \text{jlkm} \). The probability of occurrence of these haplotypes in the progeny is 0.0625, to find at least one progeny we use the formula described by Sedcole (method III) [18]: carrying a specific pollen haplotype,

\[
n = \frac{\left[2(r - 0.5) + z^2(1-q)\right] + z\left[z^2(1-q)^2 + 4(1-q)(r-0.5)\right]^{\frac{1}{2}}}{2q}
\]

Where \( n \) = total number of plants to be screened
\( r \) = required number of plants with the desired genotype
\( q \) = frequency of plants with the desired genotype
\( p \) = probability of recovering the required number of plants with the desired genotype
\( z \) = probability based on a z distribution

Therefore, we need to screen at least 56 plants in this example to be 95% confident to find one offspring carrying the AB12 haplotype. If \( \Delta 2 \) is true, it is necessary to select a haplotype to distinguish it from \( \Delta 3 \) (i.e., AD12). Using the same calculation as above, 56 individuals need to be screened to find at least one genotype carrying the AD12 haplotype with a 95% probability.
On the other hand, if tetrasomic inheritance is true for our species, it is necessary to determine, whether chromosome segregation or double reduction occurs. The probability of finding one haplotype produced by chromatid segregation is 0.2688 (the sum of the frequencies of haplotypes of light shaded boxes in Table 3.3). Using the same formula as above, to find at least one of these haplotypes in the progeny with a 95% confidence, a total of 11 individuals need to be analyzed. Furthermore, if we want to assess the occurrence of double reduction, which occurs with a frequency of 0.0256 (the sum of the frequencies of no shaded lines in Table 3.3), a total of 139 individuals need to be screened.

**Model testing**

Our program can be used to determine the most likely model underlying pollen compatibility. For example, in a cross between ABCD1234 and ABCE1235, these genotypes have one different allele at each locus. Under Model I, reciprocal crosses between these genotypes reaches 75% pollen compatibility. In contrast, under Model II, this cross is completely incompatible. Multivalent formation, chromatid segregation and double reduction during meiosis have an effect on pollen compatibility. A similar cross, in an autotetraploid species ($\beta = 1$) with chromatid segregation and double reduction ($\alpha = 1/7$), will have 71.6% pollen compatibility under Model I or 6.9% under Model II (Figure 3.6). In general, Model II is more restrictive than Model I. Model I ranges from 0% to 50% compatibility, when one different allele is present at either S or Z. In contrast, Model II has values lower than 50%, if there are fewer than three different alleles at any locus (Figure 3.7).
### Output 3: Result table

<table>
<thead>
<tr>
<th>Melosis Behavior</th>
<th>Genotype</th>
<th>Compatibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loci</td>
<td>Loci</td>
<td></td>
</tr>
<tr>
<td>G1   G2   G3   G4</td>
<td>A   B   C   D</td>
<td>G diff. (%)</td>
</tr>
<tr>
<td>0.00 0.00 0.00 0.00</td>
<td>0.00 0.00 0.00</td>
<td>75.0 0.0</td>
</tr>
<tr>
<td>ABECD1234 ABECD1235</td>
<td>ABECD1234 ABECD1235</td>
<td>75.0 0.0</td>
</tr>
</tbody>
</table>

**Figure 3.6 Program output 3.**

Summary output that show differences in pollen compatibilities between Model I and II, and disomic or tetrasomic inheritance.
Figure 3.7 Hypothesis testing of SI models.
Analysis of differences in pollen compatibility under Model I and Model II for species under non-preferential pairing. Expected pollen compatibility percentages are plotted against the number of different alleles at S and Z loci.

**Allele composition**

If meiotic parameters are known or can be calculated, but no markers are available for genotyping S and Z, it is possible to infer the number of alleles segregating in a population of a pair cross. In a plant species with tetrasomic inheritance ($\beta = 1$) and chromatid segregation/double reduction ($\alpha = 1/7$), the number of segregating alleles between the pollen donor and female plays an important role on the pollen compatibility in reciprocal crosses. For instance, if the number of total segregating alleles is four or more for at least one locus (i.e., ABCD1234 X AAAB1115), pollen compatibility is higher than 90% under Model I but only 63%
for Model II (Table 3.4). However, the distribution of these alleles in the two genotypes is important for the pollen compatibility in reciprocal crosses. If ABCD1234 is used as the pollen donor it has two and three alleles for S and Z, respectively, which differ from the female AAAB1115. In contrast, when AAAB1115 is used as the pollen donor, it has zero and one allele for S and Z, respectively, that differ from the female ABCD1234, resulting in a pollen compatibility of 47% under Model I, or 4% under model II (Table 3.3). In addition, if the number of segregating alleles between genotypes drops to two, pollen compatibility is significantly reduced. For instance, a cross between AAAB1115 X AABB1111 will result in no pollen compatibility under any Model. However, if AABB1115 is used as male, the single different allele at Z will cause an increase of pollen compatibility to 47% under Model I, or 4% under Model II (Table 3.4).

Table 3.4 Differences in reciprocal crosses between genotypes that have alleles identical by descent (IBD)

<table>
<thead>
<tr>
<th>Pollen Donor</th>
<th>Female</th>
<th>Total segregating</th>
<th>Pollen donor</th>
<th>Female alleles</th>
<th>Pollen compatibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABCD1234</td>
<td>AAAB1115</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>AAA81115</td>
<td>ABCD1234</td>
<td>4</td>
<td>5</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>AAAA1234</td>
<td>AAA81115</td>
<td>3</td>
<td>5</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>AAA81115</td>
<td>AAAAD1234</td>
<td>3</td>
<td>5</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>AAA81115</td>
<td>AAABD1134</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>AAA81115</td>
<td>AAA81115</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>AABB1114</td>
<td>AABB1115</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>AABB1114</td>
<td>AABB1114</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>AABB1111</td>
<td>AAA81115</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AABB1115</td>
<td>AABB1111</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Alleles that are IBD are in bold and underlined. First column (total number of segregating alleles) is the number of different alleles between S or Z across genotypes. The columns that follow count the number of different alleles between males and females.

The information provided by PollenCALC can be used to design experiments in order to test hypotheses related to the pairing and segregation of chromosomes during meiosis in tetraploid species with a gametophytic S-Z SI system. The program can be used if linked or functional markers for S and Z are available or not.
The genetic mechanism of SI is still unknown for some self-incompatible tetraploid grasses. Also, there are no markers available for S and Z which makes it difficult to understand and predict pollen compatibility. As described above, there are two possible models for pollen compatibility. *In vitro* or *in vivo* pollinations can be performed to determine, which of the two models most likely applies. As demonstrated, only few different alleles are required for Model I. Furthermore, under Model I, if only one different allele is found at either S or Z, pollen compatibility reaches 50%, whereas for Model II values below 50% are common when the number of different alleles is below three for any locus. Both *in vitro* and *in vivo* pollination tests provide bias estimates of pollen compatibility. For instance, current *in vitro* pollination tests rely on imaging and counting to provide estimates of pollen compatibility. *In vivo* pollination analyzed as seed set can be affected by pollen death due to environmental conditions and pre or post zygotic abortion. Nevertheless, PollenCALC can be used to design experiments in order to identify genotypes that are most informative to distinguish the two models i.e. genotypes with extreme pollen compatibilities. Eventually the use of S and Z linked markers are required for confirmation.

There are other applications for this program when markers are available. For example, genome-wide markers are used in several studies to determine preferential pairing in polyploids. In species with SI, segregation of markers linked or in linkage disequilibrium with S and Z are distorted, which can result in biased interpretation of marker segregation data [19,20]. PollenCALC can help to identify genotypes that can distinguish between types of pairing. Knowledge about pairing and inheritance in a given species makes it possible to correct mapping distances for segregation distortion due to SI in mapping experiments.
Also, PollenCALC calculates pollen haplotype frequencies of individuals in pair crosses and can be used to calculate pollen compatibility between those genotypes in reciprocal crosses. This information can also be used to maximize seed yield in hybrid breeding programs [21]. If markers for S and Z are available, they can be used in combination with our software in breeding programs. For instance in a hybrid seed production programs, PollenCALC can be used to select genotypes with specific S-Z genotypes within one heterotic group that contrast with the S-Z genotype from the individuals in the other heterotic group. If alleles are different between groups, seed yield will be maximized. Also, it is possible to determine potential progeny genotype and frequencies by combining haplotypes and frequencies. In the longer run, this program can be used to simulate allele and genotypes frequencies for S, Z, and loci genetically linked to S and Z over multiple generations. Thus, PollenCALC is a useful tool for plant geneticists and plant breeders working with tetraploid species.

CONCLUSION

PollenCALC is the only program that we are aware of that is able to calculate gamete haplotypes and pollen compatibility for tetraploid species with a gametophytic S-Z SI. The frequencies obtained can be used to determine the number of individuals needed to test different models for GSI, to optimize experimental designs, and to take segregation distortion of linked markers into account, indicating linkage disequilibrium with SI loci.

Availability and requirements

Project name: PollenCALC
Project home page: https://github.com/BerndWollenweber/PollenCALC/downloads
Operating system(s): tested under Windows XP, Windows Vista, Windows 7
Programming language: C++
Other requirements: none.
License: free non-commercial research-use license.
Any restrictions to use by non-academics: none.

**Abbreviations**

SI, Self-incompatibility; GSI, Gametophytic self-incompatibility; SSI, Sporophytic self-incompatibility

**Competing interests**

The authors declare that they have no competing interests

**Authors’ contributions**

AAA has developed the algorithms and carried out initial design for the calculation and estimation of pollen compatibility under the two proposed models. BW conceived the software, carried out its design and participated in writing the draft of the manuscript. UKS and TL have participated in writing the draft of the manuscript. All authors have read and approved the final manuscript.

**Acknowledgements**

We are grateful to RF Baker Center for Plant Breeding (Iowa State University) that supports Andrea Arias Aguirre.

**REFERENCES**

CHAPTER 4: MAPPING A NEW SOURCE OF SELF-FERTILITY IN PERENNIAL RYEGRASS

Andrea Arias Aguirre\textsuperscript{1}, Bruno Studer\textsuperscript{2}, Ursula Frei\textsuperscript{1}, Thomas Lübberstedt\textsuperscript{1}

\textsuperscript{1} Department of Agronomy, Iowa State University, Ames, Iowa
\textsuperscript{2} Institute of Agricultural Sciences, Forage Crop Genetics, ETH Zurich, 8092 Zurich, Switzerland

ABSTRACT

There is a rising interest of moving towards hybrid breeding in outcrossing species. Self-compatibility (SC), which occurs at low rates in self-incompatible species, could be used to develop inbred lines, a key requirement for hybrid breeding programs. In perennial ryegrass, the existence of SC independent from the self-incompatibility (SI) loci S and Z has been reported.

In this study, we used 98 F\textsubscript{2} individuals from a cross between a self-compatible ecotype and an individual of the VrnA mapping population that were tested with markers for two candidate linkage groups 3 and 5 as well as markers for candidate regions at the S and Z locus. We were able to determine a tentative location of a SC locus and tested the possible interaction to other loci involved in SI and SC. This information will help to fine map the SC locus, and for marker assisted selection.
INTRODUCTION

Increasing interest of moving towards hybrid breeding in several outcrossing species requires manipulation of mechanisms preventing or allowing self-fertilization. Species that are forced to outcross display a strong inbreeding depression, which suggests the possibility of exploiting heterosis in hybrid breeding. One key factor for moving from population or synthetic to hybrid breeding is the ability to develop inbred lines (Arias Aguirre et al., 2011). This is a particular challenge in self-incompatible (SI) species.

Self-compatibility (SC) is not unusual in self-incompatible species. Mutations at SI loci have been determined to be involved in self-fertility and have been mapped in rye (Secale cerealis L.) (Fuong et al., 1993; Voylokov et al., 1993; Voylokov et al., 1998). In Phalaris coerulescens, mutations at S and Z as well as an independent locus have been identified (Hayman and Richter, 1992). Finally, a recent study in switchgrass (Panicum virgatum) reported a genotype with a high rate of selfing (61.3%), suggesting presence of a major QTL controlling this trait (Liu and Wu, 2012).

In perennial ryegrass, a locus causing SC that segregates independently from S and Z has been reported (Thorogood et al., 2005). This locus (T) is located on linkage group (LG) 5. It has been speculated, whether the product of (T) prevents transcription of S or Z, or the interaction occurs at the post-transcriptional level. Thorogood et al. (2005) also found an additional effect on S at LG 1, but segregation of SS/tt genotypes did not always fit the expected compatibility. In that same study, segregation distortion on LG 7 suggested the interaction of the SC locus on this LG. However, no major significant QTL was detected in this region. In addition, there is also a high rate of segregation distortion reported in several mapping populations for LG 3 which may be involved in SC or related traits, such as flowering and vernalization response (Anhalt, 2008).
Thorogood et al. (2005) observed a 1:1 phenotypic segregation between 50% self-compatible and full self-compatible plants in a bi-parental F$_2$ population, and concluded that SC is controlled by a major gene. We have established a self-compatible perennial ryegrass genotype which is independent of the one characterized by Thorogood et al (2005). Our question was, whether SC in our population is controlled by the same major gene located in LG 5, and whether the 1:1 phenotypic ratio could also be observed in our F$_2$ population. The objectives of our study were to i) determine which gene model best describes the phenotypic segregation found in our F$_2$ population; ii) find a relationship between phenotypic data and marker data for candidate chromosomes and loci; and iii) map SC loci in a F$_2$ population obtained by crossing a potential new source of SC and an individual of the VrnA mapping population (Jensen et al., 2005).

**MATERIALS AND METHODS**

**Plant materials**

To characterize loci affecting SC, a self-compatible genotype obtained after five generations of selfing by single seed descent (with an expected average homozygosis of 96.9%) was crossed to a self-incompatible individual of the VrnA mapping population (Jensen et al., 2005). The self-compatible genotype is assumed to be homozygous for both the S and the Z locus, whereas the VrnA plant has one and two alleles segregating for Z and S, respectively (Studer, unpublished data). A single F$_1$ individual was selfed during summer 2011. 98 F$_2$ plants were vernalized for five weeks at 5 °C with a 10h light per day photoperiod in a growth chamber during winter 2011. All plants were grown in the Horticulture greenhouse at Iowa State University, Ames, IA, in a 16 h light per day photoperiod and 20 °C for flowering.
**Pollination tests**

Flowering tillers were harvested prior to anthesis. Stigma were collected on a 2% agar, 10% sucrose and 100 ppm Boric acid medium, and incubated at 25C for 24 h. Stigma were self-pollinated and incubated for 4 h at 25C. Ovaries were cut using razor blades and placed on microscope slides. A drop of staining solution (0.2% aniline blue in 2% K₃PO₄) was placed over pollinated stigma and covered with a slide. Pollinated stigma were inspected under a Zeiss Axioplan 2 Imaging UV light microscope and images were analyzed with the Axiovision software (Zeiss, Germany) at the Microscope and Imaging Facility at Iowa State University. Each plate contained at least six self-pollinated stigma. Each genotype was analyzed in two independent replications. Pollen compatibility classification was based on observations of all the stigma collected in each replication. If most of the pollen grains were uncolored and showed long pollen tubes, the pollen/stigma donor was classified as self-compatible. If about half of the pollen grains developed strongly stained but short pollen tubes, the pollen/stigma donor individual was classified as half-compatible.

**Phenotypic segregation analysis**

Chi square tests were used to evaluate, which genetic model is consistent with observed segregation data. Four phenotypic segregation models for SC were tested, assuming it is controlled by a single gene: a) additivity in which three distinguishable phenotypic classes are expected in a 1:2:1 ratio (Model I); b) dominance in which two distinct phenotypic classes are expected in a 3:1 segregation (Model II); c) homozygote class lethality in which one of the homozygote classes is completely missing and the two remaining genotypic classes are expected in a 2:1 ratio (Model III); and d) gametic selection in which one of the gametes carrying a
recessive allele is not transmitted to the next generation (either as male or female gamete) and, therefore, only two phenotypic classes are expected in a 1:1 ratio (Model IV).

**Marker data**

35 markers for LG 3 and 5 were randomly selected from EST-SSR derived markers (Studer B et al., 2010) and tested for polymorphism. In addition, three markers closely linked to Z (Manzanares et al., 2011), as well as four markers closely linked with a region segregating for S (Studer, unpublished data) were used in this study (Table 4.1).

<table>
<thead>
<tr>
<th>Marker names</th>
<th>Linkage group</th>
</tr>
</thead>
<tbody>
<tr>
<td>SBA, SBF, SBC1, SBH2</td>
<td>1a</td>
</tr>
<tr>
<td>ZB1, ZB2, ZB3</td>
<td>2b</td>
</tr>
<tr>
<td>G06-029, G05-134, G04-054, G04-098, G07-058, G01-059, G01-098, G07-071, G02-075</td>
<td>3c</td>
</tr>
<tr>
<td>G02-057, G01-045, G05-071, G03-052, G06-096, G05-065, G03096, G07-065, G05-094, G07-024</td>
<td>5c</td>
</tr>
<tr>
<td>G04-043, G01-080</td>
<td></td>
</tr>
</tbody>
</table>

a Designed from candidate region provided by Studer et al. (unpublished data)
b Studer et al. (unpublished data)
c Studer et al. (2010)

The first step of marker screening was to evaluate for presence of bands on 2% agarose gels, using F1 DNA as template. Markers with at least a single band were tested for polymorphisms in a sample of 16 individuals of the F2. Polymorphic markers were used for genotyping using high resolution melting (Studer et al., 2009). PCR reactions contained 1x LCGreen dye (Idaho Technology, Inc., UT, USA), 0.1 mM of dNTPs, 10 µM of each forward and reverse primer, 20 mM of magnesium chloride, and 20 ng of sample DNA. In addition, 15 µL of mineral oil were added to the PCR mix to avoid evaporation. PCR amplification was conducted in a PTC-200 Thermal Cycler (MJ Research Inc., Watertown, MA, USA) following the protocol by Studer et
Annealing temperatures for each primer were standardized, using gradient PCR. A final cycle of 30 s at 94°C and at 25°C was added for heteroduplex formation. The LightScanner Instrument (96-well plate format) as well as the LightScanner software Version 2 (Idaho Technology, Inc., Utah, US) were used for melting analysis. Melting temperatures used across markers ranged from 68-98°C in the auto exposure setting.

**Linkage map construction**

Twenty one markers were used to cover the two LGs 3 and 5, 9 markers for LG3 and 12 for LG5. Each marker was anchored to its corresponding LG using the consensus map for ryegrass as a reference (Studer B *et al.*, 2010). Map positions were obtained using the Haldane mapping function in the Onemap package of R (Margarido *et al.*, 2007). A two-point analysis with minimum likelihood of the odds (LOD) of 3.0 and a maximum recombination fraction of 0.5 was applied for all markers on each LG and for two polymorphic markers near the S and three polymorphic markers near the Z loci on LG 1 and 2, respectively. Marker order was obtained using the rapid chain delineation (Doerge, 1996), and the ripple algorithm (Lander, 1987) in a seven-marker window. The TRY algorithm was used to determine the most likely position of the remaining markers on each LG. MapChart v2.2 (Voorrips, 2002) was used to generate graphs for each LG.

**Phenotypic and marker segregation analysis**

Chi-square tests for goodness of fit were used to compare observed phenotypic data and expected segregation under the same four models proposed at the phenotypic level. All markers with a p-value <0.05 were considered significantly different from expected segregation ratios.
Hypothesis testing

Single gene models

We used Model IV, which was consistent with the observed data (p-value >0.05) at the phenotypic level to test for marker-trait associations using Chi square tests of independence. We tested for statistical associations between phenotypic segregation and any given marker on LG 3, and LG 5 as well as for candidate markers at the S and Z region. Markers and phenotypes were considered associated for p-values <0.05.

RESULTS

Phenotypic data

From the 98 plants obtained initially, 18 died along the course of the experiment. Those 18 plants were generally smaller and had a lower growing rate compared to the rest of the population. Some of them died before vernalization, but the majority died during or right after vernalization. A total of 80 plants survived the vernalization process, but only 68 produced flowering tillers. The number of flowering tillers ranged from one to eight among the surviving plants (data not shown). Only 45 plants were useful for pollination tests. The remaining 23 plants had fewer than two flowering tillers, presented non-dehiscent anthers or had a low pollen shedding rate, and were not included in pollination tests. Based on the assumption of a major gene controlling SC in this population, we considered the 45 remaining F2 plants sufficient for genetic analyses.

In vitro pollination tests used to assign plants to phenotypic classes showed two distinct characteristics of pollen that were used to determine SC. One is staining of the pollen tubes which were bright for inhibited pollen and light for the compatible pollen. The second is the pollen tube length that had to be at least four times larger than the pollen grain diameter. Plants with nearly 100% germinated pollen grains with long pollen tubes on self-stigma were counted as fully
compatible (Figure 4.1A). Plants that showed a majority of bright pollen grains with a pollen tube smaller than the pollen grain diameter were considered semi-compatible (Figure 4.1B). Estimation of compatibility was done accounting for non-viable pollen grains that looked unstained, raisin shaped or empty. Among the 45 plants, 22 were classified as fully compatible and 23 were classified as semi-compatible. No plants displayed a 100% SI phenotype, as did one of the parents of the mapping population (VrnA genotype), and as it is normally expected in perennial ryegrass.

Figure 4.1. In vitro pollination tests for pollen-stigma interaction.
A) Fully compatible reaction after self-pollination, B) semi-compatible reaction after self-pollination

For Models I and II, three phenotypic classes were expected. Our observations fell into only two major categories, fully and semi-compatible. Chi-square tests indicated that phenotypic
segregation data were inconsistent with single gene Models I, II (p<0.01), and III (p <0.05). In contrast, observed segregation data were consistent with Models IV (p >0.05) (Table 4.2).

Table 4.2. Observed and expected counts used in chi-square test for goodness of fit from different one gene models. P >0.05 means there is no significant difference between the observed and expected segregation.

<table>
<thead>
<tr>
<th>Phenotypic Class</th>
<th>Observed counts</th>
<th>Model I 1:2:1</th>
<th>Model II 3:1</th>
<th>Model III 2:1</th>
<th>Model IV 1:1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fully compatible</td>
<td>22</td>
<td>11.25</td>
<td>33.75</td>
<td>30</td>
<td>22.5</td>
</tr>
<tr>
<td>Half incompatible</td>
<td>23</td>
<td>22.5</td>
<td>N/A</td>
<td>15</td>
<td>22.5</td>
</tr>
<tr>
<td>Incompatible</td>
<td>-</td>
<td>11.25</td>
<td>11.25</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P&lt;0.001</td>
<td>P&lt;0.001</td>
<td>p=0.0114</td>
<td>p=0.88</td>
</tr>
</tbody>
</table>

**Linkage mapping**

Linkage mapping resulted in markers grouped into expected linkage groups compared to the consensus map as published by Studer et al. (2010). In terms of length, LG 3 is 13.1 cM larger than the reference map (123.6 cM). Similarly, the length of LG 5 is 10.4 cM larger compared to the 101.7 cM of the same LG in the reference map. Moreover, the marker order in LG 3 was slightly different compared to the reference map, especially for markers G07_071 and G02_075, which mapped distant from neighboring markers in the reference map. In contrast, for LG5 marker order is conserved for closely linked markers with a few exceptions such as G03_052, G07_065 and G05_094, which differed from their position on the reference map (Figure 4.2).

Genetic linkage maps for LG 1 and LG 2 were only created to determine marker order and distance for the candidate regions for S and Z, respectively. Their lengths were shorter than the reference map.
Figure 4.2. Genetic linkage maps for markers on linkage groups 1, 2, 3, and 5 in the F2 population. For linkage groups 1 and 2, only the genetic distances between candidate markers are given.

**Marker segregation analysis for S and Z markers**

All the candidate markers at S and Z were significantly different from a 1:2:1 expectation for codominance (p<0.05, Table 4.3). Moreover, all markers at S and Z were lacking one homozygote class, except for SBH2. This suggests that markers may be dominant inherited. However, only SBC1 and ZB2 were consistent with that mode of inheritance. The remaining markers at Z segregated consistent with model IV; the model that was consistent with the observed phenotypic segregation (see Table 4.2).

<table>
<thead>
<tr>
<th>Marker</th>
<th>Chromosome</th>
<th>Model I 1:2:1</th>
<th>Model II 3:1</th>
<th>Model III 2:1</th>
<th>Model IV 1:1</th>
</tr>
</thead>
<tbody>
<tr>
<td>SBC1</td>
<td>1</td>
<td>S</td>
<td>NS</td>
<td>NS</td>
<td>S</td>
</tr>
</tbody>
</table>
Segregation analysis for markers on LG 3 and LG 5

Close to 50% of the markers on LG 3 were significantly distorted from the expected 1:2:1 ratio for co-dominance. However, most of the markers that were considered distorted from a 1:2:1 segregation ratio fitted the dominant inheritance Model II, or the lethality Model III. Most markers showed all expected genotypic classes, some of them distorted from expectation but still present (Figure 4.3). However, there is a region of approximately 40 cM between markers G07_058 and G01_098, with one of the homozygote classes completely missing (Figure 4.3).

Table 4.4. Deviation from expected segregation ratios under four single-gene models. Significant differences are shown at $p = 0.05$ level.

<table>
<thead>
<tr>
<th>Marker</th>
<th>Model I 1:2:1</th>
<th>Model II 3:1</th>
<th>Model III 2:1</th>
<th>Model IV 1:1</th>
</tr>
</thead>
<tbody>
<tr>
<td>G06_029</td>
<td>S</td>
<td>NS</td>
<td>NS</td>
<td>S</td>
</tr>
<tr>
<td>G05_134</td>
<td>NS</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>G04_054</td>
<td>NS</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>G04_098</td>
<td>NS</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>G07_058</td>
<td>S</td>
<td>NS</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>G01-059</td>
<td>S</td>
<td>NS</td>
<td>NS</td>
<td>S</td>
</tr>
<tr>
<td>G01-098</td>
<td>S</td>
<td>NS</td>
<td>NS</td>
<td>S</td>
</tr>
<tr>
<td>G07_071</td>
<td>NS</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>G02_075</td>
<td>NS</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
</tbody>
</table>

S = significant difference
NS= non-significant difference
Figure 4.3. Segregation of polymorphic markers on linkage group 3.
Markers are shown in ascending order of their chromosome position. The three bars for each marker represent the two homozygous and the heterozygous class. The most abundant homozygous class (A) comes from the self-compatible line.

In contrast, all markers except one differed significantly on LG 5 from both the expected ratios for co-dominant (1:2:1) and dominant marker segregation (3:1) (p<0.05). Furthermore, there is a clear absence of one homozygote class for all markers except G03-096, G05-094, G07-024 and G04-043 (Figure 4.4) which behave as codominant markers. The remaining markers had a similar frequency in the other two genotypic classes.
Figure 4.4. Segregation of polymorphic markers on linkage group 5.
Markers are shown in ascending order of their chromosome position. The three bars for each marker represent the two homozygous and the heterozygous classes. The most abundant homozygous class (A) comes from the self-compatible line.

**Hypothesis testing**

Even though the candidate markers for S and Z showed clear segregation distortion from the 1:2:1 model, marker segregation was not associated with phenotypic segregation. Both markers for S and the three markers for Z were independent from the phenotypic observations (Table 4.4) (p >0.05).

Markers with only two genotypic classes on LG 3 were not consistent with observed phenotypic segregation ratios (Table 5). For markers on LG 5, a region of approximately 14 cM between the markers G05-071 and G05-065 was significantly associated with the phenotypic classification (Table 4.4).
Table 4.5. Test for independence between marker segregation and phenotypic segregation for markers in all linkage groups.

<table>
<thead>
<tr>
<th>Linkage group</th>
<th>Marker</th>
<th>Genotype</th>
<th>Half Compatible</th>
<th>Full Compatible</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SBC1</td>
<td>A</td>
<td>4</td>
<td>5</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H</td>
<td>19</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>A</td>
<td>6</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SBH2</td>
<td>H</td>
<td>4</td>
<td>5</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B</td>
<td>13</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>ZB1</td>
<td>A</td>
<td>12</td>
<td>11</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H</td>
<td>11</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ZB3</td>
<td>A</td>
<td>12</td>
<td>11</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H</td>
<td>11</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ZB2</td>
<td>A</td>
<td>7</td>
<td>7</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H</td>
<td>16</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>G07_058</td>
<td>A</td>
<td>5</td>
<td>2</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H</td>
<td>18</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td></td>
<td>G01-059</td>
<td>A</td>
<td>7</td>
<td>4</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H</td>
<td>16</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td></td>
<td>G01_098</td>
<td>A</td>
<td>7</td>
<td>5</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H</td>
<td>14</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>G02_057</td>
<td>A</td>
<td>8</td>
<td>13</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H</td>
<td>15</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>G01_045</td>
<td>A</td>
<td>8</td>
<td>13</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H</td>
<td>15</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>G05_071</td>
<td>A</td>
<td>6</td>
<td>14</td>
<td>0.03**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H</td>
<td>17</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>G03_052</td>
<td>A</td>
<td>3</td>
<td>14</td>
<td>0.002*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H</td>
<td>18</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>G06_096</td>
<td>A</td>
<td>3</td>
<td>16</td>
<td>0.0002*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H</td>
<td>20</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>G05_065</td>
<td>A</td>
<td>4</td>
<td>16</td>
<td>0.0006*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H</td>
<td>19</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>G07_065</td>
<td>A</td>
<td>9</td>
<td>14</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H</td>
<td>13</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>A</td>
<td>6</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td></td>
<td>G03_096</td>
<td>H</td>
<td>13</td>
<td>8</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B</td>
<td>4</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

P-value <0.05 show an association between marker and phenotype
DISCUSSION

In this study, we considered two linkage groups of interest for self-compatibility and also used markers for the S and Z loci to unravel interactions between these loci in relation to self-compatibility.

Population size, genetic map and segregation distortion

In our study, map lengths were slightly different than the ones obtained for the reference map (Studer B et al., 2010). The small population size obtained for this research resulted from the presence of non-dehiscent anthers, which also was observed by Thorogood et al. (2005). In addition to this limitation, our population exhibited a high mortality rate that may be caused by recessive lethality genes segregating in the population.

There have been several reports of segregation distortion for linkage group 3 in different mapping populations and it may be associated with a SC locus (Armstead, 2008). However, we were not able to statistically associate any marker in this linkage group to SC. The great amount of unexplained distortion can be due to LD between markers in LG3 and LG5. Additional analyses showed that there was a correlation between markers in both linkage groups ranging from r=0 to 0.49 and that markers G05-134, G04-054, G04-098, G07-058, G01-059 and G01-098 in LG3 were statistically associated with markers G02-057, G05-071, G03-052, G06-096, G05-065, G07-065 and G01-080 in LG5 (r=0.27 to 0.49 pvalue<0.05)

Comparative genetics of map regions involved in flowering and fertility showed that those regions were distributed over all linkage groups (Armstead, 2008). Therefore, the analysis of related traits, i.e., self-compatibility was always confounded with the effects of these regions as well as with the effect of both self-incompatibility loci. The segregation distortion observed in
both linkage groups 3 and 5 was in agreement with observations obtained in other mapping populations in *Lolium perenne* (Anhalt, 2008).

**Genetic model for self-fertility in the mapping population derived from a cross between a SC genotype and a SI individual from the VrnA mapping population**

The markers with a significant association with the phenotype (p<0.05) were located between 19.6 and 33.3 cM on chromosome 5 and were flanked by markers G01-054 and G07-065. This region of approximately 14 cM was the most likely position for the SC loci in this population. Moreover, as there were four markers in this interval (G05-071, G03-052, G06-096 and G05-065) it was possible that there was more than a single SC locus on LG5. Thus, we refer to the SC haplotype.

In a model with fully functional self-incompatibility, no progeny are expected after selfing. In contrast, in this study there was a substantial number of SC F2 plants with seed set, and there was not distortion of S or Z markers genotypes in the resulting F2. Our observations provide evidence of a single region on LG 5 controlling SC, but with more than one significant marker. The region show a dominant gene action with epistastic effects over S and Z, as there is no SI in case of presence of at least one dominant SC haplotype. We were not able to recover any genotype without at least one SC haplotype. Since up to now our SC gene or genes have not been fine mapped in any population and given the evidence of a few markers with a high correlation to the phenotype, more work is needed to narrow down the candidate genes.

Most likely the SC determinant(s) show gametophytic expression, since all progeny derived from a single F1 plant and selfed F2 genotypes showed two different phenotypic classes (half and fully fertile). In contrast, if the expression would have been sporophytic, all homozygous SCSC and heterozygous SCsc progeny would be expected to have the same genotype.
Gametes probably on the male side carrying the recessive alleles (sc) are likely less competitive compared to gametes carrying the dominant (SC) allele. This insufficiency in the gametes carrying the recessive allele could explain the underrepresentation of homozygous recessive individuals in the F₂. This distortion can be extended to neighboring markers on linkage group 5, where the SC gene or genes are located (Figure 4.5).

**Figure 4.5. Proposed gene action model.**
There is gametic selection against male gametes carrying the “sc” alleles, therefore “scsc” genotypes do not exist in F₂. As a result alleles in LD with “sc” will be distorted and they behave as 1:1 segregation at the phenotypic level.

**Comparison with previous studies**

The distortion observed in LG5 also was reported and attributed to self-compatibility in a previous study (Thorogood et al., 2005). As expected, distortion of markers in LG5 cover a large region including one of the distal ends of the linkage group. It is also noticeable that the
distortion increases closer to the SC region and decreases for markers that map further away. This extended distortion may be explained by the large LD blocks present in an F2 population. However, the significant association obtained for markers in the region between 19.6 and 33.3 cM region provides evidence that at least one major SC gene is located in this region. The position of this major SC locus in our study differs from the location of the SC locus T reported by Thorogood et al. (2005), who found a significant QTL (LOD 6.08) for a marker at 19.8 cM. Unfortunately, due to different in markers used in the previous study, it is not possible to assess the difference in position between their QTL and our significant markers. To establish confirmation, new HRM markers need to be designed or an allelism test needs to be performed to confirm whether it is the same gene.

The segregation of S and Z markers showed that there are no effects of the SI haplotypes on self-fertility. In the study by Thorogood et al. (2005), a slight evidence of interaction between a combination of marker genotypes at the SC loci and a marker in the S region was reported. However, this marker combination did not predict the phenotype a 100% of the times (Thorogood et al., 2005). With the present study, we corroborated that there is no marker combination between the linked marker for SC and the SI linked marker able to predict the phenotype. Therefore, it is possible to say that the presence of a single SF allele overrides the effect of the SI loci. Nevertheless, despite the absence of an association at the genotypic level, it is possible that the SC determinant interacts with the products of S and Z either by forming complexes or destroying them to present the effective functionality of the SI system.

In species with a single SI gene, the most common source of SC is a mutation and, therefore, a loss of function SI allele. In grasses, at least three different sources have been identified. One locus named T in Phalaris coerulescens (Hayman and Richter, 1992). In rye, there are two mutations directly related to S and Z and an additional locus in chromosome 5 that is also a
source of SC (Fuong et al., 1993). Both studies provide evidence that there are genes in other species independent of S and Z that can contribute to self-compatibility.

**Implication for forage breeding**

This study renders a potential new source of SC that can be used as a starting point to understand the interaction between SI and SC genes. The understanding of this interaction is crucial to enable the use of SC genes in a line development program.

For instance, it is possible to use SC haplotypes to develop inbred lines. However, at the end the SC haplotype would be fixed and limit the ability to use SI for hybrid seed production (Arias Aguirre et al., 2011). There are two possible alternatives for this, the first option is to self the heterozygote individuals, continuously and selected the ¼ homozygote class which is not self-compatible. This could only be possible under no lethality of this genotypic class. The second option would be to find S-Z haplotypes which may be more difficult to override than the one present in this population. Furthermore, it may be possible to find certain haplotype combinations with an environmental sensitivity, i.e., temperature that will allow using the SC haplotype for line development under certain environmental condition but would trigger a SI in a different environment. Therefore it is important to broaden the understanding of the interaction between the SC and SI loci.

The population used in this study is a potential new source SC haplotype that can be introgressed to different genetic backgrounds with different S and Z haplotypes to better understand the interaction between these loci.

**REFERENCES**


CHAPTER 5: VALIDATION OF TWO MODELS FOR SELF-INCOMPATIBILITY IN AUTOTETRAPLOID RYEGRASS USING HIGH RESOLUTION MELTING-BASED MARKERS

Andrea Arias Aguirre¹, Bruno Studer², Javier do Canto¹, Ursula Frei¹, Thomas Lübberstedt¹

¹ Department of Agronomy, Iowa State University, Ames, Iowa
² Department of Molecular Biology and Genetics, Faculty of Science and Technology, Research Centre Flakkebjerg, Aarhus University, Forsøgsvej 1, 4200 Slagelse, Denmark

ABSTRACT

*Lolium perenne* (ryegrass) is a species of major economic importance due to its multiple uses such as forage, turf, and cover crop and its potential use as a feedstock for bioenergy. Ryegrass occurs naturally as a diploid (2n=14). However, several autotetraploid cultivars were obtained via colchicine. Ryegrass displays a two-locus gametophytic self-incompatibility system (SI) that remains intact even at the tetraploid level. There are two plausible models for SI at the tetraploid level, Model I in which only one different allele in the pollen grain from the female at either S or Z locus makes pollen compatible. Model II states that any S-Z combination in the pollen grain matching the female would make pollen incompatible. The goals of this study are i) to evaluate which of the two models better explains SI in our bi-parental autotetraploid ryegrass population; ii) to provide evidence that SI maintenance was not specific to genotypes used in previous studies; and iii) to evaluate the sensitivity and efficacy of HRM genotyping in polyploid species.

The present study used 43 progeny of a cross between two autotetraploids and characterized them with three HRM-based markers that were reported to co-segregate with Z. The confirmation of the underlying model for SI in autotetraploids in combination with the use of markers as well as
automated software for pollen compatibility estimation such as PollenCALC can be of great use to perform controlled crosses in tetraploid ryegrass

**INTRODUCTION**

*Lolium perenne* (ryegrass) is an important crop species due to its multiple uses such as forage, turf, cover crop and its potential use as a feedstock for bioenergy (Arias Aguirre et al., 2011; Donnison, 2009; Studer B et al., 2010). Ryegrass occurs naturally as a diploid species (2n=14). However, autotetraploid cultivars were obtained via colchicination and released for forage production. Both diploid and autotetraploid genotypes display a gametophytic self-incompatibility system (SI) (Baumann et al., 2000; Fearon et al., 1984a; Thorogood et al., 2002). This mechanism is determined by two independent loci named S and Z, located in linkage groups 1 and 2, respectively (Thorogood et al., 2002). Both loci are involved in preventing self-pollination.

The consensus model for gametophytic SI in diploid grasses suggests that fifty percent of the pollen produced by a pollen donor is compatible to a pollen receptor, if the pollen donor genotype has one different allele at either S or Z compared to the receptor genotype (Klaas et al., 2011b; Thorogood et al., 2002; Yang, 2009; Yang et al., 2008). Additionally, if the pollen donor has precisely one different allele at each locus, S and Z, then 75% of the pollen is compatible with the pollen receptor genotype. If the number of different alleles at either locus of the pollen donor increases, pollen is 100% compatible.

In autotetraploid species the interaction between pollen and stigma alleles is more complicated due to the presence of four alleles at a locus and sixteen possible pollen haplotypes for both loci. In contrast to gametophytic SI in several dicots, where polyploidy causes breakdown of self-incompatibility, the monocot SI system remains functional in polyploids (Fearon et al., 1984a).
Two plausible models for SI have been proposed by Arias et al. (2012) and are in agreement with those tested by Fearon and Hayward (1984): Model I: if any allele at either S or Z in the pollen differs from the alleles for S or Z in the pistil, pollen is compatible (Arias Aguirre et al., 2012); Model II: if at least one allele at S and one allele at Z locus present in the pollen grain matches the pistil, then pollen is incompatible (Arias Aguirre et al., 2012).

Initial studies on the appropriate genetic model for SI in autotetraploids was addressed by in vitro pollination tests (Fearon et al., 1984a; Fearon et al., 1984c). Fearon & Hayward (1984) used pollen from diploid plants as testers against their tetraploid counterparts, and found evidence supporting Model II. However, some key questions remain to be answered. First, does the model proposed hold true when tetraploid plants are pollinated with their own diploid pollen? Does the number of alleles at each locus affect the proposed models or is there dominance? Moreover, percentage of compatibility is confounded between ploidy and the effect of SI loci since the pollen used came from a diploid parent. Finally, the accuracy of determining SI genotypes by using in vitro pollination tests is limited. Pollen viability, media, and staining can affect the assessment of in vitro pollen germination and, therefore, SI genotype inference. Besides, it is challenging to differentiate pollen compatibilities of 30% and 58%, which were expected frequencies in the cited study, and could result in miss-classification of genotypes. Thus, application of molecular markers linked to SI loci help to increase accuracy of classification and determination of compatibility percentages, and ultimately the correct model for SI in autotetraploids.

Manzanares, et al. (2011) showed that HRM-based markers were able to discriminate between different Z haplotypes in ryegrass. The markers tested in this study were developed within 200 kb in the Z region and enabled discrimination of a total of 13 haplotypes among different populations of selected half-sibs that trace back to nine founders (Manzanares et al., 2011).
The goals of this study were i) to evaluate the efficiency of HRM genotyping for discriminating different iso-allelic genotypes in autopolyploid species; and ii) to evaluate which of the two models better explains SI in our bi-parental autotetraploid ryegrass population.

**MATERIALS AND METHODS**

**Plant materials**

The genomes of the two parents, F1-30 and F1-39, of the VrnA population (Jensen et al., 2005) with defined S and Z genotypes were doubled using colchicine. The tetraploid individuals were crossed to obtain a segregating population (Figure 5.1). In this population, a total of two alleles are segregating for the S locus that are represented by letters A and B, and three alleles for the Z locus represented by numbers 1, 2, and 3. A total of 50 plants were used for this study, grown in the Iowa State University Agronomy greenhouses in Ames, IA, for 3 months. Three-month old plants were vernalized in a growth chamber for five weeks at 5C with a 10h light per day photoperiod. After vernalization, plants were grown until flowering at a 16h day length at 20C in the Iowa State University Horticulture greenhouses in Ames, IA. We used the progeny of this cross, retrospectively, to infer which model best describes SI in our autotetraploid population.
Figure 5.1 Scheme of autotetraploid population development using colchicine doubling. Alleles from S are assigned as letters and alleles for Z are assigned numbers.

The estimation of expected frequencies under the two possible models was performed, using the pollen compatibility software PollenCALC (Arias Aguirre et al., 2012). This software predicts pollen compatibility between two tetraploid genotypes. The software predicts genotypic classes, gamete frequencies, and progeny genotypes. Expected frequencies were only calculated for the AABB 2233 (female) x AABB 1122 (male) cross, because seed was only harvested on the AABB 2233 genotype. Even though there are two different alleles for Z in F1-39, they show a similar phenotype (Studer, unpublished data). Since S genotypes were identical in both parents, alleles for S were not expected to contribute to compatibility in the parental cross. Therefore, we concentrated our efforts on the segregating Z locus. A male with a Z-1122 produces three pollen genotypes but not all three genotypes are compatible to the female AABB 2233 (Table 5.1).

Table 5.1 Possible pollen genotypes from Z locus produced by male AABB1122 and compatibility to the AABB 2233 under each model

<table>
<thead>
<tr>
<th>Pollen genotype</th>
<th>Under Model I</th>
<th>Under Model II</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>12</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>22</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

We used PollenCALC to estimate the expected haplotype frequencies of Z in the gametes of both the male and female genotype under Model I and Model II (Table 5.2). Female gamete haplotypes for Z were not expected to differ under either model. Under Model I, the male AABB 1122 is 84% compatible with the female AABB 2233 under the assumption of normal
chromosome segregation. This means that only pollen grains with a Z-11 and Z-12 are expected to be compatible with the AABB 2233 female.

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Expected frequency</th>
<th>Genotype</th>
<th>Expected frequency</th>
<th>Genotype</th>
<th>Expected frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>0.16</td>
<td>12</td>
<td>0.672</td>
<td>11</td>
<td>0.168</td>
</tr>
<tr>
<td>23</td>
<td>0.68</td>
<td>11</td>
<td>0.168</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>0.16</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Pollen gametes carrying 22 would be incompatible under Model I  
** Pollen gametes carrying 12 or 22 would be incompatible under Model II

For Model II, the expectation is that only pollen carrying the Z-11 genotype are compatible with the AABB 2233 female, thus only 16.8% of the pollen from the AABB 1122 male is expected to pollinate the female.

Since it is difficult to genotype gametes, we relied on the progeny, where only compatible pollen genotypes were expected to contribute. For Model I, six genotypic classes are expected for the Z locus in the progeny resulting from two compatible pollen genotypes with three possible female gametic genotypes (Table 5.3). In contrast, under Model II, only pollen carrying a Z-11 genotype is compatible with the female and, therefore, only three genotypic classes for the Z locus are
expected in the progeny (Table 5.4). All calculations were performed under the assumption that all compatible pollen grains are viable and equally able to pollinate the female and that the female gamete is not inhibited after an incompatible reaction. Therefore, the frequencies of male gametes used in Table 2 and 3 were recalculated, weighing expected compatible pollen frequencies by the pollen compatibility under each model. For instance, for Model I the frequency of male pollen genotype Z-12 is 0.672/0.84 =0.8 (Table 5.2).

Table 5.3 Expected genotypic classes and frequencies for the Z locus tetraploid population progeny under Model I using weighed frequencies

<table>
<thead>
<tr>
<th>Male Gametes</th>
<th>Female Gametes</th>
<th>Progeny</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weighed Frequency</td>
<td>Genotype</td>
<td>Weighed Frequency</td>
</tr>
<tr>
<td><strong>0.8</strong></td>
<td>12</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.16</td>
</tr>
<tr>
<td><strong>0.2</strong></td>
<td>11</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>0.16</td>
</tr>
</tbody>
</table>

Table 5.4 Expected genotypic classes and frequencies for the Z locus tetraploid population progeny under Model II using weighed frequencies

<table>
<thead>
<tr>
<th>Male Gametes</th>
<th>Female Gametes</th>
<th>Progeny</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weighed Frequency</td>
<td>Genotype</td>
<td>Weighed Frequency</td>
</tr>
<tr>
<td><strong>1</strong></td>
<td>11</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>0.16</td>
</tr>
</tbody>
</table>

**Ploidy level analysis**

Plant tissue was collected and kept cool until chopping. Chopping was performed with a razor blade and LB01 buffer (Dolezel et al., 1989). The resulting solution was poured through a nylon mesh into 2 ml Eppendorf tubes and centrifuged at low speed for 3 minutes. Supernatant was
discarded and the pellet was resuspended using the LB01 lysis buffer. Finally, 25 µl propidium iodide stock was added to the solution, which was transferred to a flow cytometry tube. Fifty plants were analyzed by flow cytometry in the Flow Cytometry Facility at Iowa State University. A diploid and a tetraploid control were used to determine ploidy levels of the 50 plants.

**Phenotypic assays**

For 43 plants, flowering tillers were bagged with white Lawson bags for four weeks. Additional tillers were used for in vitro pollination tests in 10 plants to confirm self-incompatibility. Stigmas were collected on 2% agar containing 10% sucrose and 100 ppm Boric acid. Petri dishes containing stigma were incubated at 25C for 24 h. Stigmas were pollinated with pollen from the same plant (“self-pollinated”) and placed back in an incubator for 4 hours. Ovaries were cut using razor blade and placed on microscope slides. A drop of a staining solution, 0.2% aniline blue in 2% K$_3$PO$_4$, was placed over pollinated stigmas and covered with a cover slide. Pollinated stigmas were inspected under a Zeiss Axioplan 2 Imaging UV light microscope and images were analyzed with the Axiovision software (Zeiss, Germany) at the Microscope and Imaging Facility at Iowa State University.

**Z haplotype genotyping**

After confirming that plants in the population were self-incompatible, molecular analysis by high resolution melting (HRM) genotyping was performed (Studer et al., 2009). PCR reactions contained 1x LCGreen dye (Idaho Technology, Inc., UT, USA), 0.1 mM of dNTPs, 10 µM of each forward and reverse primer, 20 mM of magnesium chloride, and 20 ng of DNA. In addition, 15 µL of mineral oil was added to the PCR mix to avoid evaporation. PCR amplification was conducted in a PTC-200 Thermal Cycler (MJ Research Inc., Watertown, MA, USA) following
the protocol of Studer et al. (2009). A final cycle of 30 s at 94C and 30 s at 25C was added for heteroduplex formation. The LightScanner Instrument (96-well plate format) was used for HRM analysis and the LightScanner software Version 2 (Idaho Technology, Inc., Utah) was used for melting analysis. Melting temperatures ranged from 68-98C in an auto exposure setting. A total of three markers (Studer et al. 2012) were used to determine the segregation of Z in 43 tetraploid progeny as well as the male parent.

**Quantitative and Qualitative Discrimination of the two models**

The main difference between Model I and II is the number of Z genotypes expected in the progeny. Model I, which is a more permissive model would result in six genotypic classes in the progeny compared to the three expected under Model II. Unfortunately, the difference between those expected genotypes is not due to presence/absence of alleles, but rather their number. Under Model I the six expected genotypes are 1222, 1223, 1233, 1122, 1123 and 1133 (Table 5.3). From those genotypes only 1122, 1123 and 1133 are expected under Model II and at different frequencies (Table 5.4). HRM genotyping can be used as co-dominant marker system in diploid species (Studer et al., 2009), but it was not clear, whether different isoallelic autotetraploid genotypes such as 1122 and 1222 can be discriminated. To solve this problem, we added a 1:1 mixture of the diploid parents F1-30 and F1-39 with defined haplotypes for Z-12 and Z-23, respectively, which together were expected to represent one of the most frequent genotypic classes under Model I. We used a Z-12 genotype to represent the 1122 tetraploid parental class, which is expected to be the most frequent under Model II. Finally, we included a Z-13 genotype and all its pair-wise combinations to the two parental classes to have controls for all possible genotypes expected under Model I. We were able to represent all possible genotype classes, except for 1222. The purpose of these controls was to ask, whether genotypes expected under Model I but not Model II could be identified (qualitative analysis).
Independent of this qualitative analysis, we addressed the question, whether segregation among progeny genotypes was consistent with one of the two models. Once haplotypes were assigned to the corresponding expected genotypic classes, we used $X^2$ tests, to compare observed with expected genotypic frequencies. The analysis for Model I was done with five degrees of freedom and with two degrees of freedom for Model II.

RESULTS

Phenotypic assays and flow cytometry

Out of the 50 plants grown for this study, only 43 were confirmed to be tetraploids after flow cytometry. These 43 plants produced normal tillers and florets after vernalization. Florets and stigmas had a larger size compared to diploid plants. The average number of tillers and florets was not significantly different among plants (data not shown). None of the individuals produced seed after selfing, and no pollen tubes grew after using stigmas and self-pollen in in-vitro pollination tests.

Z haplotype genotyping

The three markers used for Z genotyping were polymorphic in our population. The 43 individuals showed different curves or marker genotypes under high resolution melting. No single marker was able to assign a genotype to the corresponding genotypic class. For instance, markers ZB1 and ZB2 classified all individuals into only two classes (Figures 5.2 & 5.3), whereas ZB3 discriminated three different genotypes (Figure 5.4). We found a total of four haplotypes for the Z locus among the 43 samples, which supports Model I (Table 5.5), if haplotype frequencies are not taken into consideration.
Figure 5.2 Shifted melting curves for primer ZB1 shows a clear segregation of the 43 genotypes into two categories

Figure 5.3 Shifted melting curves for primer ZB2 that similarly to ZB1 show a segregation of the population into two groups
Figure 5.4 Difference curves for primer ZB3 showed up to three different curves for the 43 genotypes in the population.

Table 5.5 Melting curves calls for markers ZB1, ZB2 and ZB3 and the respective haplotype designation

<table>
<thead>
<tr>
<th>Corresponding DNA control</th>
<th>call_ZB1</th>
<th>call_ZB2</th>
<th>call_ZB3</th>
<th>Number of individuals in this category</th>
</tr>
</thead>
<tbody>
<tr>
<td>1122</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>10</td>
</tr>
<tr>
<td>1133</td>
<td>A</td>
<td>A</td>
<td>D</td>
<td>8</td>
</tr>
<tr>
<td>1123</td>
<td>B</td>
<td>A</td>
<td>D</td>
<td>24</td>
</tr>
<tr>
<td>1223</td>
<td>B</td>
<td>B</td>
<td>D</td>
<td>1</td>
</tr>
<tr>
<td>1233</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>0</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td></td>
<td></td>
<td>43</td>
</tr>
</tbody>
</table>

Quantitative and Qualitative Discrimination of genotypic classes:

We performed a qualitative discrimination of the genotypic classes for which we used a 1:1 mix between the diploids F1-30 and F1-39 and a Z-13 sample and were able to assign all observations to corresponding genotypic classes in both models. We were able obtain DNA controls for all expected genotypic classes under Model II and all expected genotypic classes under Model I except for 1222. Using these controls, we were able to assign the 42 progeny to genotypic classes 1122, 1133, and 1123. We also found one individual in the progeny corresponding to the 1223 genotypic class. The other two genotypic classes expected under Model I were missing. Table 5.6
shows that observed segregation ratios differed significantly (p=0.001) from expected segregation under Model I, whereas the phenotypic data did not show a significant deviation from the expected frequencies under Model II.

Table 5.6 Chi Square test for hypothesis testing of Model I and Model II.

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Expected Frequency</th>
<th>Expected counts</th>
<th>Observed counts</th>
<th>Obs - Exp</th>
<th>Squared</th>
<th>Squared/expected</th>
</tr>
</thead>
<tbody>
<tr>
<td>1222</td>
<td>0.13</td>
<td>5.6</td>
<td>0</td>
<td>-5.6</td>
<td>31.36</td>
<td>5.6</td>
</tr>
<tr>
<td>1223</td>
<td>0.55</td>
<td>23.7</td>
<td>1</td>
<td>-22.7</td>
<td>515.29</td>
<td>21.7</td>
</tr>
<tr>
<td>1233</td>
<td>0.13</td>
<td>5.6</td>
<td>0</td>
<td>-5.6</td>
<td>31.36</td>
<td>5.6</td>
</tr>
<tr>
<td>1122</td>
<td>0.03</td>
<td>1.3</td>
<td>10</td>
<td>8.7</td>
<td>75.69</td>
<td>58.2</td>
</tr>
<tr>
<td>1123</td>
<td>0.13</td>
<td>5.6</td>
<td>24</td>
<td>18.4</td>
<td>338.56</td>
<td>60.5</td>
</tr>
<tr>
<td>1133</td>
<td>0.03</td>
<td>1.3</td>
<td>8</td>
<td>6.7</td>
<td>44.89</td>
<td>34.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>Expected Frequency</th>
<th>Expected counts</th>
<th>Observed counts</th>
<th>Obs- Exp</th>
<th>Squared</th>
<th>Squared/expected</th>
</tr>
</thead>
<tbody>
<tr>
<td>1122</td>
<td>0.16</td>
<td>6.72</td>
<td>11</td>
<td>4.28</td>
<td>18.32</td>
<td>2.73</td>
</tr>
<tr>
<td>1123</td>
<td>0.68</td>
<td>28.56</td>
<td>24</td>
<td>-4.56</td>
<td>20.79</td>
<td>0.73</td>
</tr>
<tr>
<td>1133</td>
<td>0.16</td>
<td>6.72</td>
<td>7</td>
<td>0.28</td>
<td>0.08</td>
<td>0.01</td>
</tr>
</tbody>
</table>

SUM 186.2
P Value <0.001

*Only 42 individuals were used in this test because the 43th has a genotype which was not consistent with any of genotypic classes expected with this model

DISCUSSION

The use of HRM-based markers is a rapid and sensitive essay for genotyping. The present study adds further proof to the data presented by Manzanares et al. (2011) and provides evidence that
HRM genotyping can be reliably performed in tetraploid ryegrass. It is important to remark that when, used in tetraploids, diploid DNA mixes are needed to assign melting curves to the corresponding genotypic classes. Once melting curves are assigned to specific genotypes, tetraploids themselves can be used as controls for genotypic classes.

Both the absence of seed upon self-pollination and the lack of pollen tube growth upon selfing in in vitro pollination test demonstrate that SI prevails at the tetraploid level in the population used in this study. The autotetraploid parents of the population used in this study are different in their genetic background from the materials used by Fearon and Haywards (1984). Our study provides thus additional evidence that SI is consistent across different induced tetraploid populations.

The male gametic frequency correction made by weighting the expected frequencies by the pollen compatibility helped to determine the expected genotypic frequencies more accurately. Two main conclusions can be drawn from our study: (1) the female is not losing receptivity after encountering an incompatible pollen grain, and (2) even minute amounts of compatible pollen are sufficient to pollinate virtually all female gametes.

In this study, the number of expected haplotypes found was misleading in terms of model testing and an additional step was required to determine the best model to explain the observed segregation. For example, the presence of four genotypic classes would be an indication of the validity of Model I, because under Model II only three classes are possible. However, to further confirm this hypothesis, discrimination between genotypic classes was needed. In autotetraploids the assignment of observed genotypes to expected genotypic classes is not as simple as in diploids. At this ploidy level, the presence of a single allele in more than one copy increases the uncertainty of assigning the correct genotypic classes. This problem is worse in case of distorted segregation, such as for markers linked to SI. The addition of an equal mixture of diploid parents DNA as well as a Z-13 genotype and all possible combinations between the three helped us to
assign observed genotypes to the corresponding genotypic classes. These DNA mixture previously were used in tetraploid potato to replicate five different biallelic combinations (De Koeyer, 2010). However, there are no reports of a successful implementation to distinguish genotypic classes with different combinations of more than two alleles. The present study provides evidence of the possibility of discriminating among genotypic classes and suggests that HRM can be used for genotyping polyploids efficiently. We were not able to discriminate genotypes 1222 and 1122, because we did not have respective diploid DNA to reconstruct genotype 1222. However, we were able to discriminate other isoallelic genotypes, such as 1223 and 1233. Even though both genotypes can form the same type of duplexes after re-annealing, the frequency of those duplexes is different between 1223 and 1233. Therefore, they are expected to show different melting curves. DNA mixtures for genotypic classes 1223 and 1233 were clearly discriminated by one of the markers (Figure 5.4), where individual 1223 was represented by a blue line, whereas the 1233 DNA control falls into the grey category. Therefore, we are confident that HRM would have also discriminated between the 1122 and 1222 genotype classes. Absence of a sixth distinct genotype from the five genotypes that we were able to reconstruct within the segregating autotetraploid population was interpreted as evidence for absence of the 1222 genotype.

Our data support Model II in which a combination of at least a single allele at S and Z in the pollen grain that matches the pistil can make pollen incompatible. Moreover, we provided direct evidence that the number of alleles in pollen does not affect the efficiency of the self-incompatibility system. The one deviating genotype in this population (1223) might have occurred by pollen contamination. Alternatively, it may be due to a leaky SI, which allows germination of incompatible pollen and further fertilization at low frequency. In tetraploids, chromatid segregation might be another possibility for occurrence of unexpected marker
genotypes. However, in the case of this specific population chromatid segregation would not result in any different pollen haplotype since all possible pollen genotypes haplotypes produced by chromatid segregation are already present under normal chromosome segregation (Z-11, and Z-22). Therefore, the only difference would have been in the frequency and not the genotypes. Another explanation may be aneuploidy in our population. However, we discard this possibility due to screening of those 43 genotypes using flow cytometry. With all this information, the most possible explanation is that an event occur between the marker and the Z allele which caused the appearance of the unexpected genotype in the progeny.

Model I is more restrictive than Model II, from an evolutionary perspective, Model II in which a single allele differing at either S or Z cause pollen compatibility would mean more pollen compatibility in the longer run. Therefore more inbreeding would be possible. Out of the four alleles in the pollen (two at S and two at Z) only one needs to be different from the female to be make pollen compatible under Model II. Instead, under Model I, pollen grains need to have one different allele at both loci in order to be compatible. Therefore, Model I ensures outcrossing and is more successful at minimizing inbreeding depression (Arias et al., 2011).

In the series of papers published in the 1980s (Fearon et al., 1984a; Fearon et al., 1984b; Fearon et al., 1984c), the authors demonstrated the validity of the S-Z combination model using pollination tests (which we call Model I). In the first manuscript of this series, the authors used pollen from a diploid plant to pollinate a related autotetraploid. In a following manuscript they used pollen from a tetraploid plant to pollinate a related diploid. In both manuscripts they used in vitro pollination tests to make inferences about parental genotypes and pollen compatibility. There are several biases introduced during in vitro pollination tests, one is the mixture of living and dead pollen, another is the time of pollen collection and finally the possibility of an unknown lethality rate due to the staining process. In addition, there is the difficulty of counting successful
pollinations in pollen grains that may be in clusters. In our research we used progeny to retrospectively make inferences about the parents by assessing the inheritance of polymorphic markers linked to the Z region. Another difference of the present study is that we used gametes from tetraploid plants to assess compatibility and we were, therefore, able to rule out dosage effects and dominance of the alleles present in the population. Although we confirm a similar model as the one described by Fearon (1984), we avoided the confounding effects of fertilizing gametes that come from plants with different ploidy levels. We were thus able to show that no dominance effects occur in male or female gametes.

Tetraploid ryegrass cultivars are part of the current germplasm pool and the understanding offered by studies in the 1980s was of great importance for cultivar development at this ploidy level. In this study, we were able to provide further supporting evidence of the S-Z model for self-incompatibility in autotetraploid ryegrass by use of linked markers. Genotyping of both SI loci is of great interest for grass breeding programs.

In order to use SI for hybrid seed production in tetraploid ryegrass, it is important to maintain sufficient diversity at S and Z to maximize seed production in a cross of two genotypes (Arias Aguirre et al., 2012). Alleles segregating for S and Z need to be as different as possible among heterotic groups to maximize seed set. HRM-based genotyping combined with the bioinformatics tools such as PollenCALC can be used for parent selection and to create or improve existing heterotic pools based on S-Z allele differences to maximize seed set in hybrid seed production programs.

REFERENCES


CHAPTER 6: General Conclusions

The objectives of this research were to analyze biological and molecular mechanisms in bioenergy grasses that can be used for establishing hybrid breeding schemes for these crops. In addition, a more refined objective was to use ryegrass to map a new source of self-fertility in ryegrass and to validate the SI model and effectiveness in tetraploid ryegrass which included the development of software for pollen compatibility calculation in tetraploids.

In an exhaustive review, different strategies for hybrid breeding were discussed (Arias Aguirre et al., 2011). The feasibility of moving towards hybrid seed production relies on the balance between seed cost for the farmer and revenue from higher performing varieties. Therefore, the primary effort needs to be in the development of heterotic pools to maximize hybrid performance for the target traits. If successful, it may be possible to offer the farmer hybrid seed that is worth the initial investment.

Nevertheless, there are other prerequisites, addressed in this thesis. The need for controlled pollination was our main focus. In polyploid grasses with S-Z SI system, the estimation of pollen compatibility can be laborious and time consuming. One of the objectives of this research was to
automate pollen compatibility calculations for allo- and autotetraploids. PollenCALC is software enabling to calculate gamete haplotypes and pollen compatibility for tetraploid species with a gametophytic S-Z SI. We discussed different uses for this tool such as calculation of S-Z haplotypic frequencies in pollen, model testing, and optimization of experimental designs (Arias Aguirre et al., 2012). However, this tool also can be used in breeding programs for the development of heterotic groups. For instance, PollenCALC can be used to determine the necessary amount of S-Z combinations needed within a group to ensure seed set and also to calculate how many different alleles and frequencies are required in the other pool in order to use SI for hybrid seed production.

To address line development when using SI as hybrid mechanism, a new source of SF was analyzed based on markers in candidate linkage groups and linked markers for the S-Z regions. For this study, we found a strong association between markers in linkage group 5 and SF. Even though we were not able to find a strong association of SF to markers in linkage group 3, we found a strong segregation distortion for markers in this linkage group, which are also in LD with markers in LG5. We were able to exclude associations between the S-Z haplotypes present in our population and SF. Finally, we observed a strong inbreeding depression and lethality rate upon selfing. The latter two issues are of great concern for hybrid breeding. This raises the question of feasibility of using SF for line development, because additional efforts will be required to develop strong inbreds with substantial seed set. Another great disadvantage is that SF genes seem to have a dominant mode of action. Thus, ultimately only $\frac{1}{4}$ of the progeny will be inbred and SI. This is why it is reasonable to explore other mechanism such as pseudo-compatibility, which would allow SI to be turned off by an environmental stimulus and turn back on when needed for hybrid seed production.
Finally, the validity of the proposed SI model for tetraploids was analyzed using HRM-markers. We analyzed and discussed the use of HRM to discriminate between expected genotypic classes in tetraploids using diploid DNA mixtures. We were successful in identifying and differentiating genotypes with similar alleles in different combinations (isoallelic). The model proposed in the 1980s was validated but new outcomes such as the use of molecular markers instead of phenotypic assays to assess SI was discussed. The use of marker technology enabled us to confirm genotypes early in the plant cycle compared to the use of phenotypic essays which require a 5 week period of vernalization and a 4-week period for flowering.

Overall, different aspects for hybrid seed production in bioenergy grasses have been addressed. We identified primary needs in terms of line and heterotic pool development. Bioenergy grasses are in their infancy in terms of hybrid breeding. Finally, in the aim of identifying mechanisms to control self-pollination, we were able to identify a self-fertility source in ryegrass and also validate the SI model in an autotetraploid population of ryegrass. Both studies are a contribution in order to move towards hybrid breeding for perennial grasses.

REFERENCES
