Genetic assessment of plant material for studies of the genetic causes to female biased sex ratios in *Salix viminalis* L.

Amela Kujović

Genetic assessment of plant material for studies of the genetic causes to female biased sex ratios in *Salix viminalis* L.

Amela Kujović

Supervisor:	Sofia Berlin Kolm, Department of Plant Biology, SLU
Assistant Supervisor:	Ann-Christin Rönnberg Wästljung, Department of Plant Biology, SLU
Examiner:	Anna Westerbergh, Department of Plant Biology, SLU

Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

Faculty of Natural Resources and Agricultural Sciences
Department of Plant Biology

Abstract

Biased sex ratios are highly common among dioecious plants. *Salix viminalis* is a dioecious tree or a woody shrub that has an overall female biased sex ratio. Most likely the bias is initiated in an early stage of development either by abortion of certain male zygotes or by genetic incompatibility leading to failure of fusion between the sperm and the egg. *Salix viminalis* has a female heterogametic or ZW sexual system, where the W represents a female-specific chromosome or region, meaning that all female offspring has one maternally inherited haplotype.

In this work, I designed primers for markers located in the sex determination region on chromosome 15. By changing different parameters of PCR and by analysing the sequences of all the individuals, the expected genotypes of the used individuals were confirmed. Thus, in the further studies the markers can be applied in offspring and their genotypes can be determined, in order to estimate the sex ratio.

Moreover, a germination test of different crosses of two populations of *S. viminalis* was done. One population had even sex ratio and other one female biased sex ratio. By comparing the germination rates between the crosses, the effect of inbreeding on germination rates in each cross was studied. Furthermore, based on the germination test the relatedness between homozygosity and white seeds phenomena was studied, since we assume that white seeds have homozygous genotype. In addition, the data which included germination rates was analysed in JMP program using ANOVA test.

The results indicate that the reasons that some PCR programs in this study gave better results using the same primers than the others may be formation of primer dimers, secondary structures and that the sex determination region has many repetitions.

The results show that there is significant difference between the germination rates between the crosses and between the catkins of each cross. Based on assumptions, in this work it is suggested that white seeds may not present homozygous individuals. Moreover, homozygosity may be related also to the black seeds or the homozygous seeds might not be even fully formed. Furthermore, it is possible that genetic incompatibility between homozygous allelic combinations have led to the differences among the germination rates of the crosses. In the further studies of the white seeds phenomena, sequencing of the genomes of the seeds is required in order to investigate the relatedness between homozygosity and white seeds.

Keywords: sex ratio, willows, dioecy, white seeds, homozygosity, inbreeding

Table of Contents

1. Introduction

Background

Salix viminalis is a shrub or a tree, which grows up to 3 meters. Seeds are formed in female clusters of flowers or catkins and are dispersed by wind. The species can propagate both vegetatively and sexually (Pucholt et al., 2016, in prep).

Willows in the *Salix* genus belong together with poplars to the *Populus* genus to the family Salicaceae and species in both genera are dioecious. Dioecy is a phenomenon where there are male catkins on one individual and female catkins on another. It is rare among flowering plants, since only 5 % of plants are dioecious (Renner, 2014) and it is a phenomenon that prevents self-pollination (Hardy et al., 2002).

Many organisms have sex chromosomes that harbour sex determination regions. Sex chromosomes can be heteromorphic (different) as in mammals or homomorphic (similar) as in willows (Pucholt et al 2015). *S. viminalis* has a sex determination locus located on chromosome 15 and the position has been determined by quantitative trait locus mapping (Pucholt et al., 2015).

Sex determination systems vary among dioecious plants, because dioecy has evolved many times from hermaphroditic ancestors (Charlesworth, 2002). *S. viminalis* has a female heterogametic system or ZW, where W represents a female-specific chromosome or region, meaning that all female offspring has one maternally inherited haplotype (Pucholt et al., 2015). In addition to female heterogamety, there is also male heterogamety XY, which is the most common sex determination system among plants (Ming et al., 2011).

S. viminalis has an overall female biased sex ratio that is likely caused by some unknown early acting genetic mechanism (Altröm-Rapaport et al., 1997). Possible genetic mechanisms can be some form of genetic incompatibility, resulting in failure of fusion of the egg and sperm or mortality of male zygotes with certain genotypes (Pucholt et l., 2016, in prep).

Genetic markers and haplotypes

A genotype is a set of alleles or forms of a gene for a specific trait (Klug et al., 2010). Since a study of a genotype of one individual cannot give any information about the position of a specific allele on a chromosome, a study of haplotypes is needed. A haplotype is a group of alleles that are inherited together and transferred from one parent to offspring.

Using specific methods, haplotypes can be determined, as it was previously done for the individuals that were used in this work (Pucholt et al., 2016, in prep). Genotyping-by-sequencing or GBS was used for determining the markers positions (Pucholt et al., 2016, in prep).

GBS is a method, used for creating a genomic DNA library, in order to find varieties between samples that are used. Firstly, by using restriction enzyme (for example ApeKI), DNA samples are digested. At restriction place of each sample specific barcode adapters are

ligated, in order to mark each sample. Moreover, PCR is performed, the products are cleaned up and Illumina sequencing is done. After that, an alignment of samples with reference genome is done and varieties between each sample are detected (Elshire et al., 2011).

The varieties between DNA sequences or genetic markers are used to discriminate between species or individuals of one species. The closer the marker is to the gene, in which we are interested, the less possibilities are there for a recombination between the gene and the marker. Thus, that kind of marker is reliable and can be used for analysis. Using markers, it is possible to discriminate between homozygous and heterozygous individuals within species (Pucholt, n.d.).

Single nucleotide polymorphism (SNP) is a polymorphism at a single base position and has typically two allelic variants among individuals or between individuals within species of a population (Klug et al., 2010). The analysis of SNPs requires two steps, discovery and detection. In this work I did a part of detection of SNPs in the crossing experiment of a specific population of *Salix viminalis*.

Moreover, in the previous study, maternally and paternally segregating markers were used in order to determine haplotypes in the S5 *Salix viminalis* population (Pucholt 2016, in prep). Maternally segregating markers are markers that are heterozygous in the mother and homozygous in the father. Paternally segregating markers are markers that are heterozygous in the father and homozygous in the mother (Pucholt 2016, in prep).

Coefficient of relatedness

Inbreeding is mating between individuals that are closely related. It is a common phenomenon in small populations as in the populations that were used in this work. The consequence of inbreeding is increased proportion of homozygotes in a population. The strength of inbreeding between individuals can be estimated by the coefficient of relatedness, which measures the expected percentage of homozygosity in the system. The value of the coefficient of relatedness is between 0 and 1. If the value is 1, alleles are identical and homozygous at every locus, which means that they are heavily inbred. Population with that value is inbreeding population. In that case, expression of recessive alleles is high (Klug et al., 2010).

Aims

In this work, crossings between individuals of two populations of *S.viminalis* (S3 and S5) were done, in order to produce populations with specific genotypes and to compare the germination rates between the crosses. In addition, I have designed haplotype specific markers that will be used to determine the haplotypes of the produced populations. Furthermore, the produced populations and the haplotype specific markers will be used in further studies about sex ratios in *S.viminalis*.

The individuals that were used in this work have different genotypes in the sex determination and the linked sex ratio locus (Table 1). The genotypes of the individuals were already determined in the previous study (Pucholt et al., 2016, in prep).

Table 1. Genotypes of the individuals used in the study

Origin	Individual	Sex	Genotype
Female parent to S3	78195	F	W-c/Z4-d
Female parent to S5	78021	F	W-c/Z3-a
Male parent to S3 & S5	81084	M	Z2-b/Z1-a
S5	031	F	W-c/Z2-b
S5	228	F	W-c/Z1-a
S5	221	M	Z2-b/Z3-a

Alleles at the sex determination locus are denoted W, Z1, Z2, Z3 or Z4 whereas alleles at the sex ratio locus are denoted a, b, c or d (Pucholt et al., 2016, in prep). It is hypothesised that homozygosity at the sex ratio locus is lethal in males. For example, males with the allelic combination Z1-a/Z3-a were missing from the population, which led to the observed female biased sex ratio. The reason may be genetic incompatibility in the sex ratio distorter locus in the male plants with the a/a combination (Pucholt et al., 2016, in prep).

In this work, I have produced new crosses with expected genotypes, which are presented in the Table 2. Based on the genotypes of the offspring different outcomes are expected. If for example, homozygosity in the sex ratio locus is a lethal male combination, S7, S8, S9 and S10 would have female biased sex ratios.

Table 2. Expected genotypes of the offspring. Genotypes with possibly lethal allelic combinations are underlined. The S6 cross might have lethal allelic combination, since it has Z4 haplotype, which might act differently (Pucholt et al., 2016, in prep).

Origin	Cross	Genotypes
S6	221x78195	Z2-b/W-c, Z2-b/Z4-d, Z3-a/W-c, Z3-a/Z4-d
S7	221x78021	Z2-b/W-c, Z2-b/Z3-a, Z3-a/W-c, Z3-a/Z3-a
S8	81084x031	Z2-b/W-c, Z2-b/Z2-b, Z1-a/W-c, Z1-a/Z2-b
S9	221x228	Z2-b/W-c, Z2-b/Z1-a, Z3-a/W-c, Z3-a/Z1-a
S10	81084x228	Z2-b/W-c, Z1-a/Z2-b, Z1-a/W-c Z1-a/Z1-a

Furthermore, I did the germination test for all the crosses in order to compare the germination rates between the crosses and to estimate the effect of inbreeding on germination rates. The crosses that had genotypes with possibly lethal allelic combinations were expected to have lower germination rates.

In addition, the species has two types of seeds, white and black ones. White seeds are light, smaller than black seeds and it is possible that a part of them that is missing, which might be endosperm. To know what exactly is missing in white seeds, further studies are required. In germination test, I have counted seeds of four groups, black non-germinated, white non-germinated, black germinated and white germinated, in order to observe if white seeds germinate or not and to calculate the percentage of white seeds within each cross. Since there might be something that is missing in white seeds and therefore none of them have germinated, I have assumed that white seeds represent the individuals with genotypes with possibly lethal allelic combinations.

2. Materials

In this experiment individuals of two populations of *S. viminalis* were used. The population S5 represents a cross between 78021 (mother) and 81084 (father) and includes offspring 228 (female), 031 (female) and 221 (male). It was planted in 2004 in a sandy soil field in Pustnäs south of Uppsala. This population has female biased sex ratio, which was confirmed in the previous study (Pucholt et al., 2016, in prep).

The population S3, which has even sex ratio (Pucholt et al., 2016, in prep), is a cross between the two individuals 78195 (mother) and 81084 (father). The selection of the parents of the population was done in 2005 (Höglund et al., 2005) and the population was studied for the first time in 2010 (Berlin et al., 2010). In this work, only parents of S3 population were included.

The shoots were collected in January 2016 and stored in a cold room at – 4 degrees for 2-3 days, until they were used for the experiment.

3. Methodology

Crossings

The greenhouse experiment was set up after the harvest of shoots of each individual. The stem cuttings of each individual were put in big flasks with tap water. There were six flasks with each of the individuals: 81084, 78195, 221, 228, 031 and 78021. After development of male flowers, pollen of each male individual was collected into glass tubes with sticks with cotton wool and stored at 4°C. After development of female flowers, the pollen was transferred to receptive parts with cotton sticks and pollinated flowers were labelled. In two to three weeks, seeds were developed and collected into labelled plastic bags before they started to spread. Plastic bags were stored at room temperature until most of the seeds were out of each catkin. The seeds were separated from catkin parts by bag shaking and stored in paper bags at -20°C.

DNA extraction

About 100g grams of young leaves of each individual were collected into microcentrifuge tubes and frozen in liquid nitrogen. Genomic DNA was extracted immediately after collecting using the manufacturer´s protocol (DNeasy Plant Mini Kit by Qiagen).

After extraction of DNA, the concentration of DNA was measured using Nanodrop 1000 Spectrophotometer and samples were stored it at -20°C.

Table 3. Measured concentrations of DNA of each individual

Individual	Concentration of DNA [ng/µL]
78195*	64.6; 93.7

78021	61.3
81084*	31.6; 146.7
031*	113; 66.1
228*	38.8; 121
221	90.3

*Concentration of DNA is measured twice in two separate samples of the individual

Primer design

In this work, four groups of markers were used. All of them were already used in the previous study and they represent the group of markers that are most related either to sex or missing of male offspring (Pucholt et al., 2016, in prep).

Table 4. Markers for each haplotype used in the study

W	Z1	Z2	Z3
S5-28686	S5-95526	S5-57526	S5-73064
S5-139809	S5-88390	S5-149560	S5-141738
S5-121281	S5-116921	S5-96142	S5-128227
S5-134746	S5-2147	S5-60574	S5-64221
S5-81006		S5-5355	S5-142200
		S5-77183	

Sequences of the markers were blasted against the *Salix purpurea* genome at Phytozome (https://phytozome.jgi.doe.gov/pz/portal.html#), in order to get positions of extended sequence in the genome (around 1000 base pairs (bp) long sequence with the sequence of the marker in the middle). The positions were used in JBrowse (Phytozome) to get extended sequences in the *Salix purpurea* genome. After that, sequences of the markers were blasted to the extended sequences. That is how the positions of markers in the genome were found. Using Primer3 (http://primer3.ut.ee), the primer pairs were designed. The information that was needed for this step is the start position of a marker, length of a primer (64 bp for all of them) and product size range (300 to 500 bp for all of them). Based on the position of primers (the marker should be in the middle between primers) and values that show the possibility of formation of secondary structures, primers were designed.

Table 5. The names of primers that were used in PCR

Number of primer pair	Left primer	Right primer	Target haplotype
1	S5-28686-ext-L	S5-28686-ext-R	W
2	S5-139809-ext-L	S5-139809-ext-R	W
3	S5-73064-ext-L	S5-73064-ext-R	Z3
4	S5-121281-ext-L	S5-121281-ext-R	W
5*	S5-95526-ext-L	S5-95526-ext-R	Z1
6	S5-57526-ext-L	S5-57526-ext-R	Z2
7*	S5-149560-ext-L	S5-149560-ext-R	Z2
8	S5-141738-ext-L	S5-141738-ext-R	Z3

9	S5-128227-ext-L	S5-128227-ext-R	Z3
10	S5-134746-ext-L	S5-134746-ext-R	W
11	S5-81006-ext-L	S5-81006-ext-R	W
12	S5-64221-ext-L	S5-64221-ext-R	Z3
13	S5-142200-ext-L	S5-142200-ext-R	Z3
14	S5-88390-ext-L	S5-88390-ext-R	Z1
15	S5-116921-ext-L	S5-116921-ext-R	Z1
16	S5-2147-ext-L	S5-2147-ext-R	Z1
17	S5-96142-ext-L	S5-96142-ext-R	Z2
18	S5-60574-ext-L	S5-60574-ext-R	Z2
19	S5-5355-ext-L	S5-5355-ext-R	Z2
20	S5-77183-ext-L	S5-77183-ext-R	Z2

*primers were excluded (3/3/2016)

PCR

PCR was done several times by changing different parameters such as the concentration of DNA, number of cycles in PCR and annealing temperature. Programs used in the experiment are shown in the table.

**Table 6.** PCR programs with different parameters

PCR parameter	PCR program										
	1	2	3	4	5	6	7	8	9	10	11
Number of cycles	20	30	30	20	30	30	30	30	30	30	30
Annealing temperature	55	55	55	55	54	55.4	57	58	58	60	60
Duration of elongation	30	30	50	50	30	30	50	50	30	30	50
dH20	18.25	18.75	18.75	22.3	22.3	22.3	22.3	22.3	22.3	22.3	22.3
10x PCR Gold Buffer (Roche)	2.5	2.5	2.5	3	3	3	3	3	3	3	3
dNTP(Thermo Scientific)	0.5	0.5	0.5	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6
MgCl2 solution (Roche)	1.5	1.5	1.5	1.8	1.8	1.8	1.8	1.8	1.8	1.8	1.8
Left primer	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Right primer	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	05	0.5	0.5
Template DNA	1	0.5	0.5	1	1	1	1	1	1	1	1
AmpliTaqGold Polymerase (Roche)	0.25	0.25	0.25	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
Volume of a sample	25	25	25	30	30	30	30	30	30	30	30

The PCR products were run on agarose gel electrophoresis. The ones that had shown clear band were sent for sequencing (including product cleaning) to Macrogen. Using SeqMan Pro, chromatograms were inspected and SNP genotypes were then confirmed using the SeqMan program in Lasergene (DNASTAR Inc., Madison, WI, USA). By identifying SNPs,

the used types of markers were confirmed, showing that they can be used also in further studies of sex determination in produced offspring. Thus, more detailed studies of biased sex ratio in *S. viminalis* can be done.
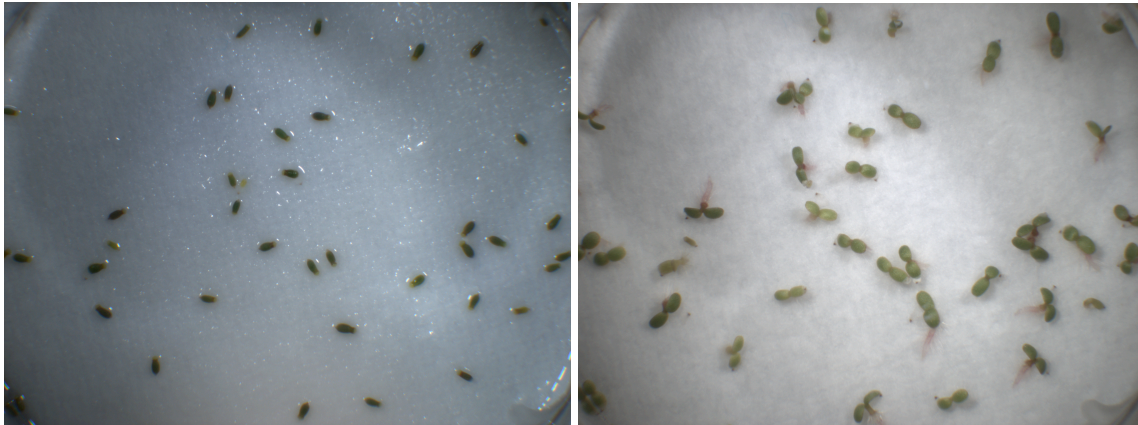
Germination test

Bags from the freezer were taken out two weeks before the start of germination test. Each cross had ten bags and each bag represented one catkin. Bags were labelled with the numbers. The experiment was set up in six-well plates. I scattered the seeds of each bag on a paper and tapped two layers of well-watered filter paper on them. Seeds have stuck on a filter paper and each bag or catkin was represented by three random samples, which were marked with letters as shown in the scheme below.
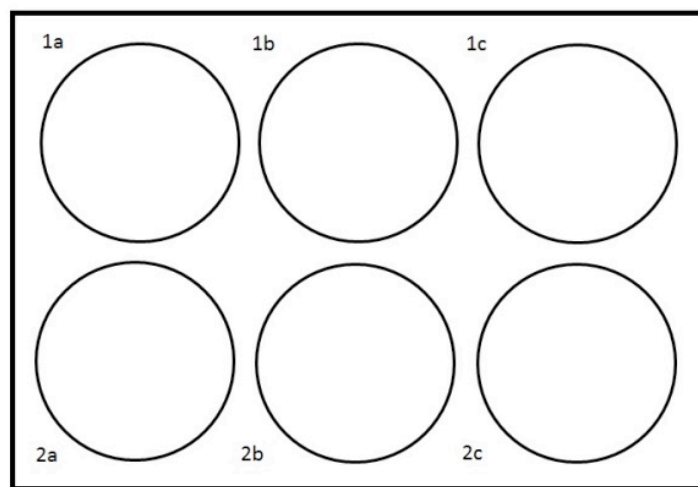
Fig 1. The scheme represents a six-well plate used for the germination test

Immediately after the preparation of a plate, the pictures were taken with microscope or camera. Seeds were put in a light room (16 hours of light, temp. light: 22°C, temp. dark: 20°C, light intensity: 110) observed for three days and watered to preserve seeds from drying out. The third day, pictures were taken again and compared to previous ones, in order to count germinated and non-germinated seeds.

After the all pictures were taken, I used Fiji (a version of ImageJ) for the analyses. There are two types of seeds, black and white. Both types were counted separately; black non-germinated, white non-germinated, black germinated and white germinated, by overlaying each picture taken after three days with a one from a first day.

Fig 2. The germination test for the 031x81084 cross (1a well). The left picture represents the first day of the germination test and the right picture the third day of the germination test.

Coefficient of relatedness

For calculating the coefficient of relatedness between offspring of each cross, I used pedigree path analysis. First, I drew a scheme of relatedness as it is presented below for the cross between 228 and 221.
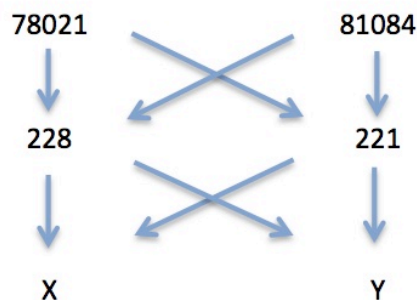
Fig 3. The scheme represents the cross between 228 and 221. The parents of both individuals are 78021 and 81084. The X and Y represent two random individuals of offspring of the cross.

In order to calculate the coefficient of relatedness in the offspring population, it is necessary to follow the formula below. The F is a coefficient of relatedness and N is a number of "arrows" in the path.

$$F = \left(\frac{1}{2}\right)^{N}$$

Thus, for the cross between 228 and 221, the calculation was done as it is shown below.

$$F = \left(\frac{1}{2}\right)^{2} + \left(\frac{1}{2}\right)^{2} + \left(\frac{1}{2}\right)^{4} + \left(\frac{1}{2}\right)^{6} + \left(\frac{1}{2}\right)^{3} - \left(\frac{1}{2}\right)^{4} = 0.75$$

$\frac{1}{2}^{2}$ - the path (X - 228 - Y) has two arrows

$\frac{1}{2}^{2}$ - the path (X - 221 - Y) has two arrows

$\frac{1}{2}^{4}$ - the path (X - 228 - 78021 - 221 - Y) has four arrows

$\frac{1}{2}^{4}$ - the path (X - 228 - 81084 - 221 - Y) has four arrows

$\frac{1}{2}^{4}$ - the path (X - 221 - 78021 - 228 - Y) has four arrows

$\frac{1}{2}^{4}$ - the path (X - 221 - 81084 - 228 - Y) has four arrows

Statistics

All the data from germination test was put in Microsoft Excel in a table, which had specific form, in order to be used in further analysis in JMP program.

Table 7. An example of a table that was used in JMP program.

			B	W	BG	WG	TG	NS	GR
S9	1	a	16	4	0	0	0	20	0
	1	b	16	4	2	0	2	22	0,090909091
	1	c	13	7	0	0	0	20	0
	2	a	19	2	0	0	0	21	0
	2	b	15	2	1	0	1	18	0,055555556
	2	c	14	3	1	0	1	18	0,055555556

The first column of the table 7 represents the name of a cross, second is a catkin (a bag) and the third is a replicate within a catkin. Furthermore, the numbers of black non-germinated (B), white non-germinated (W), black germinated (BG) and white germinated seeds (WG) were used in further calculations. Germination rate is a quotient between the number of total germinated seeds (TG) and the number of all seeds (NS).

Using JMP program, analysis of variance (ANOVA test) was done, in order to compare variation between the germination rates of all the crosses. In addition, the same was done for the variation between the germination rates of the catkins within each cross.

4. Results

Primer design and PCR

For all the individuals (78021, 78195, 81084, 221, 228, 031), I tested the primers that were created based on markers. As a result, the pictures of gel electrophoresis of only one individual are shown below, in order to compare the effects of each changed PCR parameter. The individual that I chose is 81084, since it is the father of both populations and it was used in all the programs.
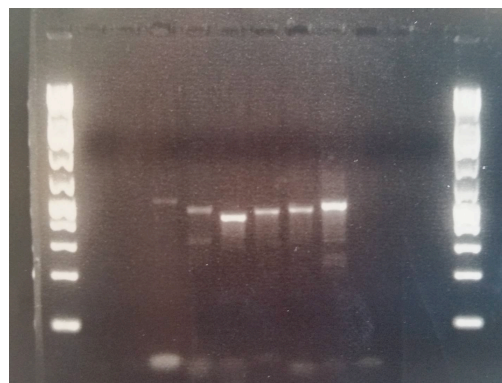

Fig 4. PCR program 1 used in 81084 individual with primers 1-7 (21.1.2016)
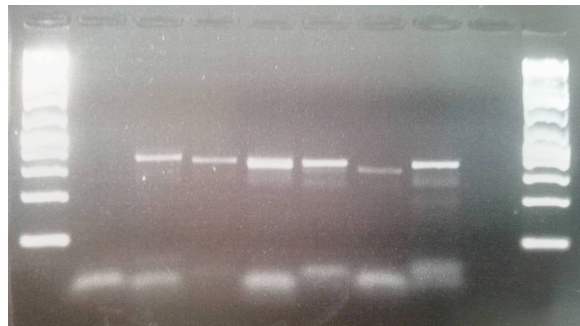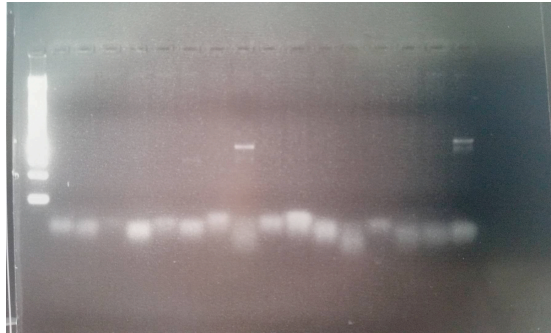

Fig 5. PCR program 2 used in 81084 individual with primers 1-7 (20.1.2016)

By comparing figures 4 and 5, I got more double bands in reactions 4, 5 and 7 by using program 2. Moreover, the product of the first primer pair is missing in the second figure. The bands of the second and third product are thicker in the figure 1. The band of the sixth product is thicker in figure 1 than in figure 2. Three products of program 1 (third, fourth and sixth) were sent for sequencing. There were no good products of program 2. The last columns of both pictures represent negative controls (no primers added).
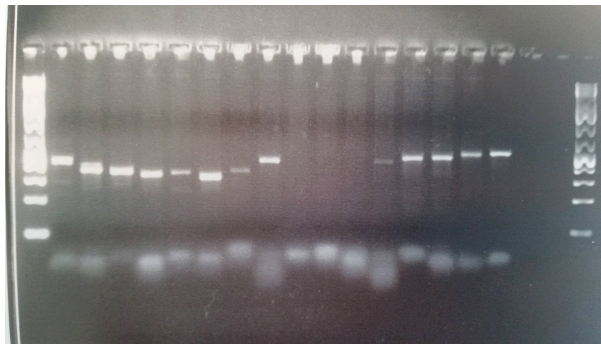

Fig 6. PCR program 3 used in 81084 individual with primers 8-17 (4.2.2016)

Using program 3, I got six products out of 10 possible. Two of them were weak and six of them were good. One of the good products was sent for sequencing. The fourth product has possible second band, which is very weak. The last column represents a negative control.

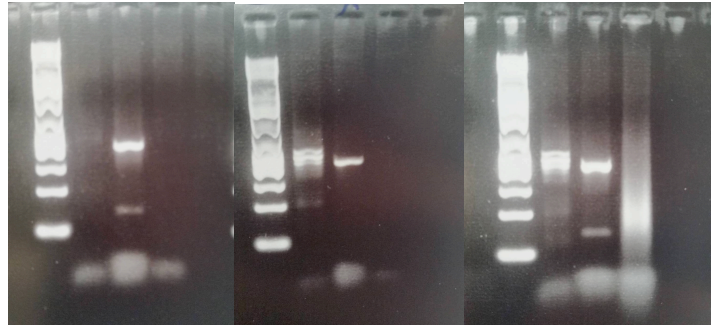**Fig 7.** PCR program 4 used in 81084 individual with primers 1-17

Fig 8. PCR program 7 used in 81084 individual with primers 1-17 (12.2.2016)

Fig 9. PCR program 8 used in 81084 individual with primers 1-17 (19.2.2016)

In PCR reactions of programs 4, 7 and 8 the same primers were used. In the program 4, none of the products was good. There were 14 empty columns and 3 bad products. In the program 7, there were 4 empty columns, 3 bad products with weak bands, 3 bad products with double bands and 7 good products. Five of latter were sent for sequencing. In program 8, there were 7 empty columns, 5 bad products with weak bands, 2 bad products with double bands and 3 good products that were sent for sequencing. The last columns of all of the figures represent negative controls.
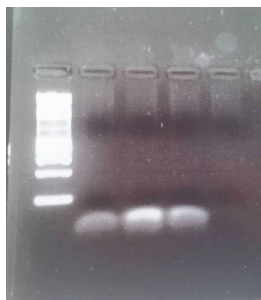
Fig 10. PCR programs 5, 6 and 9 used in 81084 individual with primers 18-20 (6.4.2016)

I used three sets of primers (18, 19 and 20) in the programs 5, 6 and 9. Using program 5, I got 2 empty columns and one bad product with double bands. In the program 6, I got one bad product with multiple bands, one empty column and one good product that was sent for sequencing. In the program 9, all of the products were bad with multiple bands. The last columns of all figures represent negative controls.
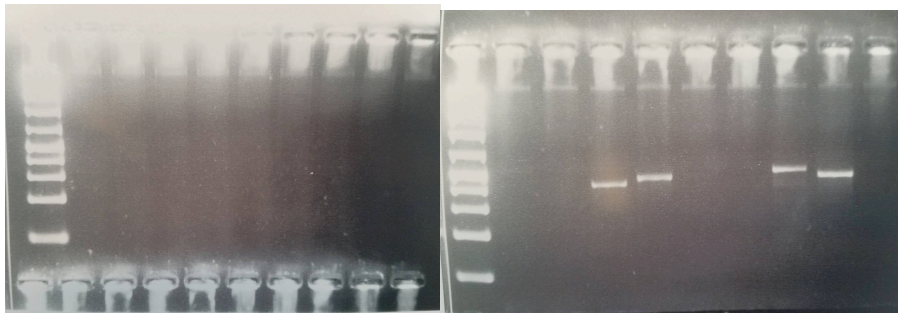
Fig 11. PCR program 10 used in 81084 individual with primers 1-17 (26.2.2016)

Using program 10, I got 4 products out of 17 possible. All of the products were good and one of them was sent for sequencing. The last column in the second figure represents a negative control.

Fig 12. PCR program 11 used in 81084 individual with primers 18-20 (5.4.2016)

In program 11, I did not get any out of three possible products. The last column represents negative control.

PCR program

Primer pair	1	2	3	4	5	6	7	8	9	10	11
1											
2											
3											
4											
5											
6											
7											
8											
9											
10											
11											
12											
13											
14											
15											
16											
17											
18											
19											
20											
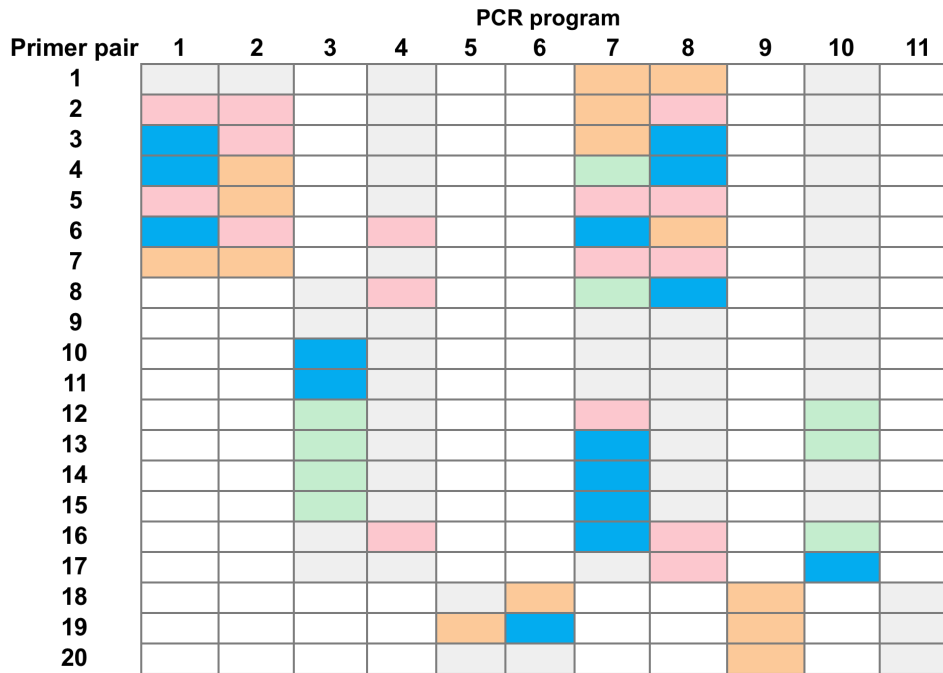
Fig 13. Quality of the PCR products; empty products (grey), bad products with weak bands (red), bad products with double bands (orange), good products with clean bands (green), good products with clean strong bands that were sent for sequencing (blue).

According to the Figure 13, the least successful programs (with the lowest number of good PCR products) were programs 2, 4, 5, 9, 10 and 11. The difference between these programs was in the volume of template DNA, annealing temperature, elongation time and number of cycles.
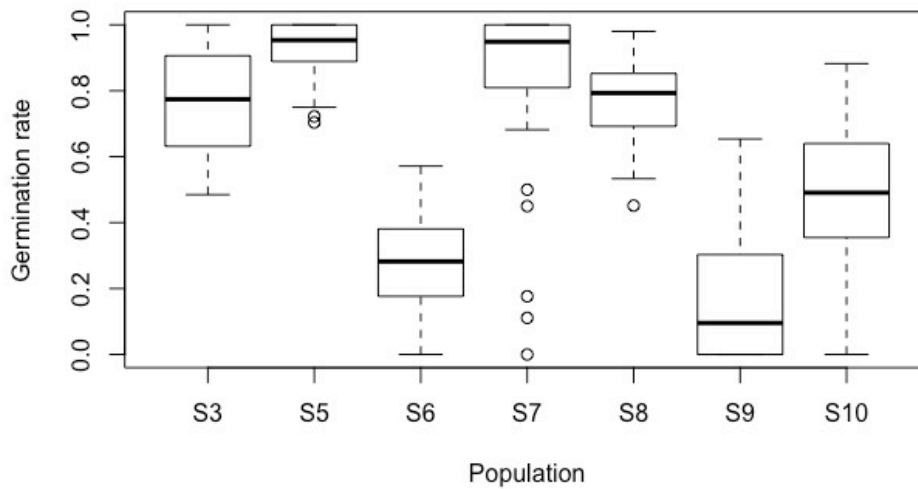
Germination test

Black non-germinated (B), white non-germinated (W), black germinated (BG) and white germinated (WG) seeds were counted, in order to calculate germination rate for each cross and to compare it between crosses. The numbers are shown in the table below.

Table 8. Numbers of each type of seeds in the all the crosses.

Origin	B	W	BG	WG	All seeds
S3	183	24	650	0	857
S5	51	2	642	0	695
S6	445	2	176	0	623
S7	111	0	512	0	623
S8	113	128	794	0	1035
S9	463	95	117	0	675
S10	329	89	342	0	760

As it is shown in the table above, the crosses S6 and S9 had the highest number of black non-germinated seeds. The cross S5 had the lowest number of black non-germinated seeds. The highest number of white non-germinated seeds was within the cross S8. In contrast, the

S7 cross did not have any of them. The cross S8 had the biggest number of black germinated seeds and the cross S6 had the lowest number of them. None of the crosses had white germinated seeds.

Fig 14. Variation in germination rates between the crosses.

The crosses S5 and S7 had the highest germination rate. In addition, in the cross S7, there were also lower values. The crosses S3 and S8 had similar range of values from 0,4 to 1. The cross S10 had the widest range of values of all the crosses (from 0,0 to 0,9) with the highest concentration around 0,5. The lowest germination rates were in the crosses S6 and S9. The cross S9 had wider range of values than the cross S6.

In addition, in JMP program the analysis of variance (ANOVA test) of germination rates of each cross was done. The null hypothesis is that there is no variance between germination rates between crosses. Moreover, the same test was done for the variance of germination rates between catkins within each cross. The null hypothesis is that there is no difference between germination rates between catkins within each cross.

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Crossing	6	6	15,325368	489,3750	<,0001*
catkin[crossing]	63	63	6,121723	18,6172	<,0001*

Fig 15. The result of the analysis of variance (ANOVA test) of germination rates between the crosses and the analysis of variance of germination rates between catkins within each cross. The p-values of both tests are very low (p<0,0001).

Finally, in JMP program LSMeans Tukey HSD (Honestly Significant Difference) test was done to compare the differences between the crosses.

Level	Least Sq Mean

Level					Least Sq Mean
s5	A				0,92367517
s7		B			0,81340462
s3		B			0,76481359
s8		B			0,76233296
s10			C		0,46547767
s6				D	0,28471105
s9					E 0,16458922

Fig 16. Differences between the germination ratios between the crosses. The crosses that have the same letter are not significantly different. The crosses that do not have the same letter are significantly different.

As it is shown in the figure above, the crosses S3, S7 and S8 have the same letter, which means that there is no significant difference between them. The crosses S5 and S10 have different letters than the already mentioned ones, but they seem to be less different than them in comparison to the rest. The S6 and the S9 are significantly different from the other crosses.

Coefficient of relatedness

A coefficient of relatedness for offspring of each cross has been calculated, in order to estimate the presence of homozygotes between crosses.

Table 9. Coefficient of relatedness for each cross.

Population	Cross	F
S3	81084x78195	0.5
S5	81084x78021	0.5
S6	221x78195	0.5
S7	221x78021	0.75
S8	81084x031	0.75
S9	221x228	0.75
S10	81084x228	0.75

The crosses S7, S8, S9 and S10 have the highest coefficients of relatedness, which means that they are the most inbred and they have the highest presence of homozygosity. The crosses with lowest values of inbreeding are S3, S5 and S6. Thus, they are the least homozygous.

5. Discussion

PCR and primer design

In primer design, many factors are important such as primer length, primer length difference, GC content, melting temperature, hairpins, dimers and primer specificity.

The optimal length of a primer is 16-28 nucleotides, including a length difference between the forward and reverse primer of no more than 3 nucleotides (Wu et al. 2004). Many primer design software tools ensure that there is no difference in length bigger than 3 nucleotides (Chuang et al. 2013). I can exclude that this factor is a reason for unsuccessful products, since I was using Primer3 software.

The ratio of appearance of the nucleotides G and C in the primer is a GC proportion, which should be in a range of 40-60 %, in order not to affect the annealing conditions of the primer (Chuang et al. 2013 and references therein). The GC proportion of all the primers that were used was in the ideal range.

The most critical factor in PCR is the melting temperature of the primers, which should between 50°C and 62°C. If a melting temperature of a primer is higher than 5°C, the PCR is unsuccessfully performed. Moreover, the difference between melting temperatures of a primer pair used in the same tube should be as low as possible, ideally 0 (Chuang et al. 2013 and references therein). In this work, all the melting temperatures of the primers were in the ideal range. In addition, it is possible that there were bigger differences between melting temperatures of the used primer pairs in some tubes than in the others, but not bigger than 5°C.

Hybridization of two primers results in a dimer and can negatively impact the PCR experiment. Dimers can show up in two forms, as cross-dimers or self-dimers. A cross-dimer is a result of hybridization of the forward and reverse primer, in contrast to a self-dimer, which is formed when two forward or two reverse primers anneal to each other (Vallone and Butler 2004). Since I cannot confirm that there are no primer dimers in the PCR products that I got, it is possible that have appeared.

Primer can anneal to itself, which results in the secondary structure or a hairpin (Vallone and Butler 2004). In Primer3 software, there are values that show the possibility of hairpin formation. The ideal value is 0, but for some of the used primers in this work, the value was higher. Thus, there is possibility that some of the products had secondary structures.

If the primers do not anneal to a specific region or if they anneal to different regions within a DNA template, the PCR experiment is not successful (Ye et al. 2012). Some of the primers in this study probably annealed to different regions, since there were multiple bands.

Finally, I can exclude the possibility that if one of the PCR parameters was standardised that it would give better results, because there are many factors, that impact the reaction success. The reasons that some PCR programs gave better results using the same primers than the others may be formation of primer dimers, secondary structures and that the sex determination region has many repetitions.

Inbreeding effect

In this study, my conclusions are based on the results of germination test and not on analysis of sequences of sex determination region of all the offspring. Thus, I can only have assumptions about the effects of homozygosity on the germination rates.

The p-value of ANOVA test for variation between the crosses is very low ($p < 0,0001$), which means that the null hypothesis is rejected. Thus, there is a big variation between the crosses, which might be the consequence of genetic causes, for example genetic incompatibility between maternal and paternal haplotypes. In addition, the p-value for the variance between catkins within each cross is the same ($p < 0,0001$), which means that there is also big variance between catkins within each cross. This can be explained by ecological causes, such as sparse or abundant pollination (Hardy et al., 2002).

The crosses S7, S8, S9 and S10 have the same coefficient of relatedness ($F = 0,75$). I have expected that these crosses will have the lowest germination rates. The S9 cross is the least successful in the germination test, so the expectations are correct. The S7, S8 and S10 have higher germination rate than the S9, so the expectations for these crosses are not correct.

The crosses S3, S5 and S6 have the coefficient of relatedness equal to 0,5. I expected that they would have the highest germination rates. The S5 has the highest germination rate, but the germination rate of the S6 is lower than the germination rates of the S7 and S8, which have higher coefficient than the S6. The reason that the S6 has very low germination rate might be that there is a new genotype in the offspring (new haplotype Z4 in the mother 78195) and at the same time the lethal homozygosity is not connected to the phenomena of non-germinated white seeds. That means that black seeds might be homozygous too.

Based on the coefficients of relatedness, the expected germination rates of the crosses are not in accordance with the results, so the impact of the inbreeding on the germination rates of the crosses cannot be estimated. Since the coefficients of the crosses S7, S8, S9 and S10 are the same and yet there are differences between their germination rates, it is possible that there is genetic incompatibility between homozygous allelic combinations (Table 2).

White seeds

According to the results of the germination test, where none of the white seeds have germinated, I assume that my hypothesis that white seeds are homozygous is correct. Homozygosity is lethal according to the previous study (Pascal 2016, in prep). In order to know what exactly is missing in the white seeds, further studies are required.

In the S3 cross there were 24 non-germinated white seeds, which was unexpected according to my hypothesis that there are no white seeds, since this is the cross that has even sex ratio according to the previous study (Pucholt in prep, 2016).

For the cross S6 I did not form any hypothesis, since according to the previous study (Pucholt 2016, in prep) it is possible that this cross has the new haplotype Z4. Thus, there may be new allelic combinations with unknown effects on the germination rate. There were 2 non-germinated white seeds out of total number of 623 seeds, which is less than 1 % of all the seeds.

For the crosses S5, S7, S8, S9 and S10 I expected that 25 % of all the seeds will be white (homozygous). The cross S5 had 2 non-germinated white seeds, which is 0,3 % of all the seeds. There were no white seeds in the cross S7. The cross S8 had totally 1035 seeds and 128 of them were white, which is 12,4 % of white seeds. The hypothesis for S10 cross is rejected, due to the fact that there are 11,7 % of white seeds in this cross.

Finally, none of the hypotheses related to the crosses S3, S5, S7, S8, S9, and S10 is confirmed. The possible explanation might be that white seeds are homozygous, but their number is not fully presented in the germination test. Moreover, it is possible that while setting up the experiment, most of the white seeds stayed in the bags, since they are lighter than black seeds.

Secondly, the explanation might be that white non-germinated seeds do not present the homozygous offspring and that there is another reason for their phenotype and low germination. Furthermore, the seeds with homozygous genotype may not be even fully formed, so they might die in an earlier stage of development, for example as zygotes or the fusion of the zygotes does not even happen.

6. Conclusion

The reasons that some PCR programs in this study gave better results using the same primers than the others may be formation of primer dimers, secondary structures and that the sex determination region has many repetitions. In terms of the germination test, based on assumptions, in this work it is shown that white seeds may not present homozygous individuals. Moreover, homozygosity may be related also to the black seeds or the homozygous seeds might not be even fully formed. In the further studies of the white seeds phenomena, sequencing of the genomes of the seeds is required in order to investigate the relatedness between homozygosity and white seeds. Furthermore, the coefficients of relatedness are not in accordance with the germination rates of the crosses, so the impact inbreeding on the germination rates of the crosses cannot be estimated. Moreover, it is possible that there is genetic incompatibility between homozygous allelic combinations, which may have led to the differences among the germination rates of the crosses.

Acknowledgement

References

Alström-Rapaport, C., Lascoux, M., & Gullberg, U. (1997). Sex determination and sex ratio in the dioecious shrub Salix viminalis L. *Theoretical and Applied Genetics*, *94*(3-4), 493-497.

Berlin S, Lagercrantz U, von Arnold S, Öst T, Rönnberg-Wästljung AC. 2010. High-density linkage mapping and evolution of paralogs and orthologs in *Salix* and Populus. *BMC Genomics*. 11:129

Charlesworth, D. (2002). Plant sex determination and sex chromosomes. *Heredity*, *88*(2), 94-101.

Chuang, L. Y., Cheng, Y. H., & Yang, C. H. (2013). Specific primer design for the polymerase chain reaction. Biotechnology letters, 35(10), 1541-1549.

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

Hardy, Ian CW, and Ian CW Hardy. Sex ratios concepts and research methods. No. 576.855 S4. 2002: 349-361

Höglund S, Larsson S, Wingsle G. 2005. Both hypersensitive and non-hypersensitive responses are associated with resistance in *Salix viminalis* against the gall midge *Dasineura marginemtorquens*. *J Exp Bot.* 56:3215-3222.

Klug, William S., and Michael R. Cummings. *Essentials of genetics*. No. Ed. 7. Prentice-Hall Inc., 2010: 148, 512-514

Ming, Ray, Abdelhafid Bendahmane, and Susanne S. Renner. "Sex chromosomes in land plants." *Annual review of plant biology* 62 (2011): 485-514.

Pucholt P, Rönnberg-Wästljung AC, Berlin S. 2015. Single locus sex determination and female heterogamety in the basket willow (*Salix viminalis* L.). Heredity. 114:575-583.

Pucholt P. n.d. Genetic markers and their application (From marker design to QTL analysis and beyond). Unpublished review.

Pucholt, P, Hallingbäck, HR, Berlin, S. 2016. Genetic incompatibility can explain female biased sex ratios in dioecious plants. Unpublished manuscript.

Renner SS. 2014. The relative and absolute frequencies of angiosperm sexual systems: Dioecy, monoecy, gynodiecy, and an updated online database. *Am J Bot.* 101:1588-1596.

Vallone, P. M., & Butler, J. M. (2004). AutoDimer: a screening tool for primer-dimer and hairpin structures. Biotechniques, 37(2), 226-231.

Wu, J. S., Lee, C., Wu, C. C., & Shiue, Y. L. (2004). Primer design using genetic algorithm. *Bioinformatics*, *20*(11), 1710-1717.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC bioinformatics*, *13*(1), 1.