Genetic and genomic characterization of the LOP locus from *Pilosella piloselloides subsp. Praealta*

A thesis

submitted in partial fulfilment

of the requirements for the Degree of

Master of Science

at

Lincoln University

by

Marion Eliane Annie Gaillard

Lincoln University

2021

Abstract of a thesis submitted in partial fulfilment of the requirements

for the Degree of Master of Science

The Genetic and genomic characterization of the LOP locus from *Pilosella piloselloides subsp. Praealta*

by

Marion Eliane Annie Gaillard

Apomixis is an asexual way of reproduction leading to the creation of clones of the mother plant. It is controlled by two independent loci in *Pilosella piloselloides subsp. Praealta* (Hawkweed): LOA (Loss of apomeiosis) and LOP (Loss of parthenogenesis), which are dominant. Apomictic plants develop one extremely competitive genotype and spread it in a colonial strategy. Only a few crops are apomictic; however, the natural creation of clones could considerably speed up the process of plant breeding and quickly stabilize new varieties adapted to climate change. *Pilosella* is one of the model plants to study apomixis. The purpose of this work was to investigate the LOP locus on a genetic and genomic level to better understand the mechanism of apomictic plants in the hope of applying it to crops. Using a panel of deletion mutants, this study made it possible to define the size of the LOP locus to about 650kb. A sequence of acceptable quality was checked for this locus using BAC isolates. Finally, the PAR gene contained in the locus for parthenogenesis in *Taraxacum* (dandelion) (Underwood *et al.*, 2020) was sequenced in *Pilosella* and several species from the Asteraceae family, indicating the level of conservation of the dominant allele compared to the recessive alleles.

Keywords: Apomixis, asexual reproduction, apospory, parthenogenesis, LOP, amplicon sequencing

Acknowledgements

First and foremost, I need to warmly thank my associate supervisor Ross Bicknell who has been the best mentor I could have ever wished for. Thank you for believing in me more than I believed in myself. Thank you for your continuous encouragement throughout and for allowing me to learn at my own pace. Thank you for trusting me, being there to answer endless questions and allowing me to make my own decisions.

Thank you for valuing each and every part of my work. Thank you for your time, patience, and kindness. You not only taught me how to think like a scientist, but you also shared your passion for science, research and apomixis! This was a fantastic experience, and this past year has and will have a significant impact on my future career and life. Thank you for pushing me over the finish line. Every day I became more and more confident in my abilities overall, not just in lab work. I'm definitely a better person now than when I entered your lab one year ago.

I will never forget what you did for me, and I wish I could pay you back for all your help. However, one day, I can pass on what I have learned from you to others who are just as passionate.

A big thank you to my supervisor Christopher Winefield for giving me time despite having a busy schedule, for always making sure I was in the best condition and for the touch of sarcasm.

Thank you to Sylvia Erasmuson for her patience, and I wish you a happy retirement spending it with family and friends, and finally enjoying more time in your garden.

Thank you to Panimar Vijayan for her supervision in the lab and Ting-Hsuan Chen for her help with the sequencing data.

Thank you to both my parents for supporting me and believing in me no matter what happens. Thank you for allowing me the freedom to study abroad on the other side of the world without hesitation and supporting me each step of the way, even though you missed me so much. Thank you for everything you did to make this possible. I am forever grateful!
Thank you to my sister "Doudou" for her love and support

Thank you to the entire Hoggets Rugby Team for amazing games and fun times, with a special thought for Billie, Caitlain and Liz.

I would also like to thank my friends Eva, Nelly, Aurélie for their incredibly loyal and invaluable friendship even after two years of being away from France. That helped me to stay strong and to continue my studies. Thank you, Caroline, for a fantastic insight into Canada snowy landscapes.

Thank you to Nicole for being my first friend in what was a new world for me.

Finally, thank you Sam, for your unconditional and priceless friendship, for the most incredible road trip and for the peg battles!

Table of Contents

List of tables

List of figures

Chapter 1 Introduction to apomixis

Apomixis (Apo- = “free from”; mixis = “mixing” in Greek) is a term that was introduced by Winkler in 1908 to describe a “substitution of sexual reproduction by an asexual multiplication process without nucleus and cell fusion.” [Translated from German] (Winkler, 1908 ; as cited in Bicknell et Koltunow, 2004).

It is believed that the first time apomixis was witnessed and described was in 1841 by John Smith. He noted that plants he imported from Australia to England had only female flowers, but they produced perfect seeds despite the absence of any male organ. In other words, the female flowers produced seeds without intervention of pollen. He was not able to precisely establish the name of the characterized plant at the time, but it was later described as *Alchornea ilicifolia* (syn. *Caelebogyne ilicifolia*).

This definition of apomixis has been refined with time to describe an alternative to sexual reproduction, which does not involve meiosis or fertilization, but results in the clonal production of a mother plant through seed. In sexual reproduction, germ cells (2n,4c) first undergo meiosis, a two-step division process that results in haploid gametes (1n,1c). The crossing over and different assortments of the homologous chromosomes occurring during meiosis lead to a unique set of DNA. Each of the male and female gametes then fuses, resulting in a zygote (2n) that is genetically different from the parents.

In apomixis, embryos are directly formed from unreduced egg cells or maternal tissues of the ovary. The chromosomal material transmitted to the progeny does not undergo any recombination event and is identical to the maternal plant (Bicknell et al., 2016). Apomixis, therefore, results in uniform populations and in cases where it can occur without a pollination partner, apomixis can facilitate the rapid colonization of new environments. Many apomictic species are known as weeds. Some examples are *Pilosella* (hawkweed)*,* *Taraxacum* (dandelion) and *Opuntia* (prickly pear).

Apomixis has been witnessed in more than 400 species within 300 genera and 40 plant families of Angiosperms (Carman, 1997) including *Pilosella piloselloides subsp. Praealta*. A common theory suggests apomixis evolved out of sexual ancestors (Holsinger, 2000) as a result of mutations in genes involved in the key stages of the sexual development program, resulting in a “short circuit of the sexual pathway” (Spillane et al., 2001). Apomixis has evolved independently many times across the different angiosperm plant families (Bicknell, 2004).

It was initially suggested that it should lead to an evolutionary ‘dead end’ (Darlington, 1939; Grant, 1981) because the loss of genotype heterogeneity would make these apomictic populations unable to adapt to environmental change.

Although it is counterintuitive, this expectation does not meet the reality in nature. One reason for this is that apomixis is seldom expressed as the only mechanism of reproduction. Often sexual and asexual reproduction coexist independently, either in different individuals within a population or both can occur in the same plant, either side by side or one is favoured over the other. In *Pilosella,* both sexual and asexual seeds form on the same plant, with the asexual seeds typically representing more than 98% of the total (Bicknell et al., 2003).

Table 1.1 Advantages and disadvantages of apomixis vs sexual reproduction.

In green the advantages, in red the disadvantages.

Apomixis	**Sexual reproduction**
Only the existing, elite genotype is multiplied Uniparental reproduction avoids the cost of maintaining males and provides an advantage in colonization (Baker, 1959 ; Baker et Stebbins, 1965).	Energy is allocated to producing offspring that may be unfit to its environment The mother plant always needs a mating partner A high amount of recombination occurs during meiosis
Deleterious mutations accumulate through generations The multiplied genotype is only performant in a stable environment	Sexual reproduction creates genetic diversity Genetic variability provides opportunities to adapt to environmental changes

As shown in the previous table, the advantages of one reproductive pathway are often the disadvantages of the alternative one and vice versa. This could explain why apomixis and sexual reproduction often coexist in nature. One theory suggests there may be a sex/apomixis switch that responds to environmental conditions in some species. Environment factors would then modify the penetrance of the traits. Indeed, an increase in the frequency of sexual ovule development (meiosis instead of apomeiosis) under stress has been observed in some species of the genus *Boechera* (Mateo de Arias, 2015).

1.1 Types of apomixis

The different forms of apomixis are distinguished according to the cell type involved in the creation of the unreduced embryo sac.

1.1.1 Sporophytic apomixis

In this type of apomixis, the female gametophyte phase is entirely bypassed (Leblanc and Mazzucato, 2001). The embryo forms directly from a somatic cell of the ovule (Fei et al., 2019).

This form of apomixis typically requires the sexual formation of an endosperm tissue, so pollination is still a requirement for seed formation. The fertilization of the meiotic embryo sac is, therefore, still required for the development of the clonal embryo.

1.1.2 Gametophytic apomixis

When we talk about apomixis, we usually refer to gametophytic apomixis. The embryo comes from an unreduced egg cell, which means an egg cell that didn't undergo meiosis (was not reduced). This process is called "apomeiosis". The unreduced egg cell then differentiates to form an embryo sac by mitosis. Within this structure, an unreduced egg cell forms. The subsequent development of an embryo from this egg cell without fertilization is a process called "parthenogenesis" (Bicknell and Catanach, 2015). Gametophytic apomicts are sub-divided into two categories according to the mechanism used to avoid meiosis. These two mechanisms probably evolved independently between *Hieracium* and *Pilosella* (Mráz and Zdvořák, 2019).

Diplospory

The diploid megaspore mother cell avoids meiosis to develop an embryo sac (Tucker and Koltunow, 2009). The interruption of the meiosis is due to a failure in chromosome pairing. It leads to a mitosis-like division and the formation of an unreduced megaspore (with the same ploidy as the maternal plant). This unreduced megaspore divides, leading to the formation of an unreduced embryo sac and then an unreduced egg. Plants in the genus *Hieracium* use this type of apomixis.

Apospory

Meiosis is completely bypassed. Somatic cells of the ovule near the developing megaspore enlarge and differentiate to form aposporous initial cells. These have the maternal genotype and act as megaspore-like cells (Bicknell and Catanach, 2015). The AI's then develop into embryo sacs without the intervention of any external material (Okada et al., 2013). Plants in the genus *Pilosella* use this type of apomixis.

Aposporous apomicts may or may not require pollination for endosperm development (a sperm cell from the male gametophyte fuses selectively with an unreduced polar nucleus). *Pilosella* species typically do not require pollination and are therefore called 'autonomous apomicts' (Koltunow, 1993).

1.2 Apomixis in *Pilosella praealta*

Gregor Mendel was one of the first geneticists to be interested in *Pilosella* (formerly *Hieracium)*. Mendel chose this plant to retry the experiments he conducted on other species to validate his hypothesis on the particulate nature of inheritance. He didn't know that this plant rarely uses sexual reproduction and therefore *Pilosella* didn't show the inheritance pattern that Mendel previously found on peas. Mendel remained convinced of his findings but died with the idea that the patterns he found were only applicable to some species and could not be generalized. We now know that, unfortunately, *Pilosella* was a poor model system for Mendel's work on sexual inheritance. It has, however, recently been found that *P.praealta* possesses several useful features for the molecular study of apomixis (Bicknell et al., 2016).

Most of the species in the genus *Pilosella* are apomictic, but some sexual species exist and some sexual individuals do occur within typically apomictic species. As both asexual and sexual seed form in this system, *Pilosella* is called a "facultative apomict" (Mráz and Zdvořák, 2019). Estimates of the level of facultativeness vary between species and variety, but they are typically above 90% asexual (93%: Mráz and Zdvořák, 2019, 97%: Bicknell et al., 2003). In some species, the sexual female gametophyte development is not affected. Meiotically reduced and unreduced aposporous gametophytes co-exist in the same ovule (Drews and Koltunow, 2011). However, the initiation of apomixis in *Pilosella* usually interrupts the sexual process (Tucker et al., 2003). The adjacent sexual megaspores then degenerate during aposporous embryo sac formation.

1.2.1 *Pilosella praealta* as a model plant

Pilosella praealta is a good model plant to study apomixis because:

- It is easily cultivated
- It is easily grown and manipulated in vitro
- It has a small stature,
- It has a short generation time, and abundant seed set to facilitate rapid turnover of experimental populations.

- Both sexual and apomictic biotypes are available.
- It uses autonomous endospermy to avoid difficulties associated with pseudogamy.
- Male meiosis and pollen formation are functional in most apomictic biotypes (for transfer to sexual recipients)
- Apomixis is easily assessed and quantified in these plants

(Bicknell, 1994)

1.2.2 The genetics of apomixis in Pilosella

Apomixis in *Pilosella praealta* is controlled by two genetic loci (*LOA* and *LOP*). They were first described in a wild type individual called R35 through the use of deletion mutants (Catanach et al., 2006; Koltunow et al., 2011). Thus, the loci were named after the phenotypes of the mutants presenting a defect in the so-called loci.

The *LOSS OF APOMEIOSIS* Locus (LOA/loa). This locus is associated with the avoidance of meiosis and the differentiation of aposporous initial cells. The inheritance of the dominant *LOA* allele confers apomeiosis (meiosis does not occur). Homozygous recessive genotypes carrying *loa* alleles have the sexual biotype using meiotic egg cell formation.

The *LOSS OF PARTHENOGENESIS* locus (*LOP/lop*). This locus is associated with the avoidance of fertilization. It controls the development of the unreduced embryo and the endosperm. The *LOP* locus exhibits suppressed recombination (Catanach et al., 2006). Inheritance of the dominant *LOP* allele confers parthenogenesis (embryo development occurs spontaneously). Homozygous recessive genotypes carrying *lop* alleles have the sexual biotype requiring fertilization to initiate embryo formation.

In both cases, the dominant alleles at these loci act to overcome a specific developmental checkpoint. In the case of *LOA,* the dominant allele enables cells adjacent to the megaspore mother cell to assume a fate similar to the megaspore. They then divide, and the products undergo the specific developmental pathway of mega-gametogenesis, leading to the formation of an embryo sac.

In the case of *LOP,* the dominant allele enables the egg cell to overcome quiescence and to divide without the stimulus of fertilization.

A third locus controls the development of the endosperm, called the **LOSS OF AUTONOMOUS ENDOSPERMY** locus. There is some data to suggest that this is in linkage with *LOP,* but very little is currently understood about the structure or function of this locus in *Pilosella*.

1.2.3 Identification of apomixis genes in other systems

Pennisetum

Many of the species in the genus *Pennisetum* reproduce by apospory. Only one dominant locus is involved, but it always exists in the heterozygous state in an apomictic parent (Ozias-Akins et al., 2003). In *Pennisetum squamulatum* this locus is a large, dominant sequence called the Apospory-Specific Genomic Region (ASGR), which confers both apomeiosis and parthenogenesis (Ozias-Akins et al., 1998). *BabyBooM-Like* (*PsASGR-BBML*) genes located within this locus have been discovered and present similarity to *BABY BOOM* (*BBM*) genes of Arabidopsis and Brassica, which are involved in embryogenesis (Boutilier et al., 2002).

A study showed that the expression of one of these genes called *BBM1* is initially specific to the male allele but is subsequently biparental. However, in rice (*Oryza sativa*), this gene induces embryo formation without fertilization when it is only expressed in the egg cell. This study also claimed that “asexual propagation without genetic segregation could be engineered in a sexually reproducing plant”, although this was not actually achieved. (Khanday et al., 2019).

Taraxacum

Apomixis in *Taraxacum* is obligate meiotic diplospory, which has been described as “the type that is most similar to sexual reproduction among apomixis systems”(Asker and Jerling, 1992; as cited in Majeský et al., 2012). Two unlinked dominant apomixis loci have been described, controlling: 1. The avoidance of meiotic reduction (*DIPLOSPOROUS-Dip*) and 2. Parthenogenesis of the embryo (*PARTHENOGENESIS-Par*)” (Van Dijk et al., 2009).

Boechera

Boechera is the model plant for studying apomixis in the Brassicaceae family. Apomixis in *Boechera* is gametophytic apomixis (Böcher, 1951) where both apospory and diplospory can occur (Carman et al., 2019). One candidate apomixis gene has been identified in *Boechera* and is called APOLLO (apomixis-linked locus). This gene shows increased expression in apomictic ovaries compared to sexual ovaries (Corral et al., 2013).

Poa

In the case of *Poa*, there is evidence that parthenogenesis can segregate independently of aposporous embryo sac development (Albertini et al., 2001).

A complex model has been described for *Poa pratensis* that includes five single, unlinked genes: the Apospory initiator (*Ait*) gene, the Apospory preventer (*Apv*) gene, a Megaspore development (*Mdv*) gene, the Parthenogenesis initiator (*Pit*) gene, and the Parthenogenesis preventer (*Ppv*) gene (Matzk et al., 2005).

1.3 Use of apomixis in plant breeding

Apomixis has long been avoided by plant breeders in a quest for new cultivars as it resulted in no change between generations, so it opposed genetic gain. Now it is realized that it could be very valuable in crops if it could be controlled.

Every year, farmers around the world buy new seeds whose valuable characteristics (resistance to a pathogen, resistance to drought, high productivity…) have been stabilized using F1 hybrid technology. Unfortunately, this genetic work is compromised with each new generation because the events of meiosis (division of genetic material) and fertilization (supply of external genetic material) generate new genotypes. As a result, F2 generations tend to be much more variable and to have lower average yields than F1 generations, so farmers are encouraged to buy more F1 hybrid seed rather than save seed from a previous crop.

One of the reasons why we study apomixis is that if we applied this mechanism to crops, during the breeding process, the avoidance of recombination would mean that valuable traits would be retained through the generations. This would empower poor farmers to raise their own high yielding hybrid seeds, increasing production levels in countries that currently do not have the resources to buy hybrid seeds every year. It may also reduce the cost and time of crop breeding as it does not require inbred line production or mechanisms of pollination control such as cytoplasmic sterility and fertility restoration.

Reducing the cost and the time needed for plant breeding will also help breeders to respond more rapidly to changing ecological circumstances.

The idea of a “conditional apomixis” has emerged in the last decades where reproduction could be temporally switched to sexual reproduction for breeding, then back to apomixis for seed multiplication (Spillane et al., 2004). To achieve this, apomictic crops would need to retain an ability to reproduce sexually so that they can be bred, and some form of inducible gene action would be required. Second, ideally, male organ development would be blocked to prevent the horizontal transfer of apomixis genes into wild relatives. Finally, these crops should have autonomous endosperm development so pollination is not essential for crop production (Kandemiṙand Saygili, 2015).

Table 1.2 Advantages and disadvantages of apomictic crops.

In green the advantages, in red the disadvantages

Advantages	**Disadvantages**
Reduced cost of breeding	Sexual reproduction is still needed for breeding
Reduced time per breeding cycle	Horizontal transfer to weeds
Elite genotypes in two generations	Require endosperm development
Hybrid seeds in almost all crops	Reduction of genetic diversity

1.3.1 Conversion of sexual crops to apomictic crops

Wide crosses with apomictic wild relatives

Apomixis is present in some wild species but rarely in crops. Only some crops species such as Citrus, berries, guayule and several forage grasses have been observed to use apomixis as the main method of reproduction. There is, however, evidence of apomixis in several wild relatives to the cultivated crops (Hanna, 1991). Traces of apomixis are also believed to remain in grain sorghum and pearl millet (Bashaw, 1980). The transfer of apomixis from wild relatives may therefore be possible for some crops, but even in cases where this occurs, attempts at the transfer of apomixis have not been successful so far (Fiaz et al., 2021).

Mutations in apomixis related genes

Mutants capable of meiosis-independent embryo sacs development

Two mechanisms have been reported to artificially induce apomeiosis by mutations that convert meiosis into a mitosis-like division. The first mechanism involves the mutation of the gene *Osd1* which ultimately results in the elimination of the second division in meiosis in *Arabidopsis*. When associated with two other mutations that eliminate recombination and modify chromatid segregations, the triple mutant genotype is called *MiMe*. (Fiaz et al., 2021; Kandemiřand Saygili, 2015). In the second mechanism, a mutation in the *Arabidopsis SWI1* gene leads to apomeiosis and diploid egg formation, but only at a very low level (Ravi et al., 2008). Both of these mechanisms need a way of eliminating the paternal chromosome since the seeds they produce have higher ploidy levels than the species.

Mutants capable of fertilization-independent embryo development

Some apomixis candidate genes have been described to be implied in the control of embryo development, but only a few are promising.

The PsASGR-BabyBoom-Like Gene. The transfer of the *PsASGR-BABYBOOM*-like gene (*PsASGR-BBML*) enables parthenogenesis in the monocotyledonous crops maize and rice (Conner et al., 2017) and the dicotyledonous species tobacco (Zhang et al., 2020).

The “Salmon System” in Wheat**.** The transfer of the nucleus from the sexual wheat line *Salmon* into the cytoplasm of two varieties of *Aegilops* resulted in male sterile plants with the ability for autonomous embryo development (Matzk et al., 1995).

The RWP-RK Domain (*RKD*) is responsible for keeping the egg cell quiescent in the absence of fertilization. The knockdown of this single gene domain leads to the egg cell division without fertilization (Rövekamp et al., 2016).

Genetic transformation

Genetic transformation is a potential approach for the transfer and introgression of apomixis into crops. However, induction of apomixis with gene-editing strategies requires prior in-depth knowledge of apomictic genes and genetic variation that naturally exist in apomictic populations.
The cloning of the LOA (Kotani et al., 2014) and LOP (current studies) loci in *P.praealta* is progressing.

Henderson et al. (2020) have introduced a *Cas9* construct in an apomictic *Pilosella* (formerly *Hieracium)* species via *Agrobacterium*-mediated leaf disk transformation. The efficiency of CRISPR/Cas9 editing to target the endogenous *PHYTOENE DESATURASE* (*PDS*) gene was evaluated and gave promising results, opening a way for the use of gene editing in apomixis research and identification of apomixis genes.

AIMS OF THE THESIS STUDY

If we knew more about the genes that control apomixis in a model system we would be better equipped to engineer it into crops. The *LOA* and *LOP* loci of *Pilosella* have been mapped through ESTs to chromosomal regions, and the structure of *LOA* has been reported.

1. The first aim was to extend a BAC tiling path within the *LOP* locus
2. The second aim was to fine map the *LOP* region. Using the sequenced BAC tiling path and a genome draft, markers were designed and tested to define the region of suppressed recombination around the *LOP* locus.
3. The third aim was to examine whether the different forms of reproduction conferred by different alleles at *LOP* (sexual and parthenogenic) influence patterns of allelic diversity in the genus *Pilosella.*

Chapter 2 General Materials and Methods

2.1 A note on taxonomy

Devising a taxonomic scheme that adequately describes and categorizes the diversity present within an apomictic complex is a difficult task. The genus *Hieracium*, as instated by Linnaeus in 1753, has been the subject of several major revisions. At the time of Mendel's work (Mendel, 1869), all species were included within the genus Hieracium, but subsequently, the genus was subdivided into two genera; *Hieracium* and *Pilosella*. Most of the plants used in the current study are now assigned to the genus *Pilosella*. At different times individual clones have also been renamed. The model plant used in this study, R35, has previously been named *Hieracium caespitosum* (Catanach et al., 2006), *Hieracium praealtum* (Bicknell et al., 2016) and *Pilosella praealta* (Bräuning et al., 2018) in accordance with previous naming conventions. The Global Compositae Database currently lists it under the name *Pilosella piloselloides subsp. Praealta*. For simplification the name *Pilosella praealta* is used in the text.

2.2 Plant material

2.2.1 Wild types and mutants

The plant material used for this study were generated for previous studies (Catanach et al., 2016) and maintained throughout the years. The wild type tetraploid "R35" was extensively used throughout the study. The diploid apomict "PH70" was derived from R35 as a meiotic product, so it just contains alleles from R35 but has half the nuclear DNA. This individual PH70 has both the *LOA* and *LOP* alleles at the appropriate loci and acts as an apomict like the wild type. A genome sequence based on PH70 DNA was used for the mapping work. Mutants of R35 lacking the *LOP* locus were also used in the allelic diversity study along with wild types and mutants of other species of *Pilosella* and *Hieracium* (see chapter 5).

2.2.2 Detection of apomixis

The detection of apomixis for this study was performed by Dr Ross Bicknell. "Detection of apomixis on the basis of morphological indicators requires careful observation of the progeny of individual plants by a person who is thoroughly familiar with the mode of pollination of the parent and the variability expected among its sexually produced offspring" (Bashaw, 1980). In *P. praealta,* apomixis can be demonstrated by severing the top off an immature floral bud. This removes the anthers and stigmas but leaves the ovaries intact. If seed still forms, then the plant has elements of apomixis. It is recommended that the seed be sown and the progeny tested for ploidy as this further qualifies the degree of apomixis, separating plants that have full apomixis (for example a tetraploid will yield tetraploid seedlings) from those that have only a functional copy of *LOP* but not *LOA* (in which a tetraploid would yield diploid seedlings). This method is simple, quick, repeatable and quantifiable.

2.3 DNA extraction

Approximatively 100mg of fresh leaves were harvested from each plant and placed in a test tube containing glass beads and snap-frozen using liquid nitrogen. Each tube was vortexed for 6 seconds using a Silamat S5 shaker to grind the plant material to a fine powder. DNA was extracted using the reagents and manufacturer's protocol in the DNeasy® Plant Mini Kit 15 (Qiagen, Hilden, Germany).

Genomic DNA was quantified using an Invitrogen Qubit™ 4 Fluorometer.

2.4 Sequencing

The region containing the *LOP* locus has been sequenced using Oxford Nanopore and Illumina technologies. The data were combined using different assembly tools to polish the data and create the most accurate sequence.

2.4.1 3.4.1 Illumina

Illumina is a next-generation technology where DNA is synthesized using a modified DNA polymerase that incorporates fluorescently labelled dNTPs into a DNA template.
Illumina sequencing was performed by Dr Ross Bicknell's team prior to this study. This technology provided **short reads** of 25-100bp that were be assembled into contigs to create a map of the genome of the apomict *P. praealta* individual PH70. The level of base-calling accuracy of Illumina technology is high (**99.9%**). However, the sequences coming from this technology were subject to random errors, especially at the overlapping regions of the contigs, and assembly was complicated in regions rich in repeat sequences.

2.4.2 Oxford Nanopore

Oxford Nanopore Technology is a fourth-generation DNA sequencing technology based on ion current variations caused by the passage of single strands of DNA molecules through a tiny protein channel (Nanopore).

This high throughput technology produced **long reads** that offered a better coverage of the LOP locus. The samples were processed rapidly, and the results were displayed in real-time. However, the accuracy is lower than Illumina (**98%**), and the data were subject to systemic errors.

2.5 Genome assembly and identification of the *LOP* locus in *P. praealta*

Following the sequencing, assembly programs were used to reconstruct the *P. praealta* PH70 genome. Two assemblers were used: SHASTA and FLYE. Both of these assemblers create a 'hybrid assembly' merging the data from the two alleles at LOP within the diploid PH70. Consequently, the assembly used was not accurate for either allele, but it was valuable for developing markers and estimating genetic distances. This work was conducted by my supervisor Dr Chris Winefield in support of this study.

2.5.1 SHASTA/FLYE

The strength of two technologies, SHASTA and FLYE, were combined to create a more accurate assembly. SHASTA used the Oxford Nanopore data, and FLYE used Illumina data.

One contig within the SHASTA assembly (“1486 BLAST hit”) aligned to the known BAC sequence at the dominant allele of *LOP*. Contig 1486 SHASTA was BLAST back against the FLYE assembly with a 100% match).

This contig was used for mapping the deletion mutants (Catanach et al., 2016) and for the research of allelic diversity in this study.

Chapter 3 BAC library screening: BAC walking

3.1 Introduction

One of the aims of this study was the genetic characterization of the *LOP* locus of *P.praealta*. For this to succeed, the mapping data needed to be as precise as possible so that studies of genome structure, organization and function on which it was to be based were also accurate. Sequence quality in particular, has a significant impact on downstream analysis.

The original genome map for *P. praealta* was built mainly from short-read data from next-generation sequencing technologies. The read length was around 50 - 300 bp. Although NGS technologies have the advantage of a much higher throughput with dramatically lower cost, they also have the disadvantages of higher error rates compared to Sanger sequencing. The first errors appear at the nucleotide level with the base calling (substitutions, insertions or deletions). Each base is assigned a phred-like quality score as a by-product. At the technology level, there are also specific error patterns associated with each genome sequencer (Victoria Wang et al., 2012). At the assembly level, mistakes happen during the computation in genome assembly: missed overlaps leave gaps in the assembly, while false overlaps create ambiguous maps or create connections between regions that should not be associated (Palmer et al., 2010).

To improve the quality of the sequence, a genetic map can be integrated with a physical map. Bacterial Artificial Chromosomes (BACs) accept inserts up to 300 Kb in length. The use of BACs containing genes of interest is an advantageous method for obtaining sequences from specific regions of interest. One of the inserts from a BAC can cover several sequencing reads and thus polish the sequence to facilitate the construction of DNA libraries to analyze genomic structure.

The bacterial artificial chromosome (BAC) cloning vector is a circular plasmid based on the E. coli fertility factor (F-factor) replicon. It is difficult to clone and maintain large intact DNA in bacteria. However, the F factor controls its own replication. It contains oriS and repE genes, which allow unidirectional replication, and parA and parB genes, which maintain its low copy number (Shizuya et al., 1992). The F-factor also has genes that regulate DNA synthesis so that its copy number is kept at a low level; and genes that regulate the partition into the daughter cells after E. coli divides.

Table 3.1 Advantages and disadvantages of BAC system over YAC or mammalian artificial chromosomes

advantages	disadvantages
Stability of insert propagation over multiple generations Large inserts Handy manipulation Accurate endogenous gene expression	Unanticipated changes in the cell phenotype Recombinant BAC construct can cost time More easily degraded

This chapter is about the identification of specific BAC clones corresponding to genomic sequences of the *LOP* locus. One BAC clone was identified in a previous study, but it didn't cover the entire *LOP* locus. The aim of the library screening was to find clones corresponding to two marker probes, originally designed against lettuce genome sequence.

3.1.1 BAC library

In a previous study, a BAC (Bacterial Artificial Chromosome) library was created by Dr Ross Bicknell's team from DNA of the wild type tetraploid apomict R35 (unpublished data). The library contains approximately 258,048 colonies (672 Petri dishes containing 16x24=384 colonies), each carrying one piece of *Pilosella piloselloides var praealta* wild type "R35" DNA. The average insert size in the library is approximately 141kb. BAC libraries usually have several times coverage of the genome (Shizuya et al., 1992). The fold genome coverage is inversely proportional to the average size of the insert. The larger the size of the insert, the fewer clones that are required. The higher the coverage, the better the data quality. In this study, the whole genome of R35 has a 5-fold coverage.

3.1.2 BAC pooling

For a faster identification of the BAC colonies of interest, a compression of the 258,048 colonies into one super-pool was done in a previous study (unpublished data) according to this process:

The library consisted of 672, 384-well plates. Each plate contains 384 x 141kb = 54Mbp (0.75% of 4x).

The 672 384-well plates were pooled into seven 96-well plates, where each well contained the equivalent of a 384 BAC clones. These were referred to as the 'pool plates'. Each pool plate contains 54Mbp x 96 = 5,184Mbp (72% of 4x)

Seven pool plates were then reunited in 7 rows of a new super-pool plate. In this unique plate, each horizontal row contained the equivalent of one of the 7 pool plates (see figure 1). Each row of the super pool well contains 12 x 54Mbp = 1,512Mbp (9% of 4x) and each column contains 8 x 54Mbp = 432Mbp (6% of 4x).

The position of a well in a super-pool plate gave an equivalence of one row of one of the 7 pool plates, and the position of a well on a pool plate gave the address of the source BAC colony plate.

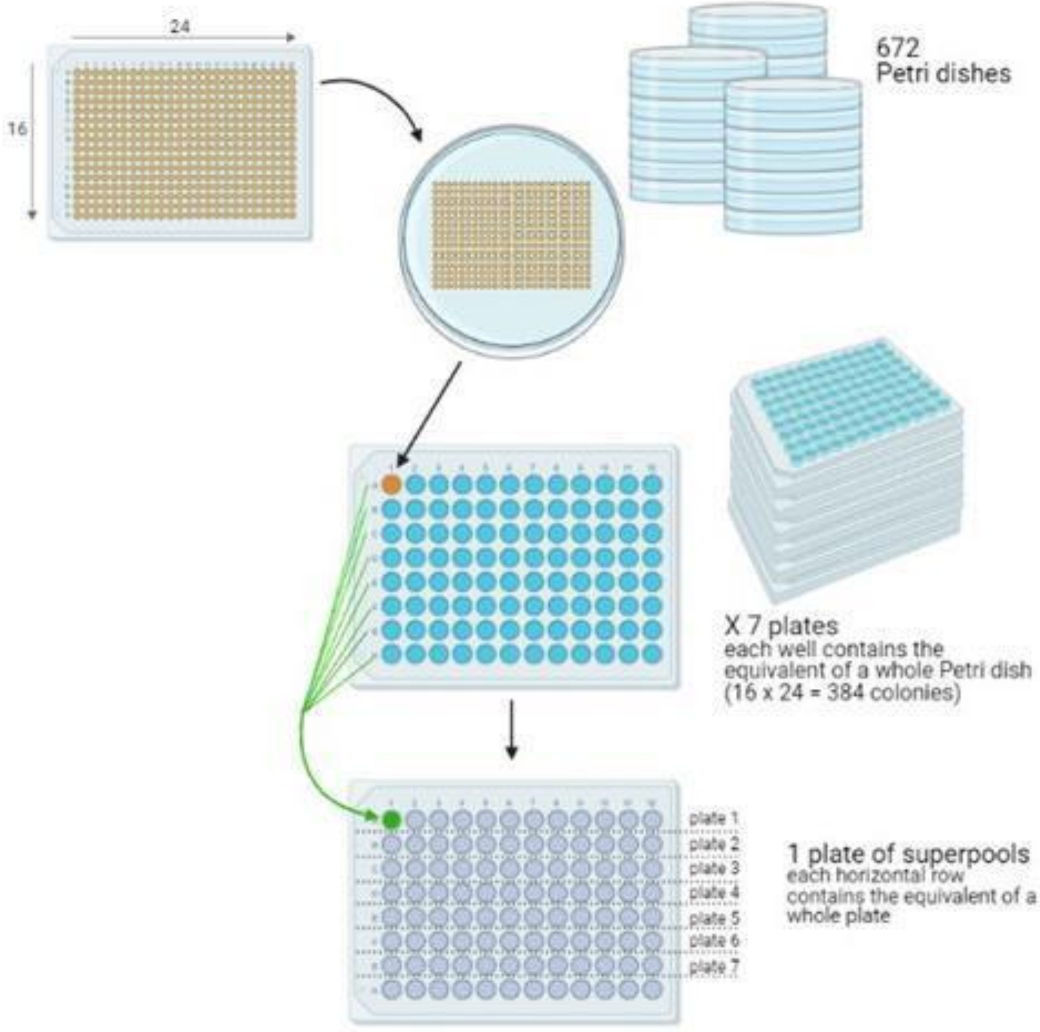

24
16
672
Petri dishes
X 7 plates
each well contains the equivalent of a whole Petri dish (16 x 24 = 384 colonies)
plate 1
plate 2
plate 3
plate 4
plate 5
plate 6
plate 7
1 plate of superpools
each horizontal row contains the equivalent of a whole plate

Figure 3.1 Schematic display of the BAC pooling

3.1.3 Bac screening

The approach used was PCR-based screening, going back from the super-pool to the clone containing the DNA sequence of interest. All PCR amplifications were performed on a Bio-Rad T100 Thermal Cycler. Once pooled, clones containing sequences of interest were located by identifying the subset of pools containing the corresponding markers. The pooling technique saved both time and materials. Instead of doing a PCR for each of the 258,048 colonies, the best candidate BAC colony plates were identified with 180 PCR reactions (84 super-pool reactions followed by 96 pool column reactions). An electrophoresis gel was run to check if the amplification was successful and identify the candidate BAC colony plates. A DNA fragment analyzer partially automated the process, and it also allowed discernment of different bands that varied by only a few base pairs. This was particularly valuable for the markers with allelic product size differences that would not have been apparent on a gel. Twelve super pools for each primer were chosen to go back to the seven pool plates. The electrophoresis gel identified 3 to 4 candidate column pools, which was confirmed by the fragment analyzer. The numbers of the plates and the columns gave an equivalence for each of the 672 BAC library plates. Once a BAC library plate was identified it was replicated onto an agar plate. Colonies were then picked using pipette tips. To avoid contamination, the columns were split into two halves (A-H and I-P).

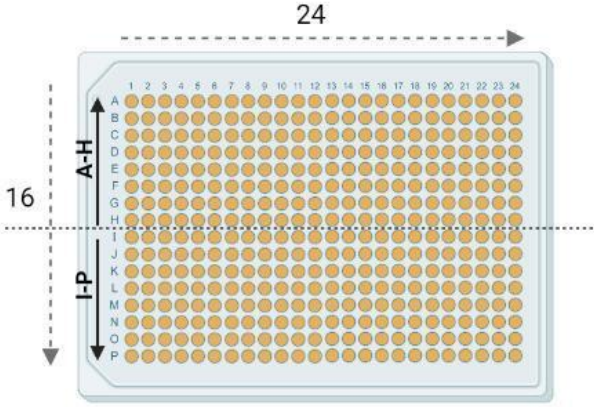


Figure 3.2 Plan showing the process to collect pools of 8 BAC colonies (half columns).

Picked colonies were placed in a template plate with 50uL of water, boiled up at 95 degrees for 10 minutes, cooled to 4 degrees then spun for 2 minutes at 4000 rpm. Two microliters of this supernatant was used as a template in the subsequent PCR reaction. This final step typically required a further 40 reactions (24 column streaks and 16 well reactions) to test and identify the colony of interest. Some savings were made because several candidates often resulted from the first set of 84 super-pool reactions, and this step only needed to be conducted once for each primer set.

Markers

The super-pool plates were tested using primers to the markers Lsat46 and 44/45 (Table 4). Genomic DNA of the wild type R35 was used as the positive control. The marker Lsat46 is co-dominant, identifying 3 of the 4 alleles at *LOP* in R35. Allele 1 is the dominant allele at LOP that confers parthenogenesis. Alleles 2, 3 and 4 are recessive alleles that encode for the *lop* alleles. The Lsat46 primers clearly identified alleles 1 and 4 by product size differences (177bp for allele 1 and 174 for allele 4), while products of alleles 2 and 3 are of a similar size (176bp) and could not be separated for this marker (see figure 3.3). Marker 44/45 is a dominant marker with a 330bp product for allele 1 at LOP. For both markers, the PCR conditions were:

Step 1: 95^{o}C: 3 minutes

Step 2: 95^{o}C: 30 seconds

Step 3: 60^{o}C: 30 seconds

Step 4: 72^{o}C: 1 minute
Repeat steps 2-4 a total of 40 times

Step 5: 72^{o}C: 5 minutes

Hold: 8^{o}C:

Table 3.2 Details of the two markers used for the BAC screening

Marker	Forward primer	Reverse primer	Predicted allele product sizes*
Lsat 46	GAGAGAAGATTGCCATCAGAACC	GAATTAAATAGYGGATGGAAGAGG	177 (1), 176 (2 & 3), 174 (4)
44/45	TCTTGGCCACCGCCAACAAGGTGA	ACTGGCCCG CGCGGATGTCT	330 (1), 327 (2&3), 339 (4)

** Alleles in parentheses*

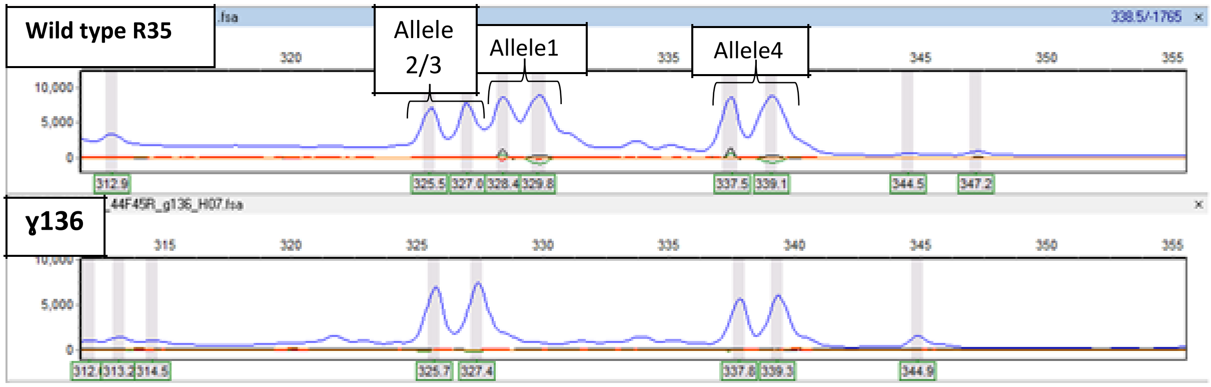


Figure 3.3 **Fragment analyzer data for the marker 44/45 with the wild type R35 as a positive control and ɣ136 as a negative control, used to assess the candidates**

The dominant allele (called allele 1) can be found using deduction: R35 is tetraploid and possess 4 alleles (One dominant allele 1 + alleles 2 and 3 on top of each other and allele 4). ɣ136 is a deletion mutant that only possesses recessive alleles (2 or 3, and 4)

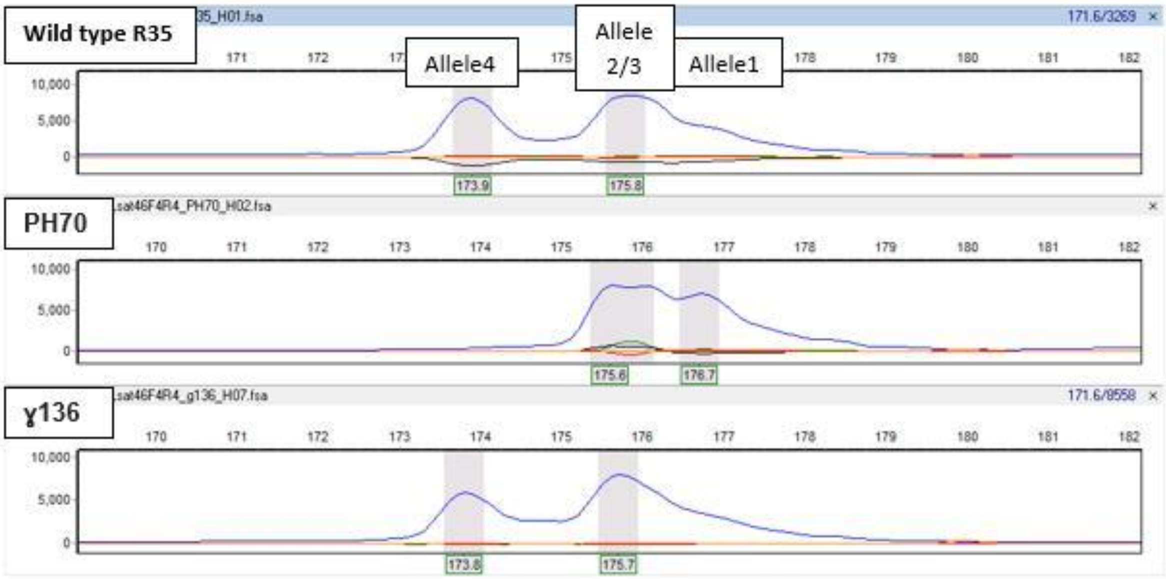


Figure 3.4 **Fragment analyzer data for the marker Lsat46 with the wild type R35 and the polyploid PH70 as positive controls and ɣ136 as a negative control.**

The dominant allele (called allele 1) can be found using deduction: R35 is tetraploid and possess 4 alleles (One dominant allele 1 + alleles 2 and 3 on top of each other and allele 4). Allele 1 is only expressed as a shoulder in R35 but is a clear peak in its diploid derivative PH70. ɣ136 is a deletion mutant that only possess recessive alleles (2 or 3, and 4)

3.2 Results and discussion

3.2.1 Lsat46: Co-dominant Marker (for Alleles 1, 2&3, 4) at *LOP*

Step 1 – super-pool

84 reactions were necessary to amplify each well of the super-pool. The remaining 12 wells left in the 94-well PCR plate were used for controls: R35 (alleles 1,2,3,4), PH70 (1,3), ɣ179 (2,3,4), ɣ115 (2,3,4), ɣ138 (2,3,4), ɣ136 (2,3,4), PH511 (1,2), PH544 (1,3), PH410 (1,3), PH191 (1,2), PH325 (1,2) and No Template Control.

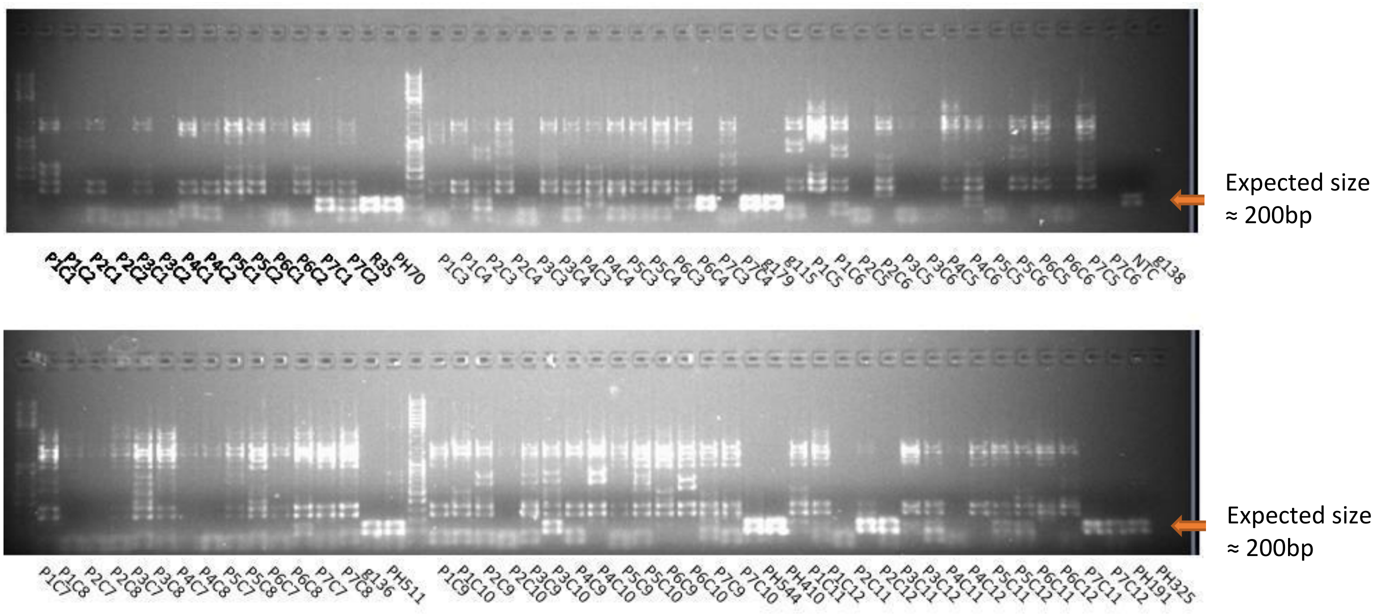
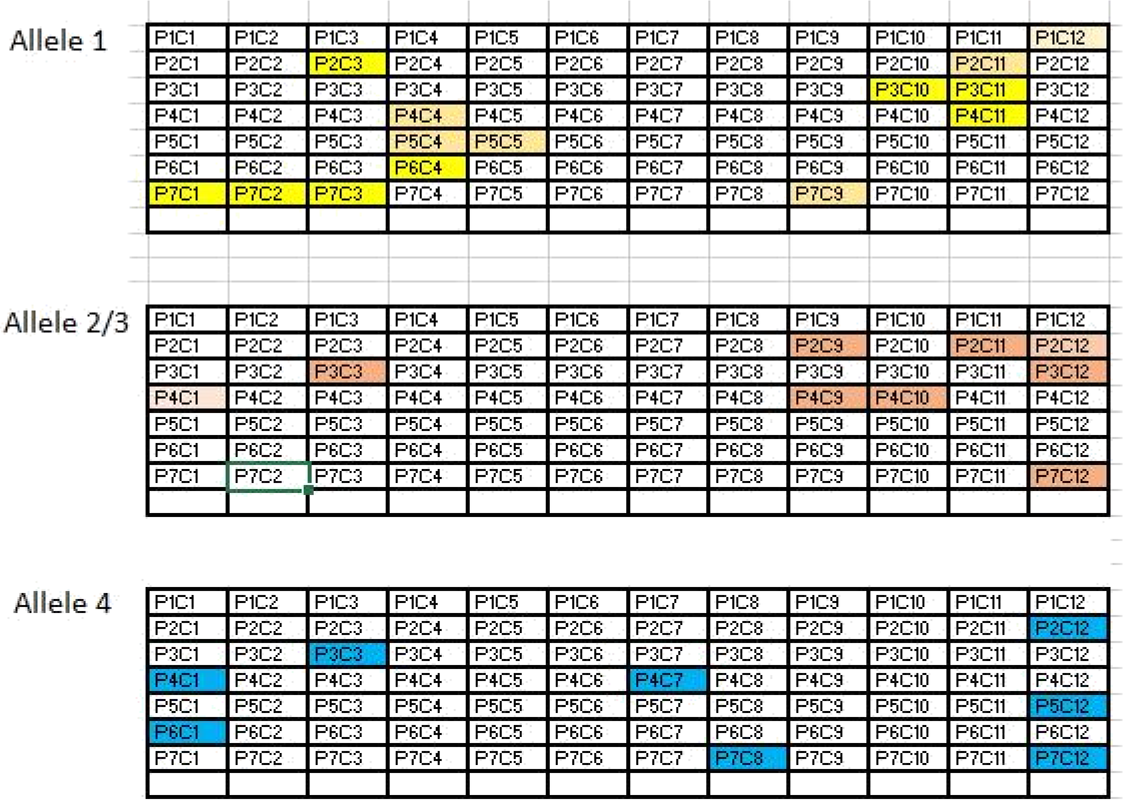


Figure 3.5 Electrophoresis gel of the PCR products from the super-pool with the marker Lsat46: Plate/column pools

Each well was identified with the number of plate and column from which they came from. P: plate; C: column
A strong and compact band around 200bp is an indicator of the potential presence of the desired insert

Super-pools with BAC colonies that contain the sequence of interest gave an intense amplification. Strong amplifications were seen in wells P7C1, P7C3, P3C10, P2C12, P3C11, P7C12. Analysis by the fragment analyzer was used to confirm the PCR product size carried by each of the candidate column pools, which might contain a BAC colony of interest. From this point, attempts to isolate BACs for each allele were treated separately (except alleles 2 and 3 which could not be separated).



Allele 1

P1C1	P1C2	P1C3	P1C4	P1C5	P1C6	P1C7	P1C8	P1C9	P1C10	P1C11	**P1C12**
P2C1	P2C2	**P2C3**	P2C4	P2C5	P2C6	P2C7	P2C8	P2C9	P2C10	**P2C11**	P2C12
P3C1	P3C2	P3C3	P3C4	P3C5	P3C6	P3C7	P3C8	P3C9	**P3C10**	**P3C11**	P3C12
P4C1	P4C2	P4C3	**P4C4**	P4C5	P4C6	P4C7	P4C8	P4C9	P4C10	**P4C11**	P4C12
P5C1	P5C2	P5C3	**P5C4**	**P5C5**	P5C6	P5C7	P5C8	P5C9	P5C10	P5C11	P5C12
P6C1	P6C2	P6C3	**P6C4**	P6C5	P6C6	P6C7	P6C8	P6C9	P6C10	P6C11	P6C12
P7C1	**P7C2**	**P7C3**	P7C4	P7C5	P7C6	P7C7	P7C8	**P7C9**	P7C10	P7C11	P7C12

Allele 2/3

P1C1	P1C2	P1C3	P1C4	P1C5	P1C6	P1C7	P1C8	P1C9	P1C10	P1C11	P1C12
P2C1	P2C2	P2C3	P2C4	P2C5	P2C6	P2C7	P2C8	**P2C9**	P2C10	**P2C11**	**P2C12**
P3C1	P3C2	**P3C3**	P3C4	P3C5	P3C6	P3C7	P3C8	P3C9	P3C10	P3C11	**P3C12**
P4C1	P4C2	P4C3	P4C4	P4C5	P4C6	P4C7	P4C8	**P4C9**	**P4C10**	P4C11	P4C12
P5C1	P5C2	P5C3	P5C4	P5C5	P5C6	P5C7	P5C8	P5C9	P5C10	P5C11	P5C12
P6C1	P6C2	P6C3	P6C4	P6C5	P6C6	P6C7	P6C8	P6C9	P6C10	P6C11	P6C12
P7C1	P7C2	P7C3	P7C4	P7C5	P7C6	P7C7	P7C8	P7C9	P7C10	P7C11	**P7C12**

Allele 4

P1C1	P1C2	P1C3	P1C4	P1C5	P1C6	P1C7	P1C8	P1C9	P1C10	P1C11	P1C12
P2C1	P2C2	P2C3	P2C4	P2C5	P2C6	P2C7	P2C8	P2C9	P2C10	P2C11	**P2C12**
P3C1	P3C2	**P3C3**	P3C4	P3C5	P3C6	P3C7	P3C8	P3C9	P3C10	P3C11	P3C12
P4C1	P4C2	P4C3	P4C4	P4C5	P4C6	**P4C7**	P4C8	P4C9	P4C10	P4C11	P4C12
P5C1	P5C2	P5C3	P5C4	P5C5	P5C6	P5C7	P5C8	P5C9	P5C10	P5C11	**P5C12**
P6C1	P6C2	P6C3	P6C4	P6C5	P6C6	P6C7	P6C8	P6C9	P6C10	P6C11	P6C12
P7C1	P7C2	P7C3	P7C4	P7C5	P7C6	P7C7	**P7C8**	P7C9	P7C10	P7C11	**P7C12**

Figure 3.6 Identification of the potential pool candidates for each allele with the marker Lsat46: Plate/column pools

Colored squares represent positive results. Saturated toned squares for allele 1 and 2/3 are the results with the highest confidence. For allele 4, all the results are supposedly reliable.

Step 2 – pools

For each of the 8 super pools for allele 1 and alleles 4, 7 pool plates were tested. A well a the pool plate corresponds to a BAC library plate.

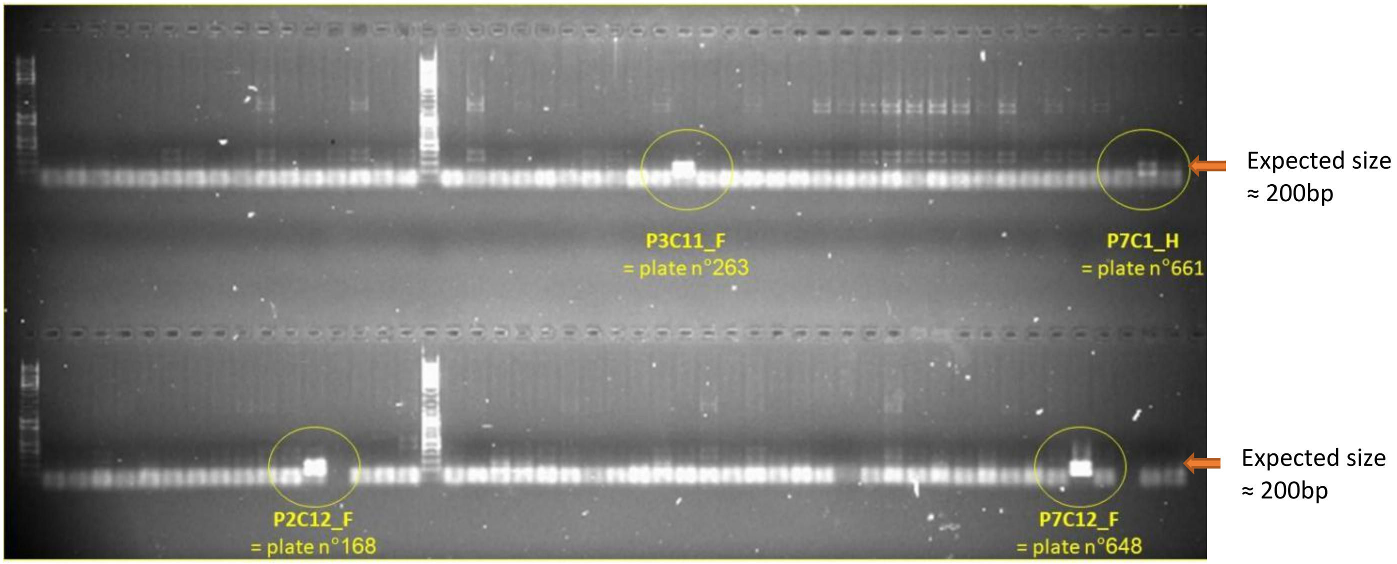
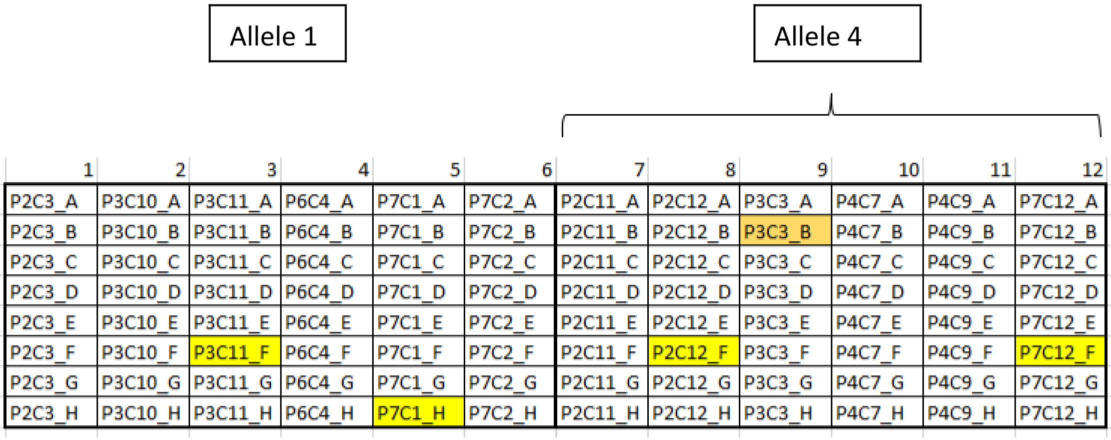


Figure 3.7 Electrophoresis gel of the PCR products from the pool plates with the marker Lsat46.

The four candidates that have been found are marked in yellow. As other candidates did not contain the insert, it is likely that primer dimers formed during the PCR leading to the lower bands on the electrophoresis gel.

Four BAC plate candidates were identified, two for allele 1 and two for allele 4, corresponding to plates 263, 661, 168, 648.


Allele 1
Allele 4

1	2	3	4	5	6	7	8	9	10	11	12
P2C3_A	P3C10_A	P3C11_A	P6C4_A	P7C1_A	P7C2_A	P2C11_A	P2C12_A	P3C3_A	P4C7_A	P4C9_A	P7C12_A
P2C3_B	P3C10_B	P3C11_B	P6C4_B	P7C1_B	P7C2_B	P2C11_B	P2C12_B	P3C3_B	P4C7_B	P4C9_B	P7C12_B
P2C3_C	P3C10_C	P3C11_C	P6C4_C	P7C1_C	P7C2_C	P2C11_C	P2C12_C	P3C3_C	P4C7_C	P4C9_C	P7C12_C
P2C3_D	P3C10_D	P3C11_D	P6C4_D	P7C1_D	P7C2_D	P2C11_D	P2C12_D	P3C3_D	P4C7_D	P4C9_D	P7C12_D
P2C3_E	P3C10_E	P3C11_E	P6C4_E	P7C1_E	P7C2_E	P2C11_E	P2C12_E	P3C3_E	P4C7_E	P4C9_E	P7C12_E
P2C3_F	P3C10_F	P3C11_F	P6C4_F	P7C1_F	P7C2_F	P2C11_F	P2C12_F	P3C3_F	P4C7_F	P4C9_F	P7C12_F
P2C3_G	P3C10_G	P3C11_G	P6C4_G	P7C1_G	P7C2_G	P2C11_G	P2C12_G	P3C3_G	P4C7_G	P4C9_G	P7C12_G
P2C3_H	P3C10_H	P3C11_H	P6C4_H	P7C1_H	P7C2_H	P2C11_H	P2C12_H	P3C3_H	P4C7_H	P4C9_H	P7C12_H

Figure 3.8 Identification of the potential pool candidates for each allele with marker Lsat46
Saturated toned squares are the results with the highest confidence.

Table 3.3 Detail of the four candidates BAC plates identified for the marker Lsat46

Address	allele	plate
P3C11_F	1	263
P7C1_H	1	661
P2C12_F	4	168
P7C12_F	4	648

Step 3 - Petri dishes

For each positive pool result, the colonies on the corresponding BAC plate were tested to isolate the colony containing the identified sequence. The Petri dishes 263 and 168 did not give any result. One BAC colony of interest was identified in plate 661, half column (A to H) number 3. One BAC colony of interest was potentially identified in plate 648 in two positions, respectively half column (I to P) number 13 or half column (A-H) number 22. As only one BAC can be present in one plate, one or the other must be due to a contamination.

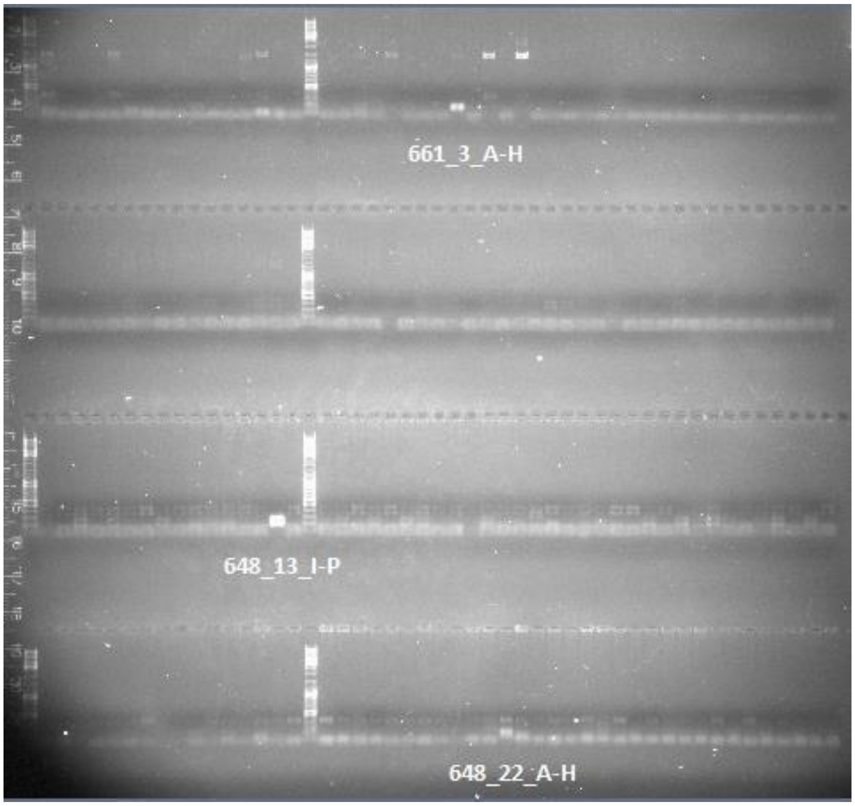


Figure 3.9 Electrophoresis gel of the PCR products from the tested BAC colonies with the marker Lsat46.
A strong band around 200bp is likely to represent the desired insert while lower weak bands are likely to represent primer dimers.

Each of these 3 candidates were investigated. A PCR plate was set with 3 x 8 colonies.

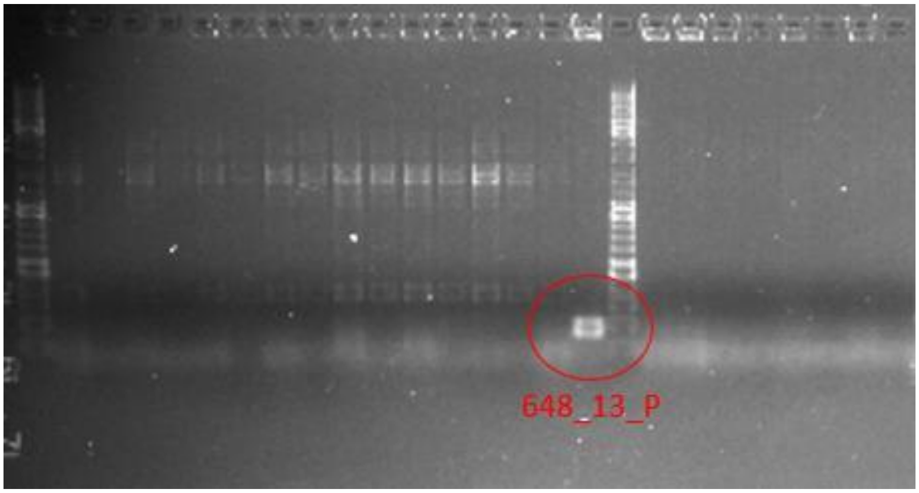


Figure 3.10 Electrophoresis gel of the PCR products from the targeted half columns of the Petri dish with the marker Lsat46.

The plate 661 did not give any amplification. However, a BAC colony of interest was successfully identified in the plate 648 despite a contamination.

Final confirmation

At the final step, the three candidates which remained that could contain the marker Lsat46 were tested using the genetic analyzer. R35 was used as a positive control. The colony 648-13-P (Petri dish n° 648, column n°13, line P) contained a DNA fragment that corresponded to the Lsat46 marker at allele 4. Considering the low level of amplification (see figure 3.11) of the two other colonies, the results appear to be due to contamination.

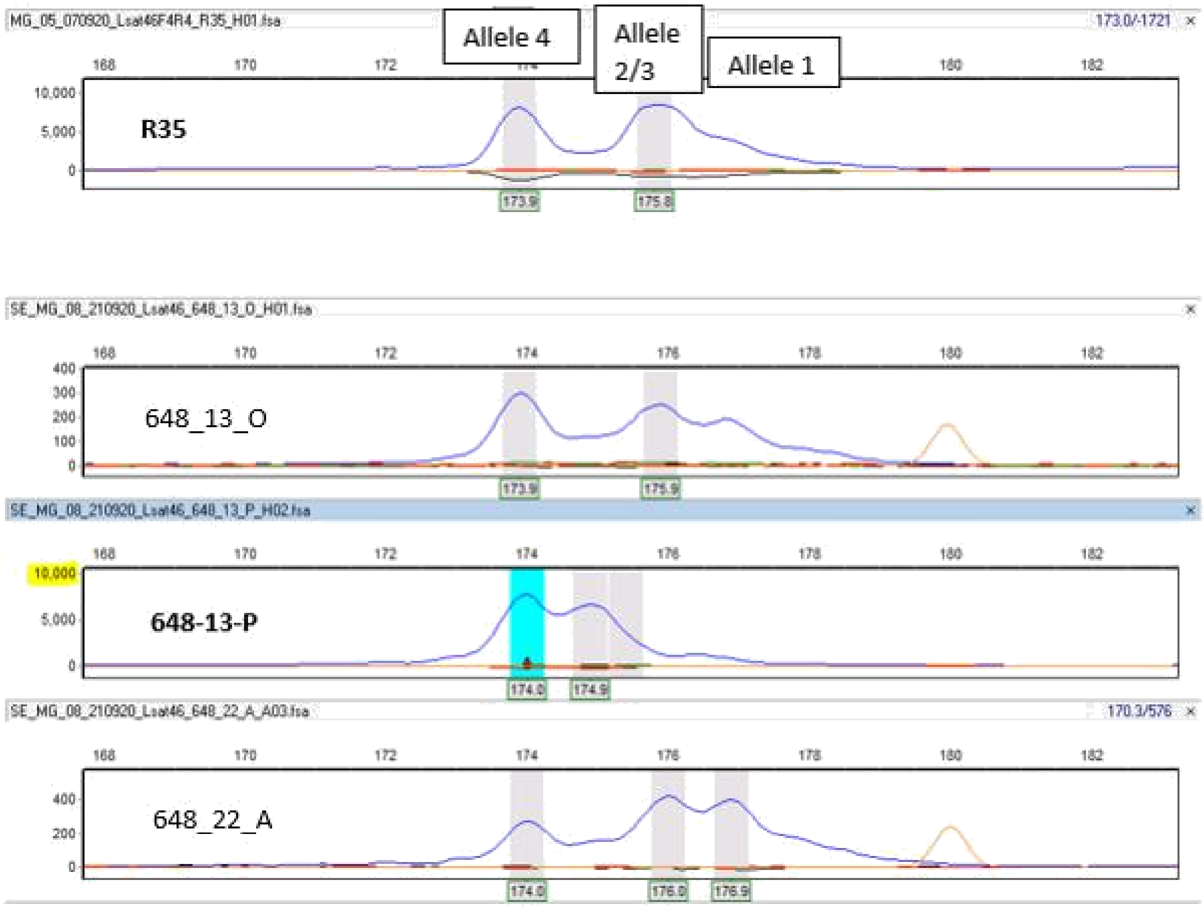


Figure 3.11 Detail of the fragment analyzer data for 3 BAC candidates and R35 as a positive control
The allele 1 is represented by a "shoulder" next to allele 2/3
The level of amplification can be seen on the y axis and reveal that only the candidate 648_13_P is relevant while the two others must be due to contamination.
This figure also indicates that the insert contained in the candidate 648-13-P is likely to correspond to the allele 4

3.2.2 Marker 44/45: Co-dominant Marker (for Alleles 1, 2&3, 4) at *LOP*

Step 1 – super-pool

84 reactions were necessary to amplify each well of the super-pool. The remaining 12 wells left in the 94-well PCR plate were used for controls: R35 (alleles 1,2,3,4), PH70 (1,3), ɣ179 (2,3,4), ɣ115 (2,3,4), ɣ138 (2,3,4), ɣ136 (2,3,4), PH511 (1,2), PH544 (1,3), PH410 (1,3), PH191 (1,2), PH325 (1,2) and No Template Control.

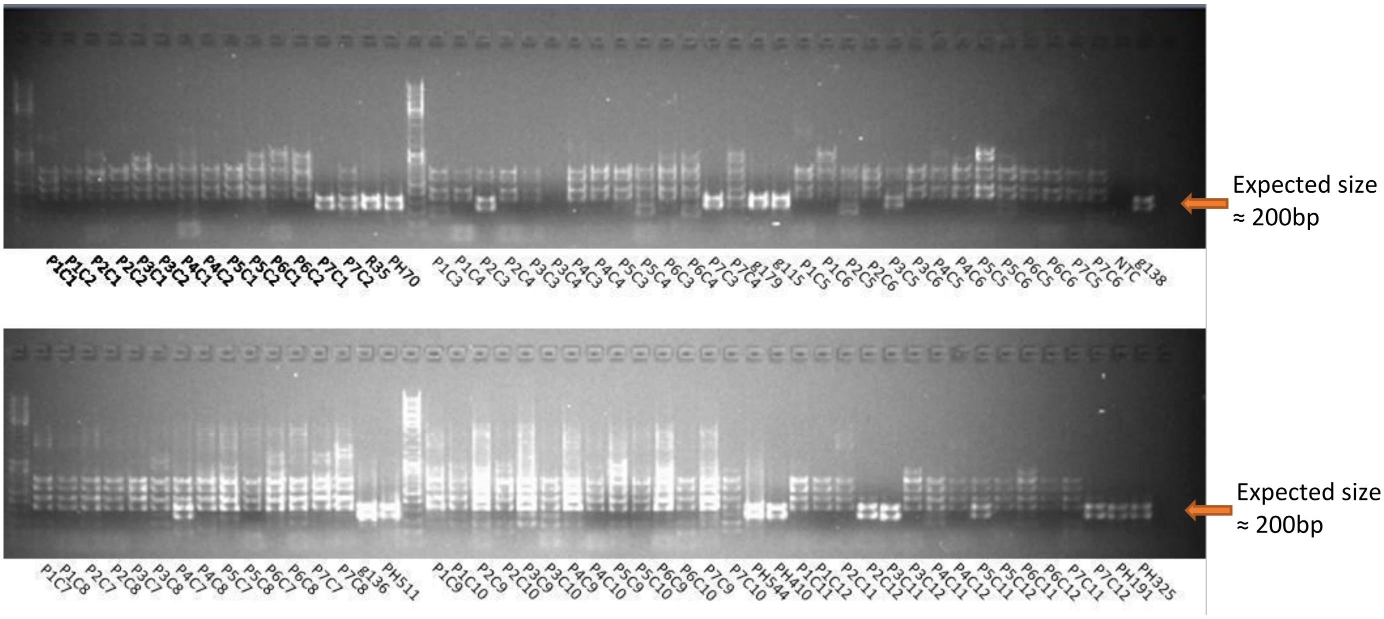
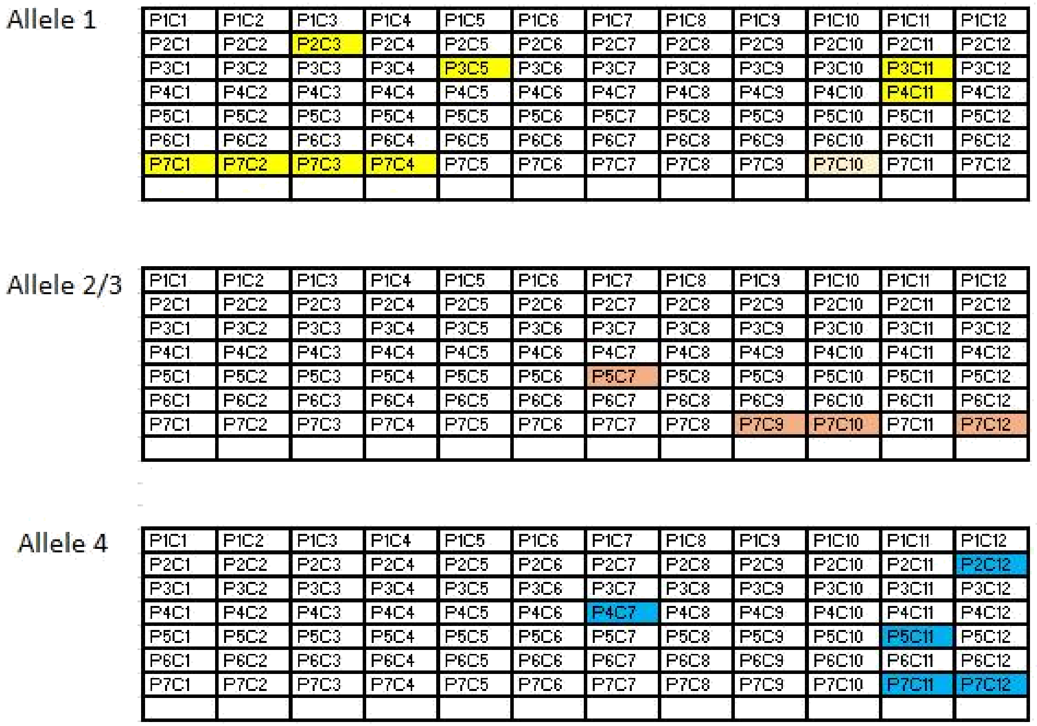


Figure 3.12 Electrophoresis gel of the PCR products from the super-pool with the marker 44/45.
Each well was identified with the number of the plate and column from which they came from. P: plate; C: column
A strong and compact band around 200bp is an indicator of the potential presence of the desired insert

Here we can clearly identify P7C1, P7C2, P2C3, P7C3, P2C12, P3C11, P5C11, P7C12 as amplifying very well. This result was then confirmed by the fragment analyzer and the alleles involved were identified.



Allele 1

P1C1	P1C2	P1C3	P1C4	P1C5	P1C6	P1C7	P1C8	P1C9	P1C10	P1C11	P1C12
P2C1	P2C2	P2C3	P2C4	P2C5	P2C6	P2C7	P2C8	P2C9	P2C10	P2C11	P2C12
P3C1	P3C2	P3C3	P3C4	P3C5	P3C6	P3C7	P3C8	P3C9	P3C10	P3C11	P3C12
P4C1	P4C2	P4C3	P4C4	P4C5	P4C6	P4C7	P4C8	P4C9	P4C10	P4C11	P4C12
P5C1	P5C2	P5C3	P5C4	P5C5	P5C6	P5C7	P5C8	P5C9	P5C10	P5C11	P5C12
P6C1	P6C2	P6C3	P6C4	P6C5	P6C6	P6C7	P6C8	P6C9	P6C10	P6C11	P6C12
P7C1	P7C2	P7C3	P7C4	P7C5	P7C6	P7C7	P7C8	P7C9	P7C10	P7C11	P7C12

Allele 2/3

P1C1	P1C2	P1C3	P1C4	P1C5	P1C6	P1C7	P1C8	P1C9	P1C10	P1C11	P1C12
P2C1	P2C2	P2C3	P2C4	P2C5	P2C6	P2C7	P2C8	P2C9	P2C10	P2C11	P2C12
P3C1	P3C2	P3C3	P3C4	P3C5	P3C6	P3C7	P3C8	P3C9	P3C10	P3C11	P3C12
P4C1	P4C2	P4C3	P4C4	P4C5	P4C6	P4C7	P4C8	P4C9	P4C10	P4C11	P4C12
P5C1	P5C2	P5C3	P5C4	P5C5	P5C6	P5C7	P5C8	P5C9	P5C10	P5C11	P5C12
P6C1	P6C2	P6C3	P6C4	P6C5	P6C6	P6C7	P6C8	P6C9	P6C10	P6C11	P6C12
P7C1	P7C2	P7C3	P7C4	P7C5	P7C6	P7C7	P7C8	P7C9	P7C10	P7C11	P7C12

Allele 4

P1C1	P1C2	P1C3	P1C4	P1C5	P1C6	P1C7	P1C8	P1C9	P1C10	P1C11	P1C12
P2C1	P2C2	P2C3	P2C4	P2C5	P2C6	P2C7	P2C8	P2C9	P2C10	P2C11	P2C12
P3C1	P3C2	P3C3	P3C4	P3C5	P3C6	P3C7	P3C8	P3C9	P3C10	P3C11	P3C12
P4C1	P4C2	P4C3	P4C4	P4C5	P4C6	P4C7	P4C8	P4C9	P4C10	P4C11	P4C12
P5C1	P5C2	P5C3	P5C4	P5C5	P5C6	P5C7	P5C8	P5C9	P5C10	P5C11	P5C12
P6C1	P6C2	P6C3	P6C4	P6C5	P6C6	P6C7	P6C8	P6C9	P6C10	P6C11	P6C12
P7C1	P7C2	P7C3	P7C4	P7C5	P7C6	P7C7	P7C8	P7C9	P7C10	P7C11	P7C12

Figure 3.13 Identification of the potential pool candidates for each allele with marker 44/45

Colored squares represent positive results. Saturated toned squares for allele 1 are the results with the highest confidence. For allele 2/3 and 4, all the results are supposedly reliable.

Step 2 – pools

12 super pools were chosen for allele 1 and allele 4, and we went back to the 7 pool plates. Each well in a pool plate corresponds to a BAC library plate.

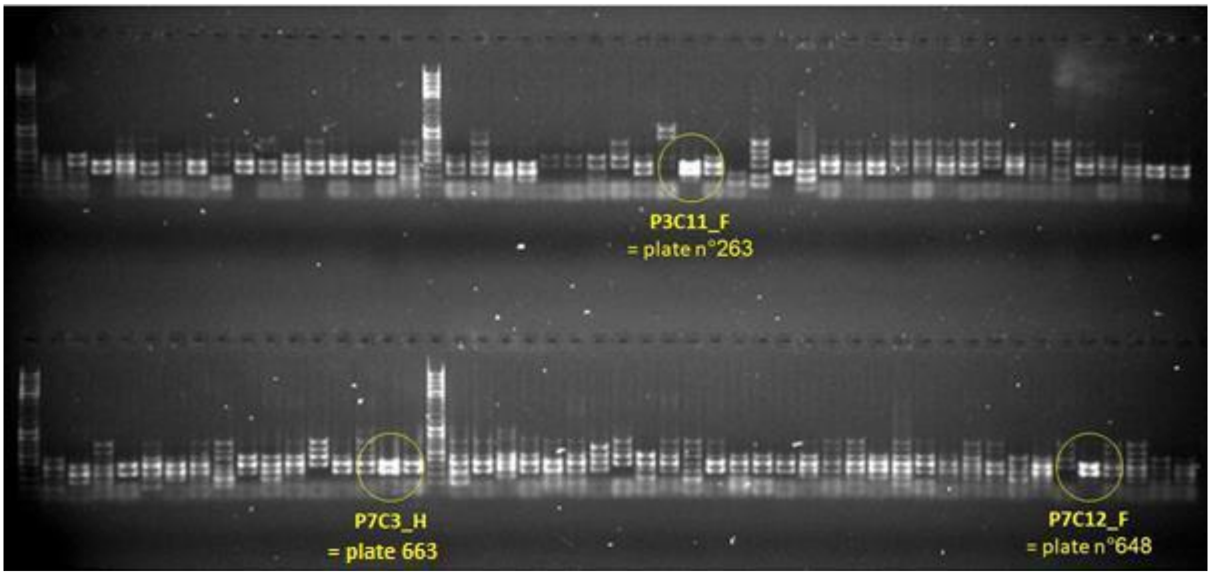


Figure 3.14 Electrophoresis gel of the PCR products from the pool plates with the marker 44/45.

The three candidates that have been found are circled in yellow.

Three BAC plate candidates were then identified, 2 for allele 1, and 1 for allele 4, corresponding to plates 263, 663, 648.

Table 3.4 Detail of the three candidates BAC plates identified for the marker 44/45.

Address	allele	plate
P3C11_F	1	263
P7C3_H	1	663
P7C12_F	4	648

A BAC colony with the 44/45 marker was already known for plate 648, so only plates 263 and 663 were further investigated.

Step 3 - Petri dishes

Plate 263 did not give any further amplification. However, for the plate 663, half row I-P was very clear. In plate 663, column 19, row I-P the band is likely to be a high molecular weight contamination.

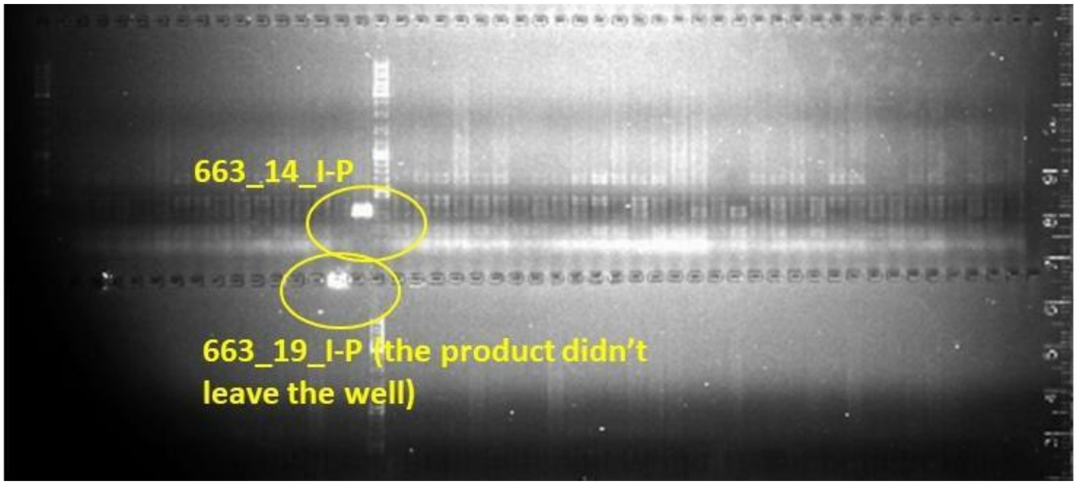


Figure 3.15 Electrophoresis gel of the PCR products from the targeted half columns of the Petri dish with the marker 44/45.

Further investigation was made for both candidates. 663_19_I-P did not give any results. However, BAC 663_14_J was identified as carrying sequence for allele 4 at LOP.

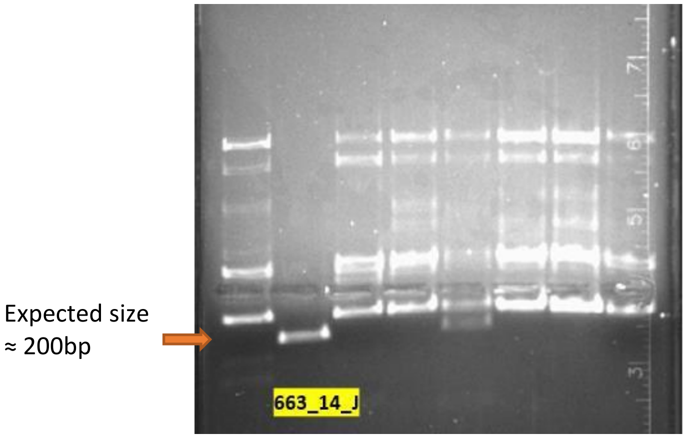


Figure 3.16 Electrophoresis gel of the PCR products from the targeted half columns of the Petri dish with the marker 44/45.

In conclusion, two new BAC clones were identified by this work. One contained the Lsat46 marker in Allele 1 at the *LOP* locus and another contained the 44/45 marker in Allele 4 at the *LOP* locus.

Chapter 4 Mapping

4.1 Introduction

4.1.1 Deletion mapping

Several statements justify the choice of mutagenesis over classical genetic mapping techniques as a tool for mapping apomixis loci.

Genetic mapping relies on genetic polymorphism between individuals to map genes for phenotypic traits. However, apomictic loci are often found to be associated with highly repetitive chromosomal regions where recombination is suppressed (Grossniklaus et al., 2001; Ozias-Akins and Van Dijk, 2007; Hand et al., 2014). Genetic mapping is also not suitable for regions of low recombination rates ("recombination cold spots") because it wouldn't meet the adequate resolution needed for map-based cloning and the mapping of the present genes (Kumar and Jain, 2018). Indeed, "radiation hybrids usually have much more deletions (or breaks) per chromosome compared to the number of recombination events observed in genetic mapping populations" (Kumar et al., 2012 as cited in Kumar and Jain., 2018), which offers a better resolution.

Furthermore, the creation of a radiation-treated mapping population is in itself a quicker and cheaper process than a genetic mapping approach. For *P. praealta,* seeds of an apomictic plant were subjected to gamma irradiation using a ^{60}Co source. Regenerant plants that were found to be stable for a mutant phenotype were then tested for components of apomixis (Catanach et al. 2006, Bicknell et al., 2001). The most time-consuming process was the selection of the most informative lines.

The use of radiation can create a panel of deletion mutants with enough variability for mapping in a system without any prior genetic markers (Catanach et al., 2006). Moreover, "clonal reproduction makes apomicts very suitable to deletion mapping" (Van Dijk 2009), and apomictic traits have a dominant inheritance.

Previous studies showed that apomixis is controlled by two independent loci associated with apomeiosis and parthenogenesis in *P.praealta* (Catanach et al. 2006). The current study focused just on parthenogenesis and therefore the mapping of the *LOP* locus associated with parthenogenesis.

Pilosella apomicts are facultative apomicts. As sexual reproduction is not completely suppressed in *Pilosella*, a capacity to form a percentage of progeny via the sexual pathway remains.

“Off-types” coming from partial sex regularly occur in test populations and often show a different ploidy level than the apomict parents (Bicknell et al., 2003; Bicknell and Koltunow, 2004; Fehrer et al., 2007).

Facultative apomixis has also been revealed to exist in natural populations of the group of *Rosaceae* (Paule et al., 2011) and *Ranunculaceae* (Cosendai and Hörandl, 2010). There is a very tight link between mode of reproduction and ploidy (Mráz and Zdvořák, 2019). Checking the ploidy level of one plant is an easy way to get back to its reproduction mode.

4.1.2 Markers

Simple Sequence Repeats (SSR) and SCAR markers were chosen for this mapping studies. Simple-sequence repeats (SSRs) are short tandem repeated motifs for which the number of repeats at a given locus can vary (Tautz, 1989). Sequence Characterized Amplified Regions (SCARs) are markers designed from sequenced regions containing allelic polymorphisms.

SSR markers have many advantages over other molecular markers (Powell et al., 1996), they are:

- Genetically co-dominant (≠ presence/absence).
- Multi-allelic (discerning between the alleles at a locus)
- Abundant (there are a lot of repeated DNA sequences in apomixis-associated genomic regions)
- Widely dispersed across the genome
- High reproducible
- Easily and automatically scored

The main value of SCAR markers is that they are designed to analyze a specific location in the chromosome. They are therefore highly specific, but may not always be co-dominant.

4.2 Material and methods

4.2.1 Previous studies

Creation of the mutants collection (Catanach et al., 2006)

A collection of gamma-induced deletion mutants has been created as a mapping resource to identify markers associated with LOA and LOP loci for *Pilosella praealta*.

The experiment started with around 5,000 seeds coming from a single apomictic clone, treated by exposure to a ^{60}Co source. The seed was then sown, and any plants that lost their ability to reproduce by apomixis were selected. In the end, Catanach et al. developed a unique set of 79 viable *Pilosella praealta* apomixis mutants, including 24 with loss of apomeiosis, 30 with loss of parthenogenesis, and 25 that had lost both traits (Catanach., 2006). This provided an opportunity to establish the roles that these loci play in seed development and to evaluate interactions between apomixis and sexual reproduction.

The descriptive nomenclature for natural off-types has been used to characterize the mutants.

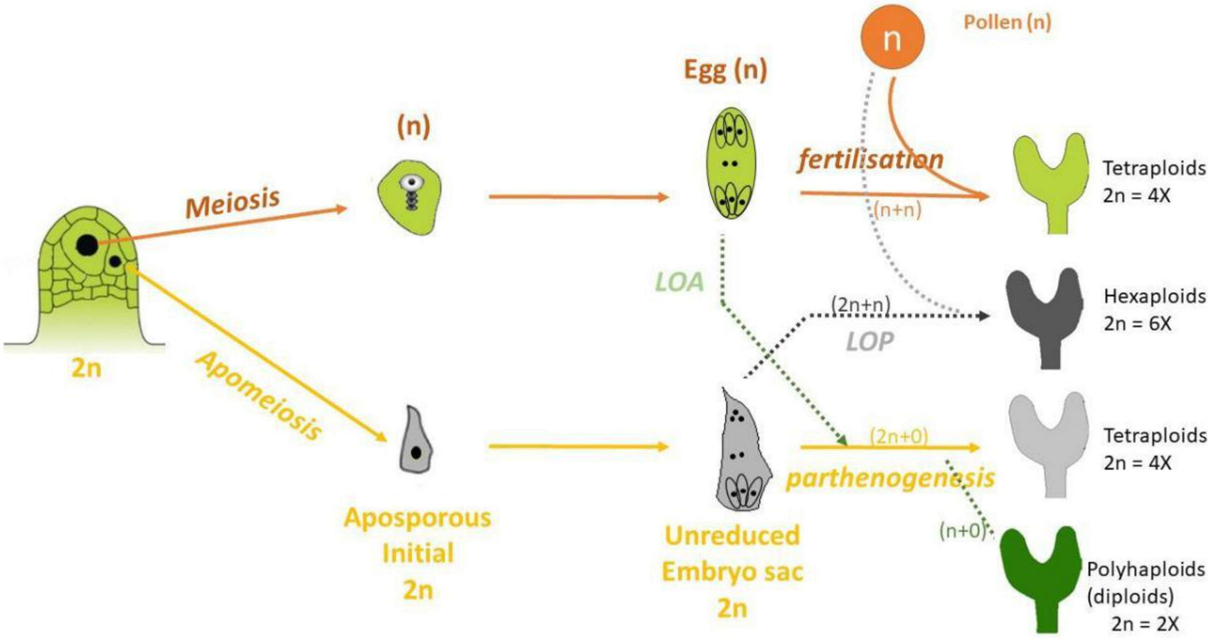


Figure 4.1 Possible pathways for the production of seed and its mutants forms.

(n+n) : loa/lop : sexual reproduction (meiosis + fertilization), gives 2n = 4X
(2n+0) : LOA/LOP : apomixis (avoidance of meiosis + avoidance of fertilization), gives 2n = 2X
(2n + n) : loa/LOP mutants : fertilization of a 2n gamete with a reduced male gamete n, gives 2n = 6X
(n+0) : LOA/lop mutants : formation of a seedling from a reduced, unfertilized (n) gamete, gives 2n = 2X

LOP mutants

In these mutants, the fertilization of an unreduced egg cell leads to an increase in ploidy (2n+n). As the mutant phenotype is 'Loss of parthenogenesis', the mutants were described as 'LOP mutants' and the locus as the '*P. praealta LOP* locus'. They still contain the LOA locus. In the current study, LOP mutants were used to detect mutation breakpoints near the LOP locus. All LOP mutants lack the LOP locus, so markers were used to test when the section of lost genomic DNA began.

LOA mutants

In these mutants, the parthenogenic development of a reduced egg cell leads to a reduction in ploidy (1n+0). As the mutant phenotype is 'Loss of apomeiosis', the mutants were described as 'LOA mutants' and the locus as the '*P. praealta LOA* locus'. They still contain the LOP locus.

Polyhaploids

Polyhaploids (see figure 17) are a useful resource to explore the LOP locus. As they are diploids, they contain only two copies of the LOP locus, which simplifies the mapping work. Unlike true apomixis that produces only clones, polyhaploids result from a meiotic process providing a valuable genomic diversity. Also, since fertilization doesn't take place in the development of polyhaploids, the meiosis only rearranges maternal alleles in these plants. The polyhaploids used were generated from an individual of *Pilosella piloselloides var praealta R35,* (commonly called R35) which was initially obtained from Dijon, France (Catanach et al., 2006). All polyhaploids have the LOP locus as it is necessary for their formation. Markers were used to detect and position recombination breakpoints flanking the LOP locus in R35.

AFLP markers

Amplified fragment length polymorphisms (AFLP) mapping was used in previous studies to distinguish the two loci implied in apomixis in *Pilosella praealta*. Comparisons of AFLP profiles identified clusters of markers linked to *LOA* and *LOP* loci that were present in the apomict but lost in the mutants (Catanach *et al.*, 2006).

4.2.2 SSR Marker design and PCR amplification

SSR markers were designed against the genomic reference sequence using Geneious to detect tandem repeats of di-nucleotides (AT, AG, CA, GT, CT, or CG). Primers were made to the conserved flanking sequence. The expectation was that each allele carries a different number of repeats, but flanked regions remained conserved enough to ensure the correct region was targeted. For each of the primers, the presence of polymorphism is assessed, as well as the position relative to the LOP locus (using the deletion mutants and polyhaploids).

The parameters for designing the primers were set as follows: primer length with ≈22 nucleotides, GC content of 40–60%, and optimum annealing temperature of at least 60°C. No self-dimerization. No off-target. amplification product size of:

1st set: 24 primers pairs producing fragments of 500-700bp

2nd set: 21 primers pair producing fragments of 250-400bp

3rd set: 10 primers pairs producing fragments of 200-400bp

4th set: 12 primers pairs producing fragments of 150 – 450 bp (not SSR)

SSR amplification was performed using 6 µL of Kapa 3g PCR mix, 0,9 µL of each primer (10 µM) (forward and reverse), 0,1 µL KAPA3G plant TAQ 0.25 unit, and 2 µL genomic DNA templates. PCR amplification was run using a T100 Thermal Cycler (BioRad, Hercules, California, USA) with the following program: 94°C for 3 min, followed by 45 cycles of 94°C for 30 s, then 30 s at the annealing temperature 60°C, 72°C for 1 min, and a final extension at 72°C for 30 min. PCR products were separated on 1% agarose gel and photographed under UV light. A fragment analyzer was then used to confirm SSR polymorphisms in amplified genomic DNA fragments and to precisely measure the size of the polymorphism.

1st set: 24 primers pairs producing fragments of 500-700bp

After a trial run of 24 pairs of SSR primers, 7 of them produced clear bands, stable amplification and potential polymorphisms for further analysis.

2nd set: 21 primers pair producing fragments of 250-400bp

It was realized that the target product's size was too big with the previous primers set to observe polymorphism in the fragment analyzer, so the size was reduced to 250-400 bp.

After a trial run, 11 primer pairs were selected.

3rd set: 10 primers pairs producing fragments of 200-400bp

Another approach has been used for this third set. All 10 of the primers were run through the fragment analyzer using a small testing set of templates. Only the ones showing clear differential pattern were selected for use with the extended template set.

4th set: 12 primers pairs producing fragments of 150 – 450 bp (not SSR)

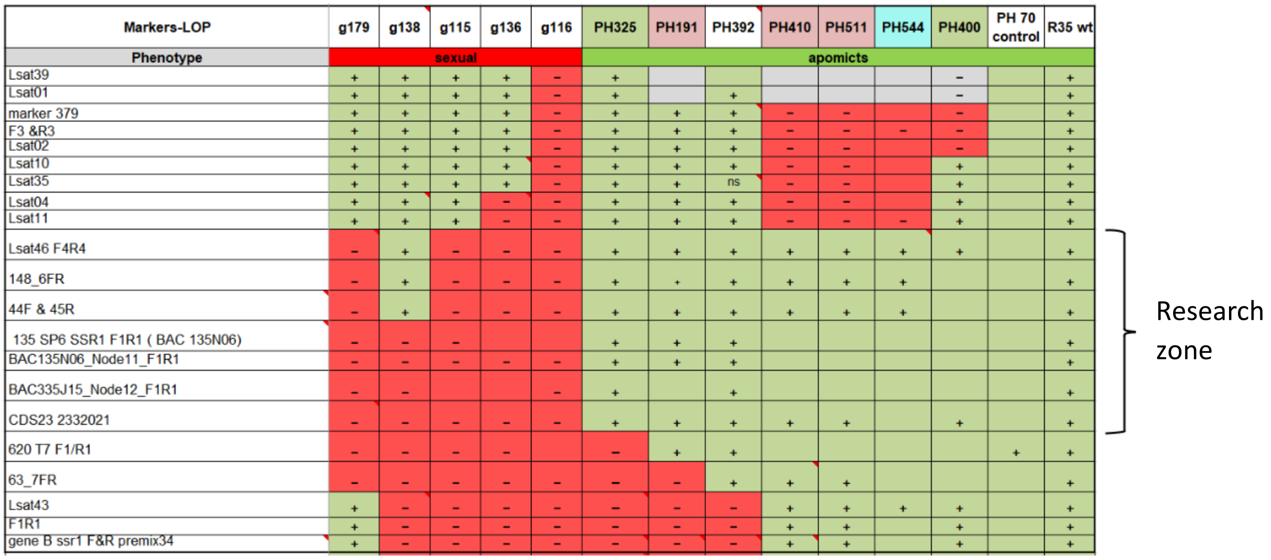
As we approached the limit of the LOP locus, it became too difficult to design more SSR primers. The last primer set was developed using an automatic function of Geneious for designing primers within a constrained sequence range. A set of 12 of the primers was run through the fragment analyzer against a small testing set of templates. Only the ones showing a clear differential pattern were selected for use with the extended template set.

4.3 Results and discussion

Prior to this research, the region of the LOP locus was determined to be between the Lsat11 and 620T7 F1/R1 markers, an interval of 1.4 Mb. This estimate was based on marker positions against the SHASTA assembled Illumina sequence data. The positions of the markers are indicated in Table 4.1.

Table 4.1 State of the mapping at LOP prior to this study. The markers are ordered according to their position on the genome. The markers are ordered according to their position on the genome.

In green: expression of a marker for a particular individual at a precise position
In red: absence of expression of a marker for a particular individual at a precise position



Markers-LOP	g179	g138	g115	g136	g116	PH325	PH191	PH392	PH410	PH511	PH544	PH400	PH 70 control	R35 wt
Phenotype	sexual					apomicts								
Lsat39	+	+	+	+	−	+		🟩				−	🟩	+
Lsat01	+	+	+	+	−	+		+				−	🟩	+
marker 379	+	+	+	+	−	+	+	+	−	−	🟥	−	🟩	+
F3 &R3	+	+	+	+	−	+	+	+	−	−	−	−	🟩	+
Lsat02	+	+	+	+	−	+	+	+	−	−	🟥	−	🟩	+
Lsat10	+	+	+	+	−	+	+	+	−	−	🟥	+	🟩	+
Lsat35	+	+	+	+	−	+	+	ns	−	−	🟥	+	🟩	+
Lsat04	+	+	+	−	−	+	+	+	−	−	🟥	+	🟩	+
Lsat11	+	+	+	−	−	+	+	+	−	−	−	+	🟩	+
Lsat46 F4R4	−	+	−	−	−	+	+	+	+	+	+	+	🟩	+
148_6FR	−	+	−	−	−	+	+	+	+	+	+	🟩	🟩	+
44F & 45R	−	+	−	−	−	+	+	+	+	+	+	🟩	🟩	+
135 SP6 SSR1 F1R1 (BAC 135N06)	−	−	−	🟥	🟥	+	+	+	🟩	🟩	🟩	🟩	🟩	+
BAC135N06_Node11_F1R1	−	−	−	−	−	+	+	+	🟩	🟩	🟩	🟩	🟩	+
BAC335J15_Node12_F1R1	−	−	🟥	🟥	−	+	🟩	+	🟩	🟩	🟩	🟩	🟩	+
CDS23 2332021	−	−	−	−	−	+	+	+	+	+	🟩	+	🟩	+
620 T7 F1/R1	−	−	−	−	🟥	−	+	+	🟩	🟩	🟩	🟩	+	+
63_7FR	−	−	−	−	−	−	−	+	+	+	🟩	🟩	🟩	+
Lsat43	+	−	−	−	−	−	−	−	+	+	+	+	🟩	+
F1R1	+	−	−	−	−	−	−	−	+	+	🟩	+	🟩	+
gene B ssr1 F&R premix34	+	−	−	−	−	−	−	−	+	+	🟩	+	🟩	+

Research
zone

As all polyploids arise from the action of LOP, they should all have a functional copy of this locus. The diagram only illustrates the polyhaploids found to have a recombinant breakpoint close to LOP. Over 2,000 polyhaploids were screened for recombination, so the seven listed polyhaploids all have breakpoints within 1cM of the locus. However, as recombination is suppressed in this region, a centiMorgan is still greater than 1Mbp at this locus.

One recombination breakpoint was determined to be between markers CDS23_32021 and 620T7 F1/R1 because one polyploid did not exhibit the 620T7 marker, but they all exhibited the CDS23_32021 marker. The CDS23_32021 and 620T7 F1/R1 markers are 95kb apart, indicating that the most proximal recombination breakpoint on this side of LOP lies in that interval. Consequently, most of the research effort was placed on determining the closest recombination breakpoint on the other side of LOP.

All of the deletion mutants had lost a large part of the LOP region, and they all exhibited a sexual-like phenotype. Some almost certainly had lost the entire locus for parthenogenesis, but LOP may also be a complex locus containing many genes, and some mutants may have just lost a critical part, but not all of it. g138, for example, has a breakpoint well inside of the interval found in all of the polyhaploids (Table 4.2), and it has a sexual phenotype. It is not clear if this means that the whole of LOP lies in the deleted region of g138 or that only a critical part of it does. As the deletion mutants have lost LOP, they were used as negative controls throughout the study.

The first set of primers tested gave two markers: 445_AG and 606_CT (see appendixes A.1 and A.2). 445_AG was clear enough but did not provide any more information than we already had from previous studies as it was very close to Lsat11. 606_CT was unclear, the PCR amplification failed for PH511 and there was a doubt for PH410.

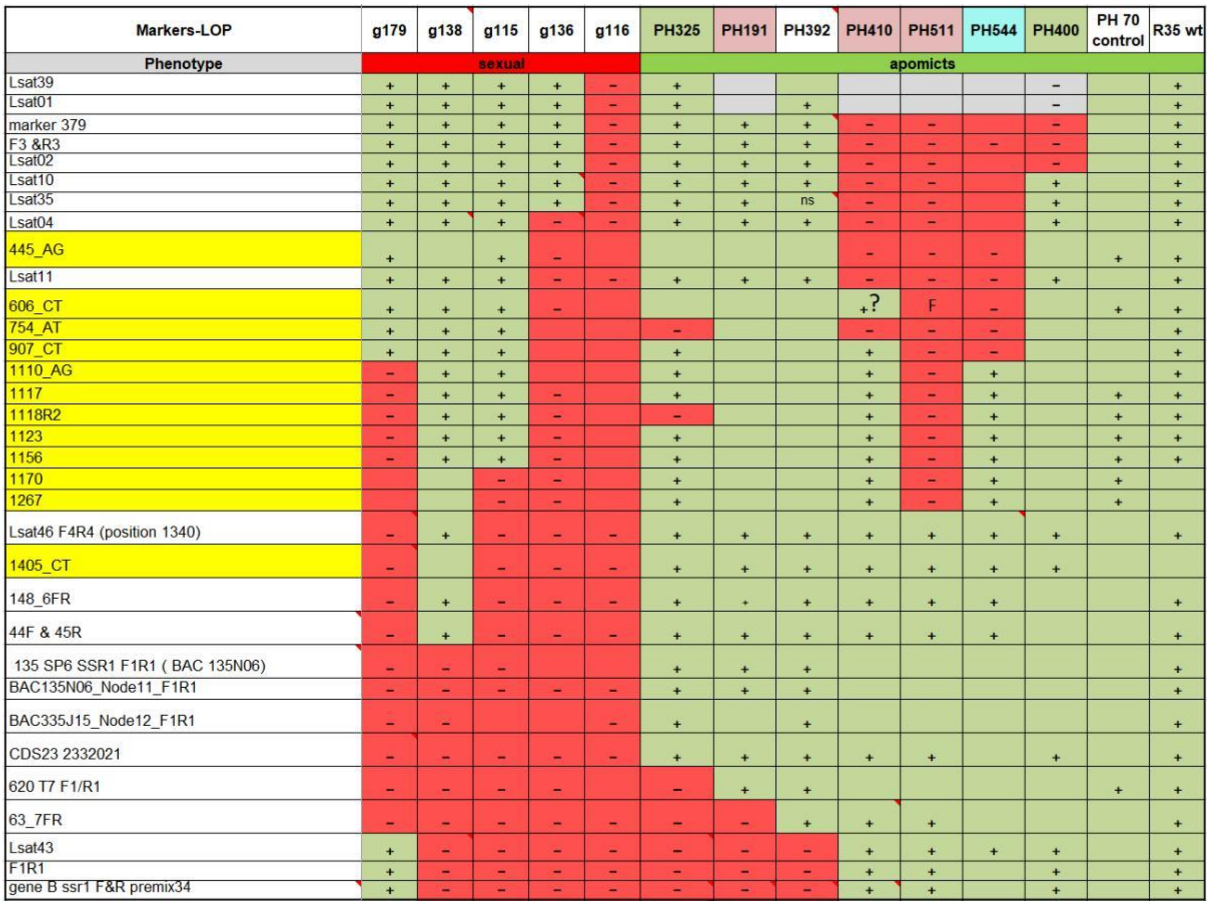
The second set gave three markers: 754_AT; 907_CT; 1110_AG (see appendixes A.3, A.4 and A.5). 907_CT eliminated one of the polyploids (PH410) because it was only expressed in PH511 and PH544.

The third set gave four markers: 1117, 118R2, 1123, 1156 (see appendixes A.6; A.7; A.8 and A.9) and the final set gave two more markers: 1170 and 1267 (see appendixes A.10 and A.11).

After the data from the four primer sets was collated, it was found that the recombination breakpoint in the final polyhaploid (number 511) was positioned between markers 1267 and Lsat46 (Table 4.2). The interval between these markers is 70kb.

Table 4.2 Mapping of the LOP locus of Pilosella praealta

In green: presence of a marker for a particular individual at a precise position
In red: absence of a marker for a particular individual at a precise position
In yellow: markers developed in this study



Markers-LOP	g179	g138	g115	g136	g116	PH325	PH191	PH392	PH410	PH511	PH544	PH400	PH 70 control	R35 wt
Phenotype	sexual					apomicts								
Lsat39	+	+	+	+	−	+						−		+
Lsat01	+	+	+	+	−	+		+				−		+
marker 379	+	+	+	+	−	+	+	+	−	−		−		+
F3 &R3	+	+	+	+	−	+	+	+	−	−	−	−		+
Lsat02	+	+	+	+	−	+	+	+	−	−		−		+
Lsat10	+	+	+	+	−	+	+	+	−	−		+		+
Lsat35	+	+	+	+	−	+	+	ns	−	−		+		+
Lsat04	+	+	+	−	−	+	+	+	−	−		+		+
445_AG	+		+	−					−	−	−		+	+
Lsat11	+	+	+	−	−	+	+	+	−	−	−	+		+
606_CT	+	+	+	−					+?	F	−		+	+
754_AT	+	+	+			−			−	−	−			+
907_CT	+	+	+			+			+	−	−			+
1110_AG	−	+	+			+			+	−	+			+
1117	−	+	+	−		+			+	−	+		+	+
1118R2	−	+	+	−		−			+	−	+		+	+
1123	−	+	+	−		+			+	−	+		+	+
1156	−	+	+	−		+			+	−	+		+	+
1170			−	−		+			+	−	+		+	
1267			−	−		+			+	−	+		+	
Lsat46 F4R4 (position 1340)	−	+	−	−	−	+	+	+	+	+	+	+		+
1405_CT	−		−	−	−	+	+	+	+	+	+	+		
148_6FR	−	+	−	−	−	+	+	+	+	+	+			+
44F & 45R	−	+	−	−	−	+	+	+	+	+	+			+
135 SP6 SSR1 F1R1 (BAC 135N06)	−	−	−			+	+	+						+
BAC135N06_Node11_F1R1	−	−	−	−	−	+	+	+						+
BAC335J15_Node12_F1R1	−	−			−	+		+						+
CDS23 2332021	−	−	−	−	−	+	+	+	+	+		+		+
620 T7 F1/R1	−	−	−	−		−	+	+					+	+
63_7FR	−	−	−	−	−	−	−	+	+	+				+
Lsat43	+	−	−	−	−	−	−	−	+	+	+	+		+
F1R1	+	−	−	−	−	−	−	−	+	+		+		+
gene B ssr1 F&R premix34	+	−	−	−	−	−	−	−	+	+		+		+

In conclusion, the mapping using recombination breakpoints in the polyhaploids set the estimated limits of the LOP locus between markers 620 T7 F1/R1 and 1267, an interval of 647kb. If the deletion breakpoint of g138 is taken into account, and assuming that LOP is a simple locus, this would imply that the locus lies between markers 620 T7 F1/R1 and 44F/45R, an interval of 527kb.

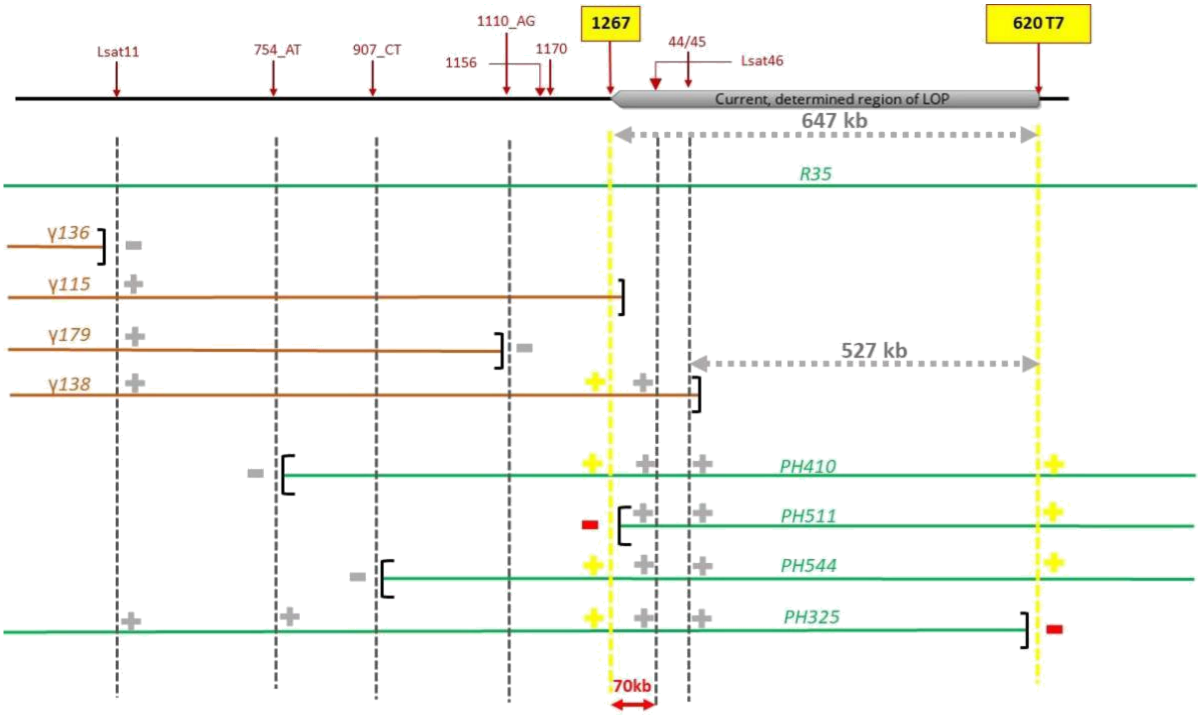


Figure 4.2 Detail of the markers and their position around the LOP locus

In brown: the deletion mutants
In green: the polyploids

1.1.1.1 Cross between a *P.praealta* deletion mutant and *P.aurantiaca*

In a previous study (unpublished), a cross was made between a deletion mutant of *P.praealta* (g138) and an apomictic tetraploid of another species (*P.aurantiaca*). The purpose of this cross was to identify the dominant allele at *LOP* in *P.aurantiaca* as it was reasoned that all of the apomictic progeny should have this allele, and all the non-apomictic progeny should not have it. For this to work, a co-dominant marker was needed that could be used to identify the different alleles of *LOP* in this population. The SSR marker 1170, developed in this study, appears very valuable for this use because it is very clear, polymorphic, co-dominant, and it is close to the *LOP* locus (see Appendix A.10).

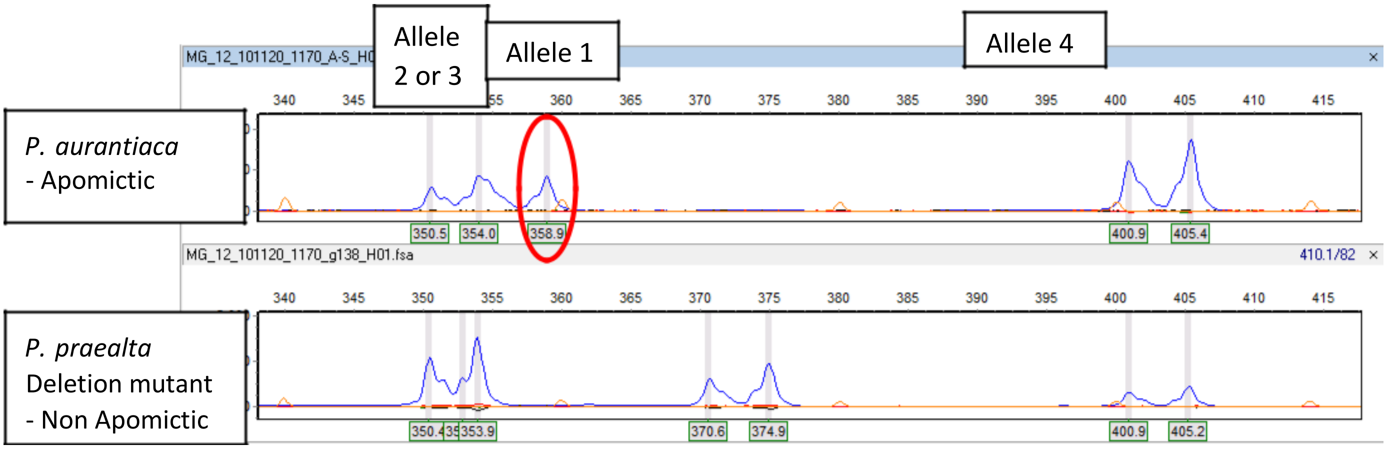


Figure 4.3 Detail of the fragment analyzer data for P.aurantiaca and the P.praealta deletion mutant g138 for the marker 1170 near the LOP locus.

The peak linked to the dominant allele at LOP (Allele 1) is circled in red. The allele numbering system was based on work in a previous study.

The peak representing the LOP allele for *P. aurantiaca* was found by process of elimination. This peak should be present in the data from all of the apomictic individuals and absent from the data from all of the non-apomictic individuals. Nine offspring from this cross were investigated (6 apomicts and 5 non-apomicts) using the fragment analyzer.

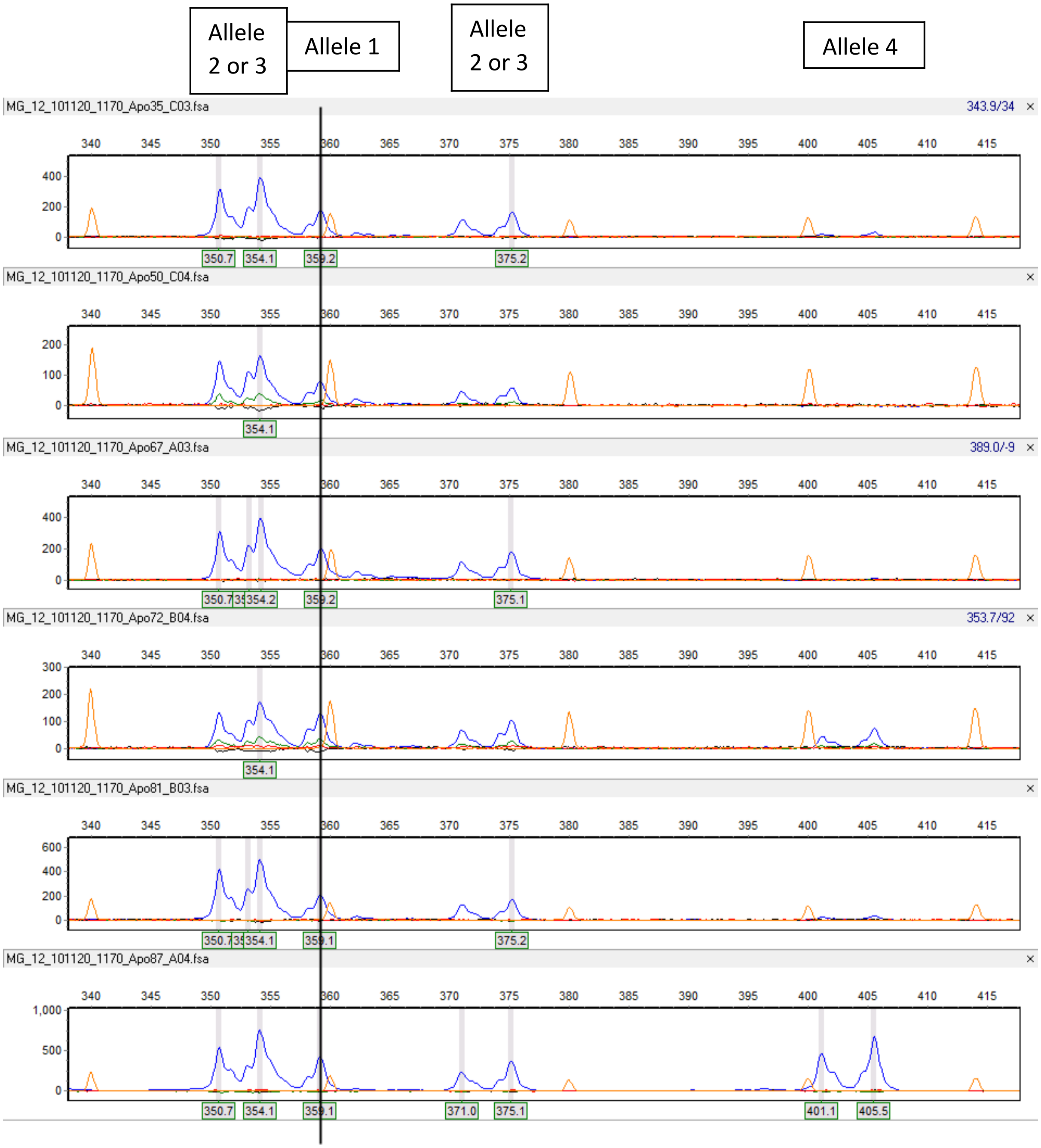


Figure 4.4 Fragment analyzer data for the apomictic offspring from a cross between the P. praealta deletion mutant g138 and P. aurantiaca, using the SSR marker 1170.

The dashed line represents the dominant allele at LOP (Allele 1). The allele numbering system was based on work in a previous study.

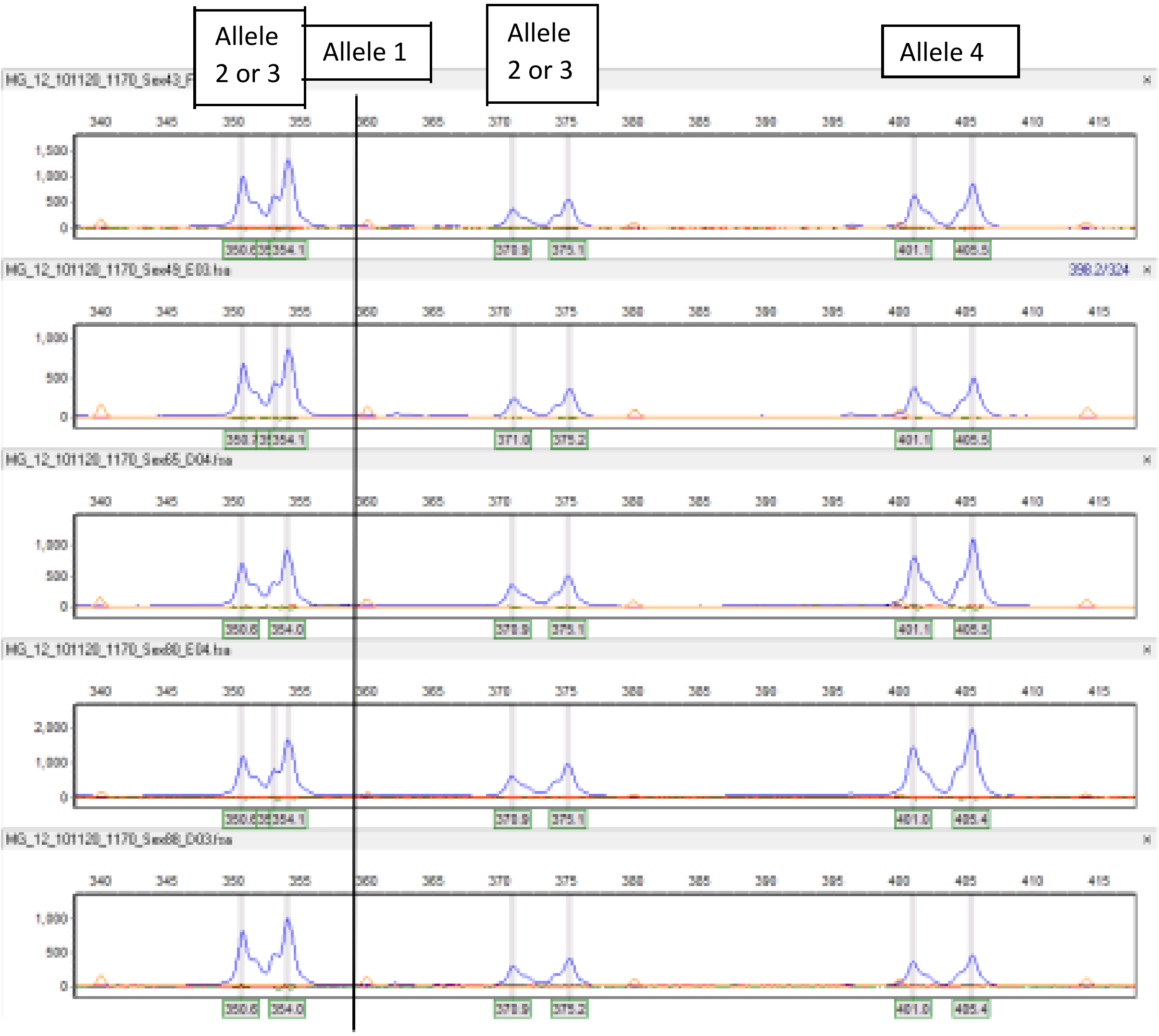


Figure 4.5 Fragment analyzer data for the non-apomictic offspring from a cross between the P. praealta deletion mutant g138 and P. aurantiaca, using the SSR marker 1170.

The dashed line represents the expected position of the dominant allele at LOP (Allele 1).

In conclusion, these results demonstrate that the 1170 marker is well suited for identifying the dominant allele at *LOP* in the species *P. aurantiaca*. It also confirms that the experimental phenotypic evaluations (apomict and non-apomict) were right as it brings genetic evidence to support the observed difference between the apomictic and non-apomictic plants in this population.

Chapter 5 Allelic diversity

5.1 Introduction

Apomixis can be seen as an aberrant way of reproduction in a world ruled by sexual reproduction. Apomixis is however known in over 400 taxa of angiosperms, many of which are very successful competitors in their habitats, so it is clear that apomixis must confer some ecological advantages for the trait to have persisted in so many taxa over evolution time. Sexual reproduction is particularly efficient in ensuring adaptability to rapidly changing environments thanks to the diversity created by recombination in each generation. Crow and Kimura (1965) noted that sexual reproduction would be beneficial over apomixis under certain conditions, such as a large population size subjected to a frequent occurrence of mutations. However, apomixis is a preferential mode of reproduction over sexual reproduction when the environment is stable (Tucker et al., 2009). Elite apomictic genotypes are cloned, and they multiply to settle down and often dominate in their environment. Where pollination is not required, this also provides a successful colonization strategy (Baker, 1959 ; Baker et Stebbins, 1965). This advantage also applies in environments where pollinators are scarce, such as mountains and high latitudes, and apomicts are particularly common in these places. The phenomenon of the unequal distribution of apomixis is called “geographical parthenogenesis” (Hörandl et al., 2008).

It has been described that asexual lineages accumulate deleterious mutations, and these can contribute to the ultimate extinction of a genotype (a phenomenon called ‘Müller’s ratchet’) (Müller, 1932). However, residual sexuality typically persists in apomictic populations (Van Dijk, 2009) and this acts against the effects of Müller’s ratchet. Sexuality occurs in the form of entirely sexual individuals within populations dominated by apomicts, and also as facultative apomixis where some of the seeds forming on a plant are sexually derived while others are produced by apomixis. Furthermore, most apomicts still produce reduced pollen even if fertilization is not required in the apomictic process. The result of all these mechanisms is that new genotypes do constantly arise in these populations to replace obsolete ones and some recombination does take place to ensure change over time.

According to the “Fisher-Muller” model, sexual reproduction is a paradox. It costs energy, and the beneficial traits it might create can be quickly lost through the generations due to a large amount of recombination. This disadvantage of sexual reproduction is usually referred to as the “cost of males” (Smith, 1971) or “the cost of meiosis” (Williams, 1975). Darlington (1939) proposed that apomixis would provide an escape from sterility, even though Darlington considered this would be “an escape into a blind alley of evolution”.

It is now acknowledged that apomictic plants clearly have found a way to thrive probably because "apomixis preserves beneficial combinations of unlinked alleles in every generation for as long as apomictic genotypes persist in the population" (Sailer et al., 2020).

Allelic sequence divergence (ASD) is a phenomenon where alleles at a locus diverge in sequence over time. It occurs at most loci in plants that are open-pollinated because sexual reproduction homogenises allelic combinations within a population. It is believed to be repressed in apomictic plants as a result of repressed meiotic recombination (Corral et al., 2009). The existence of residual sexuality in apomicts, however, means this hypothesis is probably too simplistic because recombination is occasionally happening in these plants, as described above. That said, there are some parts of the genome of an apomict that are likely to be unaffected by residual sexuality, and they are the loci that control apomixis itself. These loci, which typically act in a dominant manner, must be intact for apomixis to occur. When they are inherited, the plant is an apomict and therefore reproduces clonally. If they are not inherited, or they are broken by recombination, then sexual reproduction (the recessive state) occurs. In apomictic populations, therefore, we predict that dominant alleles conferring apomixis will remain very similar (low levels of ASD) yet recessive alleles that confer sexuality will have higher levels of ASD. This hypothesis has never been tested before experimentally.

This chapter is about the sequencing of 34 individuals and different species of the *Asteraceae* family for three genes. One is part of the LOP locus, the three others being highly conserved genes placed near the LOP locus. The purpose of this study is to highlight the allelic diversity among the apomict and sexual species and discover how the alleles are distributed depending on their similarity from an evolutionary point of view. The technology used for the sequencing is the Nanopore MinION sequencer. The sequencing data were then used to build a tree.

5.1.1 Nanopore sequencing (MinION sequencer)

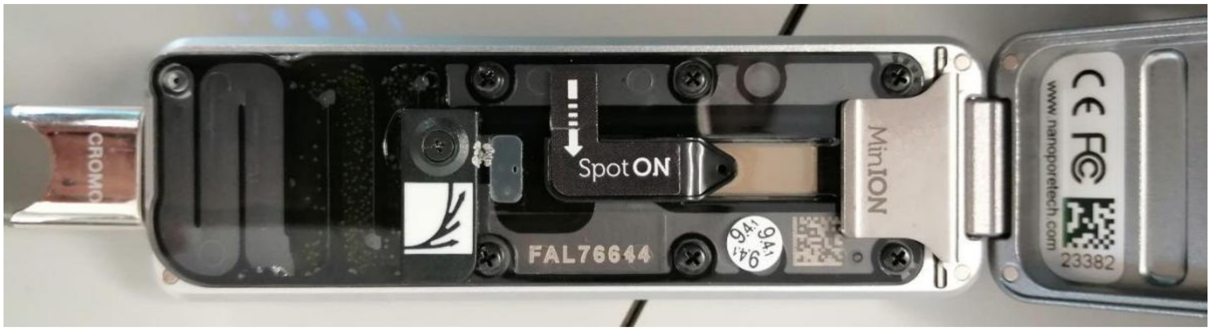


Figure 5.1 Nanopore MinION sequencer

Nanopore MinION is a pocket-size device commercialized by Oxford Nanopore Technologies. It is a long-read sequencing technology that allows a good contiguous assembly of a de novo genome. This technology can sequence long reads up to 100+ kilobase (Jain et al., 2018).

DNA molecules pass through a protein nanopore, leading to changes in the ionic current across a membrane (Bowden et al., 2019).

Table 5.1 Advantages and disadvantages of the use of Nanopore sequencer

Advantages	Disadvantages
Direct sequencing of nucleic acids allowing the detection of DNA modification Real-time data acquisition and analysis Ability to generate long reads	Requires quality DNA High cost Low throughput High insertion/deletion error rates associated with base-calling

(Tyler et al., 2018)

5.2 Material and methods

5.2.1 Plant material and DNA isolation

The plants used in the study are listed in Table 5.2. A total of 34 individuals were tested for each of the four genes. The ploidy of each plant was tested using flow cytometry to determine the number of alleles expected. Deletion mutants and segregants were used to determine which alleles represented the dominant allele associated with apomixis (designated allele 1). DNA was isolated using the method described in the General Materials and Methods section.

Table 5.2 List of the individuals sequenced with the abbreviation used in the rest of the thesis, their phenotype and ploidy.

Name	**Abreviation**	**phenotype**	**ploidy**
Pilosella piloselloides var praealta R35	R35	Apomictic	4x
Pilosella aurantiaca	A-S	Apomictic	4x
Pilosella onegensis	On	Sexual	2x
Pilosella officinarum (Caen)	P-C	Sexual	3x
Pilosella caespitosa (Tekapo)	C-T	Apomictic	4x
Pilosella lactucella	Au	Apomictic	5x
Hieracium policheae	Po	Apomictic	5x
Hieracium lepidulum	Le	Apomictic	5x
Hieracium murorum	Mu	Apomictic	5x
Pilosella cymosa	Cy	Apomictic	5x
Pilosella officinarum (Dijon)	P-D	Apomictic	4x
Pilosella piloselloides D36	D36	Apomictic	4x
Pilosella glacialis G45	G45	Apomictic	5x
Pilosella LP	Pi-LP	Apomictic	4x
Hieracium Lactucella	La		4x
Hieracium pelletarianum	L2	Sexual	2x
D2-10	D2-10	Apomictic	2x
Taraxacum officinale	To	Apomictic	3x
Hypochoeris radicata	Hr	Sexual	?
Pilosella glacialis deletion mutant γ5	G45-5	Sexual	5x
Pilosella glacialis deletion mutant γ57	G45-57	Sexual	5x
Pilosella glacialis deletion mutant γ83	G45-83	Sexual	5x
Pilosella glacialis deletion mutant γ110	G45-110	Sexual	5x
Pilosella piloselloides var praealta deletion mutant	γ116	Sexual	4x
Hybrids (*P.aurantiaca* x *P.piloselloides praealta* γ138)	Apo 50	Apomictic	6x
Hybrid A-S x γ138	Apo 72	Apomictic	6x
Hybrid A-S x γ138	Apo 87	Apomictic	6x
Hybrid A-S x γ138	Apo 67	Apomictic	6x
Hybrid A-S x γ138	Apo 35	Apomictic	6x
Hybrid A-S x γ138	Apo 81	Apomictic	6x
Hybrid A-S x γ138	Sex 49	Sexual	6x
Hybrid A-S x γ138	Sex 65	Sexual	6x

The case of G45-110

G45-110 is a genetically transformed mutant that integrated a T-DNA containing elements of the Activator/Dissociator (Ac/Ds) transposons of maize. It had lost apomixis but continued to form seeds sexually. Indeed, aposporous initials were seen to begin differentiation but failed to divide. The meiotic apparatus in this plant continued to develop normally, and a compensatory increase in sexuality was observed (Ross Bicknell et al., 2001).

5.2.2 Candidate genes

Four genes were strategically chosen for amplicon sequence comparison according to their position compared to the LOP locus. One gene, *PpPAR*, can induce egg cell division without fertilization and was first described in Taraxacum (Underwood et al., 2020), which is also part of the Asteraceae family. It appears to be placed right in the middle of the LOP locus. It is believed to be involved in the expression of parthenogenesis. *Feronia* encodes a plasma-membrane-localized receptor kinase involved in growth and development. It maps to a locus approximately 4Mbp to one side of *PpPAR*. *Eif3e* encodes a highly conserved eukaryotic translation initiation factor, mapping to a locus approximately 4Mbp to the other side of *PpPAR*. The rRNA ITS sequence was used as a control for the taxonomic relationships between these plants. Ribosomal RNA genes are highly conserved and widely explored over species and the Asteraceae family. Lettuce is one of the rare Asteraceae that is also a crop, which makes it the member of this family that might be the most observed. The sequence used as a reference for this study was from *Lactuca seriola,* which is very close relative to *Lactuca sativa*; the synteny is high, especially for the ITS interval (around 100% similarities).

Although all four genes were amplified, barcoded and sequenced for all samples, only rRNA ITS and the PAR gene were analysed for ASD.

5.2.3 Primer selection

Several primers located around each target gene have been tested. The cheapest and quickest way to assess the validity of a primer in the first place is to run an electrophoresis gel. The first PCR was run according to protocol described in the next sub chapter with 40 cycles to ensure a clear amplification. The main criteria are:

- A clear and strong band without a smear to guarantee a clear result
- Consistency of the strength of the band throughout as many samples as possible

A first selection can then be made according to these criteria. The reduced set of primers is tested on a gel again with 25 cycles to verify that the PCR products are amplified. If the gel gives good results, the amplicons are then tested on a fragment analyzer to check their purity.

Eif3e

Primers for the *Eif3e* gene were developed in previous studies (McGee, 2013).

Feronia

Ten primers were tested for the *Feronia* gene with a test set of eight templates, including a positive control, negative control and six different species. Four primers were chosen for further investigation because they had clear bands, only one product or at least very distinct products. A new PCR was run with an extended test set of 24 DNA samples.

rRNA

Ribosomal primers have been widely studied, especially in *Arabidopsis* and are therefore recognized to be reliable primers. Five pairs of primers were chosen from the literature associated with their PCR protocol. They were directly run for an extended test set of 30 DNA samples, and all gave excellent results. The primer pair chosen was ITS5A/ITS4 (Downie and Katz-Downie, 1996; White et al., 1990)

PpPAR

Ten primers around the PpPAR gene generated using Geneious were tested with a test set of 5 DNA samples. Two primers were selected and tested with an extended set of 61 DNA templates. The ones that gave the strongest bands were chosen for the studies.

5.2.4 Amplicons barcoding

Sequencing is expensive, and it is common to limit the number of runs to limit costs. Barcodes, small specific sequences of around ten bases, are attached to each amplicon to facilitate multiplex sequencing of numerous samples and genes in the same run.

Barcode sequences are encoded at the termini of linear DNA through PCR protocols. The products are then pooled together for the sequencing. Each sample can then be de-multiplexed and sequences assigned to their original sample based on their barcode sequences.

For this study, 136 amplicons had to be tested (34 individuals/species tested for four genes). It was easy to identify each of the four genes because the primers used were different. However, it was necessary to be able to identify the 34 different plants used.

Instead of creating 136 barcodes for each gene + species pair, it was decided to develop 34 species barcodes and attach them to previously tagged primers (see figure 5.2). Three consecutive PCR reactions were set up to obtain amplicons with barcodes attached, resulting in a total of 408 reactions. The first PCR reaction amplified the selected genomic DNA region (not represented in the figure). The second PCR attach used tagged primers to add common tags to the amplicons. The last PCR attached the sample-specific barcodes using the tag sequence.

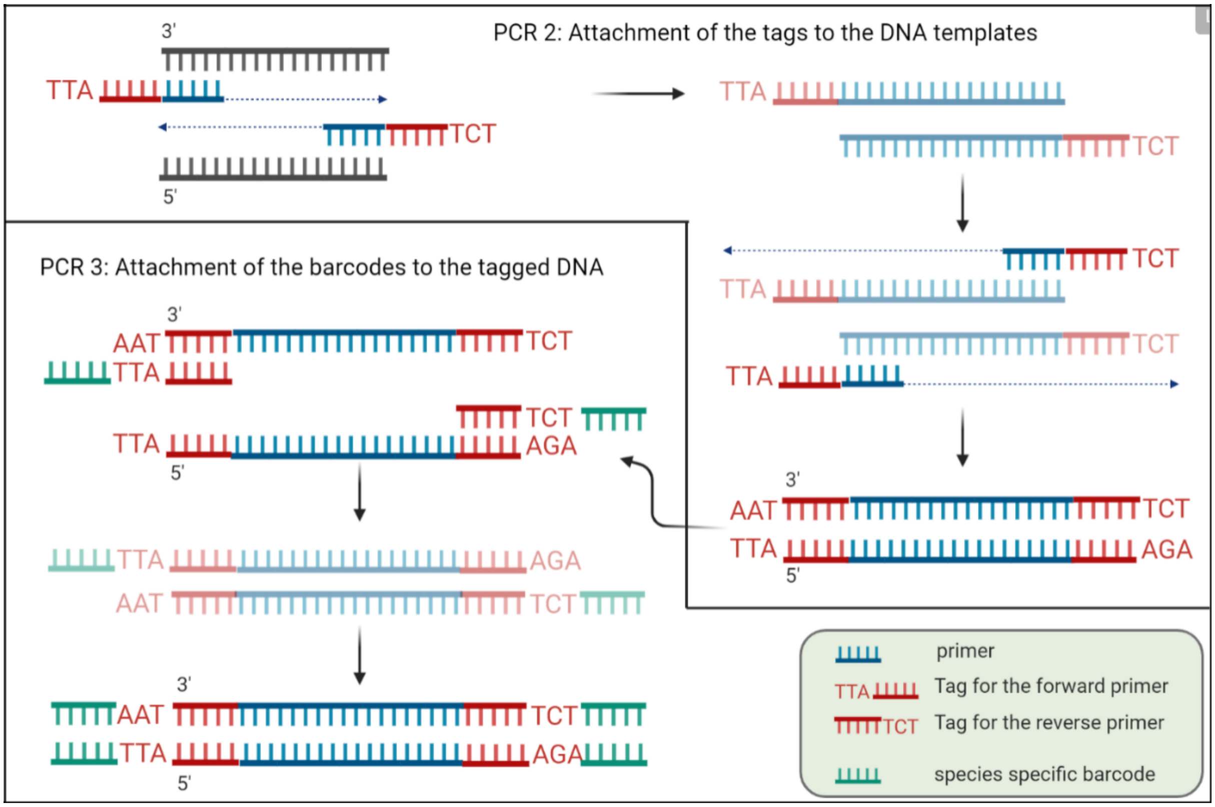

PCR 2: Attachment of the tags to the DNA templates
3'
5'
TTA
TCT
AAT
AGA
PCR 3: Attachment of the barcodes to the tagged DNA
primer
Tag for the forward primer
Tag for the reverse primer
species specific barcode

Figure 5.2 Detail of the PCR protocols to attach the customized barcodes to the tagged primers

PCR 1: Standard PCR to amplify a selected genomic region with the usual primers.

PCR 2: PCR with tagged primers. The primers were designed with an extra sequence, called a “tag” at their end that doesn’t match the genomic DNA. As the PCR cycles progress, the amplicons formed are longer, and the Tag sequence is replicated in the complementary strand thanks to the free dNTPs. In the end, a new tagged DNA template is obtained to be used in the following PCR protocol. The same forward primer tags and reverse primer tags were used for all 4 reaction sets, but the gene-specific parts of the primers were different.

PCR 3: PCR with barcoded primers. The primers consist of the Tag + the species-specific barcode attached to it. The tag part of the barcoded primers matches the tag part of the tagged DNA’s **complementary strand**. As the PCR cycles progress, the amplicons formed are longer, and the barcode sequence is replicated in the complementary strand thanks to the free dNTPs. In the end, a barcoded DNA sequence is obtained. All barcoded primers ended in a common 5bp sequence (GGTCA for forward primers and ACTGG for reverse primers). The common sequence allows for pore loading in Oxford Nanopore sequencing, which means the first 5bp are seldom correctly discerned.

Table 5.3 List of the customised barcoded primers

forward primer+tag	5'-3	reverse primer+tag	3'-5'
ITS5A_F+tag	TTATTGCTTAAGAATACGCGTAGGGAAGGAGAAGTCGTAACAAGG	ITS4-R+tag	CGTATAGTTATTCGCCTCCTCCCTCACTTCTGTGCCTCGGTCT
GeneX-2ndexon_F1+tag	TTATTGCTTAAGAATACGCGTAGACAGTATAACCTTCAGATGCTCAAT	GeneX-3rdexon_R1+tag	AACCAGGTCTTGTCTATCTTCGCCCTCACTTCTGTGCCTCGGTCT
Feronia Consen F1+tag	TTATTGCTTAAGAATACGCGTAGCARCAMACSMTGGGTGATCRGTGG	Feronia L.s ExEx F4+tag	MSCSAGAACAGACTTACCCGCCCTCACTTCTGTGCCTCGGTCT
ConsenHpPAR_F5+tag	TTATTGCTTAAGAATACGCGTAGGTTAAGATCCGGCACTTTAAACCG	ConsenHpPAR_R2+tag	CCAAAATTTGCAACCGTACTTTACCCCCTCACTTCTGTGCCTCGGTCT

species	species barcode	full forward primer 5'-3'		full reverse primer 3'-5'
G45	TTTGTTCGTTCGGCTCGCGTGA	GGTCATTTGTTCGTTCGGCTCGCGTGATTATTGCTTAAGAATACGCGTAG		CCCTCACTTCTGTGCCTCGGTCTAGTGCGCTCGGCTTGCTTGTTTACTGG
G45-5	TCTTTGGTTATCTAGCTGTATGA	GGTCATCTTTGGTTATCTAGCTGTATGATTATTGCTTAAGAATACGCGTAG		CCCTCACTTCTGTGCCTCGGTCTAGTATGTCGATCTATTGGTTTCTACTGG
G45-32	TGGCAGTGTATTGTTAGCTGGT	GGTCATGGCAGTGTATTGTTAGCTGGTTTATTGCTTAAGAATACGCGTAG		CCCTCACTTCTGTGCCTCGGTCTTGGTCGATTGTTATGTGACGGTACTGG
G45-45	TCCATTACACTACCCTGCCTCT	GGTCATCCATTACACTACCCTGCCTCTTTATTGCTTAAGAATACGCGTAG		CCCTCACTTCTGTGCCTCGGTCTTCTCCGTCCCATCACATTACCTACTGG
G45-83	AGAATTGTGGCTGGACATCTGT	GGTCAAGAATTGTGGCTGGACATCTGTTTATTGCTTAAGAATACGCGTAG		CCCTCACTTCTGTGCCTCGGTCTTGTCTACAGGTCGGTGTTAAGAACTGG
G45-110	AGAATTGCGTTTGGACAATCAGT	GGTCAAGAATTGCGTTTGGACAATCAGTTTATTGCTTAAGAATACGCGTAG		CCCTCACTTCTGTGCCTCGGTCTTGACTAACAGGTTTGCGTTAAGAACTGG
R35	CGCATCCCCTAGGGCATTGGTGT	GGTCACGCATCCCCTAGGGCATTGGTGTTTATTGCTTAAGAATACGCGTAG		CCCTCACTTCTGTGCCTCGGTCTTGTGGTTACGGGATCCCCTACGCACTGG
g116	TCCCTGAGACCCTTTAACCTGTGA	GGTCATCCCTGAGACCCTTTAACCTGTGATTATTGCTTAAGAATACGCGTAG		CCCTCACTTCTGTGCCTCGGTCTAGTGTCCAATTTCCCAGAGTCCCTACTGG
D36	GCGACCCATACTTGGTTTCAG	GGTCAGCGACCCATACTTGGTTTCAGTTATTGCTTAAGAATACGCGTAG		CCCTCACTTCTGTGCCTCGGTCTGACTTTGGTTCATACCCAGCGACTGG
D-PH3	TCAAGAGCAATAACGAAAAATGT	GGTCATCAAGAGCAATAACGAAAAATGTTTATTGCTTAAGAATACGCGTAG		CCCTCACTTCTGTGCCTCGGTCTTGTAAAAAGCAATAACGAGAACTACTGG
D2.10	TGGCTCAGTTCAGCAGGAACAG	GGTCATGGCTCAGTTCAGCAGGAACAGTTATTGCTTAAGAATACGCGTAG		CCCTCACTTCTGTGCCTCGGTCTGACAAGGACGACTTGACTCGGTACTGG
C-T	CAGTGCAATGATATTGTCAAAGC	GGTCACAGTGCAATGATATTGTCAAAGCTTATTGCTTAAGAATACGCGTAG		CCCTCACTTCTGTGCCTCGGTCTCGAAACTGTTATAGTAACGTGACACTGG
A-S	CTAGACTGAAGCTCCTTGAGG	GGTCACTAGACTGAAGCTCCTTGAGGTTATTGCTTAAGAATACGCGTAG		CCCTCACTTCTGTGCCTCGGTCTGGAGTTCCTCGAAGTCAGATCACTGG
Apo87	TCGAGGAGCTCACAGTCT	GGTCATCGAGGAGCTCACAGTCTTTATTGCTTAAGAATACGCGTAG		CCCTCACTTCTGTGCCTCGGTCTTCTGACACTCGAGGAGCTACTGG
Apo81	GCTGACTCCTAGTCCAGGGCTC	GGTCAGCTGACTCCTAGTCCAGGGCTCTTATTGCTTAAGAATACGCGTAG		CCCTCACTTCTGTGCCTCGGTCTCTCGGGACCTGATCCTCAGTCGACTGG
Apo35	TAGCAGCACATAATGGTTTGTG	GGTCATAGCAGCACATAATGGTTTGTGTTATTGCTTAAGAATACGCGTAG		CCCTCACTTCTGTGCCTCGGTCTGTGTTTGGTAATACACGACGATACTGG
Apo67	TCACAGTGAACCGGTCTCTTT	GGTCATCACAGTGAACCGGTCTCTTTTTATTGCTTAAGAATACGCGTAG		CCCTCACTTCTGTGCCTCGGTCTTTTCTCTGGCCAAGTGACACTACTGG
Apo72	CATTGCACTTGTCTCGGTCTGA	GGTCACATTGCACTTGTCTCGGTCTGATTATTGCTTAAGAATACGCGTAG		CCCTCACTTCTGTGCCTCGGTCTAGTCTGGCTCTGTTCACGTTACACTGG
Apo50	AACACACCTATTCAAGGATTCA	GGTCAAACACACCTATTCAAGGATTCATTATTGCTTAAGAATACGCGTAG		CCCTCACTTCTGTGCCTCGGTCTACTTAGGAACTTATCCACACAAACTGG
Sex88	CAGTGCAATAGTATTGTCAAAGC	GGTCACAGTGCAATAGTATTGTCAAAGCTTATTGCTTAAGAATACGCGTAG		CCCTCACTTCTGTGCCTCGGTCTCGAAACTGTTATGATAACGTGACACTGG
Sex65	GCCCCTGGGCCTATCCTAGAA	GGTCAGCCCCTGGGCCTATCCTAGAATTATTGCTTAAGAATACGCGTAG		CCCTCACTTCTGTGCCTCGGTCTAAGATCCTATCCGGGTCCCCGACTGG
Sex49	TAGCAGCACGTAAATATTGGCG	GGTCATAGCAGCACGTAAATATTGGCGTTATTGCTTAAGAATACGCGTAG		CCCTCACTTCTGTGCCTCGGTCTGCGGTTATAAATGCACGACGATACTGG
Hr	TACGTCATCGTTGTCATCGTCA	GGTCATACGTCATCGTTGTCATCGTCATTATTGCTTAAGAATACGCGTAG		CCCTCACTTCTGTGCCTCGGTCTACTGCTACTGTTGCTACTGCATACTGG
P-D	TCAGTGCATCACAGAACTTTGT	GGTCATCAGTGCATCACAGAACTTTGTTTATTGCTTAAGAATACGCGTAG		CCCTCACTTCTGTGCCTCGGTCTTGTTTCAAGACACTACGTGACTACTGG
P-C	TGAGGGGCAGAGAGCGAGACTTT	GGTCATGAGGGGCAGAGAGCGAGACTTTTTATTGCTTAAGAATACGCGTAG		CCCTCACTTCTGTGCCTCGGTCTTTTCAGAGCGAGAGACGGGGAGTACTGG
Pi-LP	TGGCAGTGTCTTAGCTGGTTGT	GGTCATGGCAGTGTCTTAGCTGGTTGTTTATTGCTTAAGAATACGCGTAG		CCCTCACTTCTGTGCCTCGGTCTTGTTGGTCGATTCTGTGACGGTACTGG
Cy	CACCCGTAGAACCGACCTTGCG	GGTCACACCCGTAGAACCGACCTTGCGTTATTGCTTAAGAATACGCGTAG		CCCTCACTTCTGTGCCTCGGTCTGCGTTCCAGCCAAGATGCCCACACTGG
On	AGAGCTTAGCTGATTGGTGAAC	GGTCAAGAGCTTAGCTGATTGGTGAACTTATTGCTTAAGAATACGCGTAG		CCCTCACTTCTGTGCCTCGGTCTCAAGTGGTTAGTCGATTCGAGAACTGG
La	CTGGGAGAGGGTTGTTTACTCC	GGTCACTGGGAGAGGGTTGTTTACTCCTTATTGCTTAAGAATACGCGTAG		CCCTCACTTCTGTGCCTCGGTCTCCTCATTTGTTGGGAGAGGGTCACTGG
L2	CTGGGAGAAGGCTGTTTACTCT	GGTCACTGGGAGAAGGCTGTTTACTCTTTATTGCTTAAGAATACGCGTAG		CCCTCACTTCTGTGCCTCGGTCTTCTCATTTGTCGGAAGAGGGTCACTGG
Au	TGTAAACATCCTACACTCAGCT	GGTCATGTAAACATCCTACACTCAGCTTTATTGCTTAAGAATACGCGTAG		CCCTCACTTCTGTGCCTCGGTCTTCGACTCACATCCTACAAATGTACTGG
Mu	TCAGGCTCAGTCCCCTCCCGAT	GGTCATCAGGCTCAGTCCCCTCCCGATTTATTGCTTAAGAATACGCGTAG		CCCTCACTTCTGTGCCTCGGTCTTAGCCCTCCCCTGACTCGGACTACTGG
Le	AGCAGCATTGTACAGGGCTATCA	GGTCAAGCAGCATTGTACAGGGCTATCATTATTGCTTAAGAATACGCGTAG		CCCTCACTTCTGTGCCTCGGTCTACTATCGGGACATGTTACGACGAACTGG
Po	AAGCCCTTACCCCAAAAAGTAT	GGTCAAAGCCCTTACCCCAAAAAGTATTTATTGCTTAAGAATACGCGTAG		CCCTCACTTCTGTGCCTCGGTCTTATGAAAAACCCCATTCCCGAAACTGG
To	TTATAAAGCAATGAGACTGATT	GGTCATTATAAAGCAATGAGACTGATTTTATTGCTTAAGAATACGCGTAG		CCCTCACTTCTGTGCCTCGGTCTTTAGTCAGAGTAACGAAATATTACTGG

PCR conditions for eif3e using KAPA2G Robust Hotstart Ready mix

The PCR started with 95°C denaturing for 3 minutes, followed by 25 cycles (reduced to 20 cycles for primer+Tag and primer+Barcodes) starting with 30 seconds at 95°C, 60°C for 30 seconds, followed by a 1 minute extension cycle at 72°C. The PCR ended with a 72°C for 3 minutes.

PCR conditions for rRNA using KAPA2G Robust Hotstart Ready mix

The PCR protocol was inspired by the protocol of Funk et al., (2012). It started with 95°C denaturing for 5 minutes, followed by a touchdown PCR starting with 45 seconds at 95°C, with a variable annealing temperature (60 - 55°C) for 1 minute, which decreased by 1 °C per cycle, followed by a 1 minute extension cycle at 72°C. The touchdown step was followed by 25 cycles (20 cycles for primer+Tag and primer+Barcodes) of 95°C for 45 seconds, 54°C for 1 minute, finishing with a 72°C extension cycle for 1 minute. The PCR ended with a 72°C for 7 minutes.

PCR conditions for Feronia using KAPA2G Robust Hotstart Ready mix

The PCR protocol started with 95°C denaturing for 5 minutes, followed by a touchdown PCR starting with 45 seconds at 95°C, with a variable annealing temperature (63 - 58°C) for 45 seconds, which decreased by 1 °C per cycle, followed by a 1 minute extension cycle at 72°C. The touchdown step was followed by 25 cycles (20 cycles for primer+Tag and primer+Barcodes) of 95°C for 45 seconds, 57°C for 45 seconds, finishing with a 72°C extension cycle for 1 minute. The PCR ended with a 72°C for 5 minutes.

PCR conditions for PpPAR

To improve the amplification of the long fragment associated with allele 1, the elongation time has been increased.

The PCRs for the first sequencing run were realized using KAPA2G Robust Hotstart Ready mix according to the following instructions:

The PCR protocol started with 95 °C denaturing for 5 minutes, followed by a touchdown PCR starting with 1 minute at 95 °C, with a variable annealing temperature (61 - 58 °C) for 1 minute, which decreased by 1 °C per cycle, followed by a 2 minutes extension cycle at 72 °C. The touchdown step was followed by 25 cycles (reduced to 20 cycles for primer+Tag and primer+Barcodes) of 95 °C for 1 minute, 57 °C for 30 seconds, finishing with a 72 °C extension cycle for 2 minutes. The PCR ended with a 72 °C for 10 minutes.

An additional step was added in the case of the *PpPAR* gene because of the presence of two fragments of very different sizes, representing different alleles.

The smallest fragment tended to be amplified more often than the longest one leading to a disequilibrium in the width of the bands on the electrophoresis gel. The long fragments were often lost over the PCRs.

To prevent that phenomenon, it was decided to band-stab the gel after the first step (technique according to Bjourson and Cooper., 1992) to isolate the amplicons of different lengths.

The PCRs for the second sequencing run were realized using a much more sensitive and efficient kit, RepliQa Hifi toughMix, according to the following instructions:

The PCR protocol started with 98°C denaturing for 5 minutes, followed by a touchdown PCR starting with 15 seconds at 95°C, with a variable annealing temperature (61 - 56°C) for 5 seconds, which decreased by 1 °C per cycle, followed by a 15 seconds extension cycle at 68°C. The touchdown step was followed by 25 cycles (20 cycles for primer+Tag and primer+Barcodes) of 98°C for 15 seconds, 55°C for 15 seconds, finishing with a 68°C extension cycle for 15 seconds. The PCR ended with a 68°C for 5 minutes.

5.2.5 Sequencing, sequence polishing and bioinformatics

The PCR products were quantified using an Invitrogen Qubit™ 4 Fluorometer and adjusted to make one single tube with all the amplicons at the same concentration. They were then purified and concentrated to a final reaction volume of 50ul using Kapa beads.

After purification, 136 fmol samples of each library were sequenced on a MinION R9.4.1 flow cell according to the manufacturer's protocol.

The sequencer ran for 24h to reach the required quantity of data necessary to generate accurate *de novo* assemblies while at the same time maximizing the number of samples sequenced.

Data analysis

Following base-calling, the raw data was collated in the fastq format, a text document which, in addition to the fasta format, contains the sequence and reading quality of each nucleotide. All sequence reads were then categorized according to the species barcodes. The ordered dataset was then processed using a python script (called Adapter.py) within the software PORECHOP to remove the adapters and barcodes.

Further data analysis proceeded using Geneious. The cleaned data were ordered by species/variety, with four genes for each plant. Due to time constraints only two genes, *HpPAR* and rRNA, were analysed in this study.

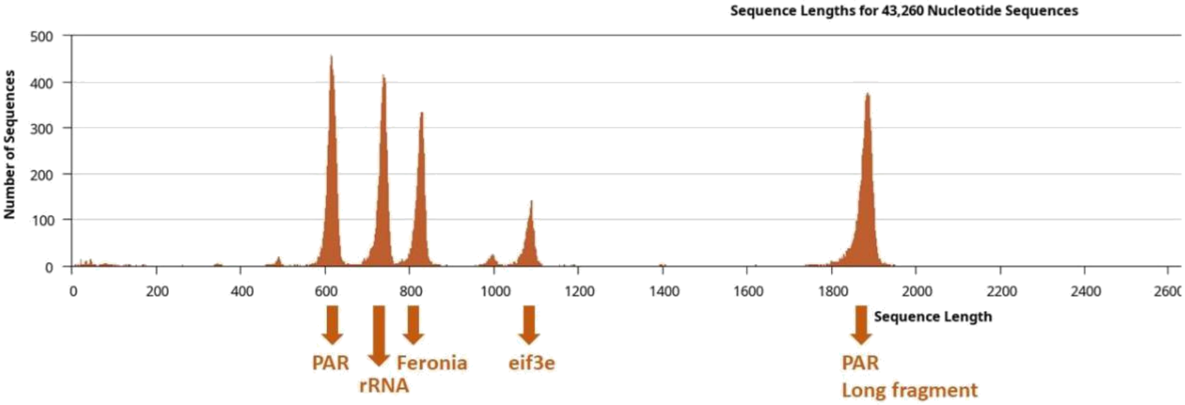


Figure 5.3 Oxford Nanopore nucleotide sequences profile for P.cymosa showing the reads ordered by sequence length with the gene associated.

Isolation of PpPAR gene

For each variety, the reads were sorted by length to allow the selection of those with the correct fragment size. The dominant allele for PAR was expected to have an amplicon size of approximately 1890bp, and the recessive alleles were expected to have an amplicon size of approximately 620bp. Typically, more than 10,000 sequences were available for each plant/gene/allele size combination. The data were trimmed along the edges to focus on nucleotides sequences with the average expected size for each plant/gene/allele size combination.

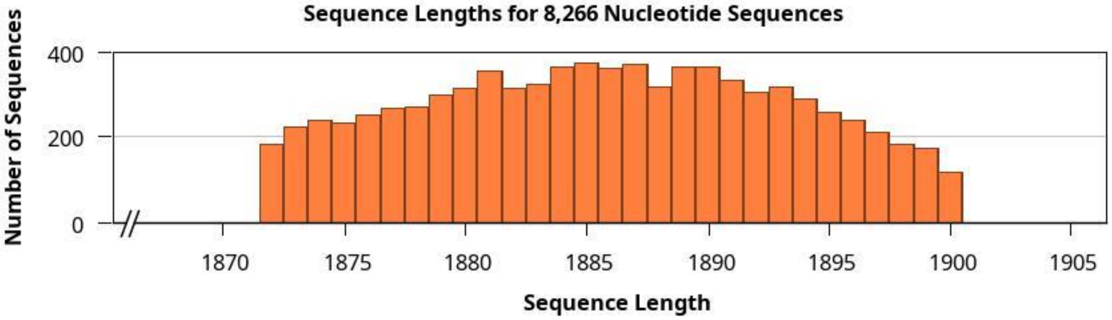


Figure 5.4 PpPAR fragments for the dominant allele, trimmed to exclude sequences that were either very long or very short.

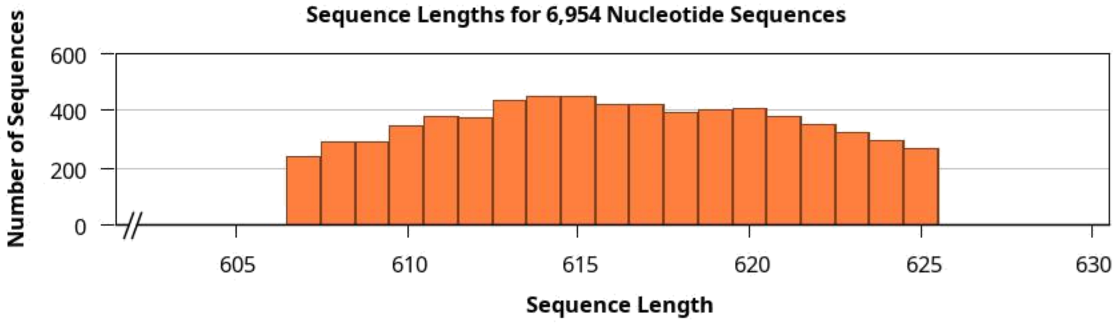


Figure 5.5 PpPAR fragments for the recessive alleles, trimmed to exclude sequences that were either very long or very short.

Mapping to reference

The reads were then aligned to the reference genome for the *Piloselloides praealta LOP* locus using the following instructions:

Sensitivity: Medium sensitivity/Fast

Fine tuning: Iterate up to five times (Recommended) Maps reads to the consensus from the previous iteration

This function only allows a few mismatches between the read and the reference. A mismatch is referred to in Geneious as a single nucleotide polymorphism (SNP). This detection of variations is taken as an advantage for the next step.

Finding Variations/SNPs:

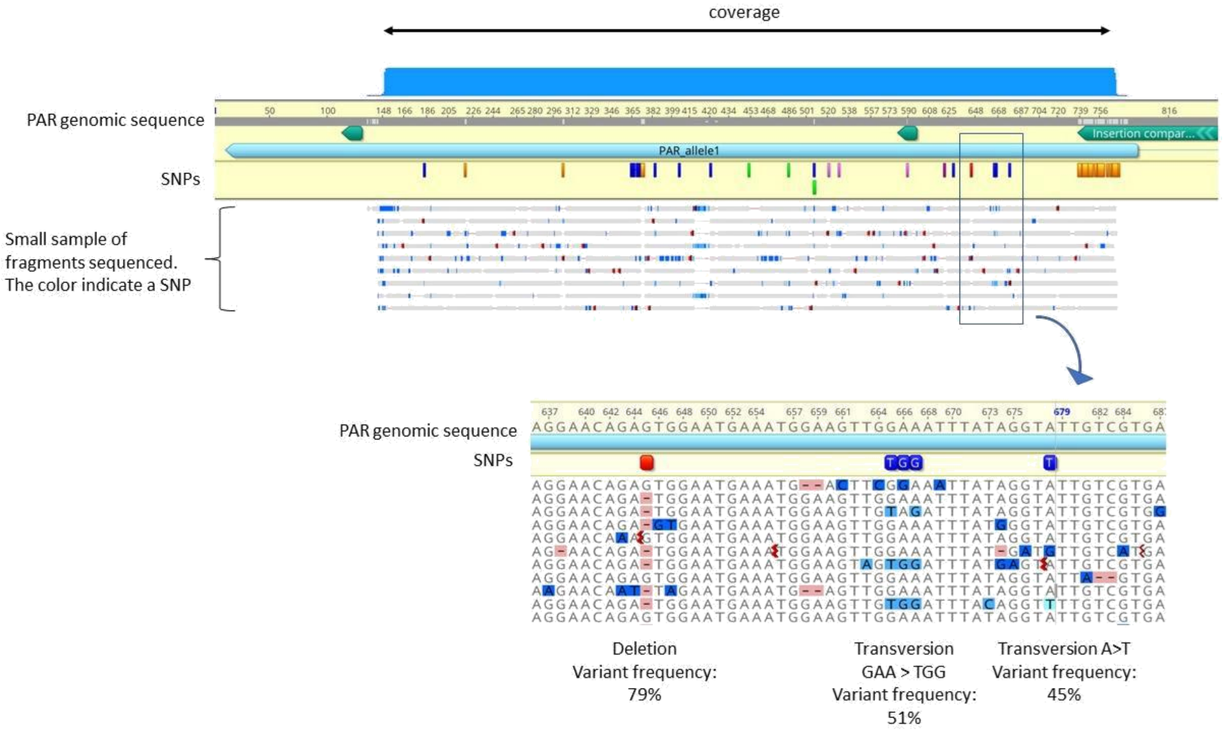
SNPs are a valuable tool to assess allelic diversity. The genome from which the reads were sequenced differs from the reference genome to which the reads are compared.

SNPs are detected using the following instructions:

Minimum coverage: 60

Minimum Variant Frequency: 0.25 (A variation is not considered a consistent SNP if its frequency is under 25% of the total)

Maximum Variant P-value: 10^{-3} (0.1% to see variant by chance)



```
AGGAACAGAGTGGAATGAAATGGAAGTTGGAAATTTATAGGTATTGTCGTGA

AGGAACAGAGTGGAATGAAATG--ACTTCGGAAATTTATAGGTATTGTCGTGA
AGGAACAGA-TGGAATGAAATGGAAGTTGGAAATTTATAGGTATTGTCGTGA
AGGAACAGA-TGGAATGAAATGGAAGTTGTAGATTTATAGGTATTGTCGTGG
AGGAACAGA-GTGAATGAAATGGAAGTTGGAAATTTATGGGTATTGTCGTGA
AGGAACAAAGTGGAATGAAATGGAAGTTGGAAATTTATAGGTATTGTCGTGA
AG-AACAGA-TGGAATGAAATGGAAGTTGGAAATTTAT-GATCTTGTCATGA
AGGAACAGA-TGGAATGAAATGGAAGTAGTGGATTTATGAGTATTGTCGTGA
AGGAACAGAGTGGAATGAAATGGAAGTTGGAAATTTATAGGTATTA--GTGA
AAGAACAAT-TAGAATGAAATG--AGTTGGAAATTTATAGGTATTGTCGTGA
AGGAACAGA-TGGAATGAAATGGAAGTTGTGGATTTACAGGTTTTGTCGTGA
AGGAACAGAGTGGAATGAAATGGAAGTTGGAAATTTATAGGTATTGTCGTGA
```

Deletion
Variant frequency:
79%

Transversion
GAA > TGG
Variant frequency:
51%

Transversion A>T
Variant frequency:
45%

Figure 5.6 Example of the Geneious window for the detection of SNPs.

Each SNP was assigned a variant frequency which could be used to sort the data. The flow cytometry data previously indicated the ploidy for each species. If, for example, the species studied was a tetraploid, we expected to find four alleles. The valuable and believable SNPs then have frequencies of approximately 25%, 50% or 75%. SNPs with a low frequency are more likely to represent base-calling errors so were ignored. On the contrary, SNPs with a frequency of over 90% are likely to exist in all the alleles.

In the case of a tetraploid, the easiest way to proceed was first to target SNPs with a 50% frequency to separate the data in 2 sets and then the ones with a 50% frequency again for both groups. To do so, one SNP with the desired frequency was strategically chosen. 2 new files are created with each one of the two possibilities for this SNP (same as reference or variant), and the operation was repeated for each file.

A good way to be convinced of the sorting reliability was to choose a SNP that combined the best with other SNPs with the same frequency.

A few individuals were left out for the analysis, specifically the segregating population from the cross between a *P.praealta* deletion mutant and *P.aurantiaca*. This cross is of greatest value in assigning alleles associated with the dominant allele at LOP. For *PpPAR* this allele sequence was clear in the sequence data set due to the different amplicon sizes.

Generating consensus sequence :

Once the data was sorted into a number of files corresponding to the number of alleles of the species, a consensus sequence was generated for each allele in Geneious using the assembly function and the following instructions:

Highest quality (60%): Bases matching at least 60% of total adjusted chromatogram quality (Recommended)

Assign Quality: Total (Sum quality of contributing bases minus non-contributing bases
Call Sanger heterozygotes >50%

5.2.6 Tree construction

Trimming of allele LOP

Before building the trees, the data from the *HpPAR* gene needed to be further adjusted. Since the amplified gene comprised of allelic fragments of different lengths, the dominant LOP allele sequence was cut to the same length as the recessive alleles.

To do so, the LOP allele of each species was aligned with one of the recessive alleles using the Pairwise Alignment (aligns a small number of similar length sequences such as genes) with the following instructions:
Global alignment type: Global alignment with free end gaps
Cost Matrix: 65% similarity

This helped to find the right-sized motif in the paired file that could be found again in the original file. The allele LOP is then trimmed at each end of the motif.

Geneious Tree Builder

Three trees have been built for the studied genes, two trees for *HpPAR* and one for rRNA using the Tree Builder function of Geneious using the following instructions:
Alignment type: Global alignment with free end gaps
Cost matrix: 65% similarity
Genetic distance Model: Tamura-Nei
Tree Buil Method: Neighbor-joining
Outgroup: No outgroup
Gap open penalty: 12

Gap extension penalty: 3

Published sequences from additional species from the *Asteraceae* family were looked for to improve the tree. Indeed, the entire genome of some crops from the *Asteraceae* family has been sequenced without all the functions associated with their genome being known.

A BLASTN search was made against the LOP allele of the *PpPAR* gene to find other *Asteraceae* with the same pattern.

Two species partially matched the sequence: *Helianthus annuus* (for a putative gene coding E3 ubiquitin-protein ligase LIN-1) and *Lactuca sativa* (for an uncharacterized gene called LSAT_8X112340).

The FASTA files were downloaded and aligned against the genomic reference with 65% similarity. The region of interest was located and copied. The copied motif was found on the original file and chopped at the right size to be later integrated into the tree building with the following change:

Cost matrix: 51% similarity

At 65% similarity, it was impossible to compute the genetic distance for these sequences because at least one pair of sequences (Ongense_3 and *Lactuca sativa*) did not overlap (or had only ambiguities in common) in the alignment.

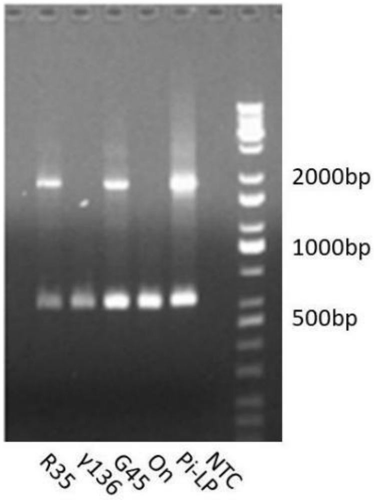
The three trees were presented with the shape that was considered to be the most indicative.
For clarity, the LOP allele was called 1 in all cases and the recessive alleles were annotated from 2 to 6.

5.3 Results/discussion

5.3.1 PpPAR

Primer selection and amplification

The electrophoresis gel of the PpPAR gene revealed the presence of two sizes of amplicons of respectively 600 bp and 1.8kbp.



R35 wild type: apomictic (positive control)

γ136 LOP deletion mutant: sexual (negative control)

G45: apomictic

P. onegensis: sexual

Pilosella officinarum (Lewis pass): Apomictic

No Template Control

Figure 5.7 Electrophoresis gel of six samples for PpPAR gene shows two sizes of amplicons.

For γ136 and P. onegensis, only one PCR product was amplified. For R35, G45 and Pilosella officinarum two sizes of PCR products were amplified.

The gel shows one band corresponding to a small fragment around 600bp for every species and a longer fragment of around 1.8kbp for only some of them.

It has been observed that the presence of this additional long fragment matches with the apomictic phenotype.

Indeed, sexual plants possess only copies of the short fragment, while apomictic plants possess at least one copy of the longer fragment. It is now known that a type II transposon is placed in front of the PAR gene in the case of the dominant allele at LOP and appears to have a role in parthenogenesis.

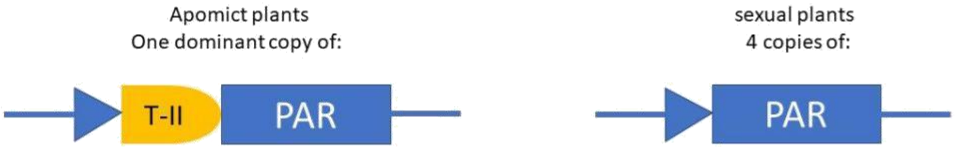


Figure 5.8 Illustration of the PpPAR gene in Pilosella piloselloides var praealta (4x).

It is present in 4 identical copies in the tetraploid sexual plants. In apomictic plants, one copy is present in a longer fragment corresponding to the gene preceded by a transposon.

Theory of epigenetic regulation

Transposons are DNA sequences that can "jump" from one place to another in the genome, and by doing so, they can modify the expression of a gene. For a long time, they were considered as "junk DNA", but their influence on the genome has probably played a significant role in the evolution (Muñoz-López and García-Pérez, 2010).

"In higher plants, the activation or movement of transposable elements (TEs) covers a significant proportion of the genome, influencing the evolution of the genome, alternation in gene expression, and frameshift mutations" (Fiaz et al., 2021). The role of retrotransposons in apomictic development has also been proposed. A retrotransposon in *Cenchrus ciliaris* was shown to have differential activity in apomictic and sexual plants (Rathore et al., 2020).

Characterization and role of PpPAR transposon

Inverted repeat sequences can be found in this transposon sequence, which would tend to confirm that this transposon belongs to the type II class of transposons as this class is characterized by a transposase gene flanked by Terminal Inverted Repeats. The repeat sequences at *PpPAR* indicate that this transposon forms part of the hAT transposon superfamily. This superfamily is widely distributed among eukaryotic kingdoms and is mainly found in genomic fragments. This is an ancient family of transposable elements that probably appeared at the early stages of plant-animal-fungi separation (Rubin et al., 2001).

Data analysis of the PpPAR gene

The sequencing of *PpPAR* gave between 20,000 and 40,000 reads per amplicon. The analysis of the sequencing data showed that there was indeed only one copy of the longer fragment, representing the dominant allele (Allele 1).

The analysis of a mutant form of G45 (G45-110) indicated only three allele could be determined yet the wild type apomict G45 had 4 alleles determined. The missing allele in G45-110 was the dominant Allele 1, which is consistent with this mutant lacking this sequence and therefore being sexual. The dominant allele was also missing in a deletion mutant of G45 that is also sexual (G45-5). One surprising result was that G45 was recorded as a pentaploid using flow cytometry but only four allele sequences were found. We suspect that one recessive allele exists in 2 identical copies in this plant.

The number of reads for G45-45 and G45-83 were too low (only 5 for G45-45 and 16 reads for G45-83) to reach any conclusions about these plants. A similar problem occurred with *Pilosella caespitosa* (C-T) where only 4 reads were recorded.

For the species *Hieracium lepidulum (Le), Hieracium policheae (Po), Hieracium murorum (Mu)* only four alleles at PAR were discovered yet they were all were recorded as being pentaploids by flow cytometry. Interestingly, although they are all apomicts no allele was found that had a transposon in the promoter of the PAR gene. These species all belong to the genus *Hieracium*, a sister genus to *Pilosella*. The conspicuous lack of sequence for one allele at *PAR* may mean that the primers used (which were designed to *Pilosella* sequence) may not have amplified the dominant allele in these plants. More work is needed to test this possibility.

Hieracium pelletarianum (L2) is a sexual diploid, but no distinct polymorphism could be observed, indicating that either only a single allele amplified or that both the alleles present were the same in sequence. *Taraxacum officinale* (To) and *Hypochoeris radicata* (Hr) did not give any consistent data for HpPAR. The hypothesis is that the PCR did not amplify well enough in these plants with the primers used.

Tree construction for the PpPAR gene

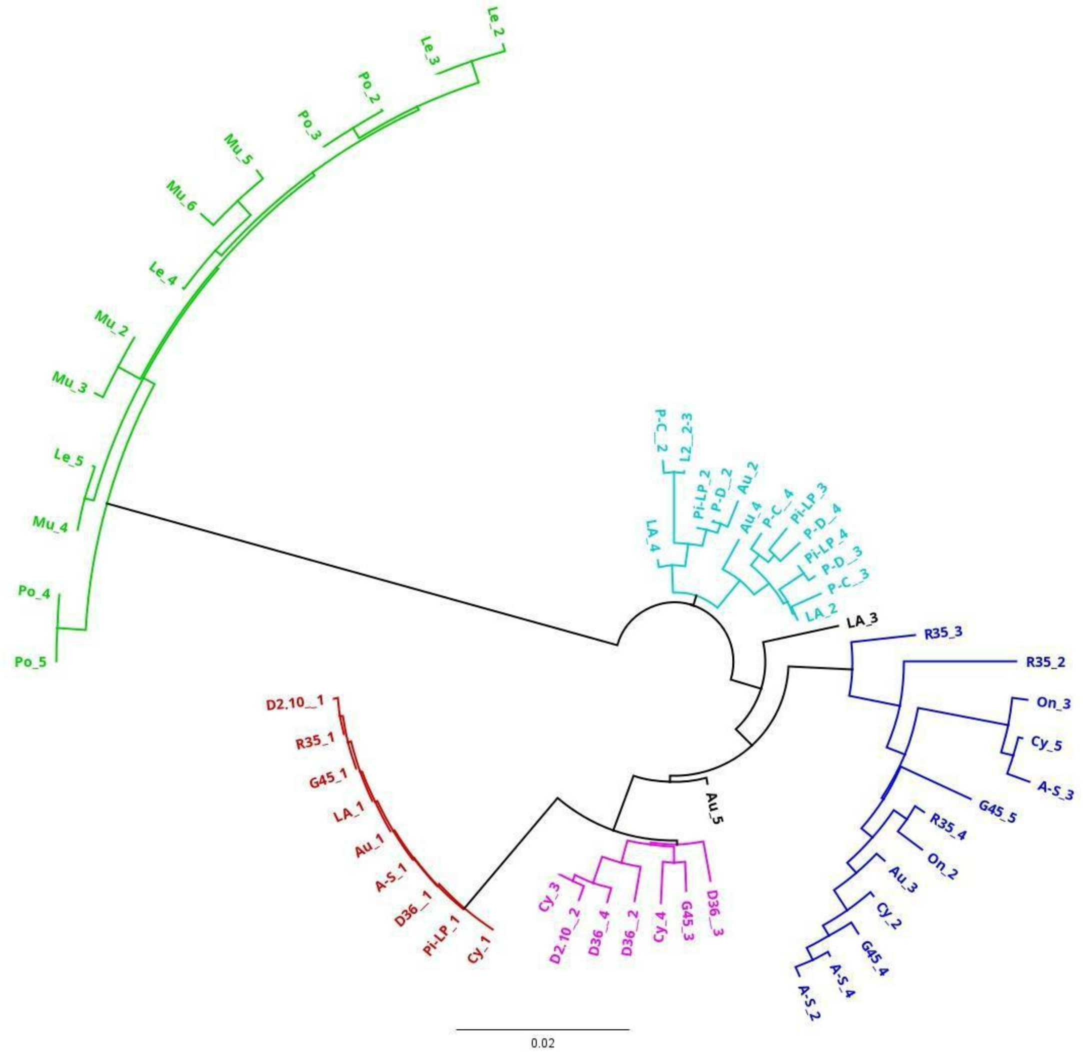


Figure 5.9 Cladistic tree of PpPAR in 16 species using a circular graphic.

The evolutionary tree showed five distinct clusters. The LOP allele (allele 1) of the different plants fall together in a single clade while the other alleles at PpPAR are mixed between the different clades. This supports the hypothesis that dominant alleles at LOP have a common ancestry, and they are less prone to ASD than the other alleles at the locus as they never exist in sexual biotypes. All known dominant copies of PAR are numbered ‘1’. All other copies are randomly assigned numbers greater than 1.

Here is another representation including two species of the same family, *Lactuca sativa* and *Helianthus annuus* which emphasizes how close some species are to each other compared to these two distant relatives.

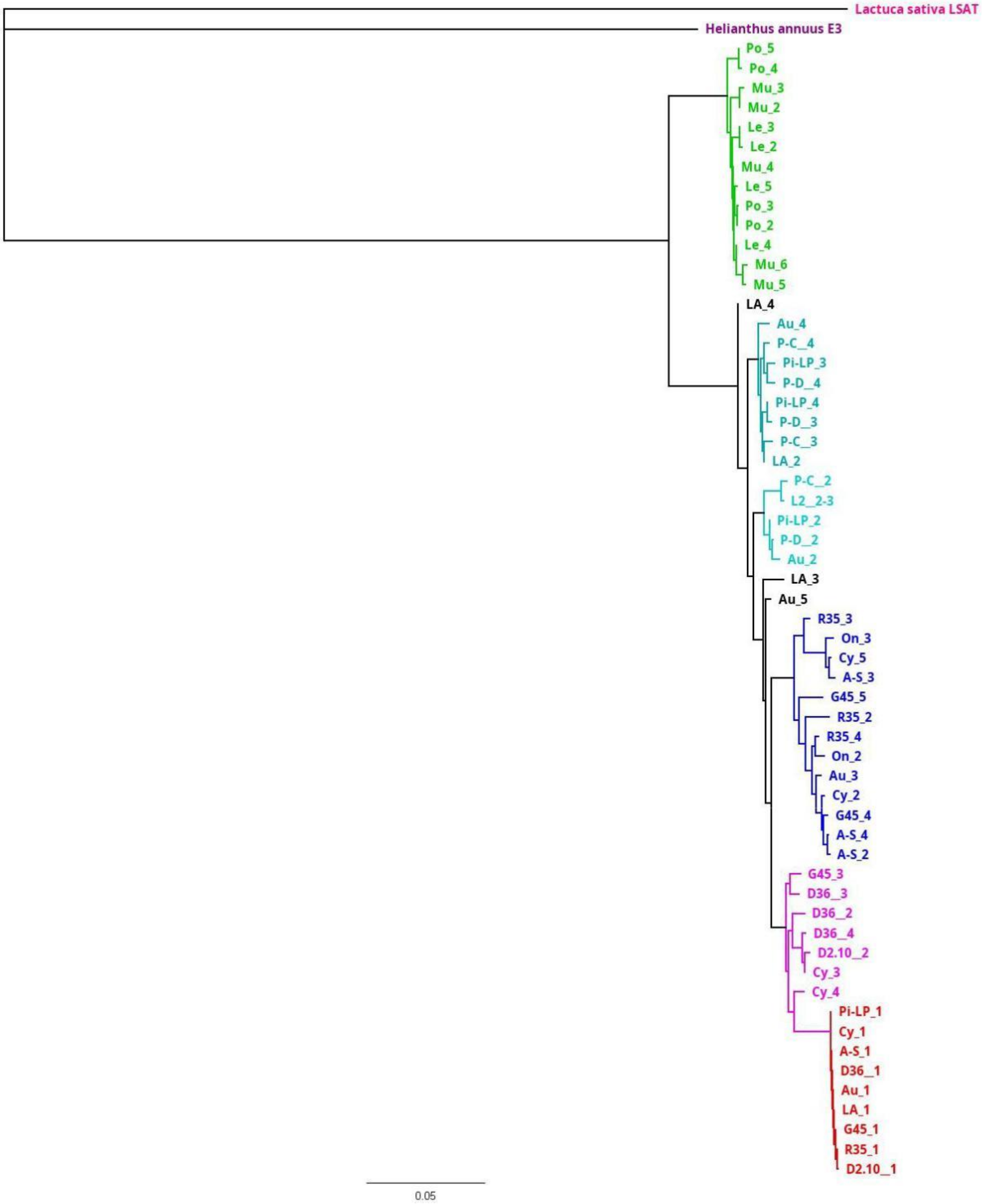


Figure 5.10 Cladistic tree of PpPAR in 16 species using a straight-line graphic and rooted with two related species (*Lactuca sativa* and *Helianthus annuus*).

As expected the more distantly related species *Lactuca sativa* and *Helianthus annuus* fall apart from the other *Pilosella* and *Hieracium*. This representation shows that the species from the genus *Hieracium* (*Hieracium policheae* (Po), *Hieracium Lepidulum* (Le) and *Hieracium Murorum* (Mu)) cluster together (in green), separated from the plants in genus *Pilosella*. *Lactuca sativa* and *Helianthus annuus* fall apart the groups.

5.3.2 rRNA

Primer selection and amplification

The amplification for the ribosomal gene was clear for all of the species and most of the primer sets. The primers ITS5A and ITS4 were chosen for further use.

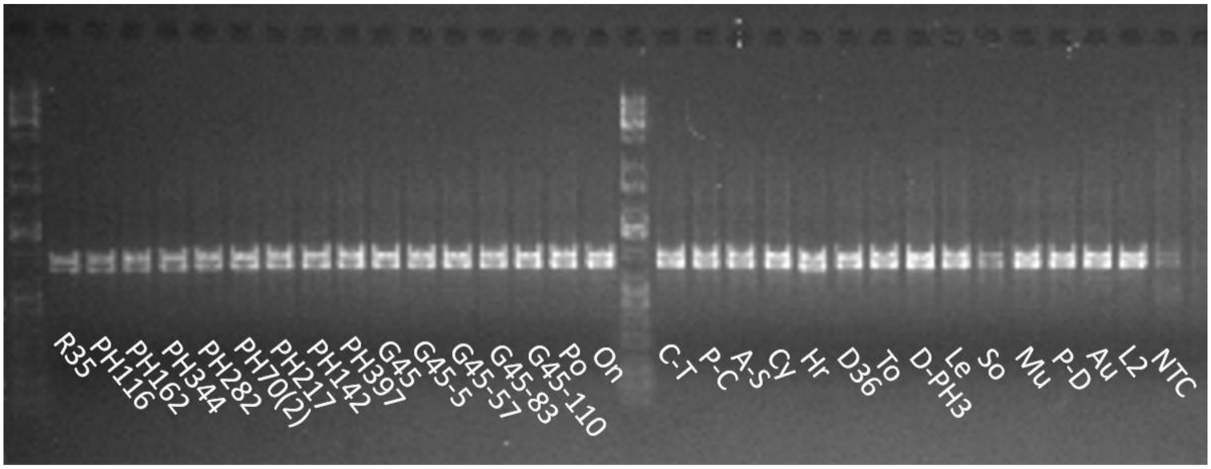


Figure 5.11 Electrophoresis gel for the rRNA gene with primers ITS5A/ITS4 x 30 DNA templates.

Only one band was expected, but this band is probably broadened due to a poor dye.

Data analysis of the rRNA gene

The sequencing of the rRNA ITS amplicons typically gave around 20,000 per amplicon. The analysis was quite easy and quick as there was one one allele for each species. This was expected as rRNA genes are typically very similar in species and it was confirmed by the data where SNPs gave mostly variant frequencies higher than 90%. The genome reference used in this study was that of *Lactuca serriola* in which rRNA ITS sequence is only present in one copy.

Tree construction for the rRNA gene

The tree for rRNA, which is a highly conserved gene, gives an indication of the natural branching of the different species.

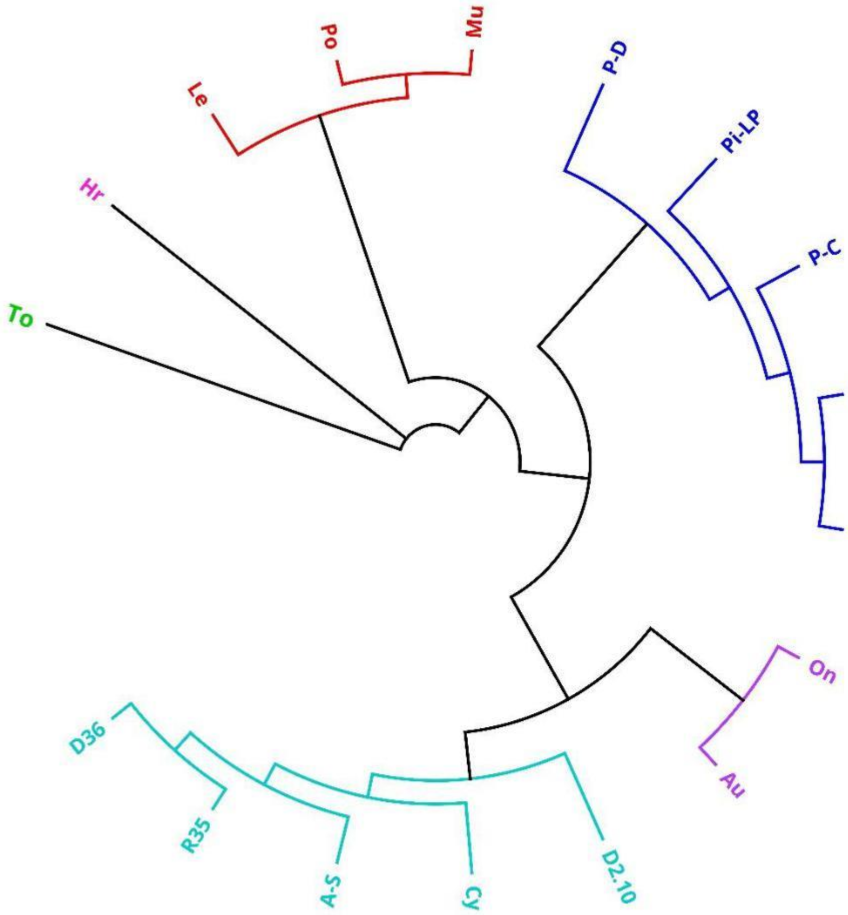


Figure 5.12 Phylogenetic tree for rRNA x 17 species.

It was interesting to see that *Hieracium policheae* (Po), *Hieracium Lepidulum* (Le) and *Hieracium Murorum* (Mu) separated out. They are all close relatives so, unsurprisingly segregate together, but they also separated from Pilosella species, which confirms the clear observations made with the *PAR* gene.

The *Pilosella* sequences are separated into three clusters:

The plants *Pilosella officinarum*: *Pilosella officinarum (Caen)* (P-C), *Pilosella officinarum (Dijon)* (P-D) and *Pilosella Lewis Pass* (Pi-LP) were predictably in the same cluster because they are from the same species although they are geographically very different. However, the presence of *Hieracium pelletarianum* (L2) and *Hieracium lactucella* (La) in the same group is surprising.

This group also appeared in the *PAR* tree with only one difference: both *PpPAR* alleles of *Pilosella lactucella* (Au) fall in this group.

Pilosella onegensis (On) and *Pilosella lactucella* (Au) fall together but are actually not that far from the previously described cluster.

Pilosella piloselloides var praealta R35 (R35), *Pilosella piloselloides D36 (D36), D2-10* (D2-10), *Pilosella aurantiaca* (A-S), *Pilosella cymosa* (Cy) form a third group. In the PpPAR tree, they were broken down into two groups together with *Pilosella onegensis* (On).

Hypochoeris radicata (Hr) and *Taraxacum officinale* (To) are species apart as expected as they are only distantly related.

Note that since the rRNA genes keep consistent in plant species, ploidy level doesn’t play any role in the positions of the different species in the tree. For example, *Pilosella onegensis* (diploid) and *Pilosella lactucella* (pentaploid) are closely clustered.

5.3.3 General comments

Additional experiments would be required to extend this phylogenetic tree model.

The entirety of the G45 mutants, *Hieracium lepidulum (Le), Hieracium policheae (Po)* and *Hieracium murorum (Mu)* need to be tested for PAR again, because we missed the LOP fragment for several of them. I would recommend using the Hifi enzyme for the PCR that is more precise and then trying to spot the band on the gel. If it is identifiable, a sample of amplicons could be extracted by band-stabbing and amplified by themselves by PCR. If the band is not identifiable, extraction of the appropriate part of the gel and the purification of the amplicons is worth considering.

Pilosella caespitosa (C-T), *Hieracium pelletarianum* (L2), *Taraxacum officinale* (To) and *Hypochoeris radicata* (Hr) need to be tested again for the four genes because the number of reads given by the sequencer was insufficient.

Finally further studies of the two other genes *Feronia* and *eif3e* should also bring more consistency to the tree model.

Chapter 6 General Discussion

The discovery of apomixis and perspectives

It has been a long time since John Smith first observed apomixis in 1841. However, the study of apomixis has lagged behind sexual reproduction because few major crops use it. Also, breeders have long shunned apomixis because the production of clones leads to a lack of genetic diversity, and that makes the job of plant improvement harder. However, the study of apomixis has increased in recent decades with the potential application of conditional apomixis to improve the plant breeding process. Apomixis brings promising perspectives in plant breeding and it carries the hope of food security in a context of global warming where it is more than ever necessary to accelerate the production of adapted varieties. The introgression of apomixis could make it possible to create high performance, true-breeding hybrid lines in a single generation. This would save considerable time. The absence of recombination would make it possible to fix the traits without losing the hybrid vigor, so apomixis would also limit the costs of maintaining parental lines and avoid the need for male sterility.

Evolutionary biology of apomixis

Understanding the molecular mechanisms of apomixis will assist our understanding of the evolutionary consequences of this trait. It has long been noted that, in plants and animals, sexual and asexual biotypes tend to have different ranges. This phenomenon is called "geographic parthenogenesis" (Vandel, 1928). Stebbins (1950) noted that autonomous reproductive systems seem to be preferentially present in unfavorable climates and/or where unreliable pollinator services are available. The tendency of *Pilosella* to use apomixis when the population is isolated or in a colonization mode was first reported by Gadella (1991). Similarly, it has been shown that apomictic species are more strongly represented in stable environments, for example, at high altitudes with slight environmental variation. Under these conditions, sexual reproduction represents an unnecessary energy cost and is often a slower process for producing seeds than cloning.

In the mid-1970s, evolutionary biologists began to study the paradox of sex (van Dijk and van Damme, 2000). Indeed, sexual reproduction, although being the most widespread mode of reproduction and considered the most efficient, contains disadvantages such as the cost of maintaining the males. From this point of view, apomixis might appear to be a much simpler process since it often involves only one parent. As genetic tools have improved, more and more new wild species have been identified that use apomixis. More than 400 apomictic species are now known, and this is clearly an underestimate of its distribution. It is now clear that apomixis has evolved independently in different clades of flowering plants (Carman 1997; Whitton et al. 2008).

This represents a case of Convergent evolution, the appearance of a trait in different species that were not inherited from a common ancestor. Typically, convergent evolution results from similar, optimized adaptive responses in different taxa to a shared selection pressure. Again, this strongly suggests that apomixis offers advantages over sexual reproduction in some environments.

Origins and transmission of apomixis

Carman showed the existence of a strong correlation between gametophytic apomixis and polyploidy in more than 126 genera (Carman 1997). Polyploidy can lead to genome disruption and it is therefore a significant mechanism driving speciation among angiosperms. Odd ploidy levels (e.g., triploidy, pentaploidy) have been reported to be linked with the presence of apomixis, although tetraploidy is the most common ploidy level among apomicts (Asker and Jerling 1992). It is commonly agreed that apomixis derives from sexual reproduction by disrupting the developmental pathways of seed formation. However, apomixis can originate from mutations or from hybridization between divergent sexual species.

Carman (1997) put forward an hypothesis that climatic fluctuations of the Pleistocene could have modified the areas of distribution of plants and this brought together different ecotypes which then hybridized. Ecotypes with putatively-precocious embryo sac development would have had the opportunity to form polyploids with ecotypes with putatively-delayed embryo sac development. Asynchronous female development in these polyploids would then have resulted in an efficient apomixis system.

Another model of how apomixis might arise was postulated by Nogler (1994). He noted that a failure in meiosis would result in a difference in ploidy in the offspring through the action of unreduced gametes. A mutation implying a non-disjunction of the germline chromosomes in meiosis would create diploid gametes with 2n chromosomes (instead of n). After fusion between two such gametes, we will obtain a 4n zygote.

Another mutation would then be needed in the parthenogenesis locus because the avoidance of meiosis in itself wouldn’t produce a viable product. Indeed, the individual would otherwise ‘polyploidize itself out of existence’ due to the exponential increase of ploidy level with the fertilization of each generation” (Stebbins, 1950).

Apomixis in *P. piloselloides var. Praealta*

P.piloselloides var. Praealta is a facultative apomict that uses apospory, one of the two gametophytic apomixis pathways. In gametophytic apomixis, the embryo is formed from an embryo sac-like structure that didn't undergo meiosis and is not fertilized. Two dominant loci control apomixis in *P.piloselloides var. Praealta*. The *LOSS OF APOMEIOSIS* (*LOA*) locus is involved in creating an embryo sac derived from a somatic cell of the ovule without meiosis. The *LOSS OF PARTHENOGENESIS* (*LOP*) locus is involved in the spontaneous formation of an embryo from the embryo sac without fertilization.

Before this study, previous work in Ross Bicknell's team (unpublished) estimated the size of the *LOP* locus in *P.piloselloides* to be 1.4Mb. This study made it possible to reduce the known sequence of the *LOP* locus in this plant to an interval of 647kb. No recombination event was observed within this interval in a sample of 2,000 recombinant plants, meaning the resolution of mapping was less than 0.2cM. This work, therefore, refines the search area for genes involved in parthenogenesis and it more precisely defines the exact location of the *LOP* locus in the genome. In addition, the isolation of BACs within the *LOP* locus provided a tool for checking the accuracy of the haplotype-phased genome assembly. It is indeed essential for future studies, to work on a sequence that is as close as possible to the real one.

PAR is a gene placed in the middle of the *LOP* locus. Discovered in *Taraxacum*, it turned out to be also present in *P.piloselloides*. As in *Taraxacum* the dominant allele is recognizable by an additional sequence upstream of the gene, corresponding to an ancient transposon. The transposon was present in all of the apomictic plants studied in the current study and absent from all sexual plants studied, without exception. This includes mutants of wild type apomicts that were converted to sexual phenotypes. Now that the sequence of PAR is known, it would be interesting to integrate this transposon upstream of a recessive gene in a sexual plant to check if apomixis is induced. This operation can be done by using *Agrobacterium tumefaciens* that has shown to be effective in *P.piloselloides*.

At least six other genes remain to be studied in the *LOP* locus of *P.piloselloides*. The genes can be deduced thanks to recognizable Open Reading Frames (ORF) patterns and from mapped transcript reads. A transcript database is available for this plant to progress this work (Bräuning et al. 2018). A panel of transgenic *P. piloselloides* could be created using *Agrobacterium*-mediated techniques, and screened using Green Fluorescent Protein (GFP) as a marker. By combining the GFP marker with promoter elements of the identified genes at *LOP,* the expression of those genes could be inferred by the patterns of GFP fluorescence.

In particular, GFP monitoring provides knowledge of the tissue, timing and subcellular localization of protein products and therefore gives an indication of the function of a gene. *Agrobacterium tumefaciens* could also be used to knock out candidate genes at LOP using gene editing and the role of that gene on embryo development could then be assessed.

A recent study showed the efficiency of CRISPR/Cas9 gene editing in *P. piloselloides*, which would facilitate this work (Henderson et al., 2020).

Allelic variability at *PpPAR*

When part of a population separates to colonize another environment, it only includes a sample from the pool of alleles available in the mother population. This is called the founder effect. These populations contain less genetic diversity and are more sensitive to genetic drift. Apomicts are often good colonizers as they are less dependent on compatible partners and fertilizing insects, so the founder effect is expected to be particularly common amongst these plants. This is partly why some authors have described apomixis as an evolutionary dead-end, because successive founder events will lead to genetic impoverishment (Darlington 1939). The lack of recombination in apomictic plants also prevents them from getting rid of deleterious mutations that accumulate in a lineage, a mechanism known as Muller's ratchet (Muller, 1932). This is also expected to reduce the viability of apomictic lineages over time. However, apomixis, far from being an evolutionary dead-end, seems to have a persistence that defies the laws of nature. One reason for this may be that, although individual lineages (genotypes) may degrade due to accumulated deleterious mutations, apomicts typically exist in populations of many clones, so as one genotype becomes obsolete, new ones emerge to replace it. Another result of the suppression of recombination around apomixis loci is that these loci are often regions rich in repeat sequences (Kotani et al., 2014; Conner et al., 2015). This can then lead to hemizygosity and further compound the reproductive isolation of apomicts. The repeat-rich nature of apomixis loci was confirmed in this study at *LOP* by the mapping work. It made it difficult at times to design primers that were specific to the sequence.

In *P. piloselloides* apomixis and sexual reproduction are not entirely independent pathways. Apomixis in aposporous *Pilosella* (formerly *Hieracium)* seems to be superimposed on the sexual path, where apomixis redirects the fate of cells with gametic potential. The continued existence of sexual reproduction in *Pilosella* populations means that genetic exchange is still occurring and that cross-over between alleles will influence the patterns of allelic variation in the population (Hand and Koltunow, 2014). This effect, however, is not likely to occur between dominant alleles for LOP as they lie in a region of non-recombination (Chapter 4).

Consequently, we hypothesized that allelic variation would be greater in sexual alleles than in dominant alleles at *LOP*. In this study and for the first time in the literature, it was shown that dominant alleles of the *PAR* gene in eight species the genus *Pilosella* were almost identical, while the recessive alleles at the same gene were more variable amongst and within species. It was also confirmed that, despite the accumulation of mutations, the dominant *PAR* allele at *LOP* remained highly conserved, presumably because they are under strong selection.

In conclusion, the primary findings of this study were:

- Parthenogenesis *P.piloselloides var. Praealta* is controlled by the locus *LOP*, which is a region of suppressed recombination with a high level of repeat sequence.
- The *LOP* locus mapped to a 647kb interval in this species.
- A precise sequence for two alleles at the *LOP* locus was checked using BACs isolates
- The study of allelic diversity for the *PAR* gene within the *LOP* locus showed a high similarity of the dominant alleles among species for this gene and greater variability amongst recessive alleles.

References

Albertini, Emidio, Andrea Porceddu, Francesco Ferranti, Lara Reale, Gianni Barcaccia, Bruno Romano, and Mario Falcinelli. “Apospory and Parthenogenesis May Be Uncoupled in Poa Pratensis: A Cytological Investigation.” *Sexual Plant Reproduction* 14, no. 4 (December 1, 2001): 213–17. https://doi.org/10.1007/s00497-001-0116-2.

Asker, Sven E., and Lenn Jerling. *Apomixis in Plants*. Boca Raton: Routledge, (1992).

Baker, H. G. “Self-Compatibility and Establishment After ‘Long-Distance’ Dispersal.” *Evolution* 9, no. 3 (1955): 347–49. https://doi.org/10.2307/2405656.

Baker, Herbert G., and G. Ledyard Stebbins, eds. The Genetics of Colonizing Species: Proceedings of the First International Union of Biological Sciences Symposia on General Biology. First Edition. New York; *London: Academic Press*, (1965).

Bashaw, E. C. “Apomixis and Its Application in Crop Improvement.” In *Hybridization of Crop Plants*, 45–63. John Wiley & Sons, Ltd, (1980). https://doi.org/10.2135/1980.hybridizationofcrops.c3.

Barcaccia, Gianni, and Emidio Albertini. “Apomixis in Plant Reproduction: A Novel Perspective on an Old Dilemma.” *Plant Reproduction* 26, no. 3 (September 2013): 159–79. https://doi.org/10.1007/s00497-013-0222-y.

Bicknell, R. A., S. C. Lambie, and R. C. Butler. “Quantification of Progeny Classes in Two Facultatively Apomictic Accessions of Hieracium.” *Hereditas* 138, no. 1 (2003): 11–20. https://doi.org/10.1034/j.1601-5223.2003.01624.x.

Bicknell, Ross A., and Anna M. Koltunow. “Understanding Apomixis: Recent Advances and Remaining Conundrums.” *The Plant Cell* 16, no. suppl_1 (June 1, 2004): S228–45. https://doi.org/10.1105/tpc.017921.

Bicknell, Ross, and Andrew Catanach. “Apomixis: The Asexual Formation of Seed.” In *Somatic Genome Manipulation: Advances, Methods, and Applications*, edited by Xiu-Qing Li, Danielle J. Donnelly, and Thomas G. Jensen, 147– 67. New York, NY: Springer, (2015). https://doi.org/10.1007/978-1-4939-2389-2_7.

Bicknell, Ross, Andrew Catanach, Melanie Hand, and Anna Koltunow. “Seeds of Doubt: Mendel’s Choice of Hieracium to Study Inheritance, a Case of Right Plant, Wrong Trait.” *Theoretical and Applied Genetics* 129, no. 12 (December 2016): 2253–66. https://doi.org/10.1007/s00122-016-2788-x.

Bicknell, Ross, Ellen Podivinsky, Andrew Catanach, Sylvia Erasmuson, and Suzanne Lambie. “Strategies for Isolating Mutants in Hieracium with Dysfunctional Apomixis.” *Sexual Plant Reproduction* 14, no. 4 (December 1, 2001): 227–32. https://doi.org/10.1007/s00497-001-0110-8.

Bjourson, A J, and J E Cooper. “Band-Stab PCR: A Simple Technique for the Purification of Individual PCR Products.” *Nucleic Acids Research* 20, no. 17 (September 11, 1992): 4675.

Bocher, Tyge W. “CYTOLOGICAL AND EMBRYOLOGICAL STUDIES IN THE AMPHI-APOMICTIC ARABIS HOLBOELLII COMPLEX,” (1951), 65.

Boutilier, Kim, Remko Offringa, Vijay K. Sharma, Henk Kieft, Thérèse Ouellet, Lemin Zhang, Jiro Hattori, et al. “Ectopic Expression of BABY BOOM Triggers a Conversion from Vegetative to Embryonic Growth.” *The Plant Cell* 14, no. 8 (August 1, 2002): 1737–49. https://doi.org/10.1105/tpc.001941.

Bowden, Rory, Robert W. Davies, Andreas Heger, Alistair T. Pagnamenta, Mariateresa de Cesare, Laura E. Oikkonen, Duncan Parkes, et al. "Sequencing of Human Genomes with Nanopore Technology." *Nature Communications* 10, no. 1 (April 23, 2019): 1869. https://doi.org/10.1038/s41467-019-09637-5.

Bräuning, Sophia, Andrew Catanach, Janice M. Lord, Ross Bicknell, and Richard C. Macknight. "Comparative Transcriptome Analysis of the Wild-Type Model Apomict Hieracium Praealtum and Its Loss of Parthenogenesis (Lop) Mutant." *BMC Plant Biology* 18, no. 1 (September 24, 2018): 206. https://doi.org/10.1186/s12870-018-1423-1.

Carman, John G. "Asynchronous Expression of Duplicate Genes in Angiosperms May Cause Apomixis, Bispory, Tetraspory, and Polyembryony." *Biological Journal of the Linnean Society* 61, no. 1 (May 1, 1997): 51–94. https://doi.org/10.1111/j.1095-8312.1997.tb01778.x.

Carman, John G., Mayelyn Mateo de Arias, Lei Gao, Xinghua Zhao, Becky M. Kowallis, David A. Sherwood, Manoj K. Srivastava, et al. "Apospory and Diplospory in Diploid Boechera (Brassicaceae) May Facilitate Speciation by Recombination-Driven Apomixis-to-Sex Reversals." *Frontiers in Plant Science* 0 (2019). https://doi.org/10.3389/fpls.2019.00724.

Catanach, Andrew S., Sylvia K. Erasmuson, Ellen Podivinsky, Brian R. Jordan, and Ross Bicknell. "Deletion Mapping of Genetic Regions Associated with Apomixis in Hieracium." *Proceedings of the National Academy of Sciences* 103, no. 49 (December 5, 2006): 18650–55. https://doi.org/10.1073/pnas.0605588103.

Conner, Joann A., Maricel Podio, and Peggy Ozias-Akins. "Haploid Embryo Production in Rice and Maize Induced by PsASGR-BBML Transgenes." *Plant Reproduction* 30, no. 1 (March 2017): 41–52. https://doi.org/10.1007/s00497-017-0298-x.

Corral, Jose M., Marcin Piwczynski, and Tim F. Sharbel. "Allelic Sequence Divergence in the Apomictic Boechera Holboellii Complex." In *Lost Sex: The Evolutionary Biology of Parthenogenesis*, edited by Isa Schön, Koen Martens, and Peter Dijk, 495–516. Dordrecht: Springer Netherlands, (2009). https://doi.org/10.1007/978-90-481-2770-2_23.

Corral, José M., Heiko Vogel, Olawale M. Aliyu, Götz Hensel, Thomas Thiel, Jochen Kumlehn, and Timothy F. Sharbel. "A Conserved Apomixis-Specific Polymorphism Is Correlated with Exclusive Exonuclease Expression in Premeiotic Ovules of Apomictic Boechera Species1[W][OPEN]." *Plant Physiology* 163, no. 4 (December 2013): 1660–72. https://doi.org/10.1104/pp.113.222430.

Cosendai, Anne-Caroline, and Elvira Hörandl. "Cytotype Stability, Facultative Apomixis and Geographical Parthenogenesis in Ranunculus Kuepferi (Ranunculaceae)." *Annals of Botany* 105, no. 3 (March 1, 2010): 457– 70. https://doi.org/10.1093/aob/mcp304.

Crow, James F., and Motoo Kimura. "Evolution in Sexual and Asexual Populations." *The American Naturalist* 99, no. 909 (November 1, 1965): 439–50. https://doi.org/10.1086/282389.

Currin, Andrew, Neil Swainston, Mark S Dunstan, Adrian J Jervis, Paul Mulherin, Christopher J Robinson, Sandra Taylor, et al. “Highly Multiplexed, Fast and Accurate Nanopore Sequencing for Verification of Synthetic DNA Constructs and Sequence Libraries.” *Synthetic Biology* 4, no. 1 (January 1, 2019). https://doi.org/10.1093/synbio/ysz025.

Darlington, C. D. *The Evolution of Genetic Systems,*. Cambridge [England: The University Press, (1939).

Downie, Stephen R., and Deborah S. Katz-Downie. “A Molecular Phylogeny of Apiaceae Subfamily Apioideae: Evidence from Nuclear Ribosomal DNA Internal Transcribed Spacer Sequences.” *American Journal of Botany* 83, no. 2 (1996): 234–51. https://doi.org/10.1002/j.1537-2197.1996.tb12701.x.

Drews, Gary N., and Anna M.G Koltunow. “The Female Gametophyte.” *The Arabidopsis Book / American Society of Plant Biologists* 9 (December 26, 2011): e0155. https://doi.org/10.1199/tab.0155.

Ernst, Alfred. *Bastardierung als Ursache der Apogamie im Pflanzenreich: eine Hypothese zur experimentellen Vererbungs- und Abstrammungslehre*. G. Fischer, (1918).

Fehrer, Judith, Anna Krahulcová, František Krahulec, Jindrich Chrtek, Radka Rosenbaumová, and Siegfried Bräutigam. “Evolutionary Aspects in Hieracium Subgenus Pilosella,” 359–90, 2007.

Fei, Xitong, Jingwei Shi, Yulin Liu, Jinshuang Niu, and Anzhi Wei. “The Steps from Sexual Reproduction to Apomixis.” *Planta* 249, no. 6 (June 1, 2019): 1715–30. https://doi.org/10.1007/s00425-019-03113-6.

Fiaz, Sajid, Xiukang Wang, Afifa Younas, Badr Alharthi, Adeel Riaz, and Habib Ali. “Apomixis and Strategies to Induce Apomixis to Preserve Hybrid Vigor for Multiple Generations.” *GM Crops & Food* 12, no. 1 (January 2, 2021): 57– 70. https://doi.org/10.1080/21645698.2020.1808423.

Gadella, T. W. J. “Variation, Hybridization and Reproductive Biology of Hieracium Pilosella L.” *Variation, Hybridization and Reproductive Biology of Hieracium Pilosella L* 94, no. 4 (1991): 455–88.

Grant, Verne. *Plant Speciation*. *Plant Speciation*. Columbia University Press, (1981). https://www.degruyter.com/document/doi/10.7312/gran92318/html.

Grossniklaus, Ueli, Gian A. Nogler, and Peter J. van Dijk. “How to Avoid Sex: The Genetic Control of Gametophytic Apomixis.” *The Plant Cell* 13, no. 7 (July 2001): 1491–98. https://doi.org/10.1105/tpc.13.7.1491.

Hand, M. L., P. Vít, A. Krahulcová, S. D. Johnson, K. Oelkers, H. Siddons, J. Chrtek, J. Fehrer, and A. M. G. Koltunow. “Evolution of Apomixis Loci in Pilosella and Hieracium (Asteraceae) Inferred from the Conservation of Apomixis-Linked Markers in Natural and Experimental Populations.” *Heredity* 114, no. 1 (January 2015): 17–26. https://doi.org/10.1038/hdy.2014.61.

Hand, Melanie L., and Anna M. G. Koltunow. “The Genetic Control of Apomixis: Asexual Seed Formation.” *Genetics* 197, no. 2 (June 2014): 441–50. https://doi.org/10.1534/genetics.114.163105.

Hanna, WAYNE W. “11 - Apomixis in Crop Plants—Cytogenetic Basis and Role in Plant Breeding.” In *Developments in Plant Genetics and Breeding*, edited by P. K. Gupta and T. Tsuchiya, 2:229–42. Chromosome Engineering in Plants: Genetics, Breeding, Evolution, Part A. Elsevier, (1991). https://doi.org/10.1016/B978-0-444-88259-2.50015-1.

Henderson, Sam W., Steven T. Henderson, Marc Goetz, and Anna M. G. Koltunow. “Efficient CRISPR/Cas9-Mediated Knockout of an Endogenous PHYTOENE DESATURASE Gene in T1 Progeny of Apomictic Hieracium Enables New Strategies for Apomixis Gene Identification.” *Genes* 11, no. 9 (September 2020): 1064. https://doi.org/10.3390/genes11091064.

Holsinger, Kent E. “Reproductive Systems and Evolution in Vascular Plants.” *Proceedings of the National Academy of Sciences* 97, no. 13 (June 20, 2000): 7037–42. https://doi.org/10.1073/pnas.97.13.7037.

Hörandl, Elvira, Anne-Caroline Cosendai, and Eva Maria Temsch. “Understanding the Geographic Distributions of Apomictic Plants: A Case for a Pluralistic Approach.” *Plant Ecology & Diversity* 1, no. 2 (November 24, 2008): 309–20. https://doi.org/10.1080/17550870802351175.

Jain, Miten, Sergey Koren, Karen H. Miga, Josh Quick, Arthur C. Rand, Thomas A. Sasani, John R. Tyson, et al. “Nanopore Sequencing and Assembly of a Human Genome with Ultra-Long Reads.” *Nature Biotechnology* 36, no. 4 (April 2018): 338–45. https://doi.org/10.1038/nbt.4060.

Kandemir̈, Nejdet, and İbrahim̈ Saygili. “Apomixis: New Horizons in Plant Breeding.” *TURKISH JOURNAL OF AGRICULTURE AND FORESTRY* 39, no. 4 (July 23, 2015): 549–56.

Khanday, Imtiyaz, Debra Skinner, Bing Yang, Raphael Mercier, and Venkatesan Sundaresan. “A Male-Expressed Rice Embryogenic Trigger Redirected for Asexual Propagation through Seeds.” *Nature* 565, no. 7737 (January 2019): 91–95. https://doi.org/10.1038/s41586-018-0785-8.

Koltunow, AM. “Apomixis: Embryo Sacs and Embryos Formed without Meiosis or Fertilization in Ovules.” *The Plant Cell* 5, no. 10 (October 1993): 1425–37.

Koltunow, Anna M. G., Susan D. Johnson, Julio C. M. Rodrigues, Takashi Okada, Yingkao Hu, Tohru Tsuchiya, Saira Wilson, et al. “Sexual Reproduction Is the Default Mode in Apomictic Hieracium Subgenus Pilosella, in Which Two Dominant Loci Function to Enable Apomixis.” *The Plant Journal* 66, no. 5 (2011): 890–902. https://doi.org/10.1111/j.1365-313X.2011.04556.x.

Kotani, Yoshiko, Steven T. Henderson, Go Suzuki, Susan D. Johnson, Takashi Okada, Hayley Siddons, Yasuhiko Mukai, and Anna M. G. Koltunow. “The LOSS OF APOMEIOSIS (LOA) Locus in Hieracium Praealtum Can Function Independently of the Associated Large-Scale Repetitive Chromosomal Structure.” *New Phytologist* 201, no. 3 (2014): 973–81. https://doi.org/10.1111/nph.12574.

Kubat, Zdenek, Roman Hobza, Boris Vyskot, and Eduard Kejnovsky. “Microsatellite Accumulation in the Y Chromosome OfSilene Latifolia.” *Genome / National Research Council Canada = Génome / Conseil National de Recherches Canada* 51 (June 1, 2008): 350–56. https://doi.org/10.1139/G08-024.

Kumar, Ajay, and Shalu Jain. “Forward Genetics Using Radiation Hybrids (Deletion Mutants) in Plants.” *Indian Journal of Plant Physiology* 23, no. 4 (December 2018): 622–29. https://doi.org/10.1007/s40502-018-0419-z.

Kumar, Ajay, Kristin Simons, Muhammad J. Iqbal, Monika Michalak de Jiménez, Filippo M. Bassi, Farhad Ghavami, Omar Al-Azzam, et al. “Physical Mapping Resources for Large Plant Genomes: Radiation Hybrids for Wheat D-Genome Progenitor Aegilops Tauschii.” *BMC Genomics* 13, no. 1 (November 5, 2012): 597. https://doi.org/10.1186/1471-2164-13-597.

Leblanc, Olivier, and Andrea Mazzucato. “Screening Procedures to Identify and Quantify Apomixis,” (January 1, 2001).

Lokki, Juhani. “Genetic Polymorphism and Evolution in Parthenogenetic Animals.” *Hereditas* 83, no. 1 (1976): 57–63. https://doi.org/10.1111/j.1601-5223.1976.tb01570.x.

Majeský, Ľuboš, Radim J. Vašut, Miloslav Kitner, and Bohumil Trávníček. “The Pattern of Genetic Variability in Apomictic Clones of Taraxacum Officinale Indicates the Alternation of Asexual and Sexual Histories of Apomicts.” *PLOS ONE* 7, no. 8 (août 2012): e41868. https://doi.org/10.1371/journal.pone.0041868.

Mateo de Arias, Mayelyn. “Effects of Plant Stress on Facultative Apomixis in Boechera (Brassicaceae).” *All Graduate Theses and Dissertations*, May 1, (2015). https://digitalcommons.usu.edu/etd/4451.

Matzk, Fritz, Sanja Prodanovic, Helmut Bäumlein, and Ingo Schubert. “The Inheritance of Apomixis in Poa Pratensis Confirms a Five Locus Model with Differences in Gene Expressivity and Penetrance.” *The Plant Cell* 17, no. 1 (January 1, 2005): 13–24. https://doi.org/10.1105/tpc.104.027359.

McGee, Rob. “Allelism and Allele Sequence Divergence of LOP, the Locus of Parthenogenesis in the Model Apomict Hieracium Praealtum (Asteraceae),” (2013).

Mendel, Gregor. “Ueber einige aus künstlicher Befruchtung gewonnenen Hieracium Bastarde” *Verhandlungen des naturforschenden Vereines in Brünn* – translated in english by “On Hieracium-hybrids obtained by artificial fertilization” (1869).

Matzk, F., H.-M. Meyer, H. Bäumlein, H.-J. Balzer, and I. Schubert. “A Novel Approach to the Analysis of the Initiation of Embryo Development in Gramineae.” *Sexual Plant Reproduction* 8, no. 5 (September 1, 1995): 266–72. https://doi.org/10.1007/BF00229382.

Mráz, Patrik, and Pavel Zdvořák. “Reproductive Pathways in Hieracium s.s. (Asteraceae): Strict Sexuality in Diploids and Apomixis in Polyploids.” *Annals of Botany* 123, no. 2 (January 23, 2019): 391–403. https://doi.org/10.1093/aob/mcy137.

Muller, H. J. “Some Genetic Aspects of Sex.” *The American Naturalist* 66, no. 703 (March 1, 1932): 118–38. https://doi.org/10.1086/280418.

Munoz-Lopez, Martin, and Jose L. Garcia-Perez. “DNA Transposons: Nature and Applications in Genomics.” *Current Genomics* 11, no. 2 (April 1, 2010): 115–28. https://doi.org/10.2174/138920210790886871.

Nogler, G. A. “Genetics of Gametophytic Apomixis - a Historical Sketch.” *Polish Botanical Studies* 08 (1994). http://agro.icm.edu.pl/agro/element/bwmeta1.element.agro-article-180f0f3a-69c5-478b-8d73-f0557aa61890.

Okada, Takashi, Yingkao Hu, Matthew R. Tucker, Jennifer M. Taylor, Susan D. Johnson, Andrew Spriggs, Tohru Tsuchiya, Karsten Oelkers, Julio C.M. Rodrigues, and Anna M.G. Koltunow. “Enlarging Cells Initiating Apomixis in Hieracium Praealtum Transition to an Embryo Sac Program Prior to Entering Mitosis.” *Plant Physiology* 163, no. 1 (2013): 216–31.

Ohno, Susumu. *Evolution by Gene Duplication*. Berlin Heidelberg: Springer-Verlag, (1970). https://doi.org/10.1007/978-3-642-86659-3.

Ozias-Akins, Peggy, Joann A. Conner, Shailendra Goel, Zhenbang Chen, Yukio Akiyama, and Wayne W. Hanna. “Genomic Structure of the Apomixis Locus in Pennisetum.” In *Plant Biotechnology 2002 and Beyond: Proceedings of the 10th IAPTC&B Congress June 23–28, 2002 Orlando, Florida, U.S.A.*, edited by Indra K. Vasil, 515–18. Dordrecht: Springer Netherlands, (2003). https://doi.org/10.1007/978-94-017-2679-5_106.

Ozias-Akins, Peggy, Dominique Roche, and Wayne W. Hanna. “Tight Clustering and Hemizygosity of Apomixis-Linked Molecular Markers in Pennisetum Squamulatum Implies Genetic Control of Apospory by a Divergent Locus That May Have No Allelic Form in Sexual Genotypes.” *Proceedings of the National Academy of Sciences* 95, no. 9 (April 28, 1998): 5127–32. https://doi.org/10.1073/pnas.95.9.5127.

Ozias-Akins, Peggy, and Peter J. van Dijk. “Mendelian Genetics of Apomixis in Plants.” *Annual Review of Genetics* 41, no. 1 (December 2007): 509–37. https://doi.org/10.1146/annurev.genet.40.110405.090511.

Palmer, Lance E., Mathaeus Dejori, Randall Bolanos, and Daniel Fasulo. “Improving de Novo Sequence Assembly Using Machine Learning and Comparative Genomics for Overlap Correction.” *BMC Bioinformatics* 11, no. 1 (January 15, 2010): 33. https://doi.org/10.1186/1471-2105-11-33.

Pamilo, Pekka. “Heterozygosity in Apomictic Organisms.” *Hereditas* 107, no. 1 (1987): 95–101. https://doi.org/10.1111/j.1601-5223.1987.tb00272.x.

Paule, Juraj, Timothy F. Sharbel, and Christoph Dobeš. “Apomictic and Sexual Lineages of the Potentilla Argentea L. Group (Rosaceae): Cytotype and Molecular Genetic Differentiation.” *TAXON* 60, no. 3 (2011): 721–32. https://doi.org/10.1002/tax.603008.

Powell, Wayne, Gordon C. Machray, and Jim Provan. “Polymorphism Revealed by Simple Sequence Repeats.” *Trends in Plant Science* 1, no. 7 (July 1, 1996): 215–22. https://doi.org/10.1016/1360-1385(96)86898-1.

Pupilli, Fulvio, and Gianni Barcaccia. “Cloning Plants by Seeds: Inheritance Models and Candidate Genes to Increase Fundamental Knowledge for Engineering Apomixis in Sexual Crops.” *Journal of Biotechnology* 159, no. 4 (June 2012): 291–311. https://doi.org/10.1016/j.jbiotec.2011.08.028.

Rathore, Priyanka, Soom Nath Raina, Suresh Kumar, and Vishnu Bhat. “Retro-Element Gypsy-163 Is Differentially Methylated in Reproductive Tissues of Apomictic and Sexual Plants of Cenchrus Ciliaris.” *Frontiers in Genetics* 0 (2020). https://doi.org/10.3389/fgene.2020.00795.

Ravi, Maruthachalam, Mohan P. A. Marimuthu, and Imran Siddiqi. "Gamete Formation without Meiosis in Arabidopsis." *Nature* 451, no. 7182 (February 2008): 1121–24. https://doi.org/10.1038/nature06557.

Rövekamp, Moritz, John L. Bowman, and Ueli Grossniklaus. "Marchantia MpRKD Regulates the Gametophyte-Sporophyte Transition by Keeping Egg Cells Quiescent in the Absence of Fertilization." *Current Biology: CB* 26, no. 13 (July 11, 2016): 1782–89. https://doi.org/10.1016/j.cub.2016.05.028.

Rubin, Eitan, Gila Lithwick, and Avraham A Levy. "Structure and Evolution of the HAT Transposon Superfamily." *Genetics* 158, no. 3 (July 1, 2001): 949–57. https://doi.org/10.1093/genetics/158.3.949.

Sailer, Christian, Jürg Stöcklin, and Ueli Grossniklaus. "Dynamics of Apomictic and Sexual Reproduction during Primary Succession on a Glacier Forefield in the Swiss Alps." *Scientific Reports* 10, no. 1 (May 19, 2020): 8269. https://doi.org/10.1038/s41598-020-64367-9.

Savidan, Yves, ed., J G Carman, ed., T Dresselhaus, and ed. "The Flowering of Apomixis : From Mechanisms to Genetic Engineering," n.d., 257.

Sharbel, Timothy F, Marie-Luise Voigt, José María Corral, Thomas Thiel, Alok Varshney, Jochen Kumlehn, Heiko Vogel, and Björn Rotter. "Molecular Signatures of Apomictic and Sexual Ovules in the Boechera Holboellii Complex." *The Plant Journal* 58, no. 5 (June 1, 2009): 870–82. https://doi.org/10.1111/j.1365-313x.2009.03826.x.

Shizuya, H., B. Birren, U. J. Kim, V. Mancino, T. Slepak, Y. Tachiiri, and M. Simon. "Cloning and Stable Maintenance of 300-Kilobase-Pair Fragments of Human DNA in Escherichia Coli Using an F-Factor-Based Vector." *Proceedings of the National Academy of Sciences* 89, no. 18 (September 15, 1992): 8794–97. https://doi.org/10.1073/pnas.89.18.8794.

Smith, John. "XXXII. Notice of a Plant Which Produces Perfect Seeds without Any Apparent Action of Pollen." *Transactions of the Linnean Society of London* os-18, no. 4 (August 1, 1841): 509–12. https://doi.org/10.1111/j.1095-8339.1838.tb00200.x.

Smith, J. Maynard. "Group Selection." *The Quarterly Review of Biology* 51, no. 2 (June 1976): 277–83. https://doi.org/10.1086/409311.

Spillane, C., A. Steimer, and U. Grossniklaus. "Apomixis in Agriculture: The Quest for Clonal Seeds." *Sexual Plant Reproduction* 14, no. 4 (December 1, 2001): 179–87. https://doi.org/10.1007/s00497-001-0117-1.

Spillane, Charles, Mark D. Curtis, and Ueli Grossniklaus. "Apomixis Technology Development—Virgin Births in Farmers' Fields?" *Nature Biotechnology* 22, no. 6 (June 2004): 687–91. https://doi.org/10.1038/nbt976.

Stebbins, G. Ledyard. *Variation and Evolution in Plants*. *Variation and Evolution in Plants*. Columbia University Press, (1950). https://www.degruyter.com/document/doi/10.7312/steb94536/html.

Tautz, Diethard. "Hypervariability of Simple Sequences as a General Source for Polymorphic DNA Markers." *Nucleic Acids Research* 17, no. 16 (August 25, 1989): 6463–71. https://doi.org/10.1093/nar/17.16.6463.

Tucker, Matthew R., Ana-Claudia G. Araujo, Nicholas A. Paech, Valerie Hecht, Ed D. L. Schmidt, Jan-Bart Rossell, Sacco C. de Vries, and Anna M. G. Koltunow. “Sexual and Apomictic Reproduction in Hieracium Subgenus Pilosella Are Closely Interrelated Developmental Pathways.” *The Plant Cell* 15, no. 7 (July 1, 2003): 1524–37. https://doi.org/10.1105/tpc.011742.

Tucker, Matthew R., Anna M. G. Koltunow, Matthew R. Tucker, and Anna M. G. Koltunow. “Sexual and Asexual (Apomictic) Seed Development in Flowering Plants: Molecular, Morphological and Evolutionary Relationships.” *Functional Plant Biology* 36, no. 6 (June 1, 2009): 490–504. https://doi.org/10.1071/FP09078.

Tyler, Andrea D., Laura Mataseje, Chantel J. Urfano, Lisa Schmidt, Kym S. Antonation, Michael R. Mulvey, and Cindi R. Corbett. “Evaluation of Oxford Nanopore’s MinION Sequencing Device for Microbial Whole Genome Sequencing Applications.” *Scientific Reports* 8, no. 1 (July 19, 2018): 10931. https://doi.org/10.1038/s41598-018-29334-5.

Underwood, Charles Joseph, Diana RIGOLA, Peter Johannes VAN DIJK, Rik Hubertus Martinus OP DEN CAMP, Michael Eric SCHRANZ, and Catharina Adriana VIJVERBERG. Gene for parthenogenesis. *World Intellectual Property Organization WO2020239984A1*, (December 3, 2020). https://patents.google.com/patent/WO2020239984A1/en.

Vandel, Albert. “La spanandrie, la parthénogenèse géographique et la polyploïdie (?) chez les Curculionides.” *Bulletin de la Société entomologique de France* 37, no. 17 (1932): 255–56.

Van Dijk, Peter and Jos van Damme. “Apomixis Technology and the Paradox of Sex.” *Trends in Plant Science* 5, no. 2 (February 2000): 81–84. https://doi.org/10.1016/S1360-1385(99)01545-9.

Van Dijk, Peter. “Apomixis: Basics for Non-Botanists.” In *Lost Sex: The Evolutionary Biology of Parthenogenesis*, edited by Isa Schön, Koen Martens, and Peter Dijk, 47–62. Dordrecht: Springer Netherlands, (2009). https://doi.org/10.1007/978-90-481-2770-2_3.

Van Dijk, Peter, Hans de Jong, Kitty Vijverberg, and Arjen Biere. “An Apomixis-Gene’s View on Dandelions.” In *Lost Sex: The Evolutionary Biology of Parthenogenesis*, edited by Isa Schön, Koen Martens, and Peter Dijk, 475–93. Dordrecht: Springer Netherlands, (2009). https://doi.org/10.1007/978-90-481-2770-2_22.

Victoria Wang, Xin, Natalie Blades, Jie Ding, Razvan Sultana, and Giovanni Parmigiani. “Estimation of Sequencing Error Rates in Short Reads.” *BMC Bioinformatics* 13, no. 1 (July 30, 2012): 185. https://doi.org/10.1186/1471-2105-13-185.

White, T, Tom Bruns, Steven Lee, John Taylor, M Innis, D Gelfand, and John Sninsky. “Amplification and Direct Sequencing of Fungal Ribosomal RNA Genes for Phylogenetics.” In *Pcr Protocols: A Guide to Methods and Applications,* 31:315–22, (1990).

Whitton, Jeannette, Christopher J. Sears, Eric J. Baack, and Sarah P. Otto. “The Dynamic Nature of Apomixis in the Angiosperms.” *International Journal of Plant Sciences* 169, no. 1 (January 1, 2008): 169–82. https://doi.org/10.1086/523369.

Williams, George Christopher. Sex and Evolution. (MPB-8)*, Volume 8*. *Sex and Evolution. (MPB-8), Volume 8*. Princeton University Press, 2020. https://doi.org/10.12987/9780691209920.

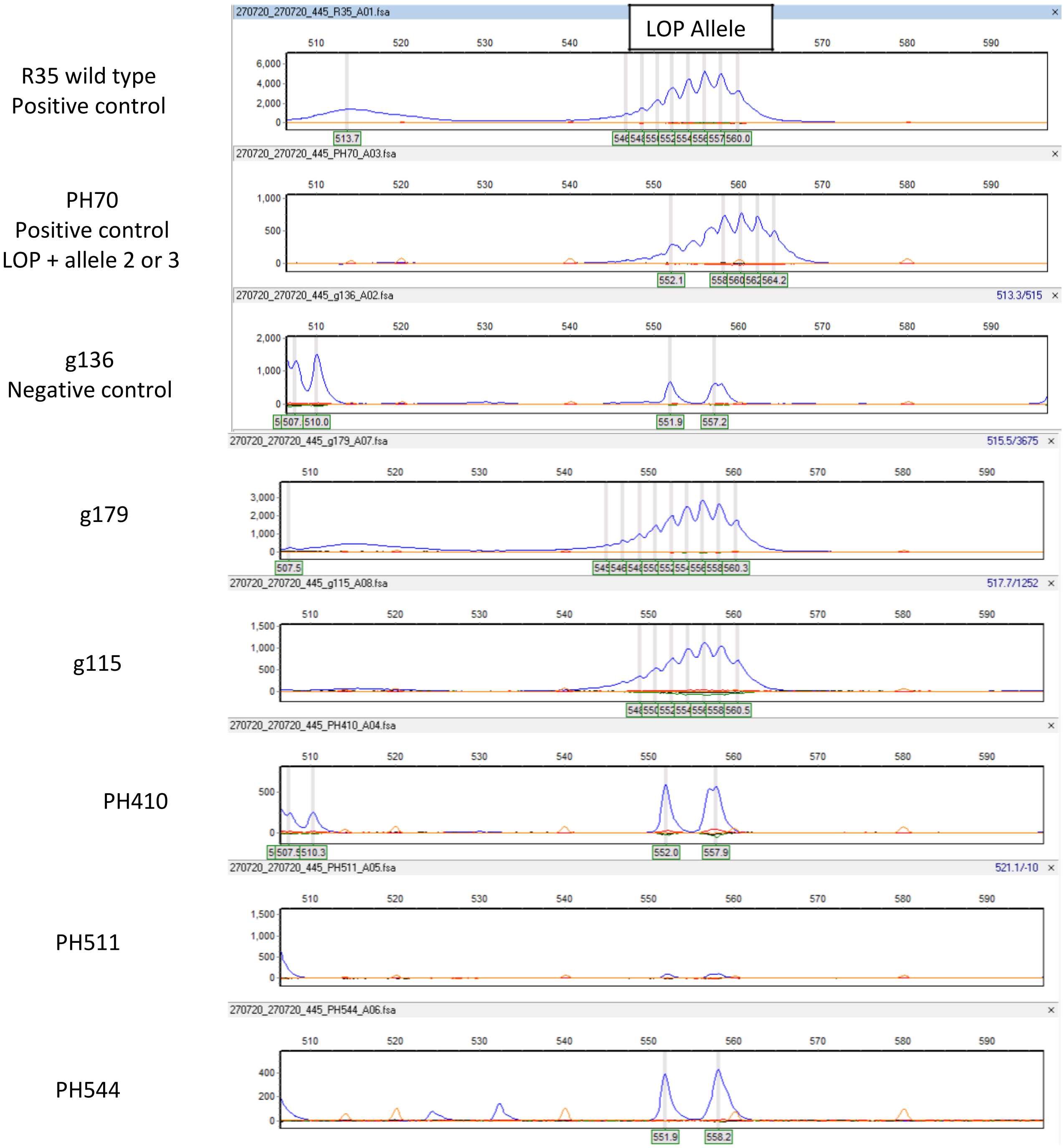
Winkler, Hans. “Parthenogenesis und Apogamie im Pflanzenreiche,” (1908). https://agris.fao.org/agris-search/search.do?recordID=US201300357120.

Appendix A

Fragment analyzer data of the markers identified in the mapping study

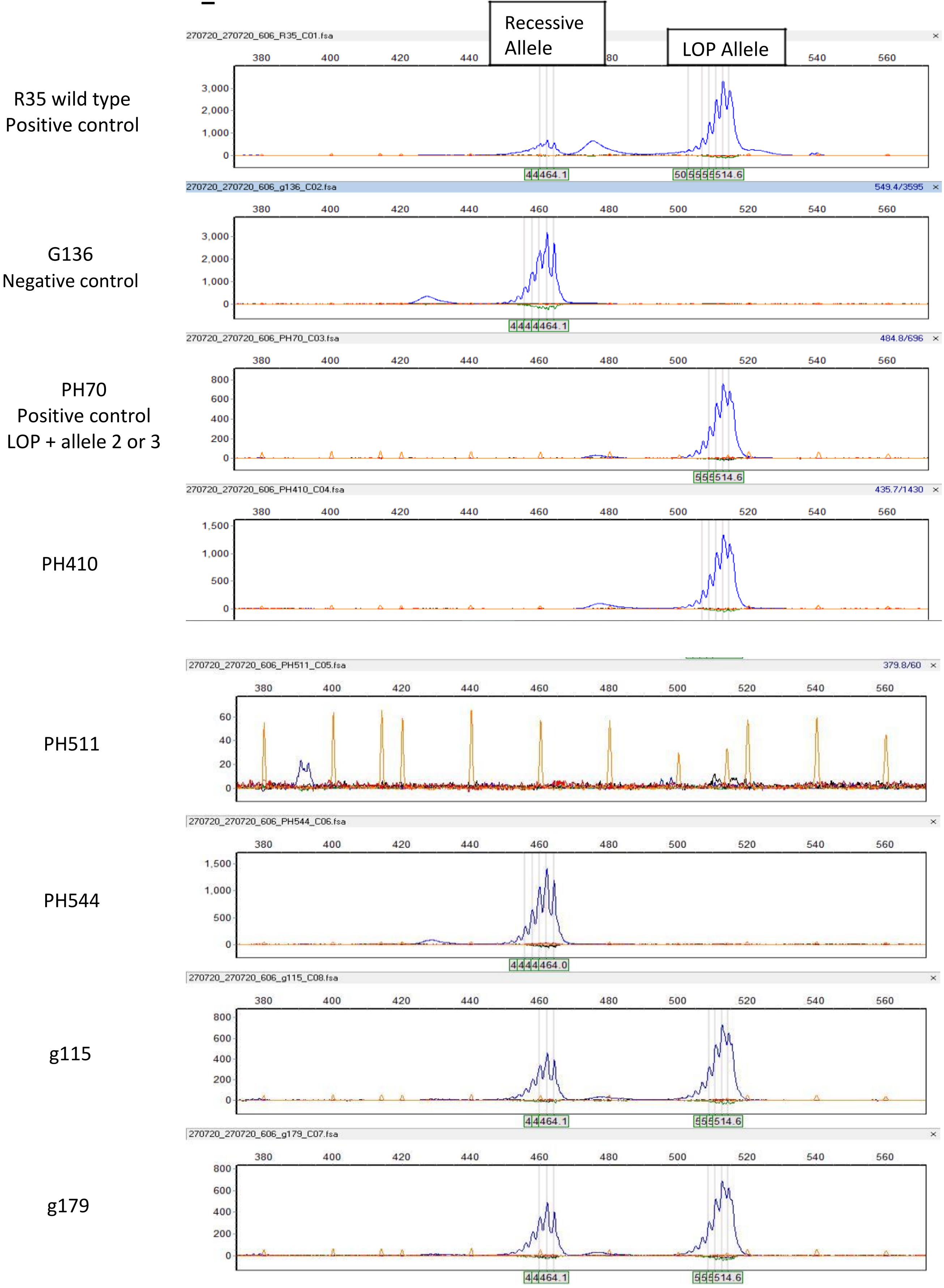
Viewed on the GeneMarker software

A.1 Marker 445_AG

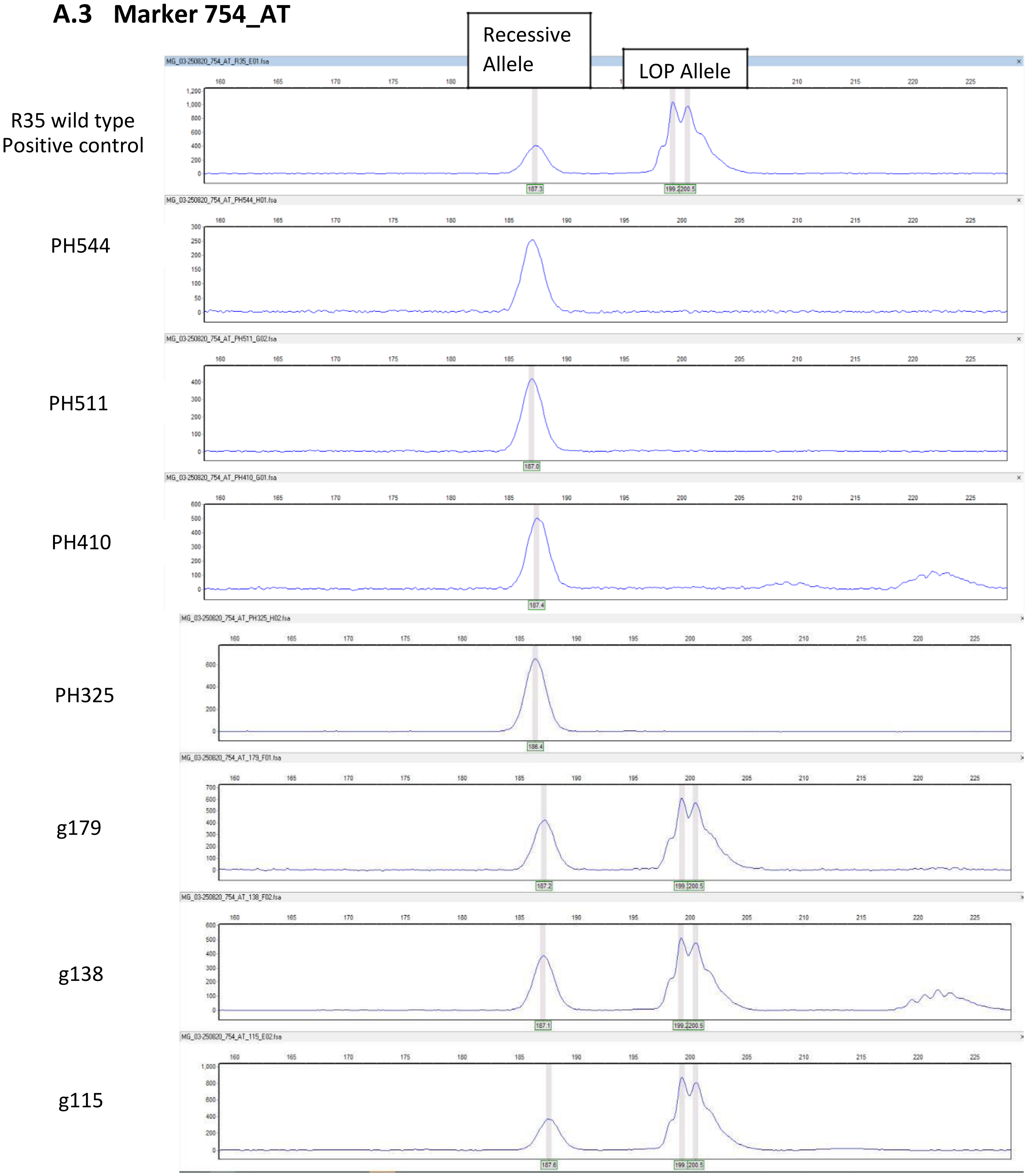


The LOP allele and another recessive allele are on top of each other. When only two peaks are visible, the LOP allele is absent.

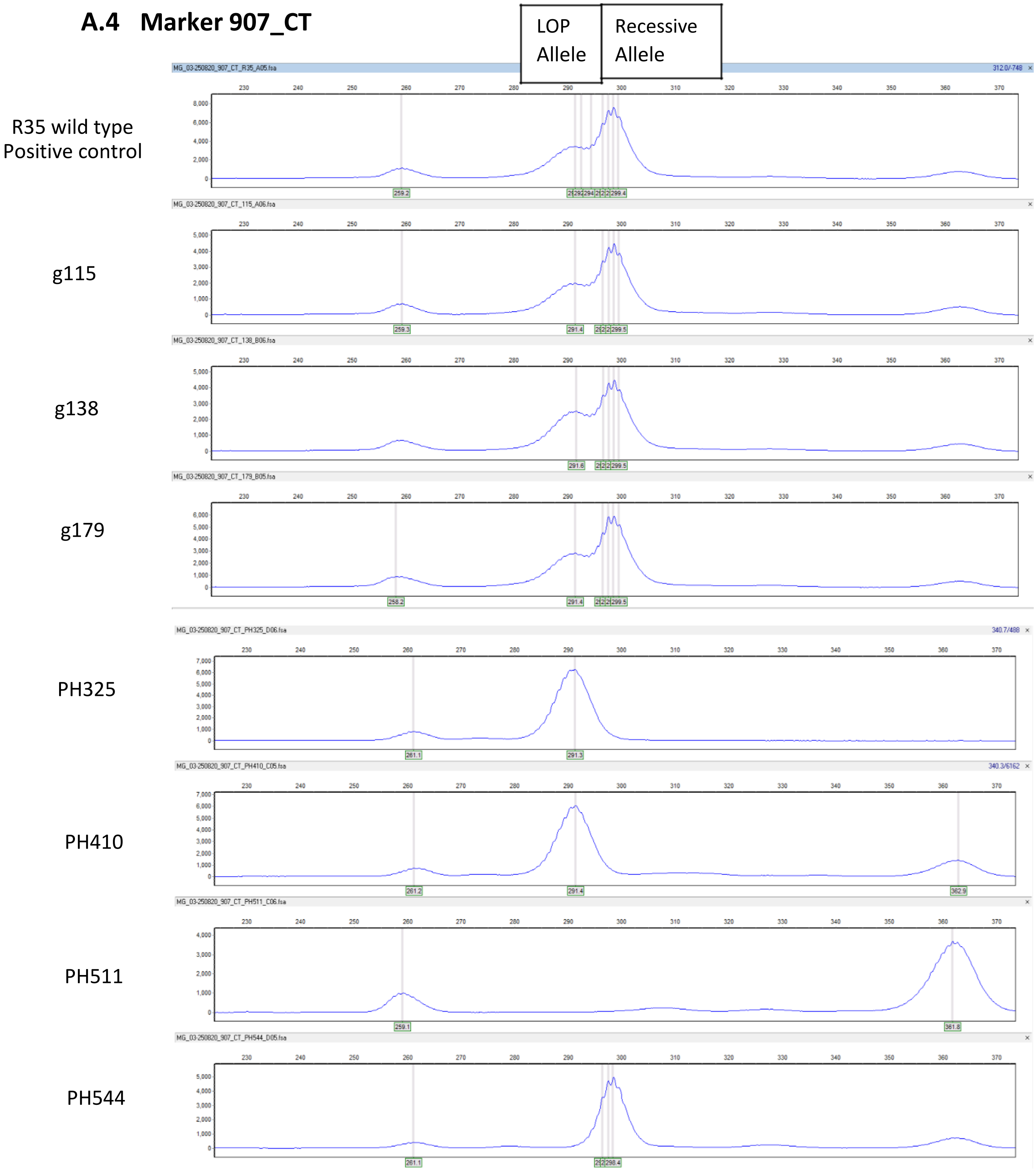
A.2 Marker 606_CT



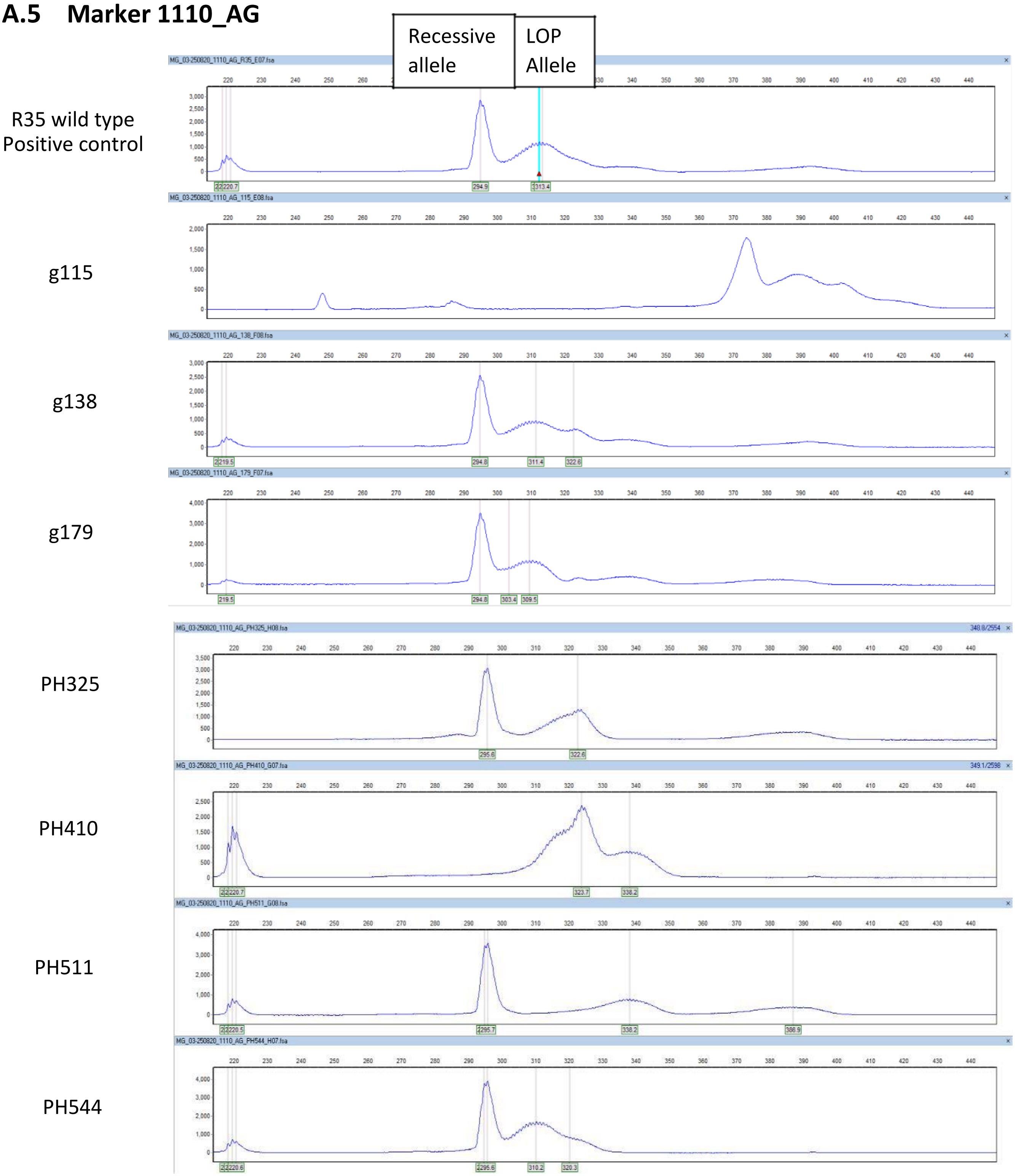
A.3 Marker 754_AT



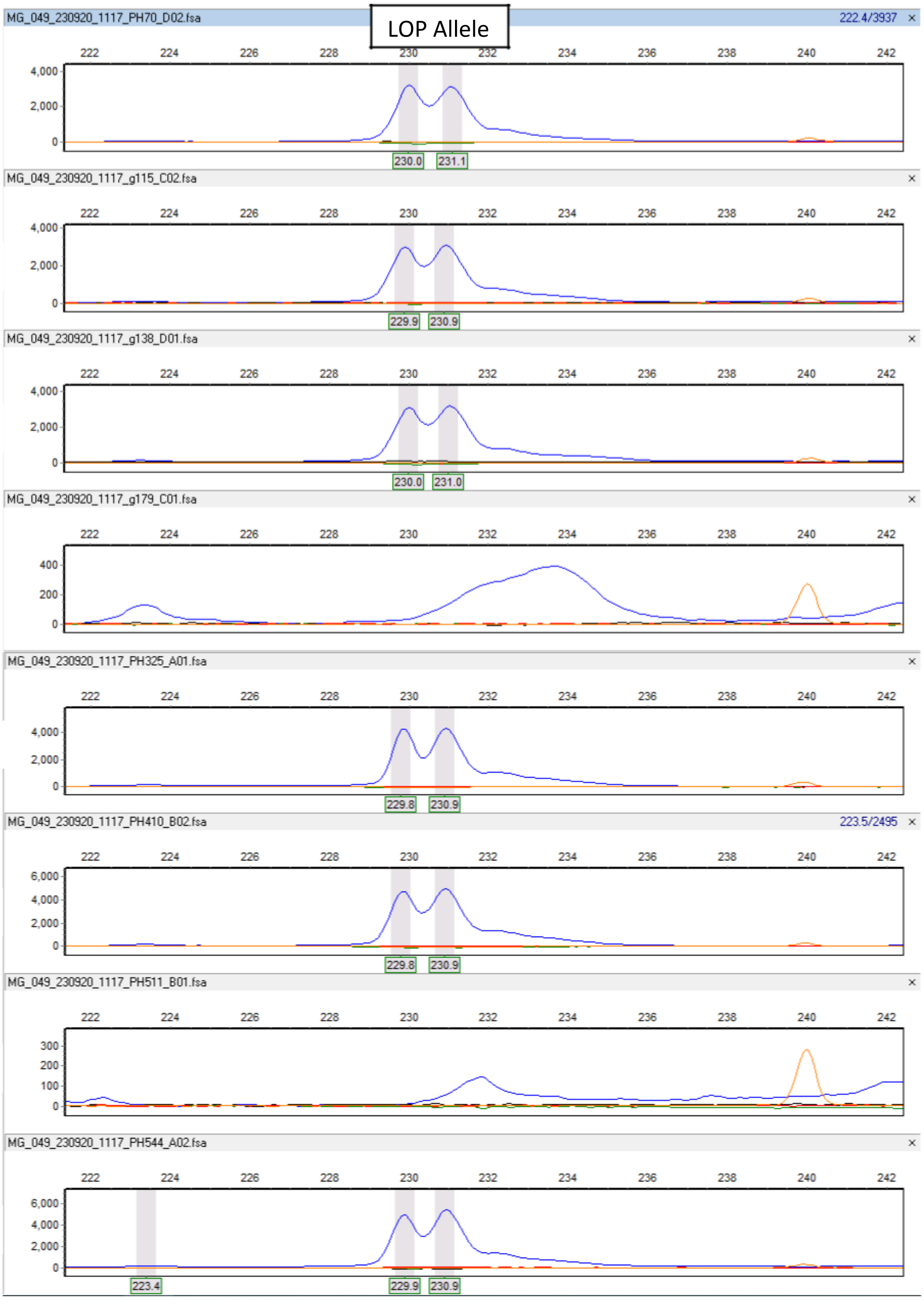
A.4 Marker 907_CT



A.5 Marker 1110_AG



A.6 Marker 1117



PH70
Positive control
LOP + allele 2 or 3

g115

g138

g179
Negative control

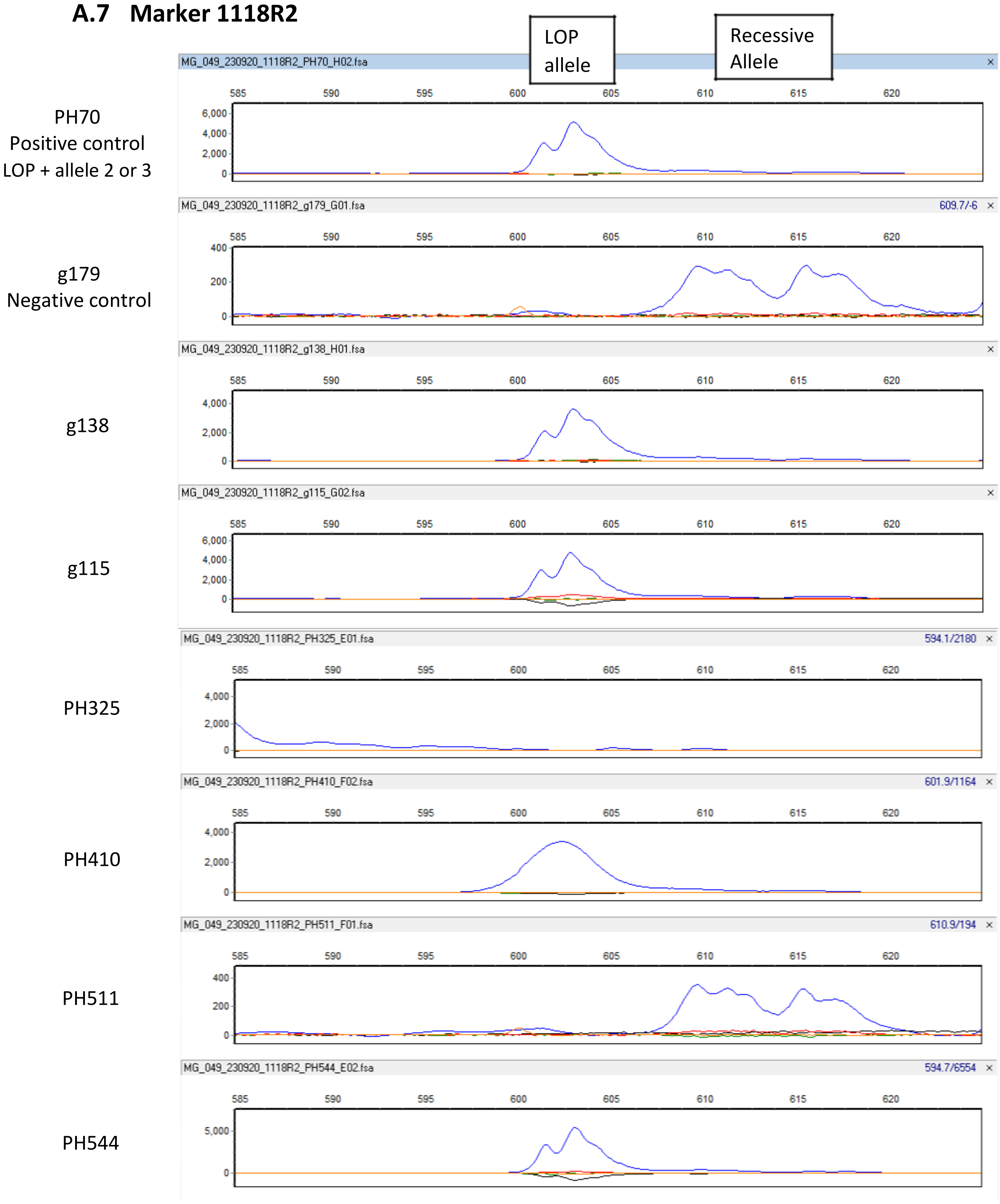
PH325

PH410

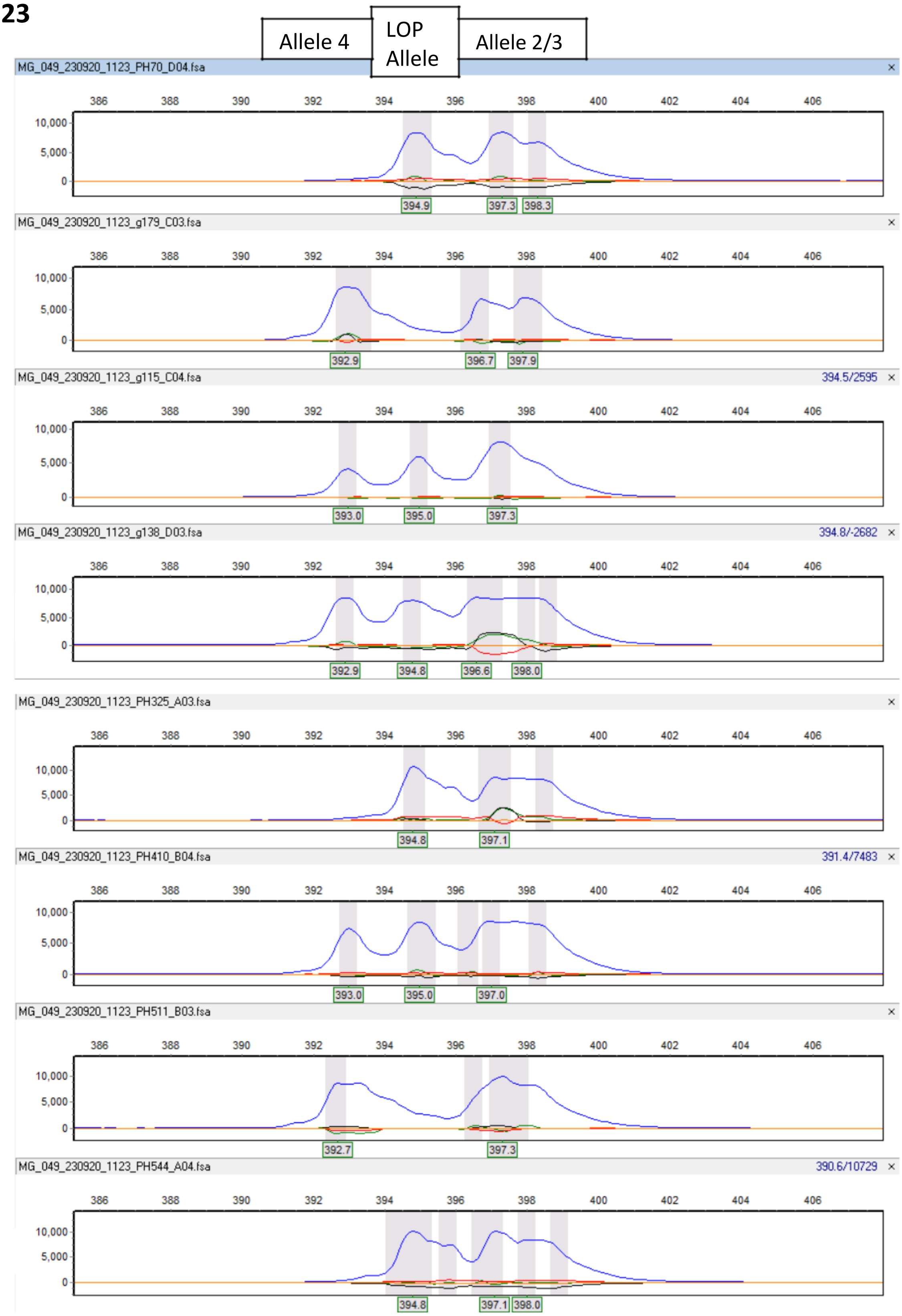
PH511

PH544

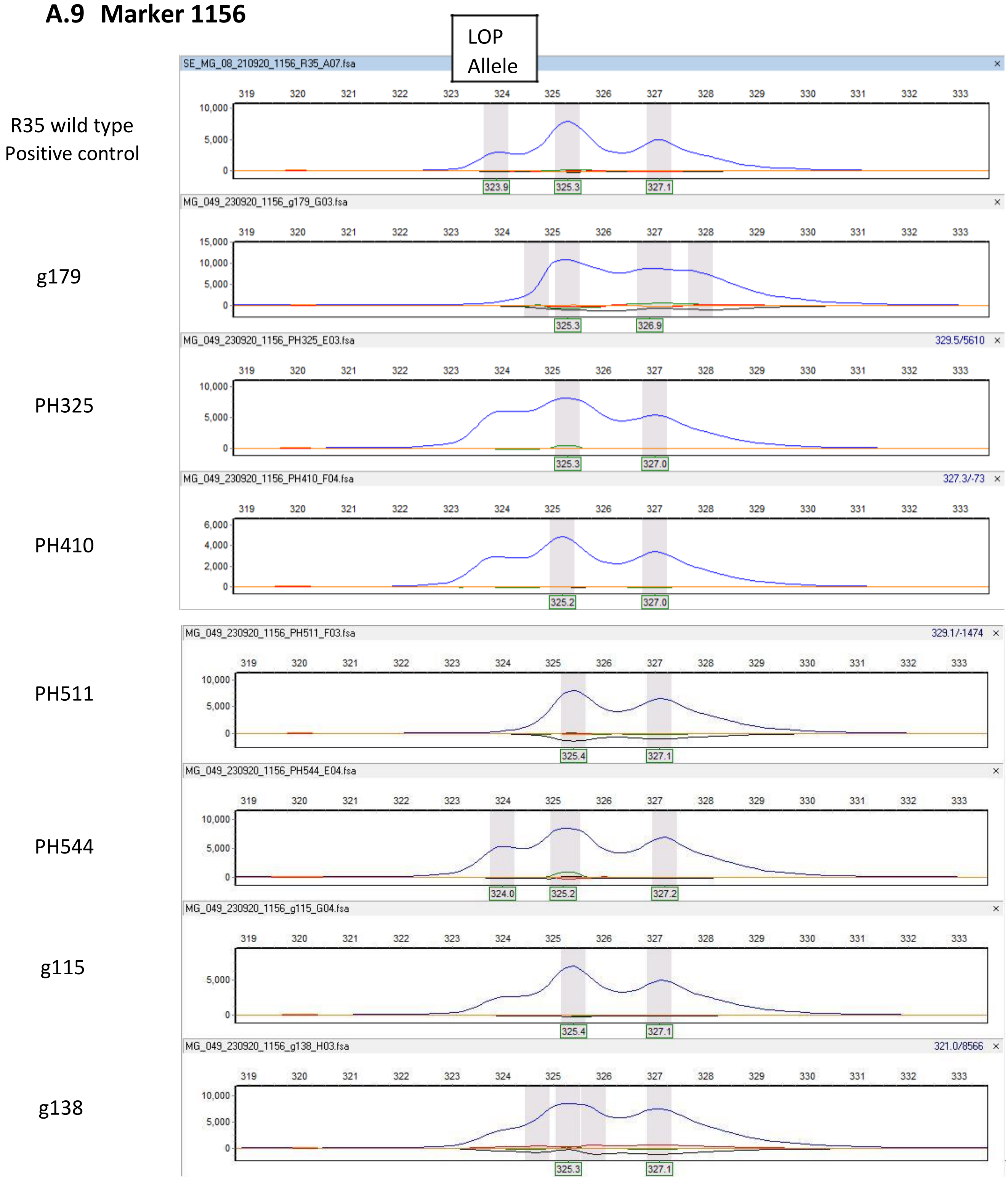
A.7 Marker 1118R2



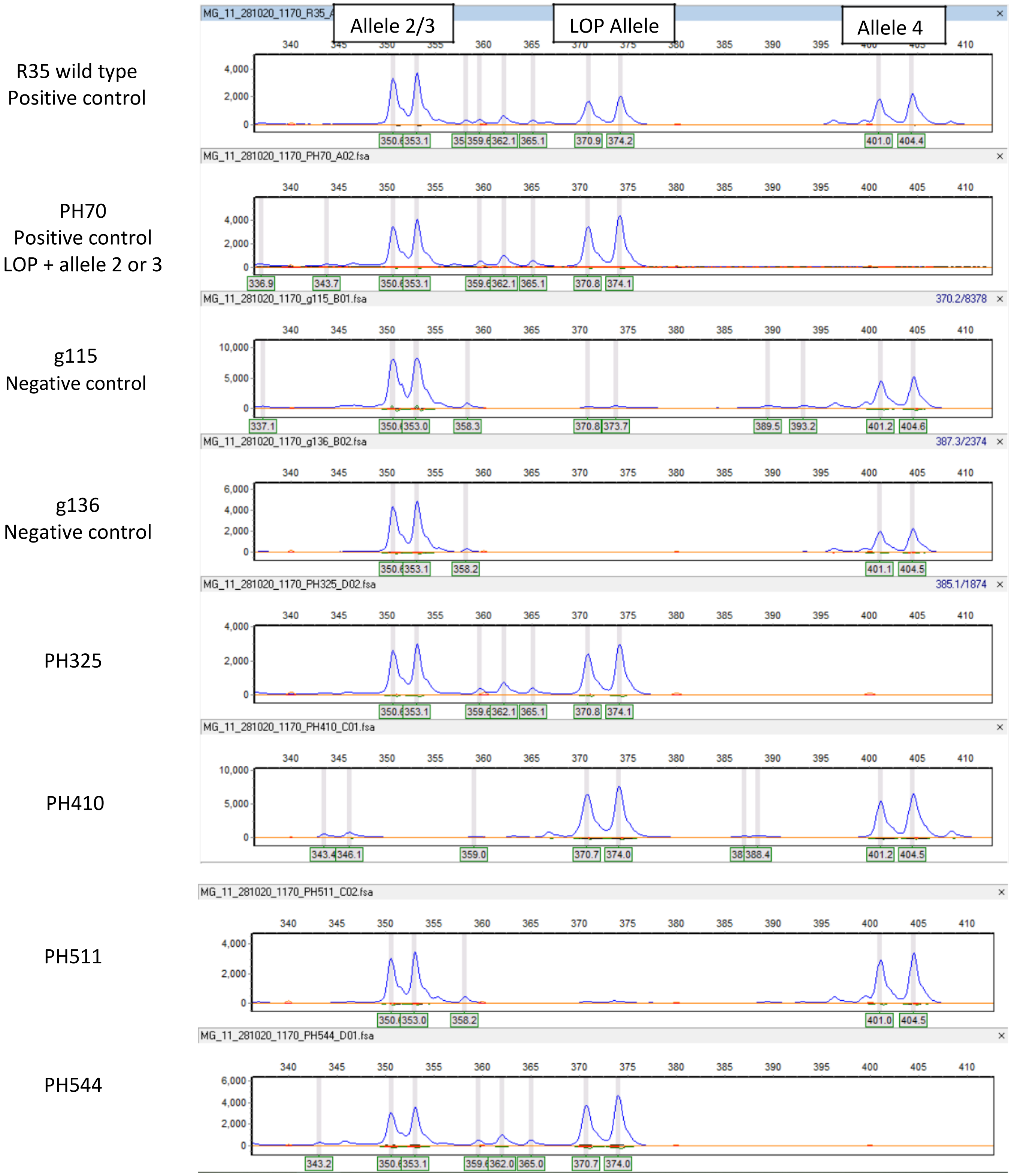
A.8 Marker 1123



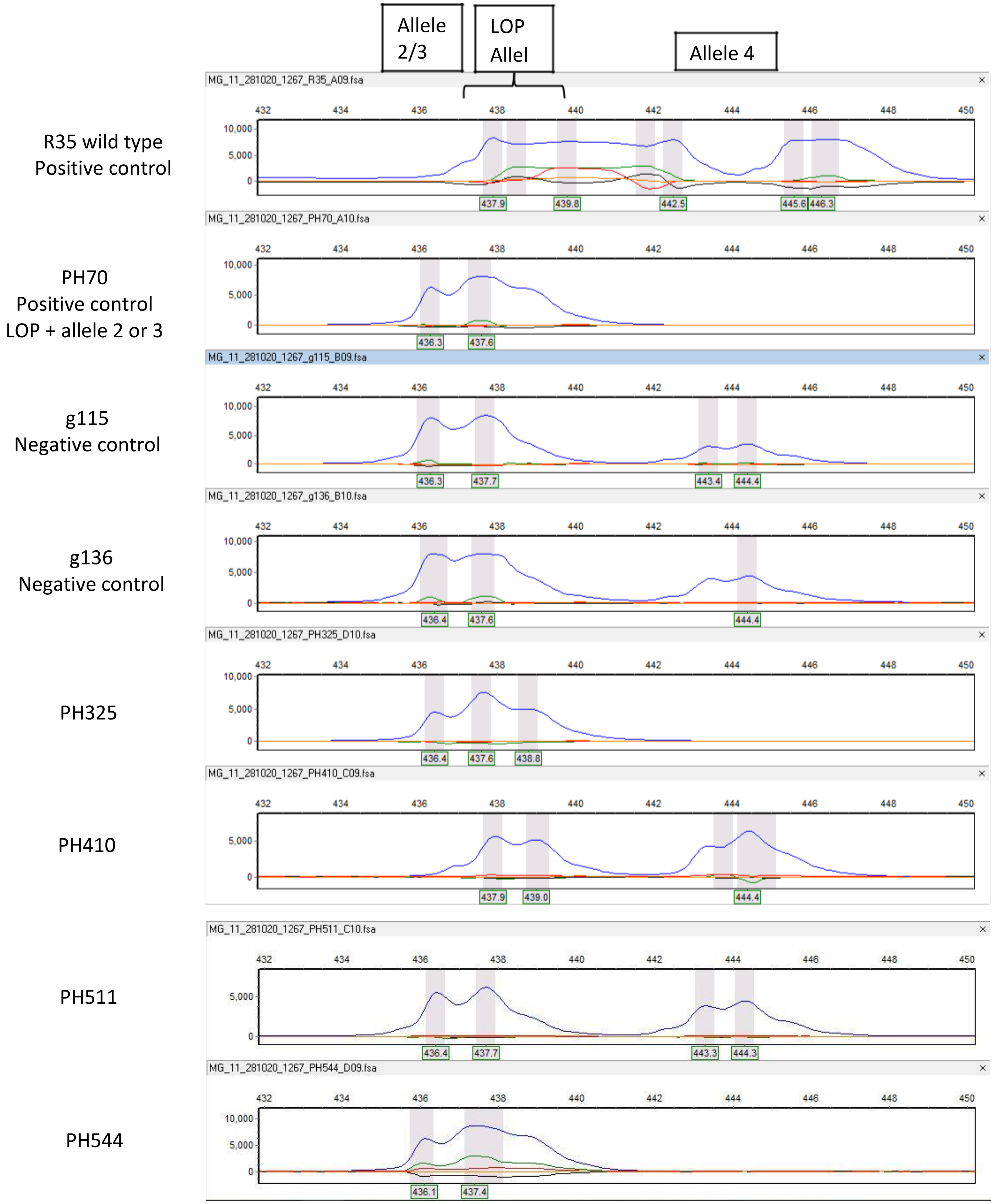
A.9 Marker 1156



A.10 Marker 1170



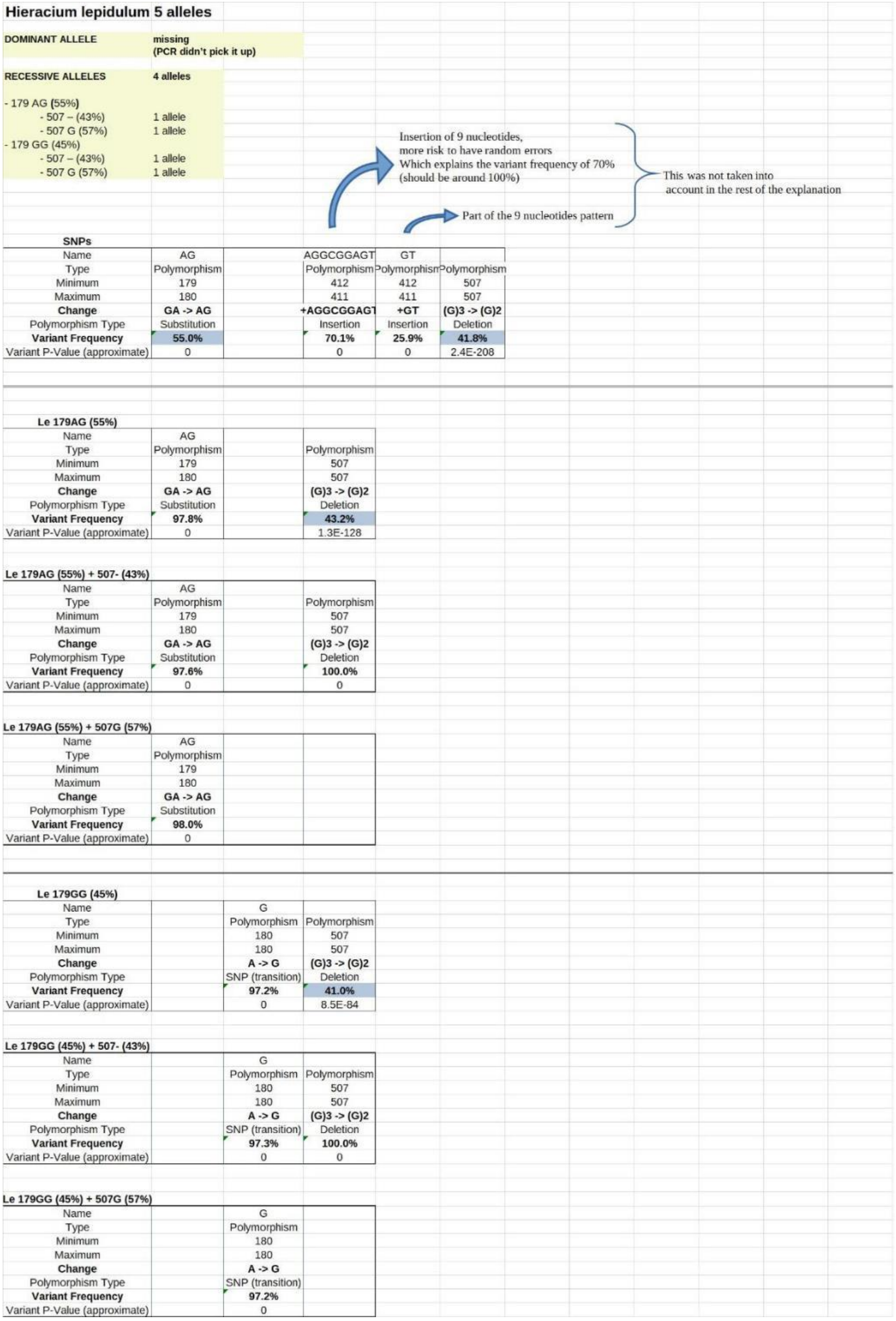
A.11 Marker 1267



Appendix B

Detail of the SNPs for each allele of the HpPAR gene for each species

Each decision step is presented, the final alleles are framed in a darker line.



Hieracium lepidulum 5 alleles

DOMINANT ALLELE	**missing (PCR didn't pick it up)**
RECESSIVE ALLELES	**4 alleles**
- 179 AG (55%)	
- 507 – (43%)	1 allele
- 507 G (57%)	1 allele
- 179 GG (45%)	
- 507 – (43%)	1 allele
- 507 G (57%)	1 allele

Insertion of 9 nucleotides,
more risk to have random errors
Which explains the variant frequency of 70%
(should be around 100%)

Part of the 9 nucleotides pattern

This was not taken into
account in the rest of the explanation

SNPs

Name	AG		AGGCGGAGT	GT	
Type	Polymorphism		Polymorphism	Polymorphism	Polymorphism
Minimum	179		412	412	507
Maximum	180		411	411	507
Change	**GA -> AG**		**+AGGCGGAGT**	**+GT**	**(G)3 -> (G)2**
Polymorphism Type	Substitution		Insertion	Insertion	Deletion
Variant Frequency	**55.0%**		**70.1%**	**25.9%**	**41.8%**
Variant P-Value (approximate)	0		0	0	2.4E-208

Le 179AG (55%)

Name	AG		
Type	Polymorphism		Polymorphism
Minimum	179		507
Maximum	180		507
Change	**GA -> AG**		**(G)3 -> (G)2**
Polymorphism Type	Substitution		Deletion
Variant Frequency	**97.8%**		**43.2%**
Variant P-Value (approximate)	0		1.3E-128

Le 179AG (55%) + 507- (43%)

Name	AG		
Type	Polymorphism		Polymorphism
Minimum	179		507
Maximum	180		507
Change	**GA -> AG**		**(G)3 -> (G)2**
Polymorphism Type	Substitution		Deletion
Variant Frequency	**97.6%**		**100.0%**
Variant P-Value (approximate)	0		0

Le 179AG (55%) + 507G (57%)

Name	AG		
Type	Polymorphism		
Minimum	179		
Maximum	180		
Change	**GA -> AG**		
Polymorphism Type	Substitution		
Variant Frequency	**98.0%**		
Variant P-Value (approximate)	0		

Le 179GG (45%)

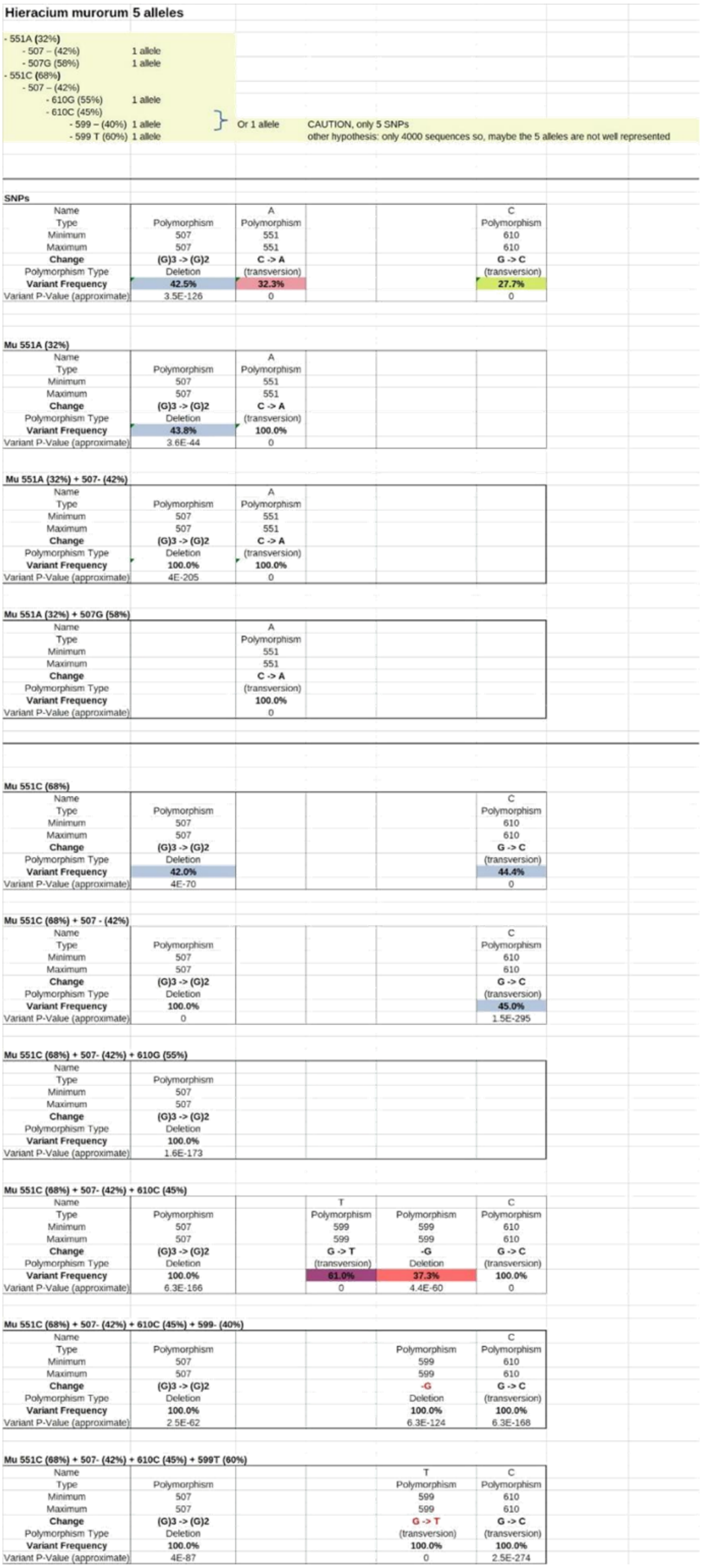
Name		G	
Type		Polymorphism	Polymorphism
Minimum		180	507
Maximum		180	507
Change		**A -> G**	**(G)3 -> (G)2**
Polymorphism Type		SNP (transition)	Deletion
Variant Frequency		**97.2%**	**41.0%**
Variant P-Value (approximate)		0	8.5E-84

Le 179GG (45%) + 507- (43%)

Name		G	
Type		Polymorphism	Polymorphism
Minimum		180	507
Maximum		180	507
Change		**A -> G**	**(G)3 -> (G)2**
Polymorphism Type		SNP (transition)	Deletion
Variant Frequency		**97.3%**	**100.0%**
Variant P-Value (approximate)		0	0

Le 179GG (45%) + 507G (57%)

Name		G	
Type		Polymorphism	
Minimum		180	
Maximum		180	
Change		**A -> G**	
Polymorphism Type		SNP (transition)	
Variant Frequency		**97.2%**	
Variant P-Value (approximate)		0	



Hieracium murorum 5 alleles

- 551A (32%)
  - 507 – (42%) 1 allele
  - 507G (58%) 1 allele
- 551C (68%)
  - 507 – (42%)
    - 610G (55%) 1 allele
    - 610C (45%)
      - 599 – (40%) 1 allele
      - 599 T (60%) 1 allele

} Or 1 allele CAUTION, only 5 SNPs

other hypothesis: only 4000 sequences so, maybe the 5 alleles are not well represented

SNPs

Name		A			C
Type	Polymorphism	Polymorphism			Polymorphism
Minimum	507	551			610
Maximum	507	551			610
Change	**(G)3 -> (G)2**	**C -> A**			**G -> C**
Polymorphism Type	Deletion	(transversion)			(transversion)
Variant Frequency	**42.5%**	**32.3%**			**27.7%**
Variant P-Value (approximate)	3.5E-126	0			0

Mu 551A (32%)

Name		A			
Type	Polymorphism	Polymorphism			
Minimum	507	551			
Maximum	507	551			
Change	**(G)3 -> (G)2**	**C -> A**			
Polymorphism Type	Deletion	(transversion)			
Variant Frequency	**43.8%**	**100.0%**			
Variant P-Value (approximate)	3.6E-44	0			

Mu 551A (32%) + 507- (42%)

Name		A			
Type	Polymorphism	Polymorphism			
Minimum	507	551			
Maximum	507	551			
Change	**(G)3 -> (G)2**	**C -> A**			
Polymorphism Type	Deletion	(transversion)			
Variant Frequency	**100.0%**	**100.0%**			
Variant P-Value (approximate)	4E-205	0			

Mu 551A (32%) + 507G (58%)

Name		A			
Type		Polymorphism			
Minimum		551			
Maximum		551			
Change		**C -> A**			
Polymorphism Type		(transversion)			
Variant Frequency		**100.0%**			
Variant P-Value (approximate)		0			

Mu 551C (68%)

Name					C
Type	Polymorphism				Polymorphism
Minimum	507				610
Maximum	507				610
Change	**(G)3 -> (G)2**				**G -> C**
Polymorphism Type	Deletion				(transversion)
Variant Frequency	**42.0%**				**44.4%**
Variant P-Value (approximate)	4E-70				0

Mu 551C (68%) + 507 - (42%)

Name					C
Type	Polymorphism				Polymorphism
Minimum	507				610
Maximum	507				610
Change	**(G)3 -> (G)2**				**G -> C**
Polymorphism Type	Deletion				(transversion)
Variant Frequency	**100.0%**				**45.0%**
Variant P-Value (approximate)	0				1.5E-295

Mu 551C (68%) + 507- (42%) + 610G (55%)

Name					
Type	Polymorphism				
Minimum	507				
Maximum	507				
Change	**(G)3 -> (G)2**				
Polymorphism Type	Deletion				
Variant Frequency	**100.0%**				
Variant P-Value (approximate)	1.6E-173				

Mu 551C (68%) + 507- (42%) + 610C (45%)

Name			T		C
Type	Polymorphism		Polymorphism	Polymorphism	Polymorphism
Minimum	507		599	599	610
Maximum	507		599	599	610
Change	**(G)3 -> (G)2**		**G -> T**	**-G**	**G -> C**
Polymorphism Type	Deletion		(transversion)	Deletion	(transversion)
Variant Frequency	**100.0%**		**61.0%**	**37.3%**	**100.0%**
Variant P-Value (approximate)	6.3E-166		0	4.4E-60	0

Mu 551C (68%) + 507- (42%) + 610C (45%) + 599- (40%)

Name					C
Type	Polymorphism			Polymorphism	Polymorphism
Minimum	507			599	610
Maximum	507			599	610
Change	**(G)3 -> (G)2**			**-G**	**G -> C**
Polymorphism Type	Deletion			Deletion	(transversion)
Variant Frequency	**100.0%**			**100.0%**	**100.0%**
Variant P-Value (approximate)	2.5E-62			6.3E-124	6.3E-168

Mu 551C (68%) + 507- (42%) + 610C (45%) + 599T (60%)

Name				T	C
Type	Polymorphism			Polymorphism	Polymorphism
Minimum	507			599	610
Maximum	507			599	610
Change	**(G)3 -> (G)2**			**G -> T**	**G -> C**
Polymorphism Type	Deletion			(transversion)	(transversion)
Variant Frequency	**100.0%**			**100.0%**	**100.0%**
Variant P-Value (approximate)	4E-87			0	2.5E-274

Hieracium policheae 5 alleles

DOMINANT ALLELE missing

RECESSIVE ALLELES

- 399 CATGCGGTC (43%)
  - 507- (44%) 1 allele
  - 507G (56%) 1 allele
- 399 CATTCGGTC (41%)
  - 507- (44%) 1 allele
  - 507G (56%) 1 allele

51 SNPs variant frequency >80%

+ SNPs

Name	C	TTGG	GG	CATGCGGTC	CATTCGGTC		A
Type	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism
Minimum	184	199	201	400	400	507	694
Maximum	184	202	202	399	399	507	694
Change	**T -> C**	**ATTT -> TTGG**	**TT -> GG**	**+CATGCGGTC**	**+CATTCGGTC**	**(G)3 -> (G)2**	**T -> A**
Polymorphism Type	(transition)	Substitution	Substitution	Insertion	Insertion	Deletion	(transversion)
Variant Frequency	**43.6%**	**42.2%**	**49.1%**	**43.2%**	**41.1%**	**41.6%**	**51.0%**
Variant P-Value (approximate)	0	0	0	0	0	8.5E-167	0

Po 399T 41%

Name	C	TTGG			CATTCGGTC		
Type	Polymorphism	Polymorphism			Polymorphism	Polymorphism	
Minimum	184	199			400	507	
Maximum	184	202			399	507	
Change	**T -> C**	**ATTT -> TTGG**			**+CATTCGGTC**	**(G)3 -> (G)2**	
Polymorphism Type	(transition)	Substitution			Insertion	Deletion	
Variant Frequency	**83.3%**	**88.3%**			**88.2%**	**41.7%**	
Variant P-Value (approximate)	0	0			0	1.8E-70	

Po 399T 41% + 507G 56%

Name	C	TTGG			CATTCGGTC		
Type	Polymorphism	Polymorphism			Polymorphism		
Minimum	184	199			400		
Maximum	184	202			399		
Change	**T -> C**	**ATTT -> TTGG**			**+CATTCGGTC**		
Polymorphism Type	(transition)	Substitution			Insertion		
Variant Frequency	**85.2%**	**89.2%**			**89.9%**		
Variant P-Value (approximate)	0	0			0		

Po 399T 41% + 507- 44%

Name	C	TTGG			CATTCGGTC		
Type	Polymorphism	Polymorphism			Polymorphism	Polymorphism	
Minimum	184	199			400	507	
Maximum	184	202			399	507	
Change	**T -> C**	**ATTT -> TTGG**			**+CATTCGGTC**	**(G)3 -> (G)2**	
Polymorphism Type	(transition)	Substitution			Insertion	Deletion	
Variant Frequency	**81.1%**	**87.3%**			**86.7%**	**100.0%**	
Variant P-Value (approximate)	0	0			0	0	

Po 399G 43%

Name			GG	CATGCGGTC			A
Type			Polymorphism	Polymorphism		Polymorphism	Polymorphism
Minimum			201	400		507	694
Maximum			202	399		507	694
Change			**TT -> GG**	**+CATGCGGTC**		**(G)3 -> (G)2**	**T -> A**
Polymorphism Type			Substitution	Insertion		Deletion	(transversion)
Variant Frequency			**98.5%**	**89.1%**		**43.9%**	**88.3%**
Variant P-Value (approximate)			0	0		6.5E-89	0

Po 399G 43% + 507G 56%

Name			GG	CATGCGGTC			A
Type			Polymorphism	Polymorphism			Polymorphism
Minimum			201	400			694
Maximum			202	399			694
Change			**TT -> GG**	**+CATGCGGTC**			**T -> A**
Polymorphism Type			Substitution	Insertion			(transversion)
Variant Frequency			**98.8%**	**88.2%**			**88.8%**
Variant P-Value (approximate)			0	0			0

Po 399G 43% + 507- 44%

Name			GG	CATGCGGTC			A
Type			Polymorphism	Polymorphism		Polymorphism	Polymorphism
Minimum			201	400		507	694
Maximum			202	399		507	694
Change			**TT -> GG**	**+CATGCGGTC**		**(G)3 -> (G)2**	**T -> A**
Polymorphism Type			Substitution	Insertion		Deletion	(transversion)
Variant Frequency			**98.2%**	**90.1%**		**100.0%**	**87.8%**
Variant P-Value (approximate)			0	0		0	0

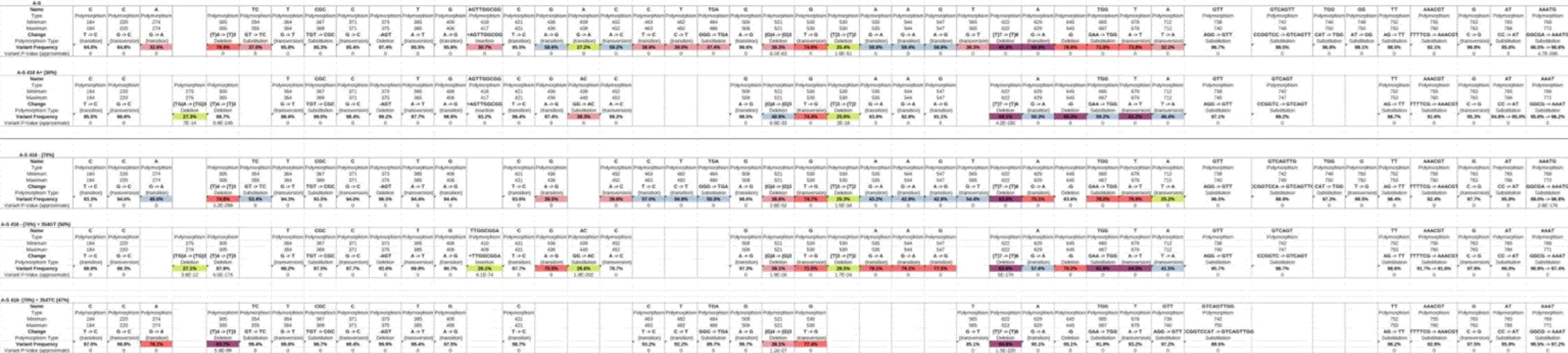
Pilosella aurantiaca **4 alleles**

DOMINANT ALLELE 1 allele

Name			A				A		A	C		C						AC
Type	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism
Minimum	275	305	477	507	521	622	716	758	853	932	1046	1214	1346	1883	1884	1991	1992	2006
Maximum	276	305	477	507	521	622	716	758	852	932	1046	1214	1346	1883	1884	1991	1992	2007
Change	(TG)4 -> (TG)3	(T)4 -> (T)3	G -> A	(G)3 -> (G)2	(G)4 -> (G)3	(T)7 -> (T)6	G -> A	(T)5 -> (T)4	(A)8 -> (A)9	T -> C	(T)4 -> (T)3	T -> C	(C)3 -> (C)2	(T)6 -> (T)5	(T)6 -> (T)5	(C)7 -> (C)6	(C)7 -> (C)6	+AC
Polymorphism Type	Deletion	Deletion	(transition)	Deletion	Deletion	Deletion	(transition)	Deletion	Insertion	(transition)	Deletion	(transition)	Deletion	Deletion	Deletion	Deletion	Deletion	Insertion
Variant Frequency	26.4%	90.3%	34.3%	39.8%	35.8%	83.6%	30.5%	28.7%	35.4%	98.7%	25.9%	31.2%	27.6%	30.1%	97.0%	99.2%	80.0%	29.0%
Variant P-Value (approximate)	4.9E-122	0	0	0	6.3E-220	0	0	1.6E-170	2.9E-12	0	1.4E-211	0	6.3E-98	7.6E-96	0	0	0	0

RECESSIVE ALLELES 3 alleles

418A+ (30%)	1 allele
418- (70%)	
354GT (50%)	1 allele
354TC (43%)	1 allele



A-S

A-S 418 A+ (30%)

A-S 418 - (70%)

A-S 418 - (70%) + 354GT (50%)

A-S 418- (70%) + 354TC (47%)

Pilosella lactucella 5 alleles

DOMINANT ALLELE 1 allele

Name							C						
Type	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism
Minimum	275	385	507	521	622	758	932	1040	1346	1880	1884	1981	1982
Maximum	276	385	507	521	622	758	932	1040	1346	1880	1884	1981	1982
Change	[illegible]	(T)4 -> (T)3	[illegible]	[illegible]	[illegible]	[illegible]	T -> C	(T)4 -> (T)3	(C)3 -> (C)2	[illegible]	[illegible]	[illegible]	[illegible]
Polymorphism Type	Deletion	Deletion	Deletion	Deletion	Deletion	Deletion	(transition)	Deletion	Deletion	Deletion	Deletion	Deletion	Deletion
Variant Frequency	25.3%	99.0%	36.4%	31.5%	61.6%	29.4%	97.8%	28.0%	29.7%	27.6%	98.7%	63.5%	73.5%
Variant P-Value (approximate)	7.9E-99	0	4.3E-189	8.4E-102	0	6.6E-103	0	5E-197	1.7E-60	2.6E-47	0	0	6.4E-308

RECESSIVE ALLELES

- 274A (26%) 1 allele
- 274G (74%)
  - 364T (54%) 1 allele
  - 364G (46%)
    - 250C (58%) 1 allele
    - 250T (42%) 1 allele

Au classified
Au 274A (26%)
Au 274G (74%) + 364T (54%)
Au 274G (74%) + 364G (46%) + 250C (58%)
Au 274G (74%) + 364G (46%) + 250T (42%)

[illegible]

SNP

[illegible]

Au 274A (26%)

[illegible]

Au 274G (74%)

[illegible]

Au 274G (74%) + 364T (54%)

[illegible]

Au 274G (74%) + 364G (46%)

[illegible]

Au 274G (74%) + 364G (46%) + 250C (58%)

[illegible]

Au 274G (74%) + 364G (46%) + 250T (42%)

[illegible]

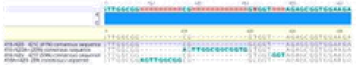
Pilosella cymosa **5 alleles**

DOMINANT ALLELE 1 allele

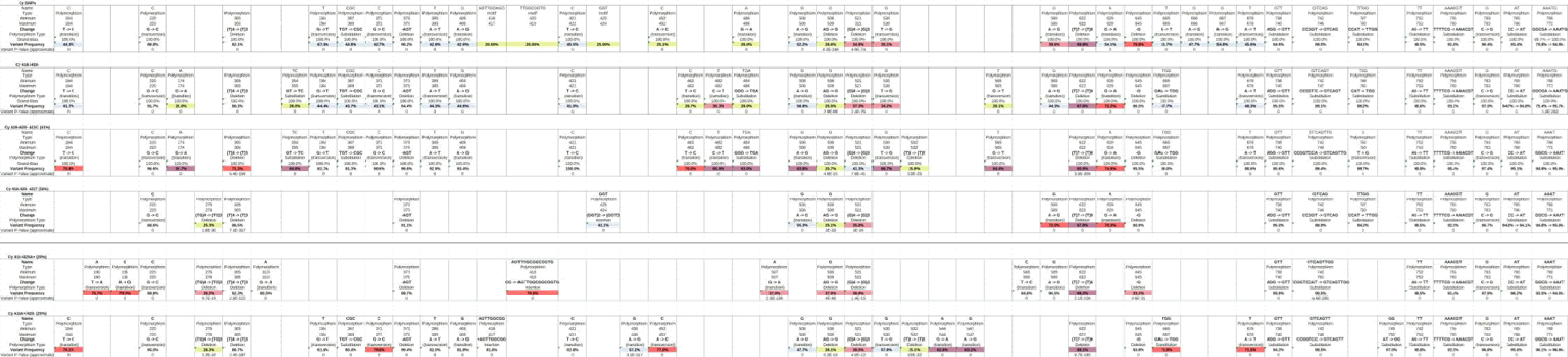
Minimum	275	305	507	521	622	758	852	1046	1274	1346	1744	1883	1884	1991	1992
Maximum	276	305	507	521	622	758	852	1046	1274	1346	1744	1883	1884	1991	1992
Change	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	A -> T	[illegible]	A -> G	[illegible]	[illegible]	[illegible]	[illegible]
Polymorphism Type	Deletion	Deletion	Deletion	Deletion	Deletion	Deletion	Deletion	Deletion	(transversion)	Deletion	(transition)	Deletion	Deletion	Deletion	Deletion
Protein Effect	Frame Shift	Frame Shift	Frame Shift	Frame Shift											
Strand-Bias	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Variant Frequency	26.8%	88.7%	36.8%	31.6%	62.9%	26.4%	59.9%	25.9%	31.2%	28.2%	63.4%	27.9%	87.1%	98.2%	79.9%
Variant P-Value (approximate)	2.9E-144	0	5.0E-274	5.9E-138	0	8.3E-120	4.6E-71	4.7E-233	0	5.1E-179	0	1.4E-67	0	0	0

RECESSIVE ALLELES

- 418-I420A+ (20%) 1 allele
- 418-I420- (20%) 1 allele
- 418-I420- (60%)
  - 421C (41%) 1 allele
  - 421T (59%) 1 allele



Cy SNPs



D2.10 **2 alleles**

DOMINANT ALLELE 1 allele

Name									T		G				
Type	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism
Minimum	275	305	507	521	622	758	852	1046	1274	1346	1744	1883	1884	1991	1992
Maximum	276	305	507	521	622	758	852	1046	1274	1346	1744	1883	1884	1991	1992
Change	**(TG)4 -> (TG)3**	**(T)4 -> (T)3**	**(G)3 -> (G)2**	**(G)4 -> (G)3**	**(T)7 -> (T)6**	**(T)5 -> (T)4**	**(A)8 -> (A)7**	**(T)4 -> (T)3**	**A -> T**	**(C)3 -> (C)2**	**A -> G**	**(T)6 -> (T)5**	**(T)6 -> (T)5**	**(C)7 -> (C)6**	**(C)7 -> (C)6**
Polymorphism Type	Deletion	Deletion	Deletion	Deletion	Deletion	Deletion	Deletion	Deletion	(transversion)	Deletion	(transition)	Deletion	Deletion	Deletion	Deletion
Variant Frequency	**25.8%**	**90.4%**	**38.0%**	**31.6%**	**61.5%**	**26.7%**	**60.5%**	**25.5%**	**31.1%**	**28.4%**	**63.3%**	**27.7%**	**87.1%**	**97.9%**	**78.9%**
Variant P-Value (approximate)	2E-160	0	0	8.8E-53	0	2.6E-186	2.1E-107	2.5E-292	0	1.1E-100	0	4.9E-86	0	0	0

RECESSIVE ALLELE 1 allele

Name	A	G	C			A		AGTTGGCGG			C	G
Type	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism
Minimum	190	196	220	275	305	310	373	418	507	521	568	589
Maximum	190	196	220	276	305	310	375	417	507	521	568	589
Change	**T -> A**	**A -> G**	**G -> C**	**(TG)4 -> (TG)3**	**(T)4 -> (T)3**	**G -> A**	**-AGT**	**+AGTTGGCGG**	**(G)3 -> (G)2**	**(G)4 -> (G)3**	**T -> C**	**A -> G**
Polymorphism Type	(transversion)	(transition)	(transversion)	Deletion	Deletion	(transition)	Deletion	Insertion	Deletion	Deletion	(transition)	(transition)
Variant Frequency	**88.1%**	**87.1%**	**88.9%**	**26.4%**	**89.8%**	**87.8%**	**94.1%**	**73.7%**	**36.2%**	**32.6%**	**91.1%**	**91.1%**
Variant P-Value (approximate)	0	0	0	5.9E-109	0	0	0	0	1.9E-207	1.8E-43	0	0

Name		GTT	T	A	T	GGT	T	AAACG	T	GCATGAAAT	G	A
Type	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism
Minimum	622	738	741	743	745	747	752	753	759	761	770	771
Maximum	622	738	741	743	745	749	752	757	759	769	770	771
Change	**(T)7 -> (T)6**	**A -> GTT**	**G -> T**	**C -> A**	**G -> T**	**CCA -> GGT**	**A -> T**	**GTTTT -> AAACG**	**C -> T**	**GCCCCGGG -> GCATGAAA**	**C -> G**	**G -> A**
Polymorphism Type	Deletion	Insertion	(transversion)	(transversion)	(transversion)	Substitution	(transversion)	Substitution	(transition)	Substitution	(transversion)	(transition)
Variant Frequency	**66.6%**	**94.8%**	**98.5%**	**94.9%**	**98.1%**	**95.7%**	**89.7%**	**92.3%**	**99.0%**	**86.5% -> 87.8%**	**77.0%**	**55.4%**
Variant P-Value (approximate)	0	0	0	0	0	0	0	0	0	0	0	6.3E-26

Pilosella piloselloides D36 **4 alleles**

DOMINANT ALLELE 1 allele

Name									T		G				
Type	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism
Minimum	275	305	507	521	622	758	852	1046	1274	1346	1744	1883	1884	1991	1992
Maximum	276	305	507	521	622	758	852	1046	1274	1346	1744	1883	1884	1991	1992
Change	(TG)4 -> (TG)3	(T)4 -> (T)3	(G)3 -> (G)2	(G)4 -> (G)3	(T)7 -> (T)6	(T)5 -> (T)4	(A)8 -> (A)7	(T)4 -> (T)3	A -> T	(C)3 -> (C)2	A -> G	(T)6 -> (T)5	(T)6 -> (T)5	(C)7 -> (C)6	(C)7 -> (C)6
Polymorphism Type	Deletion	Deletion	Deletion	Deletion	Deletion	Deletion	Deletion	Deletion	(transversion)	Deletion	(transition)	Deletion	Deletion	Deletion	Deletion
Variant Frequency	25.9%	90.5%	38.3%	31.6%	63.4%	27.8%	60.2%	26.4%	31.1%	28.0%	62.9%	28.2%	87.1%	98.1%	79.0%
Variant P-Value (approximate)	9.1E-232	0	0	4.1E-263	0	1.5E-314	1.3E-143	0	0	2.4E-130	0	3.4E-137	0	0	0

RECESSIVE ALLELES

- 629A (31%) 1 allele
- 629G (69%)
  - 528C (51%) 1 allele
  - 528T (49%) 1 allele

SNP

Name	A	G	C			A		T	AGTTGGCGG	C	T				C	G		A				GTT	GTCAGT	TGG	G	TT	AAACGT	G	AT	AAATG	G
Type	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism
Minimum	190	196	220	275	305	310	373	385	418	463	469	507		521	568	589	622	629	641	645	646	738	742	748	750	752	755	763	765	768	773
Maximum	190	196	220	276	305	310	375	385	417	463	469	507		521	568	589	622	629	644	645	650	740	747	750	750	753	760	763	766	772	773
Change	T -> A	A -> G	G -> C	(TG)4 -> (TG)3	(T)4 -> (T)3	G -> A	-AGT	A -> T	+AGTTGGCGG	T -> C	G -> T	(G)3 -> (G)2		(G)4 -> (G)3	T -> C	A -> G	(T)7 -> (T)6	G -> A	-CAGA	-G	-TGGAA	AGG -> GTT	CCGGTC -> GTCAGT	CAT -> TGG	T -> G	AG -> TT	TTTTCG -> AAACGT	C -> G	CC -> AT	GGCGA -> AAATG	A -> G
Polymorphism Type	(transversion)	(transition)	(transversion)	Deletion	Deletion	(transition)	Deletion	(transversion)	Insertion	(transition)	(transversion)	Deletion		Deletion	(transition)	(transition)	Deletion	(transition)	Deletion	Deletion	Deletion	Substitution	Substitution	Substitution	(transversion)	Substitution	Substitution	(transversion)	Substitution	Substitution	(transition)
Variant Frequency	60.9%	88.7%	90.5%	26.9%	90.5%	58.9%	95.6%	26.7%	73.5%	29.2%	30.0%	27.9%		32.0%	58.9%	29.9%	66.2%	31.1%	28.9%	73.9%	25.8%	94.8%	88.8%	95.9%	99.3%	98.2%	90.9%	96.2%	93.5% -> 93.6%	72.1% -> 94.3%	48.2%
Variant P-Value (approximate)	0	0	0	1.5E-119	0	0	0	0	0	0	0	1E-54		9.1E-37	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2.1E-37

D36 629A 31%

Name	A	G	C			A			AGTTGGCGG		T				C			A				GTT	GTCAGTTG			TT	AAACGT	G	AT	AAATG	
Type	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism		Polymorphism		Polymorphism	Polymorphism		Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism			Polymorphism	Polymorphism			Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	
Minimum	190	196	220	275	305	310	373		418		469	507		521	568		622	629	641			738	742			752	755	763	765	768	
Maximum	190	196	220	276	305	310	375		417		469	507		521	568		622	629	650			740	749			753	760	763	766	772	
Change	T -> A	A -> G	G -> C	(TG)4 -> (TG)3	(T)4 -> (T)3	G -> A	-AGT		+AGTTGGCGG		G -> T	(G)3 -> (G)2		(G)4 -> (G)3	T -> C		(T)7 -> (T)6	G -> A	-CAGAGTGGAA			AGG -> GTT	CCGGTCCA -> GTCAGTTG			AG -> TT	TTTTCG -> AAACGT	C -> G	CC -> AT	GGCGA -> AAATG	
Polymorphism Type	(transversion)	(transition)	(transversion)	Deletion	Deletion	(transition)	Deletion		Insertion		(transversion)	Deletion		Deletion	(transition)		Deletion	(transition)	Deletion			Substitution	Substitution			Substitution	Substitution	(transversion)	Substitution	Substitution	
Variant Frequency	25.8%	90.9%	92.8%	28.9%	89.5%	85.8%	97.2%		79.3%		84.5%	35.3%		33.0%	92.9%		71.4%	100.0%	80.8%			95.7%	88.0%			98.5%	91.7% -> 91.8%	96.0%	94.4% -> 94.6%	73.6% -> 95.0%	
Variant P-Value (approximate)	0	0	0	1E-58	0	0	0		0		0	9.3E-41		3.3E-30	0		0	0	0			0	0			0	0	0	0	1.2E-165	

D36 629G (69%)

Name	A	G	C			A		T	AGTTGGCGG	C			G		C	G						GTT	GTCAGTTG		GG	TT	AAACGT	G	AT	AAATG	
Type	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism			Polymorphism		Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	
Minimum	190	196	220	275	305	310	373	385	418	463		507	508	521	568	589	622			645		738	742		749	752	755	763	765	768	
Maximum	190	196	220	276	305	310	375	385	417	463		507	508	521	568	589	622			645		740	749		750	753	760	763	766	772	
Change	T -> A	A -> G	G -> C	(TG)4 -> (TG)3	(T)4 -> (T)3	G -> A	-AGT	A -> T	+AGTTGGCGG	T -> C		(G)3 -> (G)2	A -> G	(G)4 -> (G)3	T -> C	A -> G	(T)7 -> (T)6			-G		AGG -> GTT	CCGGTCCA -> GTCAGTTG		AT -> GG	AG -> TT	TTTTCG -> AAACGT	C -> G	CC -> AT	GGCGA -> AAATG	
Polymorphism Type	(transversion)	(transition)	(transversion)	Deletion	Deletion	(transition)	Deletion	(transversion)	Insertion	(transition)		Deletion	(transition)	Deletion	(transition)	(transition)	Deletion			Deletion		Substitution	Substitution		Substitution	Substitution	Substitution	(transversion)	Substitution	Substitution	
Variant Frequency	78.1%	88.3%	89.7%	28.2%	90.3%	53.1%	94.6%	39.9%	74.2%	42.0%		25.4%	25.8%	35.1%	51.0%	47.1%	61.8%			58.4%		95.2%	88.8%		97.9%	98.1%	92.3%	96.4%	94.3% -> 94.4%	72.4% -> 94.6%	
Variant P-Value (approximate)	0	0	0	1.3E-56	0	0	0	0	0	0		1.3E-11	0	1.5E-29	0	0	0			0		0	0		0	0	0	0	0	0	

D36 629G (69%) + 528C (51%)

Name	A	G	C			A			AGTTGGCGG						C	G						GTT	GTCAGTT		G	TT	AAACGT	G	AT	AAATG	
Type	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism		Polymorphism			Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism			Polymorphism		Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	
Minimum	190	196	220	275	305	310	373		418			507		521	568	589	622			645		738	742		750	752	755	763	765	768	
Maximum	190	196	220	276	305	310	375		417			507		521	568	589	622			645		740	748		750	753	760	763	766	772	
Change	T -> A	A -> G	G -> C	(TG)4 -> (TG)3	(T)4 -> (T)3	G -> A	-AGT		+AGTTGGCGG			(G)3 -> (G)2		(G)4 -> (G)3	T -> C	A -> G	(T)7 -> (T)6			-G		AGG -> GTT	CCGGTCC -> GTCAGTT		T -> G	AG -> TT	TTTTCG -> AAACGT	C -> G	CC -> AT	GGCGA -> AAATG	
Polymorphism Type	(transversion)	(transition)	(transversion)	Deletion	Deletion	(transition)	Deletion		Insertion			Deletion		Deletion	(transition)	(transition)	Deletion			Deletion		Substitution	Substitution		(transversion)	Substitution	Substitution	(transversion)	Substitution	Substitution	
Variant Frequency	80.3%	90.8%	92.3%	28.6%	90.7%	85.8%	97.5%		77.8%			38.1%		35.2%	100.0%	89.4%	68.2%			27.9%		95.5%	89.1%		99.2%	98.3%	91.5%	95.9%	93.5% -> 93.7%	74.4% -> 95.3%	
Variant P-Value (approximate)	0	0	0	4.1E-31	0	0	0		0			2E-51		3.1E-16	0	0	0			3.2E-28		0	0		0	0	0	0	0	2.1E-157	

D36 629G (69%) + 568T (49%)

Name	A	G	C					T	AGTTGGCGG	C			G									GTT	GTCAGTTG		G	TT	AAACGT	G	AT	AAATG	
Type	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism			Polymorphism	Polymorphism			Polymorphism			Polymorphism		Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	
Minimum	190	196	220	275	305		373	385	418	463			508	521			622			645		738	742		750	752	755	763	765	768	
Maximum	190	196	220	276	305		375	385	417	463			508	521			622			645		740	749		750	753	760	763	766	772	
Change	T -> A	A -> G	G -> C	(TG)4 -> (TG)3	(T)4 -> (T)3		-AGT	A -> T	+AGTTGGCGG	T -> C			A -> G	(G)4 -> (G)3			(T)7 -> (T)6			-G		AGG -> GTT	CCGGTCCA -> GTCAGTTG		T -> G	AG -> TT	TTTTCG -> AAACGT	C -> G	CC -> AT	GGCGA -> AAATG	
Polymorphism Type	(transversion)	(transition)	(transversion)	Deletion	Deletion		Deletion	(transversion)	Insertion	(transition)			(transition)	Deletion			Deletion			Deletion		Substitution	Substitution		(transversion)	Substitution	Substitution	(transversion)	Substitution	Substitution	
Variant Frequency	75.8%	85.7%	87.0%	27.9%	89.8%		91.6%	73.8%	70.5%	81.0%			50.4%	35.0%			55.1%			90.2%		94.9%	89.3%		99.3%	97.9%	93.2%	97.0%	95.0% -> 95.1%	70.7% -> 94.0%	
Variant P-Value (approximate)	0	0	0	2E-27	0		0	0	0	0			0	4.1E-15			2.3E-166			0		0	0		0	0	0	0	0	3.2E-175	

Pilosella glacialis 5 alleles

DOMINANT ALLELE	1 allele

Name							C		C				
Type	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism
Minimum	276	306	508	522	623	759	933	1347	1787	1884	1885	1992	1993
Maximum	277	306	508	522	623	759	933	1347	1787	1884	1885	1992	1993
Change	(TG)4 -> (TG)3	(T)4 -> (T)3	(G)3 -> (G)2	(G)4 -> (G)3	(T)7 -> (T)8	(T)5 -> (T)4	T -> C	(C)3 -> (C)2	G -> C	(T)8 -> (T)9	(T)9 -> (T)8	(C)7 -> (C)8	(C)7 -> (C)8
Polymorphism Type	Deletion	Deletion	Deletion	Deletion	Deletion	Deletion	(transition)	Deletion	(transversion)	Deletion	Deletion	Deletion	Deletion
Variant Frequency	26.0%	90.2%	35.9%	31.8%	49.2%	27.3%	98.7%	26.4%	38.7%	27.1%	87.7%	97.9%	77.3%
Variant P-Value (approximate)	1E-148	0	1.3E-153	9.9E-89	0	3.4E-95	0	2.8E-83	0	1.2E-94	0	0	0

RECESSIVE ALLELES	3 alleles	
- 385G (45%)	1 allele	(or 2 similar)
- 385T (55%)		
251T (46%)	1 allele	
251C (54%)	1 allele	

This SNP doesn't follow any other SNP pattern
Or doesn't enter in any combination

Reference

Name	T	C	A	G	C	T	A			T	CGC	C		T	G	AGTTGGCGG	C			AC	C	C		G	G		G		A	G				A		TGG	T	GTT	GTCAGTTGG	TT	AAACGT	G	AT	AAATG	
Type	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism			Polymorphism	Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism		Polymorphism	Polymorphism			Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	
Minimum	173	185	191	197	221	251	275		306	365	368	372	374	386	407	418	422			440	453	464		509	509	522	531		545	548			623	630	646	666	679	739	743	753	756	764	766	769	
Maximum	173	185	191	197	221	251	275		306	365	370	372	376	386	407	418	422			441	453	464		509	510	522	531		545	548			623	630	646	668	679	741	751	754	761	764	767	773	
Change	**C -> T**	**T -> C**	**T -> A**	**A -> G**	**G -> C**	**C -> T**	**G -> A**		**(T)4 -> (T)3**	**G -> T**	**TGT -> CGC**	**G -> C**	**-AGT**	**A -> T**	**A -> G**	**+AGTTGGCGG**	**T -> C**			**GG -> AC**	**A -> C**	**T -> C**		**A -> G**	**AG -> G**	**(G)4 -> (G)3**	**T -> G**		**G -> A**	**A -> G**			**(T)7 -> (T)8**	**G -> A**	**-G**	**GAA -> TGG**	**A -> T**	**AGG -> GTT**	**CCGGTCCAT -> GTCAGTTG**	**AG -> TT**	**TTTTCG -> AAACGT**	**C -> G**	**CC -> AT**	**GGCGA -> AAATG**	
Polymorphism Type	(transition)	(transition)	(transversion)	(transition)	(transversion)	(transition)	(transition)		Deletion	(transversion)	Substitution	(transversion)	Deletion	(transversion)	(transition)	Insertion	(transition)			Substitution	(transversion)	(transition)		(transition)	Deletion	Deletion	(transversion)		(transition)	(transition)			Deletion	(transition)	Deletion	Substitution	(transversion)	Substitution	Substitution	Substitution	Substitution	(transversion)	Substitution	Substitution	
Variant Frequency	**27.0%**	**58.5%**	**39.0%**	**65.6%**	**93.8%**	**26.6%**	**25.2%**		**82.7%**	**55.7%**	**55.2%**	**55.4%**	**95.4%**	**56.0%**	**59.2%**	**30.7%**	**92.0%**			**26.1%**	**29.8%**	**36.5%**		**56.1%**	**26.5%**	**35.8%**	**42.8%**		**29.7%**	**30.2%**			**40.2%**	**69.3%**	**64.2%**	**58.1%**	**58.6%**	**98.6%**	**92.3%**	**99.7%**	**94.5%**	**100.0%**	**98.3%**	**79.6% -> 95.7%**	
Variant P-Value (approximate)	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0			0	0	0		0	3.9E-156	2.4E-159	0		0	0			0	0	0	0	0	0	0	0	0	0	0	0	

G45 385G 45%

Name			A	G	C												C	GGT				C		G							C	C		A				GTT	GTCAGTTGG	TT	AAACGT	G	AT	AAATG	
Type			Polymorphism	Polymorphism	Polymorphism			Polymorphism	Polymorphism				Polymorphism				Polymorphism	Polymorphism				Polymorphism		Polymorphism		Polymorphism					Polymorphism	Polymorphism		Polymorphism				Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	
Minimum			191	197	221			276	306				374				422	425				464		509		522					616	618		630				739	743	753	756	764	766	769	
Maximum			191	197	221			277	306				376				422	425				464		509		522					617	618		630				741	751	754	761	764	767	773	
Change			**T -> A**	**A -> G**	**G -> C**			**(TG)4 -> (TG)3**	**(T)4 -> (T)3**				**-AGT**				**T -> C**	**(GGT)2 -> (GGT)3**				**T -> C**		**A -> G**		**(G)4 -> (G)3**					**+C**	**T -> C**		**G -> A**				**AGG -> GTT**	**CCGGTCCAT -> GTCAGTTG**	**AG -> TT**	**TTTTCG -> AAACGT**	**C -> G**	**CC -> AT**	**GGCGA -> AAATG**	
Polymorphism Type			(transversion)	(transition)	(transversion)			Deletion	Deletion				Deletion				(transition)	Insertion				(transition)		(transition)		Deletion					Insertion	(transition)		(transition)				Substitution	Substitution	Substitution	Substitution	(transversion)	Substitution	Substitution	
Variant Frequency			**83.7%**	**84.2%**	**97.7%**			**25.4%**	**89.2%**				**90.2%**				**84.8%**	**55.1%**				**86.6%**		**93.6%**		**32.8%**					**28.0%**	**42.6%**		**91.3%**				**98.6%**	**92.0%**	**99.7%**	**95.0%**	**100.0%**	**98.8%**	**95.8% -> 98.5%**	
Variant P-Value (approximate)			0	0	0			3.4E-183	0				0				0	0				0		0		7.5E-49					0	0		0				0	0	0	0	0	0	3.0E-322	

G45 385T 55%

Name	T	C		G	C	T	A			T	CGC	C		T	G	AGTTGGCGG	C		G	AC	C			G	G		G		A	G				A		TGG	T	GTT	GTCAGTTGG	TT	AAACGT	G	AT	AAAT	G
Type	Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism			Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism			Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism
Minimum	173	185		197	221	251	275		306	365	368	372	374	386	407	418	422		437	440	453			509	509	522	531	533	545	548			623	630	646	666	679	739	743	753	756	764	766	769	773
Maximum	173	185		197	221	251	275		306	365	370	372	376	386	407	418	422		437	441	453			509	510	522	531	533	545	548			623	630	646	668	679	741	751	754	761	764	767	772	773
Change	**C -> T**	**T -> C**		**A -> G**	**G -> C**	**C -> T**	**G -> A**		**(T)4 -> (T)3**	**G -> T**	**TGT -> CGC**	**G -> C**	**-AGT**	**A -> T**	**A -> G**	**+AGTTGGCGG**	**T -> C**		**A -> G**	**GG -> AC**	**A -> C**			**A -> G**	**AG -> G**	**(G)4 -> (G)3**	**T -> G**	**(T)3 -> (T)2**	**G -> A**	**A -> G**			**(T)7 -> (T)8**	**G -> A**	**-G**	**GAA -> TGG**	**A -> T**	**AGG -> GTT**	**CCGGTCCAT -> GTCAGTTG**	**AG -> TT**	**TTTTCG -> AAACGT**	**C -> G**	**CC -> AT**	**GGCG -> AAAT**	**A -> G**
Polymorphism Type	(transition)	(transition)		(transition)	(transversion)	(transition)	(transition)		Deletion	(transversion)	Substitution	(transversion)	Deletion	(transversion)	(transition)	Insertion	(transition)		(transition)	Substitution	(transversion)			(transition)	Deletion	Deletion	(transversion)	Deletion	(transition)	(transition)			Deletion	(transition)	Deletion	Substitution	(transversion)	Substitution	Substitution	Substitution	Substitution	(transversion)	Substitution	Substitution	(transition)
Variant Frequency	**46.3%**	**94.8%**		**49.7%**	**95.7%**	**46.4%**	**43.5%**		**77.5%**	**100.0%**	**98.8%**	**99.1%**	**100.0%**	**97.8%**	**98.3%**	**54.6%**	**97.7%**		**41.9%**	**45.8%**	**50.6%**			**44.9%**	**28.9%**	**38.3%**	**72.3%**	**25.3%**	**49.9%**	**49.7%**			**54.1%**	**51.7%**	**97.2%**	**92.2%**	**95.8%**	**98.6%**	**92.5%**	**99.7%**	**94.1%**	**100.0%**	**97.8%**	**95.5% -> 95.7%**	**87.0%**
Variant P-Value (approximate)	0	0		0	0	0	0		0	0	0	0	0	0	0	0	0		0	0	0			0	8.3E-125	2.3E-117	0	9.6E-126	0	0			0	0	0	0	0	0	0	0	0	0	0	0	0

G45 385T 55% + 251T 46%

Name	T	C		G	C	T	A			T	CGC	C		T	G	AGTTGGCGG	C			AC			A		G		G							A		TGG	T	GTT	GTCAGTTGG	TT	AAACGT	G	AT	AAAT	G
Type	Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism			Polymorphism			Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism					Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism
Minimum	173	185		197	221	251	275		306	365	368	372	374	386	407	418	422			440			509		509	522	531	533					623	630	646	666	679	739	743	753	756	764	766	769	773
Maximum	173	185		197	221	251	275		306	365	370	372	376	386	407	418	422			441			509		510	522	531	533					623	630	646	668	679	741	751	754	761	764	767	772	773
Change	**C -> T**	**T -> C**		**A -> G**	**G -> C**	**C -> T**	**G -> A**		**(T)4 -> (T)3**	**G -> T**	**TGT -> CGC**	**G -> C**	**-AGT**	**A -> T**	**A -> G**	**+AGTTGGCGG**	**T -> C**			**GG -> AC**			**G -> A**		**AG -> G**	**(G)4 -> (G)3**	**T -> G**	**(T)3 -> (T)2**					**(T)7 -> (T)8**	**G -> A**	**-G**	**GAA -> TGG**	**A -> T**	**AGG -> GTT**	**CCGGTCCAT -> GTCAGTTG**	**AG -> TT**	**TTTTCG -> AAACGT**	**C -> G**	**CC -> AT**	**GGCG -> AAAT**	**A -> G**
Polymorphism Type	(transition)	(transition)		(transition)	(transversion)	(transition)	(transition)		Deletion	(transversion)	Substitution	(transversion)	Deletion	(transversion)	(transition)	Insertion	(transition)			Substitution			(transition)		Deletion	Deletion	(transversion)	Deletion					Deletion	(transition)	Deletion	Substitution	(transversion)	Substitution	Substitution	Substitution	Substitution	(transversion)	Substitution	Substitution	(transition)
Variant Frequency	**94.1%**	**97.8%**		**96.2%**	**99.8%**	**100.0%**	**98.8%**		**83.6%**	**100.0%**	**99.2%**	**99.4%**	**100.0%**	**98.2%**	**98.5%**	**54.9%**	**97.8%**			**91.8%**			**33.0%**		**33.5%**	**38.7%**	**72.0%**	**26.1%**					**58.9%**	**94.2%**	**98.3%**	**93.6%**	**98.7%**	**98.5%**	**91.8%**	**99.7%**	**93.7%**	**100.0%**	**98.3%**	**95.1% -> 95.7%**	**89.7%**
Variant P-Value (approximate)	0	0		0	0	0	0		9.9E-130	0	0	0	0	0	0	0	0			0			0		1.9E-162	4.1E-58	0	5.1E-65					0	0	0	0	0	0	0	0	0	0	0	0	2.3E-179

G45 385T 55% + 251C 54%

Name		C			C					T	CGC	C		T	G	AGTTGGCGG	C		G		C				G		G		A	G						TGG	T	GTT	GTCAGTTGG	TT	AAACGT	G	AT	AAATG	
Type		Polymorphism			Polymorphism			Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism		Polymorphism		Polymorphism				Polymorphism	Polymorphism	Polymorphism		Polymorphism	Polymorphism			Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	
Minimum		185			221			276	306	365	368	372	374	386	407	418	422		437		453				509	522	531		545	548			623		646	666	679	739	743	753	756	764	766	769	
Maximum		185			221			277	306	365	370	372	376	386	407	418	422		437		453				509	522	531		545	548			623		646	668	679	741	751	754	761	764	767	773	
Change		**T -> C**			**G -> C**			**(TG)4 -> (TG)3**	**(T)4 -> (T)3**	**G -> T**	**TGT -> CGC**	**G -> C**	**-AGT**	**A -> T**	**A -> G**	**+AGTTGGCGG**	**T -> C**		**A -> G**		**A -> C**				**A -> G**	**(G)4 -> (G)3**	**T -> G**		**G -> A**	**A -> G**			**(T)7 -> (T)8**		**-G**	**GAA -> TGG**	**A -> T**	**AGG -> GTT**	**CCGGTCCAT -> GTCAGTTG**	**AG -> TT**	**TTTTCG -> AAACGT**	**C -> G**	**CC -> AT**	**GGCGA -> AAATG**	
Polymorphism Type		(transition)			(transversion)			Deletion	Deletion	(transversion)	Substitution	(transversion)	Deletion	(transversion)	(transition)	Insertion	(transition)		(transition)		(transversion)				(transition)	Deletion	(transversion)		(transition)	(transition)			Deletion		Deletion	Substitution	(transversion)	Substitution	Substitution	Substitution	Substitution	(transversion)	Substitution	Substitution	
Variant Frequency		**92.3%**			**97.7%**			**25.3%**	**89.8%**	**100.0%**	**98.6%**	**98.9%**	**96.9%**	**97.4%**	**98.1%**	**54.3%**	**97.7%**		**75.3%**		**93.1%**				**70.4%**	**37.9%**	**72.7%**		**98.4%**	**97.2%**			**53.7%**		**96.1%**	**91.0%**	**95.1%**	**98.8%**	**93.2%**	**98.8%**	**94.8%**	**100.0%**	**97.6%**	**74.7% -> 95.7%**	
Variant P-Value (approximate)		0			0			7.5E-44	0	0	0	0	0	0	0	0	0		0		0				0	2.8E-61	0		0	0			3.2E-319		0	0	0	0	0	0	0	0	0	5.2E-175	

Hieracium Lactucella 4 ALLELES

DOMINANT ALLELE 1 allele

Name							A	C						
Type	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism
Minimum	275	305	507	521	622	758	853	932	1046	1346	1883	1884	1991	1992
Maximum	276	305	507	521	622	758	852	932	1046	1346	1883	1884	1991	1992
Change	(TG)4 -> (TG)3	(T)4 -> (T)3	(G)3 -> (G)2	(G)4 -> (G)3	(T)7 -> (T)6	(T)5 -> (T)4	(A)8 -> (A)9	T -> C	(T)4 -> (T)3	(C)3 -> (C)2	(T)6 -> (T)5	(T)6 -> (T)5	(C)7 -> (C)6	(C)7 -> (C)6
Polymorphism Type	Deletion	Deletion	Deletion	Deletion	Deletion	Deletion	Insertion	(transition)	Deletion	Deletion	Deletion	Deletion	Deletion	Deletion
Variant Frequency	26.0%	92.1%	38.5%	33.1%	61.0%	28.0%	34.0%	98.1%	26.8%	27.5%	27.9%	86.0%	93.3%	73.0%
Variant P-Value (approximate)	4.6E-65	0	8.9E-136	7.9E-89	0	4.4E-88	0.00048	0	2.6E-135	3.1E-32	2E-35	0	0	1.4E-200

RECESSIVE ALLELES

- 253G (53%) 1 allele
- 253C (47%)
    - 484TGA (41%) 1 allele
    - 484AGA (38%) 1 allele

Name	C	G	G	TGA	A	T	G	A	G							T		AGTTGGCGG		TGA		A		G		C			T	G	G	GTT	GTCAGTT	GG	TT	AAACGT	G	AT	AAATG
Type	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism			Polymorphism		Polymorphism	Polymorphism		Polymorphism		Polymorphism		Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism
Minimum	184	196	208	220	222	250	253	256	261	276			305		373	385		416		484		486		508	521	565	622	645	665	667	687	738	742	749	752	755	763	765	768
Maximum	184	196	208	222	222	250	253	256	261	276			305		375	385		417		486		486		508	521	565	622	645	665	667	687	740	748	750	753	760	763	766	772
Change	T -> C	A -> G	A -> G	GGC -> TGA	C -> A	C -> T	C -> G	C -> A	T -> G	-G			(T)4 -> (T)3		-AGT	A -> T		+AGTTGGCGG		GGG -> TGA		G -> A		A -> G	(G)4 -> (G)3	G -> C	(T)7 -> (T)6	-G	G -> T	A -> G	A -> G	AGG -> GTT	CCGGTCC -> GTCAGTT	AT -> GG	AG -> TT	TTTTCG -> AAACGT	C -> G	CC -> AT	GGCGA -> AAATG
Polymorphism Type	(transition)	(transition)	(transition)	Substitution	(transversion)	(transition)	(transversion)	(transversion)	(transversion)	Deletion			Deletion		Deletion	(transversion)		Insertion		Substitution		(transition)		(transition)	Deletion	(transversion)	Deletion	Deletion	(transversion)	(transition)	(transition)	Substitution	Substitution	Substitution	Substitution	Substitution	(transversion)	Substitution	Substitution
Variant Frequency	80.8%	96.3%	62.6%	60.8%	26.1%	62.1%	42.5%	41.4%	42.8%	25.5%			80.8%		94.6%	38.4%		75.4%		63.8%		25.4%		29.0%	33.1%	41.6%	67.3%	61.4%	92.5%	93.1%	47.9%	95.6%	89.6%	97.8%	98.7%	91.7% -> 91.8%	96.7%	94.0%	78.2% -> 95.2%
Variant P-Value (approximate)	0	0	0	0	0	0	0	0	0	6.3E-50			0		0	0		0		0		0		0	1.5E-76	0	0	0	0	0	0	0	0	0	0	0	0	0	0

La 253G 53%

Name	C	G	G	TGA		T	G	A	G								G	AGTTGGCGG	T	TGA				G		C			T	G	G	GTT	GTCAGTT	GG	TT	AAACGT	G	AT	AAATG
Type	Polymorphism	Polymorphism	Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism				Polymorphism		Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism				Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism
Minimum	184	196	208	220		250	253	256	261				305		373		406	416	457	484				508	521	565	622	645	665	667	687	738	742	749	752	755	763	765	768
Maximum	184	196	208	222		250	253	256	261				305		375		406	417	457	486				508	521	565	622	645	665	667	687	740	748	750	753	760	763	766	772
Change	T -> C	A -> G	A -> G	GGC -> TGA		C -> T	C -> G	C -> A	T -> G				(T)4 -> (T)3		-AGT		A -> G	+AGTTGGCGG	G -> T	GGG -> TGA				A -> G	(G)4 -> (G)3	G -> C	(T)7 -> (T)6	-G	G -> T	A -> G	A -> G	AGG -> GTT	CCGGTCC -> GTCAGTT	AT -> GG	AG -> TT	TTTTCG -> AAACGT	C -> G	CC -> AT	GGCGA -> AAATG
Polymorphism Type	(transition)	(transition)	(transition)	Substitution		(transition)	(transversion)	(transversion)	(transversion)				Deletion		Deletion		(transition)	Insertion	(transversion)	Substitution				(transition)	Deletion	(transversion)	Deletion	Deletion	(transversion)	(transition)	(transition)	Substitution	Substitution	Substitution	Substitution	Substitution	(transversion)	Substitution	Substitution
Variant Frequency	86.5%	95.0%	92.3%	93.6%		96.1%	100.0%	93.2%	94.5%				81.2%		99.5%		35.8%	82.4%	36.5%	92.2%				30.6%	33.0%	48.1%	85.3%	58.1%	95.3%	95.0%	48.9%	95.3%	89.0%	98.0%	98.4%	92.6%	96.7%	94.3%	74.8% -> 95.1%
Variant P-Value (approximate)	0	0	0	0		0	0	0	0				0		0		0	0	0	0				0	2.5E-33	0	0	0	0	0	0	0	0	0	0	0	0	0	9.9E-209

La 253C (47%)

Name	C	G	G	TGA	A	T					A	T		A		T		AGTTGGCGG	C	AGA	TGA			G		C			T	G	G	GTT	GTCAGTTG	G	TT	AAACGT	G	AT	AAATG
Type	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism			Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism			Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism
Minimum	184	196	208	220	222	250			275	276	297	298	305	310	373	385		416	457	484	484			508	521	565	622	645	665	667	687	738	742	750	752	755	763	765	768
Maximum	184	196	208	222	222	250			276	276	297	298	305	310	375	385		417	457	486	486			508	521	565	622	645	665	667	687	740	749	750	753	760	763	766	772
Change	T -> C	A -> G	A -> G	GGC -> TGA	C -> A	C -> T			(TG)4 -> (TG)3	-G	G -> A	C -> T	(T)4 -> (T)3	G -> A	-AGT	A -> T		+AGTTGGCGG	G -> C	GGG -> AGA	GGG -> TGA			A -> G	(G)4 -> (G)3	G -> C	(T)7 -> (T)6	-G	G -> T	A -> G	A -> G	AGG -> GTT	CCGGTCCA -> GTCAGTTG	T -> G	AG -> TT	TTTTCG -> AAACGT	C -> G	CC -> AT	GGCGA -> AAATG
Polymorphism Type	(transition)	(transition)	(transition)	Substitution	(transversion)	(transition)			Deletion	Deletion	(transition)	(transition)	Deletion	(transition)	Deletion	(transversion)		Insertion	(transversion)	Substitution	Substitution			(transition)	Deletion	(transversion)	Deletion	Deletion	(transversion)	(transition)	(transition)	Substitution	Substitution	(transversion)	Substitution	Substitution	(transversion)	Substitution	Substitution
Variant Frequency	40.2%	79.1%	36.7%	34.6%	44.4%	35.8%			29.5%	25.5%	29.5%	41.7%	80.4%	30.0%	90.6%	66.2%		70.3%	38.8%	38.0%	41.0%			27.4%	33.4%	36.0%	69.5%	64.1%	90.1%	91.5%	47.5%	95.6%	89.1%	99.4%	98.9%	91.5% -> 91.6%	96.8%	93.8% -> 93.9%	80.5% -> 95.4%
Variant P-Value (approximate)	0	0	0	0	0	0			4.3E-17	2.6E-27	5.5E-261	0	0	0	0	0		0	0	0	0			0	2.6E-43	0	0	0	0	0	0	0	0	0	0	0	0	0	7.4E-230

La 253C 47% + 484TGA 41%

Name	C	G	G	TGA		T					A			A		T		AGTTGGCGG			TGA			G		C			T	G		GTT	GTCAGTTG	G	TT	AAACGT	G	AT	AAAT
Type	Polymorphism	Polymorphism	Polymorphism	Polymorphism		Polymorphism					Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism		Polymorphism			Polymorphism			Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism
Minimum	184	196	208	220		250					297		305	310	373	385		416			484			508	521	565	622	645	665	667		738	742	750	752	755	763	765	768
Maximum	184	196	208	222		250					297		305	310	375	385		417			486			508	521	565	622	645	665	667		740	749	750	753	760	763	766	771
Change	T -> C	A -> G	A -> G	GGC -> TGA		C -> T					G -> A		(T)4 -> (T)3	G -> A	-AGT	A -> T		+AGTTGGCGG			GGG -> TGA			A -> G	(G)4 -> (G)3	G -> C	(T)7 -> (T)6	-G	G -> T	A -> G		AGG -> GTT	CCGGTCCA -> GTCAGTTG	T -> G	AG -> TT	TTTTCG -> AAACGT	C -> G	CC -> AT	GGCG -> AAAT
Polymorphism Type	(transition)	(transition)	(transition)	Substitution		(transition)					(transition)		Deletion	(transition)	Deletion	(transversion)		Insertion			Substitution			(transition)	Deletion	(transversion)	Deletion	Deletion	(transversion)	(transition)		Substitution	Substitution	(transversion)	Substitution	Substitution	(transversion)	Substitution	Substitution
Variant Frequency	70.0%	86.8%	72.9%	71.4%		71.9%					60.7%		62.1%	64.4%	95.6%	65.4%		73.6%			97.3%			54.2%	34.8%	78.9%	69.8%	38.4%	94.7%	96.3%		96.4%	88.2%	99.6%	99.3%	91.9%	96.8%	94.3% -> 94.4%	94.8% -> 95.4%
Variant P-Value (approximate)	0	0	0	0		0					0		4.3E-252	0	0	0		0			0			0	6.1E-23	0	2.9E-255	2.7E-127	0	0		0	0	0	0	0	0	0	0

La 253C 47% + 484AGA 38%

Name		G			A							T				T		AGTTGGCGG	C	AGA									T	G	G	GTT	GTCAGTTGG		TT	AAACGT	G	AT	AAAT
Type		Polymorphism			Polymorphism							Polymorphism	Polymorphism		Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism			Polymorphism		Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism
Minimum		196			222							298	305		373	385		416	457	484			507		521		622	645	665	667	687	738	742		752	755	763	765	768
Maximum		196			222							298	305		375	385		417	457	486			507		521		622	645	665	667	687	740	750		753	760	763	766	771
Change		A -> G			C -> A							C -> T	(T)4 -> (T)3		-AGT	A -> T		+AGTTGGCGG	G -> C	GGG -> AGA			(G)3 -> (G)2		(G)4 -> (G)3		(T)7 -> (T)6	-G	G -> T	A -> G	A -> G	AGG -> GTT	CGGTCCAT -> GTCAGTTGG		AG -> TT	TTTTCG -> AAACGT	C -> G	CC -> AT	GGCG -> AAAT
Polymorphism Type		(transition)			(transversion)							(transition)	Deletion		Deletion	(transversion)		Insertion	(transversion)	Substitution			Deletion		Deletion		Deletion	Deletion	(transversion)	(transition)	(transition)	Substitution	Substitution		Substitution	Substitution	(transversion)	Substitution	Substitution
Variant Frequency		92.3%			94.6%							85.4%	78.9%		98.5%	79.8%		79.4%	91.0%	100.0%			37.6%		33.7%		69.1%	92.0%	95.9%	96.3%	95.9%	95.6%	90.1%		99.0%	91.3%	96.7%	92.8%	95.0% -> 95.8%
Variant P-Value (approximate)		0			0							0	8.8E-202		0	0		0	0	0			5.7E-26		3.2E-18		1.9E-224	0	0	0	0	0	0		0	0	0	0	0

Pilosella onegensis 2 alleles

- 482C (45%)	1 allele
- 482T (55%)	1 allele

SNPs

Name	C	C	A		TC	T	CGC	C		T	G	C	C	C	T	TGA		G		G	T		A		TGG	T	GTT	GTCAGT	TGG	TT	AAACGT	G	AT	AAATG	G
Type	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism
Minimum	184	220	274	305	354	364	367	371	373	385	406	421	452	463	482	484	507	508	521	530	565	622	629	645	665	678	738	742	748	752	755	763	765	768	773
Maximum	184	220	274	305	355	364	369	371	375	385	406	421	452	463	482	486	507	508	521	530	565	622	629	645	667	678	740	747	750	753	760	763	766	772	773
Change	T -> C	G -> C	G -> A	(T)4 -> (T)3	GT -> TC	G -> T	TGT -> CGC	G -> C	-AGT	A -> T	A -> G	T -> C	A -> C	T -> C	C -> T	GGG -> TGA	(G)3 -> (G)2	A -> G	(G)4 -> (G)3	T -> G	G -> T	(T)7 -> (T)6	G -> A	-G	GAA -> TGG	A -> T	AGG -> GTT	CCGGTC -> GTCAGT	CAT -> TGG	AG -> TT	TTTTCG -> AAACGT	C -> G	CC -> AT	GGCGA -> AAATG	A -> G
Polymorphism Type	(transition)	(transversion)	(transition)	Deletion	Substitution	(transversion)	Substitution	(transversion)	Deletion	(transversion)	(transition)	(transition)	(transversion)	(transition)	(transition)	Substitution	Deletion	(transition)	Deletion	(transversion)	(transversion)	Deletion	(transition)	Deletion	Substitution	(transversion)	Substitution	Substitution	Substitution	Substitution	Substitution	(transversion)	Substitution	Substitution	(transition)
Variant Frequency	87.5%	96.7%	94.8%	34.3%	53.3%	97.0%	93.9%	90.1%	100.0%	84.6%	95.6%	96.2%	43.1%	54.3%	55.2%	53.2%	39.5%	38.3%	37.2%	71.7%	53.2%	69.2%	55.3%	99.8%	90.4%	98.9%	95.2%	88.5%	95.8%	98.6%	92.0%	96.2%	93.7%	73.2% -> 94.8%	30.0%
Variant P-Value (approximate)	0	0	0	0.000015	0	0	0	0	0	0	0	0	0	0	0	0	1.1E-45	0	1.1E-184	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.3E-11

On 482C (45%)

Name	C	C	A		T	CGC	C		T	G	AGTTGGCGG	C	C		G			TGG	T	GTT	GTCAGT	TGG	TT	AAACGT	G	AT	AAAT
Type	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism
Minimum	184	220	274	305	364	367	371	373	385	406	416	421	452	521	530	622	645	665	678	738	742	748	752	755	763	765	768
Maximum	184	220	274	305	364	369	371	375	385	406	417	421	452	521	530	622	645	667	678	740	747	750	753	760	763	766	771
Change	T -> C	G -> C	G -> A	(T)4 -> (T)3	G -> T	TGT -> CGC	G -> C	-AGT	A -> T	A -> G	+AGTTGGCGG	T -> C	A -> C	(G)4 -> (G)3	T -> G	(T)7 -> (T)6	-G	GAA -> TGG	A -> T	AGG -> GTT	CCGGTC -> GTCAGT	CAT -> TGG	AG -> TT	TTTTCG -> AAACGT	C -> G	CC -> AT	GGCG -> AAAT
Polymorphism Type	(transition)	(transversion)	(transition)	Deletion	(transversion)	Substitution	(transversion)	Deletion	(transversion)	(transition)	Insertion	(transition)	(transversion)	Deletion	(transversion)	Deletion	Deletion	Substitution	(transversion)	Substitution	Substitution	Substitution	Substitution	Substitution	(transversion)	Substitution	Substitution
Variant Frequency	87.0%	97.0%	94.5%	38.9%	96.6%	93.4%	86.8%	100.0%	81.5%	95.8%	43.3%	97.4%	96.1%	40.3%	88.8%	67.2%	99.9%	90.6%	98.4%	94.1%	88.9%	95.3%	98.5%	93.0%	96.3%	92.4% -> 92.6%	93.6% -> 94.4%
Variant P-Value (approximate)	0	0	0	0.00000068	0	0	0	0	0	0	0	0	0	4.1E-49	0	1.2E-289	0	0	0	0	0	0	0	0	0	0	0

On 482T (55%)

Name	C	C	A		TC	T	CGC	C		T	G	C	C	T	TGA	G		G	T		A		TGG	T	GTT	GTCAGTTG	G	TT	AAACGT	G	AT	AAATG
Type	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism
Minimum	184	220	274	305	354	364	367	371	373	385	406	421	463	482	484	508	521	530	565	622	629	645	665	678	738	742	750	752	755	763	765	768
Maximum	184	220	274	305	355	364	369	371	375	385	406	421	463	482	486	508	521	530	565	622	629	645	667	678	740	749	750	753	760	763	766	772
Change	T -> C	G -> C	G -> A	(T)4 -> (T)3	GT -> TC	G -> T	TGT -> CGC	G -> C	-AGT	A -> T	A -> G	T -> C	T -> C	C -> T	GGG -> TGA	A -> G	(G)4 -> (G)3	T -> G	G -> T	(T)7 -> (T)6	G -> A	-G	GAA -> TGG	A -> T	AGG -> GTT	CCGGTCCA -> GTCAGTTG	T -> G	AG -> TT	TTTTCG -> AAACGT	C -> G	CC -> AT	GGCGA -> AAATG
Polymorphism Type	(transition)	(transversion)	(transition)	Deletion	Substitution	(transversion)	Substitution	(transversion)	Deletion	(transversion)	(transition)	(transition)	(transition)	(transition)	Substitution	(transition)	Deletion	(transversion)	(transversion)	Deletion	(transition)	Deletion	Substitution	(transversion)	Substitution	Substitution	(transversion)	Substitution	Substitution	(transversion)	Substitution	Substitution
Variant Frequency	88.7%	97.6%	96.2%	34.6%	93.0%	98.1%	94.7%	92.3%	100.0%	88.0%	95.9%	96.2%	98.5%	100.0%	98.0%	64.3%	37.0%	75.5%	95.7%	67.4%	93.5%	99.7%	91.2%	99.3%	95.6%	88.8%	99.5%	98.9%	93.0%	96.6%	95.4%	77.0% -> 95.9%
Variant P-Value (approximate)	0	0	0	1.1E-21	0	0	0	0	0	0	0	0	0	0	0	0	2.1E-70	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2.5E-253

Pilosella officinarum (Caen) **3 alleles**

- 647C (20%)	1 allele
- 647 (80%)	
- 232C (50%)	1 allele
- 232T (50%)	1 allele

Name	T	C	T	G	G	TGA	T	G	A	G								T	G	AGTTGGCGG		TGA		G				C				T	G		GTT	GT	C		AGTTGG	TT	AAACGT	G	AT	AAATG
Type	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism				Polymorphism			Polymorphism	Polymorphism	Polymorphism	Polymorphism		Polymorphism		Polymorphism	Polymorphism			Polymorphism		Polymorphism		Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism
Minimum	173	185	193	197	209	221	251	254	257	262				306			374	386	407	419		485		509	522			566		623		666	668		739	743	745		746	753	756	764	766	769
Maximum	173	185	193	197	209	223	251	254	257	262				306			376	386	407	418		487		509	522			566		623		666	668		741	744	745		751	754	761	764	767	773
Change	**C -> T**	**T -> C**	**C -> T**	**A -> G**	**A -> G**	**GGC -> TGA**	**C -> T**	**C -> G**	**C -> A**	**T -> G**				**(T)4 -> (T)3**			**-AGT**	**A -> T**	**A -> G**	**+AGTTGGCGG**		**GGG -> TGA**		**A -> G**	**(G)4 -> (G)3**			**G -> C**		**(T)7 -> (T)6**		**G -> T**	**A -> G**		**AGG -> GTT**	**CC -> GT**	**G -> C**		**GTCCAT -> AGTTGG**	**AG -> TT**	**TTTTCG -> AAACGT**	**C -> G**	**CC -> AT**	**GGCGA -> AAATG**
Polymorphism Type	(transition)	(transition)	(transition)	(transition)	(transition)	Substitution	(transition)	(transversion)	(transversion)	(transversion)				Deletion			Deletion	(transversion)	(transition)	Insertion		Substitution		(transition)	Deletion			(transversion)		Deletion		(transversion)	(transition)		Substitution	Substitution	(transversion)		Substitution	Substitution	Substitution	(transversion)	Substitution	Substitution
Variant Frequency	**64.6%**	**100.0%**	**40.6%**	**100.0%**	**100.0%**	**100.0%**	**78.9%**	**79.9%**	**80.0%**	**79.5%**				**77.1%**			**100.0%**	**39.9%**	**61.9%**	**98.1%**		**100.0%**		**38.2%**	**33.6%**			**100.0%**		**66.6%**		**100.0%**	**100.0%**		**94.6%**	**96.3%**	**78.1%**		**90.5%**	**98.7%**	**94.0%**	**98.0%**	**95.9% -> 96.0%**	**87.6% -> 96.8%**
Variant P-Value (approximate)	0	0	0	0	0	0	0	0	0	0				0			0	0	0	0		0		0	1.1E-72			0		0		0	0		0	0	0		0	0	0	0	0	0

P-C_647C (20%)

Name	T	C		G	G	TGA					A	A			A	G		T		AGTTGGCGG		TGA	G	G		G	A	C	TGT			T	G		GTT	GT	A	C	AGTTGG	TT	AAACGT	G	AT	AAAT
Type	Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism					Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism		Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism
Minimum	173	185		197	209	221					275	298		306	311	373	374	386		419		485	490	509	522	531	536	566	605	623	646	666	668		739	743	745	745	746	753	756	764	766	769
Maximum	173	185		197	209	223					275	298		306	311	373	376	386		418		487	490	509	522	531	536	566	604	623	646	666	668		741	744	745	745	751	754	761	764	767	772
Change	**C -> T**	**T -> C**		**A -> G**	**A -> G**	**GGC -> TGA**					**G -> A**	**G -> A**		**(T)4 -> (T)3**	**G -> A**	**T -> G**	**-AGT**	**A -> T**		**+AGTTGGCGG**		**GGG -> TGA**	**T -> G**	**A -> G**	**(G)4 -> (G)3**	**T -> G**	**G -> A**	**G -> C**	**(TTG)2 -> (TTG)3**	**(T)7 -> (T)6**	**-G**	**G -> T**	**A -> G**		**AGG -> GTT**	**CC -> GT**	**G -> A**	**G -> C**	**GTCCAT -> AGTTGG**	**AG -> TT**	**TTTTCG -> AAACGT**	**C -> G**	**CC -> AT**	**GGCG -> AAAT**
Polymorphism Type	(transition)	(transition)		(transition)	(transition)	Substitution					(transition)	(transition)		Deletion	(transition)	(transversion)	Deletion	(transversion)		Insertion		Substitution	(transversion)	(transition)	Deletion	(transversion)	(transition)	(transversion)	Insertion	Deletion	Deletion	(transversion)	(transition)		Substitution	Substitution	(transition)	(transversion)	Substitution	Substitution	Substitution	(transversion)	Substitution	Substitution
Variant Frequency	**92.7%**	**100.0%**		**95.9%**	**100.0%**	**100.0%**					**80.1%**	**74.7%**		**54.7%**	**84.9%**	**86.4%**	**99.3%**	**88.8%**		**82.3%**		**99.8%**	**77.1%**	**77.3%**	**34.1%**	**52.7%**	**75.8%**	**100.0%**	**52.3%**	**70.9%**	**70.2%**	**100.0%**	**100.0%**		**95.2%**	**95.0%**	**66.3%**	**30.7%**	**90.1%**	**99.3%**	**93.9%**	**97.4%**	**95.7% -> 95.9%**	**94.9% -> 96.2%**
Variant P-Value (approximate)	0	0		0	0	0					0	0		2.2E-48	0	0	0	0		0		0	0	0	1.4E-14	0	0	0	0	4.6E-176	0	0	0		0	0	0	4.6E-206	0	0	0	0	0	0

P-C_647A (80%)

Name	T	C	T	G	G	TGA	T	G	A	G								T	G	AGTTGGCGG		TGA		G				C				T	G		GTT	GTC			AGTTGG	TT	AAACGT	G	AT	AAATG
Type	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism				Polymorphism			Polymorphism	Polymorphism	Polymorphism	Polymorphism		Polymorphism		Polymorphism	Polymorphism			Polymorphism		Polymorphism		Polymorphism	Polymorphism		Polymorphism	Polymorphism			Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism
Minimum	173	185	193	197	209	221	251	254	257	262				306			374	386	407	419		485		509	522			566		623		666	668		739	743			746	753	756	764	766	769
Maximum	173	185	193	197	209	223	251	254	257	262				306			376	386	407	418		487		509	522			566		623		666	668		741	745			751	754	761	764	767	773
Change	**C -> T**	**T -> C**	**C -> T**	**A -> G**	**A -> G**	**GGC -> TGA**	**C -> T**	**C -> G**	**C -> A**	**T -> G**				**(T)4 -> (T)3**			**-AGT**	**A -> T**	**A -> G**	**+AGTTGGCGG**		**GGG -> TGA**		**A -> G**	**(G)4 -> (G)3**			**G -> C**		**(T)7 -> (T)6**		**G -> T**	**A -> G**		**AGG -> GTT**	**CCG -> GTC**			**GTCCAT -> AGTTGG**	**AG -> TT**	**TTTTCG -> AAACGT**	**C -> G**	**CC -> AT**	**GGCGA -> AAATG**
Polymorphism Type	(transition)	(transition)	(transition)	(transition)	(transition)	Substitution	(transition)	(transversion)	(transversion)	(transversion)				Deletion			Deletion	(transversion)	(transition)	Insertion		Substitution		(transition)	Deletion			(transversion)		Deletion		(transversion)	(transition)		Substitution	Substitution			Substitution	Substitution	Substitution	(transversion)	Substitution	Substitution
Variant Frequency	**57.5%**	**100.0%**	**49.6%**	**100.0%**	**100.0%**	**100.0%**	**97.8%**	**99.4%**	**100.0%**	**98.8%**				**82.7%**			**100.0%**	**27.2%**	**73.7%**	**98.0%**		**100.0%**		**34.8%**	**32.9%**			**100.0%**		**66.2%**		**100.0%**	**100.0%**		**94.5%**	**89.8%**			**90.8%**	**98.6%**	**94.1%**	**98.0%**	**96.2%**	**86.7% -> 97.1%**
Variant P-Value (approximate)	0	0	0	0	0	0	0	0	0	0				0			0	0	0	0		0		0	1E-50			0		0		0	0		0	0			0	0	0	0	0	0

P-C_647A (80%) + 232C (50%)

Name		C		G	G	TGA	T	G	A	G									G	AGTTGGCGG		TGA		G				C				T	G		GTT	GTC			AGTTGG	TT	AAACGT	G	AT	AAATG
Type		Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism			Polymorphism	Polymorphism			Polymorphism		Polymorphism	Polymorphism		Polymorphism		Polymorphism	Polymorphism			Polymorphism		Polymorphism		Polymorphism	Polymorphism		Polymorphism	Polymorphism			Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism
Minimum		185		197	209	221	251	254	257	262			277	306			374		407	419		485		509	522			566		623		666	668		739	743			746	753	756	764	766	769
Maximum		185		197	209	223	251	254	257	262			277	306			376		407	418		487		509	522			566		623		666	668		741	745			751	754	761	764	767	773
Change		**T -> C**		**A -> G**	**A -> G**	**GGC -> TGA**	**C -> T**	**C -> G**	**C -> A**	**T -> G**			**-G**	**(T)4 -> (T)3**			**-AGT**		**A -> G**	**+AGTTGGCGG**		**GGG -> TGA**		**A -> G**	**(G)4 -> (G)3**			**G -> C**		**(T)7 -> (T)6**		**G -> T**	**A -> G**		**AGG -> GTT**	**CCG -> GTC**			**GTCCAT -> AGTTGG**	**AG -> TT**	**TTTTCG -> AAACGT**	**C -> G**	**CC -> AT**	**GGCGA -> AAATG**
Polymorphism Type		(transition)		(transition)	(transition)	Substitution	(transition)	(transversion)	(transversion)	(transversion)			Deletion	Deletion			Deletion		(transition)	Insertion		Substitution		(transition)	Deletion			(transversion)		Deletion		(transversion)	(transition)		Substitution	Substitution			Substitution	Substitution	Substitution	(transversion)	Substitution	Substitution
Variant Frequency		**100.0%**		**100.0%**	**100.0%**	**100.0%**	**98.4%**	**99.8%**	**100.0%**	**99.3%**			**25.5%**	**82.4%**			**100.0%**		**92.7%**	**97.8%**		**100.0%**		**58.0%**	**33.8%**			**100.0%**		**66.6%**		**100.0%**	**100.0%**		**93.9%**	**90.0%**			**90.4%**	**98.5%**	**94.8%**	**98.0%**	**96.0%**	**91.2% -> 97.7%**
Variant P-Value (approximate)		0		0	0	0	0	0	0	0			5.8E-31	0			0		0	0		0		0	8.5E-27			0		0		0	0		0	0			0	0	0	0	0	1.0E-198

P-C_647A (80%) + 232T (50%)

Name	T	C	T	G	G	TGA	T	G	A	G								T	G	AGTTGGCGG	T	TGA						C				T	G	G	GTT	GTC			AGTTGG	TT	AAACGT	G	AT	AAAT
Type	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism			Polymorphism			Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism		Polymorphism				Polymorphism		Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism			Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism
Minimum	173	185	193	197	209	221	251	254	257	262	276			306			374	386	407	419	498	485		508				566		623		666	668	688	739	742			746	753	756	764	766	769
Maximum	173	185	193	197	209	223	251	254	257	262	277			306			376	386	407	418	498	487		508				566		623		666	668	688	741	745			751	754	761	764	767	772
Change	**C -> T**	**T -> C**	**C -> T**	**A -> G**	**A -> G**	**GGC -> TGA**	**C -> T**	**C -> G**	**C -> A**	**T -> G**	**(TG)4 -> (TG)3**			**(T)4 -> (T)3**			**-AGT**	**A -> T**	**A -> G**	**+AGTTGGCGG**	**G -> T**	**GGG -> TGA**		**(G)3 -> (G)2**				**G -> C**		**(T)7 -> (T)6**		**G -> T**	**A -> G**	**A -> G**	**AGG -> GTT**	**CCG -> GTC**			**GTCCAT -> AGTTGG**	**AG -> TT**	**TTTTCG -> AAACGT**	**C -> G**	**CC -> AT**	**GGCG -> AAAT**
Polymorphism Type	(transition)	(transition)	(transition)	(transition)	(transition)	Substitution	(transition)	(transversion)	(transversion)	(transversion)	Deletion			Deletion			Deletion	(transversion)	(transition)	Insertion	(transversion)	Substitution		Deletion				(transversion)		Deletion		(transversion)	(transition)	(transition)	Substitution	Substitution			Substitution	Substitution	Substitution	(transversion)	Substitution	Substitution
Variant Frequency	**100.0%**	**100.0%**	**83.9%**	**100.0%**	**100.0%**	**100.0%**	**97.4%**	**99.5%**	**100.0%**	**98.7%**	**25.5%**			**83.6%**			**100.0%**	**41.3%**	**60.8%**	**98.4%**	**32.0%**	**100.0%**		**28.0%**				**100.0%**		**64.7%**		**100.0%**	**100.0%**	**35.3%**	**95.5%**	**89.7%**			**92.1%**	**98.7%**	**93.1%**	**98.2%**	**96.4%**	**97.0% -> 97.3%**
Variant P-Value (approximate)	0	0	0	0	0	0	0	0	0	0	9.1E-29			0			0	0	0	0	0	0		6.6E-10				0		7.1E-263		0	0	0	0	0			0	0	0	0	0	0

Pilosella officinarum 4 alleles (Dijon)

P-D DOMINANT ALLELE 1 allele

Name	T							C					
Type	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism
Minimum	149	277	306	508	522	623	759	933	1347	1884	1885	1992	1993
Maximum	149	277	306	508	522	623	759	933	1347	1884	1885	1992	1993
Change	**G -> T**	**-G**	**(T)4 -> (T)3**	**(G)3 -> (G)2**	**(G)4 -> (G)3**	**(T)7 -> (T)6**	**(T)5 -> (T)4**	**T -> C**	**(C)3 -> (C)2**	**(T)6 -> (T)5**	**(T)6 -> (T)5**	**(C)7 -> (C)6**	**(C)7 -> (C)6**
Polymorphism Type	(transversion)	Deletion	Deletion	Deletion	Deletion	Deletion	Deletion	(transition)	Deletion	Deletion	Deletion	Deletion	Deletion
Variant Frequency	**48.5%**	**25.0%**	**89.6%**	**35.2%**	**31.1%**	**64.9%**	**28.3%**	**97.9%**	**26.9%**	**25.5%**	**86.2%**	**94.1%**	**75.0%**
Variant P-Value (approximate)	5.3E-37	1.1E-115	0	4E-128	1.8E-71	0	1.5E-100	0	5.2E-80	6.7E-20	0	0	4.4E-259

P-D RECESSIVE ALLELES

- 262T (45%) 1 allele
- 262G (55%)
  - 311G (50%) 1 allele
  - 311A (50%) 1 allele

SNPs

Name	C	G	G	T	TGA	T	G	A	G	A			A		A		T	G	AGTTGGCGG	T		TGA		G		C				T	G	GTT	GTCAG	TTGG	TT	AAACGT	G	AT	AAATG
Type	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism			Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism		Polymorphism		Polymorphism	Polymorphism	Polymorphism		Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism
Minimum	185	197	209	217	221	251	254	257	262	275			298	306	311	374	386	407	419	458		485		509	522	566		623		666	668	739	743	748	753	756	764	766	769
Maximum	185	197	209	217	223	251	254	257	262	275			298	306	311	376	386	407	418	458		487		509	522	566		623		666	668	741	747	751	754	761	764	767	773
Change	**T -> C**	**A -> G**	**A -> G**	**A -> T**	**GGC -> TGA**	**C -> T**	**C -> G**	**C -> A**	**T -> G**	**G -> A**			**G -> A**	**(T)4 -> (T)3**	**G -> A**	**-AGT**	**A -> T**	**A -> G**	**+AGTTGGCGG**	**G -> T**		**GGG -> TGA**		**A -> G**	**(G)4 -> (G)3**	**G -> C**		**(T)7 -> (T)6**		**G -> T**	**A -> G**	**AGG -> GTT**	**CCGGT -> GTCAG**	**CCAT -> TTGG**	**AG -> TT**	**TTTTCG -> AAACGT**	**C -> G**	**CC -> AT**	**GGCGA -> AAATG**
Polymorphism Type	(transition)	(transition)	(transition)	(transversion)	Substitution	(transition)	(transversion)	(transversion)	(transversion)	(transition)			(transition)	Deletion	(transition)	Deletion	(transversion)	(transition)	Insertion	(transversion)		Substitution		(transition)	Deletion	(transversion)		Deletion		(transversion)	(transition)	Substitution	Substitution	Substitution	Substitution	Substitution	(transversion)	Substitution	Substitution
Variant Frequency	**51.4%**	**93.1%**	**93.9%**	**38.0%**	**94.7%**	**55.2%**	**55.4%**	**55.3%**	**55.4%**	**35.1%**			**36.8%**	**52.5%**	**66.9%**	**96.5%**	**37.3%**	**25.2%**	**89.7%**	**47.2%**		**95.7%**		**35.7%**	**31.1%**	**96.2%**		**66.1%**		**97.5%**	**96.8%**	**94.3%**	**89.2%**	**94.5%**	**97.9%**	**92.7%**	**97.1%**	**95.4%**	**82.3% -> 95.8%**
Variant P-Value (approximate)	0	0	0	0	0	0	0	0	0	0			0	0	0	0	0	0	0	0		0		0	9.6E-63	0		0		0	0	0	0	0	0	0	0	0	0

P-D_262T (45%)

Name		G	G		TGA					A			A		A		T		AGTTGGCGG			TGA		G		C	TGT			T	G	GTT	GTCAGT	TGG	TT	AAACGT	G	AT	AAATG
Type		Polymorphism	Polymorphism		Polymorphism					Polymorphism			Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism		Polymorphism			Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism
Minimum		197	209		221					275			298	306	311	374	386		419			485		509	522	566	605	623	646	666	668	739	743	749	753	756	764	766	769
Maximum		197	209		223					275			298	306	311	376	386		418			487		509	522	566	604	623	646	666	668	741	748	751	754	761	764	767	773
Change		**A -> G**	**A -> G**		**GGC -> TGA**					**G -> A**			**G -> A**	**(T)4 -> (T)3**	**G -> A**	**-AGT**	**A -> T**		**+AGTTGGCGG**			**GGG -> TGA**		**A -> G**	**(G)4 -> (G)3**	**G -> C**	**(TTG)2 -> (TTG)3**	**(T)7 -> (T)6**	**-G**	**G -> T**	**A -> G**	**AGG -> GTT**	**CCGGTC -> GTCAGT**	**CAT -> TGG**	**AG -> TT**	**TTTTCG -> AAACGT**	**C -> G**	**CC -> AT**	**GGCGA -> AAATG**
Polymorphism Type		(transition)	(transition)		Substitution					(transition)			(transition)	Deletion	(transition)	Deletion	(transversion)		Insertion			Substitution		(transition)	Deletion	(transversion)	Insertion	Deletion	Deletion	(transversion)	(transition)	Substitution	Substitution	Substitution	Substitution	Substitution	(transversion)	Substitution	Substitution
Variant Frequency		**87.3%**	**88.0%**		**88.6%**					**76.4%**			**78.4%**	**38.2%**	**85.8%**	**92.3%**	**79.3%**		**85.2%**			**90.9%**		**70.2%**	**32.8%**	**92.5%**	**26.9%**	**65.9%**	**40.2%**	**95.5%**	**95.3%**	**94.3%**	**88.4%**	**96.2%**	**97.7%**	**93.3%**	**97.4%**	**95.5%**	**83.1% -> 95.9%**
Variant P-Value (approximate)		0	0		0					0			0	6.7E-94	0	0	0		0			0		0	1.1E-47	0	0	0	0	0	0	0	0	0	0	0	0	0	0

P-D_262G (55%)

Name	C	G	G		TGA	T	G	A	G						A			G	AGTTGGCGG	T	A	TGA				C				T	G	GTT	GTCAG	TTGG	TT	AAACGT	G	AT	AAATG
Type	Polymorphism	Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism			Polymorphism		Polymorphism	Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism		Polymorphism	Polymorphism		Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism
Minimum	185	197	209		221	251	254	257	262			277		306	311	374		407	419	458	476	485	508		522	566		623		666	668	739	743	748	753	756	764	766	769
Maximum	185	197	209		223	251	254	257	262			277		306	311	376		407	418	458	476	487	508		522	566		623		666	668	741	747	751	754	761	764	767	773
Change	**T -> C**	**A -> G**	**A -> G**		**GGC -> TGA**	**C -> T**	**C -> G**	**C -> A**	**T -> G**			**-G**		**(T)4 -> (T)3**	**G -> A**	**-AGT**		**A -> G**	**+AGTTGGCGG**	**G -> T**	**G -> A**	**GGG -> TGA**	**(G)3 -> (G)2**		**(G)4 -> (G)3**	**G -> C**		**(T)7 -> (T)6**		**G -> T**	**A -> G**	**AGG -> GTT**	**CCGGT -> GTCAG**	**CCAT -> TTGG**	**AG -> TT**	**TTTTCG -> AAACGT**	**C -> G**	**CC -> AT**	**GGCGA -> AAATG**
Polymorphism Type	(transition)	(transition)	(transition)		Substitution	(transition)	(transversion)	(transversion)	(transversion)			Deletion		Deletion	(transition)	Deletion		(transition)	Insertion	(transversion)	(transition)	Substitution	Deletion		Deletion	(transversion)		Deletion		(transversion)	(transition)	Substitution	Substitution	Substitution	Substitution	Substitution	(transversion)	Substitution	Substitution
Variant Frequency	**84.0%**	**97.8%**	**98.3%**		**99.7%**	**97.8%**	**98.9%**	**98.2%**	**100.0%**			**25.3%**		**63.6%**	**49.8%**	**99.8%**		**39.1%**	**81.4%**	**73.8%**	**36.1%**	**97.3%**	**34.3%**		**30.4%**	**99.2%**		**66.8%**		**98.9%**	**98.0%**	**94.2%**	**98.8%**	**94.3%**	**98.1%**	**91.7%**	**98.5%**	**94.5%**	**88.5% -> 95.5%**
Variant P-Value (approximate)	0	0	0		0	0	0	0	0			1.1E-100		0	0	0		0	0	0	0	0	1.2E-95		7.6E-53	0		0		0	0	0	0	0	0	0	0	0	0

P-D 262G 55% + 311G 50%

Name	C	G	G		TGA	T	G	A	G										AGTTGGCGG	T		TGA			C					T	G	GTT	GTCAGT	TGG	TT	AAACGT	G	AT	AAATG
Type	Polymorphism	Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism		Polymorphism			Polymorphism		Polymorphism			Polymorphism	Polymorphism		Polymorphism	Polymorphism		Polymorphism	Polymorphism		Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism
Minimum	185	197	209		221	251	254	257	262		276			306		374			419	458		485	508		522	566		623		666	668	739	743	749	753	756	764	766	769
Maximum	185	197	209		223	251	254	257	262		277			306		376			418	458		487	508		522	566		623		666	668	741	748	751	754	761	764	767	773
Change	**T -> C**	**A -> G**	**A -> G**		**GGC -> TGA**	**C -> T**	**C -> G**	**C -> A**	**T -> G**		**(TG)4 -> (TG)3**			**(T)4 -> (T)3**		**-AGT**			**+AGTTGGCGG**	**G -> T**		**GGG -> TGA**	**(G)3 -> (G)2**		**(G)4 -> (G)3**	**G -> C**		**(T)7 -> (T)6**		**G -> T**	**A -> G**	**AGG -> GTT**	**CCGGTC -> GTCAGT**	**CAT -> TGG**	**AG -> TT**	**TTTTCG -> AAACGT**	**C -> G**	**CC -> AT**	**GGCGA -> AAATG**
Polymorphism Type	(transition)	(transition)	(transition)		Substitution	(transition)	(transversion)	(transversion)	(transversion)		Deletion			Deletion		Deletion			Insertion	(transversion)		Substitution	Deletion		Deletion	(transversion)		Deletion		(transversion)	(transition)	Substitution	Substitution	Substitution	Substitution	Substitution	(transversion)	Substitution	Substitution
Variant Frequency	**84.3%**	**97.8%**	**98.1%**		**99.6%**	**97.9%**	**99.2%**	**98.4%**	**100.0%**		**25.5%**			**62.6%**		**99.7%**			**79.9%**	**74.7%**		**96.6%**	**34.3%**		**30.8%**	**98.9%**		**67.5%**		**99.0%**	**98.2%**	**94.2%**	**88.8%**	**95.2%**	**98.3%**	**91.6%**	**96.2%**	**94.2%**	**81.5% -> 95.4%**
Variant P-Value (approximate)	0	0	0		0	0	0	0	0		1.8E-51			1.4E-290		0			0	0		0	1.7E-46		2.4E-28	0		0		0	0	0	0	0	0	0	0	0	3.5E-283

P-D 262G 55% + 311A 50%

Name	C	G	G		TGA	T	G	A	G						A			G	AGTTGGCGG	T	A	TGA				C				T	G	GTT	GTCA	GTTGG	TT	AAACGT	G	AT	AAATG
Type	Polymorphism	Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism					Polymorphism	Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism		Polymorphism	Polymorphism		Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism
Minimum	185	197	209		221	251	254	257	262					306	311	374		407	419	458	476	485	508		522	566		623		666	668	739	743	747	753	756	764	766	769
Maximum	185	197	209		223	251	254	257	262					306	311	376		407	418	458	476	487	508		522	566		623		666	668	741	746	751	754	761	764	767	773
Change	**T -> C**	**A -> G**	**A -> G**		**GGC -> TGA**	**C -> T**	**C -> G**	**C -> A**	**T -> G**					**(T)4 -> (T)3**	**G -> A**	**-AGT**		**A -> G**	**+AGTTGGCGG**	**G -> T**	**G -> A**	**GGG -> TGA**	**(G)3 -> (G)2**		**(G)4 -> (G)3**	**G -> C**		**(T)7 -> (T)6**		**G -> T**	**A -> G**	**AGG -> GTT**	**CCGG -> GTCA**	**TCCAT -> GTTGG**	**AG -> TT**	**TTTTCG -> AAACGT**	**C -> G**	**CC -> AT**	**GGCGA -> AAATG**
Polymorphism Type	(transition)	(transition)	(transition)		Substitution	(transition)	(transversion)	(transversion)	(transversion)					Deletion	(transition)	Deletion		(transition)	Insertion	(transversion)	(transition)	Substitution	Deletion		Deletion	(transversion)		Deletion		(transversion)	(transition)	Substitution	Substitution	Substitution	Substitution	Substitution	(transversion)	Substitution	Substitution
Variant Frequency	**83.9%**	**97.8%**	**98.6%**		**99.9%**	**97.8%**	**98.6%**	**98.1%**	**100.0%**					**64.6%**	**100.0%**	**99.9%**		**73.4%**	**83.4%**	**73.2%**	**65.5%**	**98.2%**	**34.1%**		**29.6%**	**99.5%**		**66.3%**		**98.8%**	**97.8%**	**94.2%**	**88.4%**	**92.5%**	**97.7%**	**92.0%**	**96.8%**	**94.8%**	**79.2% -> 95.6%**
Variant P-Value (approximate)	0	0	0		0	0	0	0	0					0	0	0		0	0	0	0	0	4.4E-47		1.5E-23	0		0		0	0	0	0	0	0	0	0	0	6.9E-269

Pilosella officinarum **4 alleles**
(Lewis Pass)

DOMINANT ALLELE 1 allele

Name							A	C						
Type	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism
Minimum	275	305	507	521	622	758	853	932	1046	1346	1883	1884	1991	1992
Maximum	276	305	507	521	622	758	852	932	1046	1346	1883	1884	1991	1992
Change	**(TG)4 -> (TG)3**	**(T)4 -> (T)3**	**(G)3 -> (G)2**	**(G)4 -> (G)3**	**(T)7 -> (T)6**	**(T)5 -> (T)4**	**(A)8 -> (A)9**	**T -> C**	**(T)4 -> (T)3**	**(C)3 -> (C)2**	**(T)6 -> (T)5**	**(T)6 -> (T)5**	**(C)7 -> (C)6**	**(C)7 -> (C)6**
Polymorphism Type	Deletion	Deletion	Deletion	Deletion	Deletion	Deletion	Insertion	(transition)	Deletion	Deletion	Deletion	Deletion	Deletion	Deletion
Variant Frequency	**26.6%**	**89.9%**	**38.6%**	**32.3%**	**61.7%**	**28.0%**	**33.4%**	**98.0%**	**26.7%**	**28.0%**	**27.1%**	**85.8%**	**93.7%**	**74.1%**
Variant P-Value (approximate)	9.5E-194	0	0	5.4E-215	0	7.1E-240	0.000034	0	0	5.3E-96	5.6E-76	0	0	0

RECESSIVE ALLELES

- 250C (56%) 1 allele
- 250T (44%)
  - 310G (47%) 1 allele
  - 310A (53%) 1 allele

Very messy data
unsure of the choice for the 3rd allele

SNPs

Name	C	G	G		TGA	T	G	A	G	A	A		A		T		AGTTGGCGG	T		TGA		G		C			T	G	GTT	GTCAGT	TGG	G	TT	AAACGT	G	AT	AAAT	G
Type	Polymorphism	Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism		Polymorphism	Polymorphism		Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism
Minimum	184	196	208		220	250	253	256	261	274	297	305	310	373	385		418	457		484		508	521	565	622		665	667	738	742	748	750	752	755	763	765	768	772
Maximum	184	196	208		222	250	253	256	261	274	297	305	310	375	385		417	457		486		508	521	565	622		665	667	740	747	750	750	753	760	763	766	771	772
Change	**T -> C**	**A -> G**	**A -> G**		**GGC -> TGA**	**C -> T**	**C -> G**	**C -> A**	**T -> G**	**G -> A**	**G -> A**	**(T)4 -> (T)3**	**G -> A**	**-AGT**	**A -> T**		**+AGTTGGCGG**	**G -> T**		**GGG -> TGA**		**A -> G**	**(G)4 -> (G)3**	**G -> C**	**(T)7 -> (T)6**		**G -> T**	**A -> G**	**AGG -> GTT**	**CCGGTC -> GTCAGT**	**CAT -> TGG**	**T -> G**	**AG -> TT**	**TTTTCG -> AAACGT**	**C -> G**	**CC -> AT**	**GGCG -> AAAT**	**A -> G**
Polymorphism Type	(transition)	(transition)	(transition)		Substitution	(transition)	(transversion)	(transversion)	(transversion)	(transition)	(transition)	Deletion	(transition)	Deletion	(transversion)		Insertion	(transversion)		Substitution		(transition)	Deletion	(transversion)	Deletion		(transversion)	(transition)	Substitution	Substitution	Substitution	(transversion)	Substitution	Substitution	(transversion)	Substitution	Substitution	(transition)
Variant Frequency	**41.9%**	**85.2%**	**85.5%**		**84.3%**	**44.1%**	**43.0%**	**42.2%**	**43.0%**	**36.4%**	**35.6%**	**45.4%**	**65.2%**	**93.2%**	**34.6%**		**74.2%**	**36.3%**		**89.0%**		**27.1%**	**32.6%**	**89.5%**	**69.1%**		**93.0%**	**94.5%**	**95.3%**	**89.2%**	**95.5%**	**99.5%**	**98.4%**	**91.5%**	**96.8%**	**94.2%**	**94.5% -> 94.8%**	**82.9%**
Variant P-Value (approximate)	0	0	0		0	0	0	0	0	0	0	1.8E-291	0	0	0		0	0		0		0	5E-135	0	0		0	0	0	0	0	0	0	0	0	0	0	0

Pi LP 250C 56%

Name		G	G	T	TGA					A	A		A		T		AGTTGGCGG			TGA		G		C			T	G	GTT	GTCAGTTG			TT	AAACGT	G	AT	AAATG	
Type		Polymorphism	Polymorphism	Polymorphism	Polymorphism					Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism		Polymorphism			Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism			Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	
Minimum		196	208	216	220					274	297	305	310	373	385		418			484		508	521	565	622	645	665	667	738	742			752	755	763	765	768	
Maximum		196	208	216	222					274	297	305	310	375	385		417			486		508	521	565	622	645	665	667	740	749			753	760	763	766	772	
Change		**A -> G**	**A -> G**	**A -> T**	**GGC -> TGA**					**G -> A**	**G -> A**	**(T)4 -> (T)3**	**G -> A**	**-AGT**	**A -> T**		**+AGTTGGCGG**			**GGG -> TGA**		**A -> G**	**(G)4 -> (G)3**	**G -> C**	**(T)7 -> (T)6**	**-G**	**G -> T**	**A -> G**	**AGG -> GTT**	**CCGGTCCA -> GTCAGTTG**			**AG -> TT**	**TTTTCG -> AAACGT**	**C -> G**	**CC -> AT**	**GGCGA -> AAATG**	
Polymorphism Type		(transition)	(transition)	(transversion)	Substitution					(transition)	(transition)	Deletion	(transition)	Deletion	(transversion)		Insertion			Substitution		(transition)	Deletion	(transversion)	Deletion	Deletion	(transversion)	(transition)	Substitution	Substitution			Substitution	Substitution	(transversion)	Substitution	Substitution	
Variant Frequency		**74.8%**	**74.4%**	**36.5%**	**72.2%**					**63.9%**	**64.3%**	**38.4%**	**73.2%**	**83.9%**	**66.5%**		**69.0%**			**80.3%**		**53.9%**	**35.1%**	**84.4%**	**67.8%**	**33.0%**	**92.0%**	**93.5%**	**96.7%**	**88.7%**			**98.6%**	**93.8%**	**97.6%**	**95.1%**	**77.9% -> 95.9%**	
Variant P-Value (approximate)		0	0	0	0					0	0	3.6E-63	0	0	0		0			0		0	5.4E-44	0	0	0	0	0	0	0			0	0	0	0	7.7E-227	

Pi LP 250T 44%

Name	C	G	G		TGA	T	G	A	G				A			G	AGTTGGCGG	T	A	TGA				C			T	G	GTT	GTCAGTTG		G	TT	AAACGT	G	AT	AAATG	
Type	Polymorphism	Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism			Polymorphism	Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	
Minimum	184	196	208		220	250	253	256	261			305	310	373		406	418	457	475	484	507		521	565	622		665	667	738	742		750	752	755	763	765	768	
Maximum	184	196	208		222	250	253	256	261			305	310	375		406	417	457	475	486	507		521	565	622		665	667	740	749		750	753	760	763	766	772	
Change	**T -> C**	**A -> G**	**A -> G**		**GGC -> TGA**	**C -> T**	**C -> G**	**C -> A**	**T -> G**			**(T)4 -> (T)3**	**G -> A**	**-AGT**		**A -> G**	**+AGTTGGCGG**	**G -> T**	**G -> A**	**GGG -> TGA**	**(G)3 -> (G)2**		**(G)4 -> (G)3**	**G -> C**	**(T)7 -> (T)6**		**G -> T**	**A -> G**	**AGG -> GTT**	**CCGGTCCA -> GTCAGTTG**		**T -> G**	**AG -> TT**	**TTTTCG -> AAACGT**	**C -> G**	**CC -> AT**	**GGCGA -> AAATG**	
Polymorphism Type	(transition)	(transition)	(transition)		Substitution	(transition)	(transversion)	(transversion)	(transversion)			Deletion	(transition)	Deletion		(transition)	Insertion	(transversion)	(transition)	Substitution	Deletion		Deletion	(transversion)	Deletion		(transversion)	(transition)	Substitution	Substitution		(transversion)	Substitution	Substitution	(transversion)	Substitution	Substitution	
Variant Frequency	**85.1%**	**95.0%**	**94.9%**		**93.7%**	**100.0%**	**97.8%**	**96.7%**	**93.4%**			**61.4%**	**53.2%**	**99.1%**		**41.4%**	**81.6%**	**69.2%**	**36.8%**	**96.1%**	**35.5%**		**35.2%**	**94.1%**	**68.5%**		**94.7%**	**96.2%**	**96.0%**	**88.6%**		**99.6%**	**98.2%**	**92.2%**	**97.2%**	**95.1% -> 95.2%**	**79.3% -> 95.6%**	
Variant P-Value (approximate)	0	0	0		0	0	0	0	0			3.5E-271	0	0		0	0	0	0	0	2.5E-62		7E-25	0	0		0	0	0	0		0	0	0	0	0	1.3E-257	

Pi LP 250T 44% + 310G 47%

Name	C	G	G		TGA	T	G	A	G								AGTTGGCGG	T		TGA				C			T	G	GTT	GTCAGTTG		G	TT	AAACGT	G	AT	AAAT	
Type	Polymorphism	Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism			Polymorphism		Polymorphism			Polymorphism	Polymorphism		Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	
Minimum	184	196	208		220	250	253	256	261			305		373			418	457		484	507		521	565	622		665	667	738	742		750	752	755	763	765	768	
Maximum	184	196	208		222	250	253	256	261			305		375			417	457		486	507		521	565	622		665	667	740	749		750	753	760	763	766	771	
Change	**T -> C**	**A -> G**	**A -> G**		**GGC -> TGA**	**C -> T**	**C -> G**	**C -> A**	**T -> G**			**(T)4 -> (T)3**		**-AGT**			**+AGTTGGCGG**	**G -> T**		**GGG -> TGA**	**(G)3 -> (G)2**		**(G)4 -> (G)3**	**G -> C**	**(T)7 -> (T)6**		**G -> T**	**A -> G**	**AGG -> GTT**	**CCGGTCCA -> GTCAGTTG**		**T -> G**	**AG -> TT**	**TTTTCG -> AAACGT**	**C -> G**	**CC -> AT**	**GGCG -> AAAT**	
Polymorphism Type	(transition)	(transition)	(transition)		Substitution	(transition)	(transversion)	(transversion)	(transversion)			Deletion		Deletion			Insertion	(transversion)		Substitution	Deletion		Deletion	(transversion)	Deletion		(transversion)	(transition)	Substitution	Substitution		(transversion)	Substitution	Substitution	(transversion)	Substitution	Substitution	
Variant Frequency	**85.4%**	**95.2%**	**94.2%**		**93.0%**	**100.0%**	**97.6%**	**96.2%**	**93.3%**			**58.9%**		**98.7%**			**80.9%**	**68.0%**		**94.1%**	**35.2%**		**36.5%**	**93.3%**	**68.0%**		**95.2%**	**96.4%**	**96.3%**	**88.7%**		**100.0%**	**98.1%**	**91.9%**	**97.2%**	**94.8%**	**94.6% -> 95.7%**	
Variant P-Value (approximate)	0	0	0		0	0	0	0	0			3.9E-111		0			0	0		0	1E-28		2E-15	0	2.8E-260		0	0	0	0		0	0	0	0	0	0	

Pi LP 250T 44% + 310A 53%

Name	C	G	G		TGA	T	G	A	G				A			G	AGTTGGCGG	T	A	TGA				C			T	G	GTT	GTCAGTTG		G	TT	AAACGT	G	AT	AAATG	
Type	Polymorphism	Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism			Polymorphism	Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	
Minimum	184	196	208		220	250	253	256	261			305	310	373		406	418	457	475	484	507		521	565	622		665	667	738	742		750	752	755	763	765	768	
Maximum	184	196	208		222	250	253	256	261			305	310	375		406	417	457	475	486	507		521	565	622		665	667	740	749		750	753	760	763	766	772	
Change	**T -> C**	**A -> G**	**A -> G**		**GGC -> TGA**	**C -> T**	**C -> G**	**C -> A**	**T -> G**			**(T)4 -> (T)3**	**G -> A**	**-AGT**		**A -> G**	**+AGTTGGCGG**	**G -> T**	**G -> A**	**GGG -> TGA**	**(G)3 -> (G)2**		**(G)4 -> (G)3**	**G -> C**	**(T)7 -> (T)6**		**G -> T**	**A -> G**	**AGG -> GTT**	**CCGGTCCA -> GTCAGTTG**		**T -> G**	**AG -> TT**	**TTTTCG -> AAACGT**	**C -> G**	**CC -> AT**	**GGCGA -> AAATG**	
Polymorphism Type	(transition)	(transition)	(transition)		Substitution	(transition)	(transversion)	(transversion)	(transversion)			Deletion	(transition)	Deletion		(transition)	Insertion	(transversion)	(transition)	Substitution	Deletion		Deletion	(transversion)	Deletion		(transversion)	(transition)	Substitution	Substitution		(transversion)	Substitution	Substitution	(transversion)	Substitution	Substitution	
Variant Frequency	**84.7%**	**94.9%**	**95.6%**		**94.2%**	**100.0%**	**97.9%**	**97.1%**	**93.5%**			**64.1%**	**100.0%**	**99.4%**		**72.9%**	**82.5%**	**69.5%**	**63.9%**	**97.9%**	**36.6%**		**34.1%**	**94.9%**	**69.0%**		**94.4%**	**95.9%**	**95.7%**	**88.7%**		**99.4%**	**98.4%**	**92.6%**	**97.2%**	**95.7% -> 95.8%**	**82.0% -> 96.2%**	
Variant P-Value (approximate)	0	0	0		0	0	0	0	0			4.9E-184	0	0		0	0	0	0	0	1.3E-37		2.3E-28	0	0		0	0	0	0		0	0	0	0	0	1.7E-132	

Pilosella piloselloides var praealta R35 4 alleles

R35_DOMINANT ALLELE 1 allele

Name							T		G				
Type	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism
Minimum	306	508	522	623	759	853	1275	1347	1745	1884	1885	1992	1993
Maximum	306	508	522	623	759	853	1275	1347	1745	1884	1885	1992	1993
Change	(T)4 -> (T)3	(G)3 -> (G)2	(G)4 -> (G)3	(T)7 -> (T)6	(T)5 -> (T)4	(A)8 -> (A)7	A -> T	(C)3 -> (C)2	A -> G	(T)6 -> (T)5	(T)6 -> (T)5	(C)7 -> (C)6	(C)7 -> (C)6
Polymorphism Type	Deletion	Deletion	Deletion	Deletion	Deletion	Deletion	(transversion)	Deletion	(transition)	Deletion	Deletion	Deletion	Deletion
Variant Frequency	78.7%	35.8%	30.2%	64.1%	26.8%	64.4%	33.0%	26.2%	60.9%	25.4%	88.7%	98.0%	78.1%
Variant P-Value (approximate)	0	1.2E-52	0.000001	0	5.3E-31	4.2E-33	0	3.6E-28	0	0.000000024	0	0	1.4E-126

R35_RECESSIVE ALLELES
- 536A (33%) 1 allele
- 536G (66%)
  - 453A (48%) 1 allele
  - 453C (52%) 1 allele

Name	C	C	T	A			T	CGC	C		T	G	AGTTGGCGG	C	C
Type	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism
Minimum	185	221	251	275	276	306	365	368	372	374	385	407	419	422	450
Maximum	185	221	251	275	277	306	365	370	372	376	385	407	419	422	450
Change	T -> C	G -> C	C -> T	G -> A	(TG)4 -> (TG)3	(T)4 -> (T)3	G -> T	TGT -> CGC	G -> C	-AGT	A -> T	A -> G	+AGTTGGCGG	T -> C	A -> C
Polymorphism Type	(transition)	(transversion)	(transition)	(transition)	Deletion	Deletion	(transversion)	Substitution	(transversion)	Deletion	(transversion)	(transition)	Insertion	(transition)	(transversion)
Variant Frequency	96.5%	97.7%	29.0%	30.5%	28.7%	77.0%	68.2%	67.3%	67.7%	99.0%	97.7%	97.8%	30.8%	96.4%	68.8%
Variant P-Value (approximate)	0	0	0	0	6.4E-33	0	0	0	0	0	0	0	0	0	0

Name	C	T	TGA		G	G		G		A	T	A	G	T	
Type	Polymorphism	Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism
Minimum	464	483	485		509	509	522	531		536	539	545	548	566	623
Maximum	464	483	487		509	510	522	531		536	539	545	548	566	623
Change	T -> C	C -> T	GGG -> TGA		A -> G	AG -> G	(G)4 -> (G)3	T -> G		G -> A	G -> T	G -> A	A -> G	G -> T	(T)7 -> (T)6
Polymorphism Type	(transition)	(transition)	Substitution		(transition)	Deletion	Deletion	(transversion)		(transition)	(transversion)	(transition)	(transition)	(transversion)	Deletion
Variant Frequency	32.3%	32.4%	31.5%		54.4%	27.0%	37.7%	75.7%		32.8%	31.4%	33.2%	34.0%	32.4%	68.8%
Variant P-Value (approximate)	0	0	0		0	4.6E-79	6.4E-69	0		0	0	0	0	0	0

Name	A		TGG	T	GTT	GTCAGT	TTGG	TGG	TT	AAACGT	G	AT	AAATG
Type	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism
Minimum	630	646	666	679	739	743	748	749	753	756	764	766	769
Maximum	630	646	668	679	741	748	750	751	754	761	764	767	773
Change	G -> A	-G	GAA -> TGG	A -> T	AGG -> GTT	CCGGTC -> GTCAGT	CCAT -> TTGG	CAT -> TGG	AG -> TT	TTTTCG -> AAACGT	C -> G	CC -> AT	GGCGA -> AAATG
Polymorphism Type	(transition)	Deletion	Substitution	(transversion)	Substitution	Substitution	Substitution	Substitution	Substitution	Substitution	(transversion)	Substitution	Substitution
Variant Frequency	64.7%	74.4%	68.1%	69.1%	94.6%	88.9%	94.7%	95.9%	98.2%	91.8%	97.0%	95.8%	79.6% -> 96.3%
Variant P-Value (approximate)	0	0	0	0	0	0	0	0	0	0	0	0	4.8E-299

R35_536A (33%)

Name	C	C					T	CGC	C		T	G	AGTTGGCGG	C	C
Type	Polymorphism	Polymorphism				Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism
Minimum	185	221				306	365	368	372	374	385	407	419	422	450
Maximum	185	221				306	365	370	372	376	385	407	419	422	450
Change	T -> C	G -> C				(T)4 -> (T)3	G -> T	TGT -> CGC	G -> C	-AGT	A -> T	A -> G	+AGTTGGCGG	T -> C	A -> C
Polymorphism Type	(transition)	(transversion)				Deletion	(transversion)	Substitution	(transversion)	Deletion	(transversion)	(transition)	Insertion	(transition)	(transversion)
Variant Frequency	97.9%	98.3%				82.3%	95.1%	94.2%	94.6%	99.5%	98.1%	98.1%	44.3%	96.3%	96.9%
Variant P-Value (approximate)	0	0				1.2E-245	0	0	0	0	0	0	0	0	0

Name					G	G		G		A	T	A	G		
Type					Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism		Polymorphism
Minimum					509	509	522	531	531	536	539	545	548		623
Maximum					509	510	522	531	531	536	539	545	548		623
Change					A -> G	AG -> G	(G)4 -> (G)3	T -> G	(T)3 -> (T)2	G -> A	G -> T	G -> A	A -> G		(T)7 -> (T)6
Polymorphism Type					(transition)	Deletion	Deletion	(transversion)	Deletion	(transition)	(transversion)	(transition)	(transition)		Deletion
Variant Frequency					72.2%	26.3%	38.3%	74.5%	25.5%	99.9%	92.9%	95.2%	94.5%		88.6%
Variant P-Value (approximate)					0	0.000012	5.5E-32	0	1.5E-22	0	0	0	0		7.5E-245

Name	A				GTT	GTCAG			TT	AAACGT	G	AT	AAAT
Type	Polymorphism	Polymorphism			Polymorphism	Polymorphism			Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism
Minimum	630	646			739	743			753	756	764	766	769
Maximum	630	646			741	747			754	761	764	767	772
Change	G -> A	-G			AGG -> GTT	CCGGT -> GTCAG			AG -> TT	TTTTCG -> AAACGT	C -> G	CC -> AT	GGCG -> AAAT
Polymorphism Type	(transition)	Deletion			Substitution	Substitution			Substitution	Substitution	(transversion)	Substitution	Substitution
Variant Frequency	91.0%	28.0%			94.3%	89.2%			97.4%	91.5%	97.4%	95.9% -> 96.0%	95.4% -> 95.9%
Variant P-Value (approximate)	0	6.1E-45			0	0			0	0	0	0	0

R35_536G (66%)

Name	C	C	T	A			T	CGC	C		T	G		C	C
Type	Polymorphism	Polymorphism	Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism		Polymorphism	Polymorphism
Minimum	185	221	251	275		306	365	368	372	374	385	407		422	450
Maximum	185	221	251	275		306	365	370	372	376	385	407		422	450
Change	T -> C	G -> C	C -> T	G -> A		(T)4 -> (T)3	G -> T	TGT -> CGC	G -> C	-AGT	A -> T	A -> G		T -> C	A -> C
Polymorphism Type	(transition)	(transversion)	(transition)	(transition)		Deletion	(transversion)	Substitution	(transversion)	Deletion	(transversion)	(transition)		(transition)	(transversion)
Variant Frequency	95.8%	97.4%	40.2%	41.8%		74.4%	55.1%	54.1%	54.5%	98.8%	97.5%	97.6%		96.5%	52.1%
Variant P-Value (approximate)	0	0	0	0		0	0	0	0	0	0	0		0	0

Name	C	T	TGA		G	G		G						T	
Type	Polymorphism	Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism						Polymorphism	Polymorphism
Minimum	464	483	485		509	509	522	531						566	623
Maximum	464	483	487		509	510	522	531						566	623
Change	T -> C	C -> T	GGG -> TGA		A -> G	AG -> G	(G)4 -> (G)3	T -> G						G -> T	(T)7 -> (T)6
Polymorphism Type	(transition)	(transition)	Substitution		(transition)	Deletion	Deletion	(transversion)						(transversion)	Deletion
Variant Frequency	46.8%	47.1%	48.0%		45.6%	27.3%	37.4%	76.3%						47.2%	68.4%
Variant P-Value (approximate)	0	0	0		0	5E-05	8.5E-58	0						0	0

Name	A		TGG	T	GTT	GTCAGT		TGG	TT	AAACGT	G	AT	AAATG
Type	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism
Minimum	630	646	666	679	739	743		749	753	756	764	766	769
Maximum	630	646	668	679	741	748		751	754	761	764	767	773
Change	G -> A	-G	GAA -> TGG	A -> T	AGG -> GTT	CCGGTC -> GTCAGT		CAT -> TGG	AG -> TT	TTTTCG -> AAACGT	C -> G	CC -> AT	GGCGA -> AAATG
Polymorphism Type	(transition)	Deletion	Substitution	(transversion)	Substitution	Substitution		Substitution	Substitution	Substitution	(transversion)	Substitution	Substitution
Variant Frequency	51.9%	97.1%	92.0%	95.8%	94.7%	88.1%		96.0%	98.7%	92.0%	96.8%	95.8%	79.9% -> 96.9%
Variant P-Value (approximate)	0	0	0	0	0	0		0	0	0	0	0	3.7E-214

R35_536G (66%) + 453A (48%)

Name	C	C	T								T	G		C	
Type	Polymorphism	Polymorphism	Polymorphism			Polymorphism				Polymorphism	Polymorphism	Polymorphism		Polymorphism	
Minimum	185	221	251			306				374	385	407		422	
Maximum	185	221	251			306				376	385	407		422	
Change	T -> C	G -> C	C -> T			(T)4 -> (T)3				-AGT	A -> T	A -> G		T -> C	
Polymorphism Type	(transition)	(transversion)	(transition)			Deletion				Deletion	(transversion)	(transition)		(transition)	
Variant Frequency	94.1%	95.8%	77.3%			77.8%				97.5%	95.8%	95.9%		95.0%	
Variant P-Value (approximate)	0	0	0			1.3E-203				0	0	0		0	

Name	C	T	TGA		G			G						T	
Type	Polymorphism	Polymorphism	Polymorphism		Polymorphism		Polymorphism	Polymorphism						Polymorphism	Polymorphism
Minimum	464	483	485		509		522	531						566	623
Maximum	464	483	487		509		522	531						566	623
Change	T -> C	C -> T	GGG -> TGA		A -> G		(G)4 -> (G)3	T -> G						G -> T	(T)7 -> (T)6
Polymorphism Type	(transition)	(transition)	Substitution		(transition)		Deletion	(transversion)						(transversion)	Deletion
Variant Frequency	96.2%	95.5%	93.7%		97.5%		36.8%	77.7%						93.3%	87.3%
Variant P-Value (approximate)	0	0	0		0		1.5E-29	0						0	5.1E-266

Name	A		TGG	T	GTT	GTCAGT		TGG	TT	AAACGT	G	AT	AAAT
Type	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism
Minimum	630	646	666	679	739	743		749	753	756	764	766	769
Maximum	630	646	668	679	741	748		751	754	761	764	767	772
Change	G -> A	-G	GAA -> TGG	A -> T	AGG -> GTT	CCGGTC -> GTCAGT		CAT -> TGG	AG -> TT	TTTTCG -> AAACGT	C -> G	CC -> AT	GGCG -> AAAT
Polymorphism Type	(transition)	Deletion	Substitution	(transversion)	Substitution	Substitution		Substitution	Substitution	Substitution	(transversion)	Substitution	Substitution
Variant Frequency	93.5%	97.2%	91.6%	95.5%	95.0%	88.9%		96.4%	98.8%	92.1%	96.7%	95.3%	96.0% -> 96.3%
Variant P-Value (approximate)	0	0	0	0	0	0		0	0	0	0	0	0

R35_536G (66%) + 453C (52%)

Name	C	C		A			T	CGC	C		T	G	AGTTGGCGG	C	C
Type	Polymorphism	Polymorphism		Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism
Minimum	185	221		275		306	365	368	372	374	385	407	419	422	450
Maximum	185	221		275		306	365	370	372	376	385	407	419	422	450
Change	T -> C	G -> C		G -> A		(T)4 -> (T)3	G -> T	TGT -> CGC	G -> C	-AGT	A -> T	A -> G	+AGTTGGCGG	T -> C	A -> C
Polymorphism Type	(transition)	(transversion)		(transition)		Deletion	(transversion)	Substitution	(transversion)	Deletion	(transversion)	(transition)	Insertion	(transition)	(transversion)
Variant Frequency	97.3%	98.9%		72.8%		71.2%	96.4%	94.9%	95.9%	100.0%	99.0%	99.2%	45.0%	97.0%	100.0%
Variant P-Value (approximate)	0	0		0		1.1E-119	0	0	0	0	0	0	0	0	0

Name				A	G		G								
Type				Polymorphism	Polymorphism	Polymorphism	Polymorphism							Polymorphism	
Minimum				509	509	522	531							623	
Maximum				509	510	522	531							623	
Change				G -> A	AG -> G	(G)4 -> (G)3	T -> G							(T)7 -> (T)6	
Polymorphism Type				(transition)	Deletion	Deletion	(transversion)							Deletion	
Variant Frequency				33.7%	33.3%	38.1%	75.0%							88.8%	
Variant P-Value (approximate)				0	3.1E-97	3.4E-33	0							4.4E-260	

Name		TGG	T	GTT	GTCAGT		TGG	TT	AAACGT	G	AT	AAAT	
Type	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism		Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	Polymorphism	
Minimum	646	666	679	739	743		749	753	756	764	766	769	
Maximum	646	668	679	741	748		751	754	761	764	767	772	
Change	-G	GAA -> TGG	A -> T	AGG -> GTT	CCGGTC -> GTCAGT		CAT -> TGG	AG -> TT	TTTTCG -> AAACGT	C -> G	CC -> AT	GGCG -> AAAT	
Polymorphism Type	Deletion	Substitution	(transversion)	Substitution	Substitution		Substitution	Substitution	Substitution	(transversion)	Substitution	Substitution	
Variant Frequency	97.0%	92.5%	98.0%	94.8%	89.3%		95.8%	98.9%	91.9%	97.0%	96.2%	97.3% -> 97.5%	
Variant P-Value (approximate)	0	0	0	0	0		0	0	0	0	0	0	