



UPPSALA
UNIVERSITET

The Evolutionary Origin of Plastid Genomic Rearrangements in *Selaginellaceae* (Lycopodiophyta)



Agnes Simon

Degree project in biology, Master of science (1 year), 2012

Examensarbete i biologi 30 hp till magisterexamen, 2012

Biology Education Centre and Department of Organismal Biology, Uppsala

University Supervisors: Niklas Wikström and Petra Korall

Abstract

The family Selaginellaceae is a monophyletic group of inconspicuous plants, with a fossil history extending back to the Carboniferous era (~350 million years). They are so called fern allies and consist of approximately 700 species, all in a single genus – *Selaginella*. To most of us, they are known as the spike-moss family. The diversity of Selaginellaceae is highest in tropical and mid-mountainous areas, but they can also be sighted by the careful observer in most habitats. Known representatives of Selaginellaceae are the Northern prickly mountain-moss in Sweden, and the commercially sold resurrection plant *S. lepidophylla*, originating in New Mexico and Mexico.

In 2007, the plastid genome of *S. uncinata* was sequenced by Tsuji *et al.* This achievement presented useful information and tools for the further investigation of the phylogenetic relationships within the family. According to the authors of the sequencing project, there were several unique genomic rearrangements in *S. uncinata* in comparison with other fern allies. The rearrangements included a transposition with a length of 17-kb, and an 20-kb long inversion.

The aim of my study was to shed light upon the evolutionary origins of the genomic rearrangements, which in turn could give hints about the internal groupings of the spike moss family. Based on the most recent categorisation of the family Selaginellaceae by Korall (2003), representative *Selaginella* species were chosen to be analysed for the presence of the rearrangements as found in *S. uncinata*. The control or outgroup, was made up by the fern allies *Huperzia selago* and *Isöetes lacustris*. The methods used for this analysis were designed to be clear and logical, and included primer design, PCR-amplification and DNA-sequencing of the supposed rearrangement areas.

In this study we found that one of the rearrangements, the transposition, was present in most of the analysed species, except for *S. selaginoides*. The other rearrangement, the inversion, was only present in a few species, which according to the present phylogeny are closely related. We can thereby predict that neither rearrangements are common genetic characters for Selaginellaceae, rather that they have emerged through the evolutionary development of the family.

Further research is needed to confirm and validate our findings. This could be done by better and broader sampling and the design of more effective primers for the examined regions.

Table of contents

1 Introduction	4
1.1 <i>A presentation of the family Selaginellaceae</i>	4
1.2 <i>Phylogenetic studies of the family</i>	5
1.3 <i>Recent sequencing of the whole chloroplast genome of Selaginella uncinata</i>	8
1.4 <i>Aim of the study</i>	8
2 Materials and methods	
2.1 <i>Choice of taxa</i>	8
2.2 <i>Design of PCR-based techniques for the detection of rearrangements and primer design</i>	10
2.3 <i>DNA extraction, PCR-amplification and sequencing</i>	11
2.4 <i>Sequence analysis</i>	12
3 Results	12
4 Discussion	13
4.1 <i>Possible phylogenetic implications</i>	14
4.2 <i>Evidence of the transposition in most examined species</i>	14
4.3 <i>Absence of the inversion in most examined species</i>	14
4.4 <i>Phylogenetic implications of a distinct subgroup and sources of uncertainty</i>	15
4.5 <i>Expect the unexpected – Conclusions</i>	15
4.6 <i>Future considerations</i>	16
4.7 <i>Up-to-date findings</i>	16
5 Acknowledgements	16
6 Literature cited	17

1 Introduction

1.1 A presentation of the family Selaginellaceae

The family Selaginellaceae is an ancient family of fern allies wherein all taxa carry the genus name of *Selaginella*. Together with the families of Lycopodiaceae and Isoetaceae, the three families constitute the Lycophytes or Lycopodiopsida. This group in itself is very important with much to narrate about the history and evolution of land plants. The Lycophytes is a recognised monophyletic clade. Lycopods are seedless, vascular plants. Common traits of living species are the presence of microphylls (comparably small leaves with usually a singular vein) and an herbaceous habit. Secondary growth in this phylum can be observed in the fossil record during the Carboniferous age, although extensive woodiness seems to have been lost about 248 million years ago at the end of the Paleozonic period (Raven *et al.*, 2003).

Lycophytes are richly represented in the fossil record and seem to be among the first plants to have colonised territories on land. They appear from the Upper Silurian period (~425 million years) (Kenrick and Crane, 1997) and reach the peak of their rein during the Carboniferous period (Raven *et al.*, 2003; Korall, 2003). Evident ancestors of Selaginellaceae are first encountered in the fossil record during the Carboniferous era in tropical wetland floras. Little seems to have changed since this period in the morphology of the family (Korall and Kenrick, 2002). Many characters, as the terminaly ending fertile parts, the dichotomous branches, the microphyllic leaves, the characteristic megaspores and the organisation of leaves on both isophyllous and anisophyllous species, are highly similar to the *Selaginella* species in the fossil record (Thomas, 1997). Isoetaceae is typically recognised as a sister group to Selaginellaceae (Korall, 2003) and in contrast to Lycopodiaceae the two former families are heterosporous, having both micro- and megaspores (Raven *et al.*, 2003). Until relatively recently the relationships among and particularly within each family, have not been studied in extensively. With the aid of constantly developing molecular research methods, these relationships have become somewhat clearer during the last decade (Wikström and Kenrick, 1997; Rydin and Wikström, 2002; Korall and Kenrick, 2004).

In present time the cosmopolitan family of Selaginellaceae is comprised of about 700 species (Korall, 2003). This figure is somewhat questionable, as often is the case in accessing the total number of species within groups, where morphological classification of individuals might be difficult (personal comment, Petra Korall).

Maximum diversity of the family can be found in tropical rain forests from lowland to mid-montane areas, although the distribution of Selaginellaceae is not in any way limited by definite ecological, climatic or light factors. Members can be encountered as well in temperate, subarctic as in subtropical regions (Korall and Kenrick, 2002). Some have creeping or climbing habits on trees or rocks, but in contrast to the members of Lycopodiaceae very few are epiphytes (Korall *et al.*, 1999). The tiny, flat, frond-like branches with anisophyllous leaves (having leaves of two different sizes and shapes) of microphyllic makeup, is a characteristic that many species within the group share. A standout trait in many members is the presence of rhizophores: aerial, root-like structures growing along the branches of the plant, often at the branching of stems. These structures are situated distinctively on the lower side or on the upper side of the branches, extending toward the ground (Jermy, 1990). Blue iridescence is also common, facilitating light-harvest in shaded areas, as on the floor of tropical forests. In this environment much of the sunlight needed for photosynthesis is trapped by the abundant leafage of the crown canopy. Blue iridescence decreases the rate of reflectance of light off the leaves, and more energy will therefore be accessible for photosynthesis in the chloroplasts (Héban and Lee, 1984).

A number of Selaginellaceae members are the so-called xerophytes (Korall *et al.*, 1999). The term xerophytes denotes plants that exhibit various drought tolerant qualities/traits. These plants are adapted to survive in climates where water is scarce during prolonged periods of time

(Oxford University Press, 2004). The Selaginellaceae-xerophytes can be found in deserts, heath lands and other dry surroundings in numerous parts of the world (Korall, 2003). *S. lepidophylla* can be found in Texas, New Mexico and Mexico (Raven *et al.*, 2003), and is worth mentioning here. Similar to some other so-called resurrection plants, it is able to furl its stems and leaves to evade superfluous loss of water in times of water scarcity and to revert this action upon rehydration. *S. lepidophylla* is one of the species of the family grown commercially for ornamental purposes (Korall, 2003).

A northern specimen that Scandinavians might stumble into in lime rich and wet areas is *S. selaginoides* (Mossberg and Stenberg, 2003). This plant is isophyllous (having leaves of one shape and size) in contrast to the majority of other species within the family (Jermy, 1986; Korall and Kenrick, 2002). Its appearance is quite plain and it can be easy to miss. The plant has a creeping stem running along the ground, from where the main shoots grow upright, with or without strobili at the apex of the stems. The common name of *S. selaginoides* is “Northern or Prickly Mountain-moss” in English and “dvärglummer” in Swedish, and is the only species representing Selaginellaceae in Sweden (Krok and Almquist, 1994; Mossberg and Stenberg, 2003; Raven *et al.*, 2003).

1.2 Phylogenetic studies of the family

Early classifications of Selaginellaceae were naturally based on morphological characters, and often divided the family into many groups and subgroups. These classifications do not show much consistency. Conclusive traits for groupings of the family have mainly included the arrangement of leaves on stems; whether or not leaves and sporophylls are anisophyllous; growth pattern and the amount of megasporangia on vegetative stems. As mentioned above, the outcome of the different investigations resulted in quite variable groupings. Despite these inconsistencies, one observation seems to have persistently divided the family into two basal groups: species with either isophyllous or with anisophyllous leaves (Korall, 2003).

This is likewise the case in the framework for the family drawn up by Jermy (1986; 1990). Jermy's work, as previous classifications, was based on morphological characters. He divided the family into five subclades, three with isophyllous and two with anisophyllous microphylls, the later including the major part of species within the family.

A recent phylogenetic study of the family, based on DNA sequence data rather than morphological characters, had Jermy's work as a stepping stone (Korall *et al.*, 1999; Korall and Kenrick 2002, 2004; Korall, 2003). The aim of the study was to clarify the relationships within the family, and at least in the early stages, to determine whether earlier taxonomic groupings could be applied to a more exhaustive phylogenetic analysis of the Selaginellaceae (Korall *et al.*, 1999). The thesis (Korall, 2003) was based mainly on the *rbcL*-gene sequence of the chloroplast, but it also included the comparison of nuclear 26S rDNA regions and a thorough study of megaspore morphology. The later was based on the phylogenetic tree resulting from the analysis of the gene sequences. The final picture emerging introduces quite a few novelties and coincides to some extent with the earlier classification of the Selaginellaceae by Jermy (Korall *et al.*, 1999; Korall and Kenrick 2002, 2004; Korall, 2003) (Figure 1).

Subgenus *Selaginella*, also described by Jermy (1986; 1990), turned out to be a monophyletic clade and resolved repeatedly as sister group to all other members. It is comprised of two species: *S. selaginoides* and the Hawaiian Islands endemic *S. deflexa*. The sister clade named as the “rhizophoric” clade, is further divided into two clades denoted by Korall *et al.* (1999) simply as clade A and clade B; group A showing various support from low to strong in the studies, and is therefore discussed with more caution; group B has limited to strong support in all studies and is considered as a reasonable clade (Korall and Kenrick, 2002; Korall, 2003).

Clade A includes both isophyllous and anisophyllous species, which is contradictory to former classifications based solely on morphological characters. Clade A is further divided into

several subclades, of which a few examples are *Tetragonostachys* and *Articulatae*. The former includes the species as described by Jermy. *Articulatae* was defined in some earlier classifications as a subgroup within *Stachygynandrum*, (Jermy, 1990; Korall *et al.*, 1999; Korall and Kenrick; 2002). *Tetragonostachys* and *Articulatae* compose in the study by Korall *et al.* (1999) a less reliable clade, named the “dorsal rhizophoric clade”; where members have rhizophores growing on the upper side of branches and bend over the branch extending to the ground, instead of growing from the lower side of the branches. Although the results are a bit ambiguous, Korall *et al.* (1999) propose that *Ericetorum*, one of the five subgroups recognised by Jermy (1990), might emerge to be a monophyletic group in future studies.

Clade B is made up solely by anisophyllous species. In this clade there are no specific groupings resembling earlier classifications. Although many of the clades have high statistical support, Korall *et al.* (1999) choose not to appoint names to any of these clades. Additional research of morphological characters is needed, in order to give support and to further define the suggested subgroups. It is however proposed that the phylogeny obtained can be regarded as a starting point, which can facilitate future phylogenetic studies of Selaginellaceae (Korall and Kenrick, 2002; Korall 2003).

An additional curious discovery in the study is the surprisingly high rates of substitutions in both the chloroplast and the nucleus, compared to substitution rates in other groups of land plants. The study has also disclosed that substitutions in the *rbcL*-gene (the chloroplast genome) are varying in rate not only compared with other land plants, but also when comparing clades within Selaginellaceae. It is unfortunately unclear what this might point to, since the factors accounting for why some lineages have higher substitution rates are complex, and not yet properly understood (Korall, 2003; Korall and Kenrick, 2004).

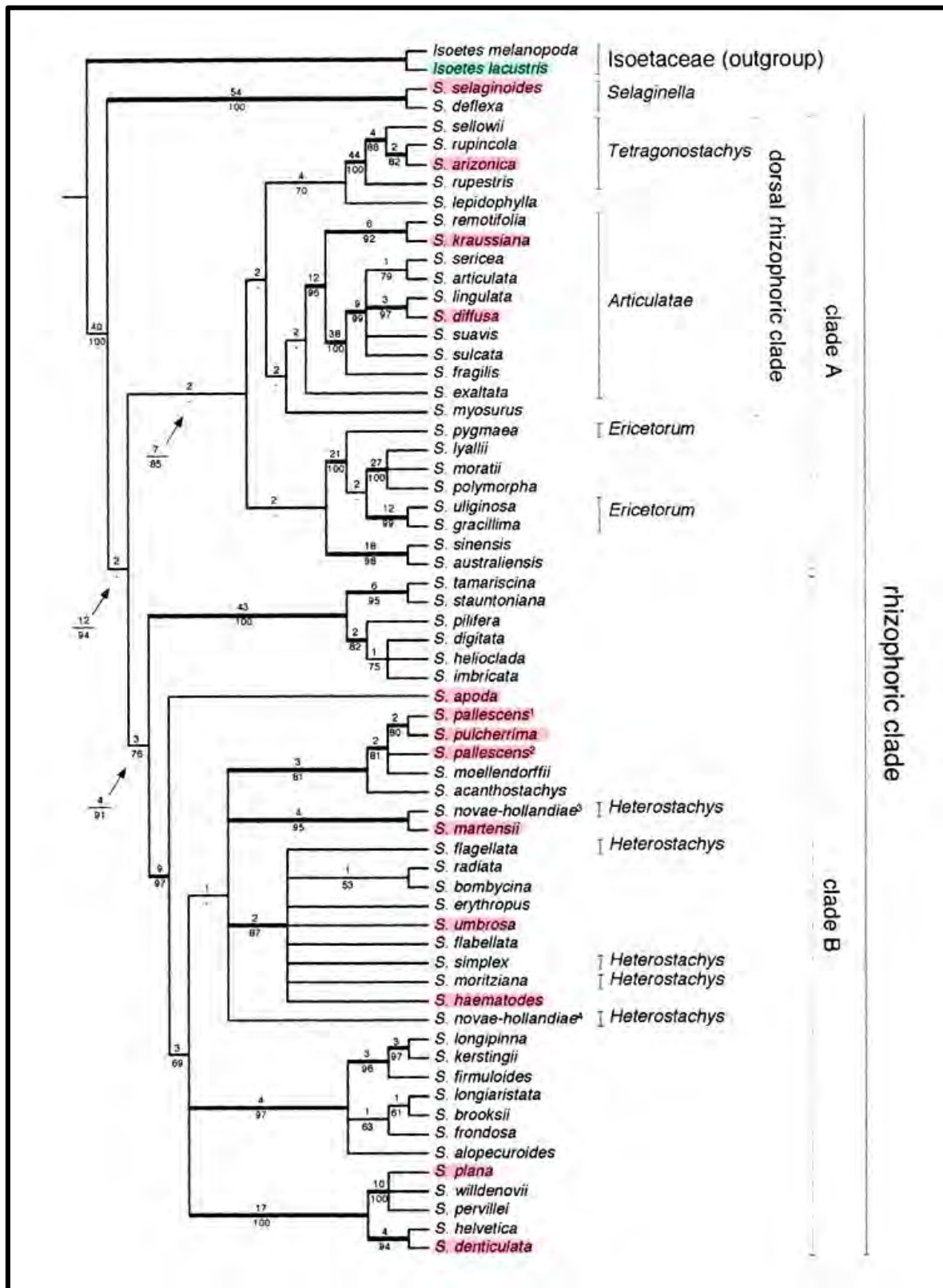


Figure 1. Phylogeny of Selaginellaceae. Korall and Taylor, 2003 (Redrawn from Korall and Kenrick, 2002). *Selaginella*, *Tetragnostachys*, *Ericetorum* and *Heterostachys* are all clades originally recognised by Jermy (1990). The species investigated in this thesis are highlighted.

1.3 Recent sequencing of the whole chloroplast genome of *Selaginella uncinata*

In the recent past, the first chloroplast genome sequencing project of a Selaginellaceae species was completed and analysed to reveal present (or absent) genes and gene order. The species in question was the “peacock spike-moss” or *Selaginella uncinata*, and the results of the project were published in February 2007 in the *Journal of Plant Research* (Tsuji *et al.*, 2007).

The chloroplast genome of *S. uncinata* was compared with the chloroplast genome of *Huperzia lucidula*, a lycophyte belonging to the family of Lycopodiaceae, and one of the few lycophytes whose whole plastid genome has been sequenced at present time. The comparison showed a novel inversion, a transposition, several gene losses and duplications of two genes in *S. uncinata*. These duplications are associated according to the article with both the inversion and the transposition. The 20-kb inversion in *S. uncinata* spans from the *psbI*-gene to the *trnC*-gene in the large single copy region (LSC). The transposition includes a 17-kb portion of the genome between *rps4* - *trnD*. This region is situated in *S. uncinata* in the small single copy region (SSC), whereas in *H. lucidula* as well as in some bryophytes, it is located in the LSC. The article argues that the transposition might be a synapomorphic feature for the Selaginellaceae family, meaning in common for all taxa. There are no definite conclusions made about the presence of the inversion in other species of the family (Tsuji *et al.*, 2007).

1.4 Aim of the study

This study investigated whether or not the transposition and the inversion in *S. uncinata*, could be regarded as shared by all taxa of the family Selaginellaceae. The use of rather simple but coherent PCR-based techniques should be sufficient to reveal if these traits are present or absent in several *Selaginella*-species. As the samplings were based on the phylogenetical framework outlined by Korall (2003), the results are put in context with the same framework to try to unfold the origin of these rearrangements.

2 Materials and Methods

2.1 Choice of taxa

Ingroup The selection of eleven members of Selaginellaceae for extraction of DNA, was designed in order to represent the major clades within the family according to the phylogeny described by Korall and Kenrick (Korall and Kenrick, 2002; Korall, 2003), moreover with consideration to material access. An additional species was included well into the laboratory phase of the project (*S. denticulata*), with the objective to clarify some of the earlier results obtained.

All specimens with the exception of *S. denticulata* and *S. selaginoides* were collected in the green house of the Department of Botany, Stockholm University with the help of Petra Korall. *S. selaginoides* was thankfully received from Niklas Wikström, who collected the specimen in Uppland, Sweden. *S. denticulata* was collected in Tenerife and identified by Emma Lundh and Gina Larsson in 2005, and the identification validated by Mats Thulin. A more informative summary of all species used in the project can be found in Table 1.

Outgroup The control or outgroup in the practical phase of the project, was comprised of *Huperzia selago* (Lycopodiaceae) and *Isoetes lacustris* (Isoetaceae), chosen for their availability and their close phylogenetic relationship with Selaginellaceae. The former was collected and identified by Niklas Wikström in Uppland, Sweden. *Isoetes lacustris* was collected and identified by Didrik Vanhoenacker in the vicinity of Tovetorp, Södermanland, Sweden.

Table 1. Species included in the project. Sources for plant material and accession numbers for sequences published in the EMBL sequence database.

Taxa	Subgenus/clade according to Korall (2002)	DNA source/voucher	EMBL Accession numbers
Ingroup			
<i>S. selaginoides</i> (L.) Link **	<i>Selaginella</i>	Wikström 2007	
<i>S. arizonica</i> Maxon	<i>Tetragonostachys</i> ; Rhizophoric clade A (dorsal rhizophoric clade)	Stockholm University Greenhouse 1998; saml. Torsten?	
<i>S. kraussiana</i> (Kunze) A. Braun	<i>Articulatae</i> ; Rhizophoric clade A (dorsal rhizophoric clade)		
<i>S. diffusa</i> (C. Presl) Spring	<i>Articulatae</i> ; Rhizophoric clade A (dorsal rhizophoric clade)	Stockholm University Greenhouse 1970/2900	
<i>S. apoda</i> (L.) Fern	Rhizophoric clade B	Stockholm University Greenhouse 1985/3872	
<i>S. pallescens</i> (C. Presl) Spring	Rhizophoric clade B	Stockholm University Greenhouse SU-x-00.1263	
<i>S. pulcherrima</i> Liebm. ex Fourn	Rhizophoric clade B	Stockholm University Greenhouse 1969/3071	
<i>S. martensii</i> Spring	Rhizophoric clade B	Stockholm University Greenhouse SU-x-00.1079	
<i>S. umbrosa</i>	Rhizophoric clade B	Stockholm University Greenhouse 1966/26201	
<i>S. haematodes</i> (Kunze) Spring	Rhizophoric clade B	Stockholm University Greenhouse 1973/13797	
<i>S. plana</i> (Desv.) Hieron	Rhizophoric clade B	Stockholm University Greenhouse 1965/6201 (Jermy 4852)	
<i>S. denticulata</i> (L.) Spring*	Rhizophoric clade B	Lundh & Larsson (V-200653) 3514111	
<i>S. uncinata</i> (Desv.) Spring	Rhizophoric clade B		AB197035
Outgroup			
<i>Adiantum capillus-veneris</i> (L.)			NC004766
<i>Huperzia lucidula</i> (Michx.) Trevis			NC006861
<i>Huperzia selago</i> (L.) C. Martius & Schrank **		Wikström 2007	
<i>Isoetes lacustris</i> L.		Vanhoenacker 2007	
<i>Marchantia polymorpha</i> L.			X04465
<i>Physcomitrella patens</i> (Hedw.) Bruch & Schimp			NC005087
All DNA was extracted from silica-dried material, except *=herbarium material; and **= fresh material			

2.2 Design of PCR-based techniques for the detection of rearrangements and primer design

The chloroplast gene order and specific parts of the genome sequence of *S. uncinata*, was compared with the chloroplast genomes of *Huperzia lucidula* (Wolf *et al.*, 2005), *Adiantum capillus-veneris* (Wolf *et al.*, 2003), *Physcomitrella patens* (Sugiura *et al.*, 2003) and *Marchantia polymorpha* (Ohyama *et al.*, 1988) (see Table 1). This was undertaken to illuminate the most interesting genomic rearrangements and to find optimal sites for the use of prime design. The most interesting and informative sites were determined to be the 20-kb inversion and the 17-kb transposition described in the introduction.

In order to detect whether or not the selected rearrangements are present in the chosen Selaginellaceae representatives, primers were designed to amplify the regions from the two outer genes of the inversion and the transposition, to the adjoining genes outside the rearrangements.

Primers were designed in pairs, based on both the gene order of *S. uncinata* and the gene order of the control species (specifically *Huperzia lucidula*). In theory, depending on which set of primer-pair yields a PCR-product, it can be determined if the rearrangement is present in a particular genome (figure 2). With exception of the *trnD*-end of the transposition, where satisfactory sequences for primer-design were not found, a total of 6 primer-pairs were designed (figure 3).

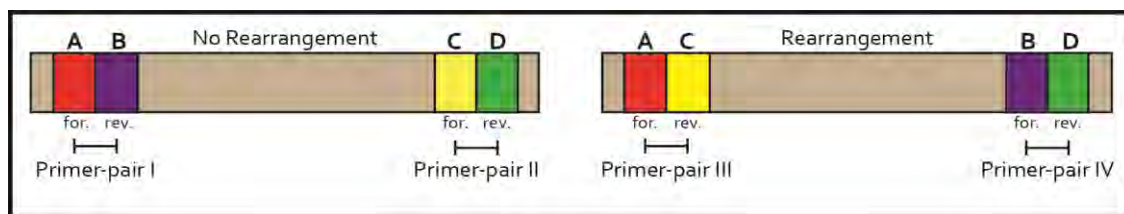


Figure 2. A schematic representation of the method for detection of the genomic rearrangements. If a product is obtained through PCR-amplification with the primer pairs 1 and/or 2, the rearrangement is not present. On the contrary, if a PCR-product is obtained with the pairs 3 and /or 4 the rearrangement is present.

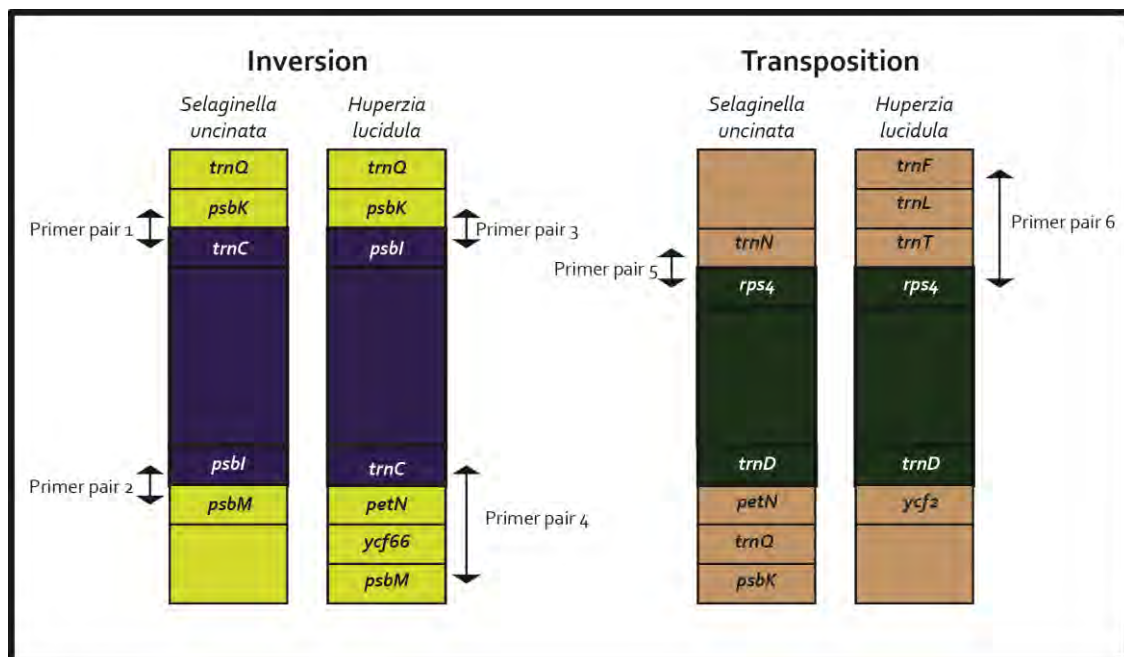


Figure 3. The gene order for the sites of the inversion can be seen on the left side, and for the transposition on the right side. Primer pair 1 consists of primers psbKF and trnCR; primer pair 2 consists of primers psbLR and psbMF; primer pair 3 consists of primers psbKF and psbLR; primer pair 4 consists of primers trnCR and psbMF; primer pair 5 consists of primers trnNF and rsp₄5R; primer pair 6 consists of primers trnFFR and rsp₄5R. See Table 2 for further description of the primers mentioned.

The designing of primers was carried out through visual alignment in the software *Se-Al v2.0a11 Carbon* (Rambaut, 2002), and viable primer-sequences were verified with the software *Amplify 3.1* (Engels, 2004). A list of the primers used is summarised in Table 2. Primer synthesis was carried out by OPERON Biotechnologies in Germany.

Table 2 Primer sequences and their position in the *S. uncinata* chloroplast genome as published in the EMBL sequence database.

Primer	Sequence	Gene position in <i>S. uncinata</i> (AB197035)	Primer sequence position in gene
<i>Primers for detection of the transposition</i>			
rps45R	CGAGGTCCTCKRTRACGGGACRT	<i>rps4</i> : 89262-89852	1-23; reverse
trnNF	GACCGCTCTACCACTGAGCYACT	<i>trnN</i> : 88180-88251	36-61; forward
trnFFR	received from Niklas Wikström	<i>trnF</i> : 41838-41910	forward
<i>Primers for detection of the inversion</i>			
psbIR	ACRARCAGCTTVAGRGTAAGCRT	<i>psbI</i> : 6153-6263	4-26; reverse
psbMF	CCYTGCAYTTATTGCYACTGCAC	<i>psbM</i> : 5509-5607	18-40; forward
psbKF	GAYCCAAYCGTGGATGTAATGCCA	<i>psbK</i> : 27725-27886	64-87; forward
trnCR	CCRGTTCRAATCYGGGTGYCGCCT	<i>trnC</i> : 25741-25811	52-75; reverse
wobbles: R=(A,G); Y=(C,T); K=(G,T); H=(A,C,T); V=(A,C,G)			

The primers paired up in various combinations should reveal if the rearrangements are present in the species included in the study (see Figure 3 for gene order and corresponding primer pairs). If PCR-products are obtained with primer pair 1 and/or with primer pair 2, but not with primer-pair 3 and /or primer-pair 4, the inversion should be present.

As to the transposition region, if a PCR-product is obtained with primer pair 5 but not with primer pair 6, the transposition should be present.

2.3 DNA extraction, PCR-amplification and sequencing

Extraction of DNA-material was done with the DNeasy Plant Kit (QIAGEN). PCR-amplification employed the following kits: PuRe Taq “Ready-To-Go”™ PCR Beads (GE Healthcare) and Phusion™ High-Fidelity PCR Master Mix (Finnzymes). Unfortunately the first kit was not satisfactory and all results were obtained using the second kit.

PCR-amplification was performed in a thermal cycler (Perkin Elmer Cetus/DNA Thermal Cycler 48) according to one of the following programs:

- 1) 98°C, 2 min. x1 followed by cycle (98°C, 0.5 min. → 56°C, 0.5 min. → 72°C, 1.5 min.) x40
- 2) 98°C, 2 min. x1 followed by cycle (98°C, 0.5 min. → 58°C, 0.5 min. → 72°C, 1.5 min.) x40
- 3) 98°C, 2 min. x1 followed by cycle (98°C, 0.5 min. → 60°C, 0.5 min. → 72°C, 1.5 min.) x40

Gel-electrophoresis was performed using agarose gel (150 ml 1xTAE + 1,5 g agarose powder) for the detection of possible PCR-products. The gel was later photographed after soaking in ethidium bromide.

Purification of PCR-products was done according to standard procedures, and DNA sequencing was carried out by MacroGen.

2.4 Sequence analysis

DNA sequences received from MacroGen were assembled using the Staden package, using Trev, Pregap4 and Gap4 (v1.6.0, 2005). The resulting consensus sequences were compared to public nucleotide databases, using BLASTN-discontiguous megablast (<http://www.ncbi.nlm.nih.gov/blast/Blast.cgi>), in order to verify that the correct region was amplified.

3 Results

Of the many PCR-products obtained, 30 were sent for sequencing after purification. The choice of DNA-fragments to be sequenced was governed partly by the quantity of the DNA in the sample after purification, also trying to make sure that there was at least one sample to be sequenced for every PCR-product obtained for all sites and for every species. Unfortunately this was easier said than done. The PCR-amplification did not always yield sufficient quantities in all sampled species and could therefore not be sent for sequencing. Furthermore, some of the sequenced fragments were of inferior quality, and could therefore not be assembled and analysed in the Staden package.

Sequences of good enough quality were compared to public nucleotide databases. This was done in order to verify that the resulting DNA-fragments, matched the expected sites of the rearrangement-regions. The results from PCR-amplifications and the public nucleotide verifications are summarised in Table 3 and 4.

Tabell 3. The table shows if products were obtained after PCR-amplification (+ = product; - = no product; NA = not applicable). From the result of the PCR-amplifications, it can be deduced whether the rearrangement is present or not in a species. Most sequences were matched against public nucleotide databases for the confirmation of their identity (Co = confirmation in public nucleotide databases). For clarification of the primer pairs, see Figure 3.

Taxa	Inversion								Transposition			
	Rearrangement				No Rearrangement				Rearr.		No Rearr.	
	Primer pair 1	Co	Primer pair 2	Co	Primer pair 3	Co	Primer pair 4	Co	Primer pair 5	Co	Primer pair 6	Co
<i>S. selaginoides</i>	-	/	-		+	N	+	N	-		+	N
<i>S. arizonica</i>	-		-		+	N	+	N	+	Y	-	
<i>S. kraussiana</i>	-		-		+	N	+	N	+	N	-	
<i>S. diffusa</i>	-		-		+	Y	+	N	-		+	N
<i>S. apoda</i>	-		-		+	Y	+	Y	+	Y	-	
<i>S. pallescens</i>	-		-		+	Y	+	Y	+	Y	-	
<i>S. pulcherrima</i>	-		-		+	Y	+	N	+	Y	-	
<i>S. martensii</i>	-		-		+	Y	+	N	+	Y	-	
<i>S. umbrosa</i>	-		-		+	Y	+	Y	+	Y	-	
<i>S. haematodes</i>	-		-		+	Y	+	Y	+	Y	-	
<i>S. plana</i>	+	N	+	Y	-		-		+	Y	-	
<i>S. denticulata*</i>	-		+	N	-		-		NA		NA	
<i>H. selago</i>	-		-		-		-		-		+	N
<i>I. lacustris</i>	-		-		-		-		-		+	Y

**S. denticulata* was tested for the presence / absence of the inversion only.

Inversion: The PCR-products amplified representing the inversion-region, indicate that the rearrangement is absent in all *Selaginella*-species, except in *S. plana* and possibly in *S. denticulata*. No PCR-amplifications for the inversion-region gave any products in the case of the control specimens, *H. lucidula* and *I. lacustris*. For *S. selaginoides* none of the PCR-amplifications produced

sufficient amount of DNA needed for sequencing. No PCR-products from *S. denticulata* were sent for sequencing, due to lack of time. All DNA-sequences obtained for *S. arizonica* were of poor quality and could not be assembled and analyzed in the Staden package, hence they could not be verified in the nucleotide database if the sequences actually were the expected ones. DNA-fragments amplified with the primers psbIR+psbK resulted in sequences of inferior quality in the case of *S. apoda*, *S. pallescens*, *S. martensii* and *S. haematodes*. The sequence for this region in *S. diffusa* matched the gene *psbK* in *S. uncinata*, and the sequence for *S. pulcherrima* corresponds to a non-coding 550 bp-region in *S. uncinata* between the genes *psbK* and *trnC*. The sequence of *S. umbrosa* matched partially with genes *psbI* and *psbK* in *S. uncinata*. PCR-products amplified with the primers psbMF+trnCR did not give sufficient DNA needed for sequencing in the case of *S. pulcherrima* and the sequence for *S. martensii* did not produce any matches in the nucleotide databases. *S. apoda*, *S. pallescens*, *S. umbrosa* and *S. haematodes* all matched partially the gene *psbM* and the whole gene *petN* connected with a 400 bp-long non-coding region in *S. uncinata*. They also matched *petN* in 99 other plants. The sequence obtained with the primers psbIR+psbMF in *S. plana* matched the region between genes *psbI-psbM* in *S. uncinata*, while the DNA-sequence amplified by trnCR+psbK was corrupt, could not be analysed with the Staden package and could therefore not be compared to the nucleotide database (Table 4).

Transposition: According to the results of the PCR-amplifications the transposition seems to be present in all *Selaginella* species, except in *S. selaginoides* and in *S. diffusa*. The transposition seems to be absent in *H. lucidula* and *I. lacustris*, as expected. None of the PCR-products amplified for *S. kraussiana* and *S. diffusa* resulted in sufficient amounts of DNA for sequencing. The sequences representing the transposition-region for *S. selaginoides* and *H. lucidula* were of poor quality and could not be assembled and analysed in the Staden package. These sequences could consequently not be compared to the public nucleotide databases for verification. The nucleotide database search results for *I. lacustris* in the nucleotide databases showed partial matches for the genes *trnF-trnL* in *S. tamariscina*, *Isoetes sp.* and 98 other various higher plants. The sequences of *S. arizonica*, *S. apoda* – *S. plana* all matched to the transposition-region in *S. uncinata* (Table 4).

Tabell 4. A summary of the results from the DNA - sequencing, assembly and analysis in the Staden package and the BLASTN-discontiguous megablast in the public nucleotide database.

Taxa	Inversion			Transposition		
	Sequenced	Analysis in Staden	BLASTN-match	Sequenced	Analysis in Staden	BLASTN-match
<i>S. selaginoides</i>	no *	-	-	yes	no ‡	-
<i>S. arizonica</i>	yes	no ‡	-			
<i>S. kraussiana</i>	yes	yes	yes ○	no*	-	-
<i>S. diffusa</i>	yes	yes	<i>psbK</i>	no*	-	-
<i>S. apoda</i>	yes	yes ‡‡	<i>psbM, petN</i>	yes	yes	<i>trnN-rps4</i>
<i>S. pallescens</i>	yes	yes ‡‡	<i>psbM, petN</i>	yes	yes	<i>trnN-rps4</i>
<i>S. pulcherrima</i>	yes***	yes	<i>psbK-trnC</i>	yes	yes	<i>trnN-rps4</i>
<i>S. martensii</i>	yes	yes ‡‡	no matches	yes	yes	<i>trnN-rps4</i>
<i>S. umbrosa</i>	yes	yes	<i>psbI, psbK, psbM, petN</i>	yes	yes	<i>trnN-rps4</i>
<i>S. haematodes</i>	yes	yes ‡‡	<i>psbM, petN</i>	yes	yes	<i>trnN-rps4</i>
<i>S. plana</i>	yes	yes ‡‡‡	<i>psbI-psbM</i>	yes	yes	<i>trnN-rps4</i>
<i>S. denticulata</i>	no	-	-	No	-	-
<i>H. selago</i>	no **	-	-	yes	no ‡	-
<i>I. lacustris</i>	no **	-	-	yes	yes	<i>trnF-trnL</i>
* = insufficient amount of DNA; ** = no PCR-product; *** = insufficient amount of DNA from primer pair 4						
‡ = poor quality; ‡‡ = sequence of primer pair 3 of poor quality; ‡‡‡ = sequence of primer 1 pair of poor quality						
○ = information missing						

4 Discussion

4.1 Possible phylogenetical implications

Tsuji *et al.* (2007) suggest that the transposition might be a character in common for all species of Selaginellaceae. They propose that the considerable rearrangements in the chloroplast genome observed for *S. uncinata* have taken place after the family's partition from Isoetaceae. This is a reasonable assumption in regard with the rearrangements considered in this study, since the transposition seems to be absent in *Isoetes lacustris* and the inversion was verified in only one species of Selaginellaceae.

4.2 Evidence of the transposition in the majority of examined *Selaginella* species

In translating the results of this study, it is quite clear that the transposition is present in all species of *Selaginella* examined for this specific rearrangement, except for *S. selaginoides*. The chloroplast regions amplified by the primers rps45R and trnNF, yielded sequences of not only the similar length, but also matched the same sequence in *S. uncinata*'s chloroplast genome (regions between 88293-89238). The matching sequences seem to correspond with an approximately 650 bp long, non-coding region between the genes *rps4-trnN* in *S. uncinata*. This is not surprising as the primers are situated at the termini of these two genes. The sequenced ends of the PCR-products were of poor quality. The consensus sequence was therefore shorter than the original PCR-product. This could explain why the matching sequence in *S. uncinata* corresponds only with a non-coding region and does not contain any part of the above-mentioned genes.

It is quite unfortunate that the PCR-products for the transposition obtained for *S. selaginoides* were of inadequate quality for verification. It would be of great interest to find out whether the transposition is absent not just in *S. selaginoides*, but also in *S. deflexa*. If this is indeed the case, it would indicate that the transposition in Selaginellaceae took place in the stem lineage of the "rhizophoric" clade. It would also support the clade *Selaginella* in being the sister clade to all other members of the family. Hence, the question whether the transposition is actually an autapomorphic feature for Selaginellaceae remains to be resolved.

4.3 Absence of the inversion in most examined species

The inversion occurs to be absent in all species of Selaginellaceae examined, except for *S. plana* and possibly in *S. denticulata*, the later in which the rearrangement could not be verified through the comparison with public nucleotide databases. In the inversion-region there were sites suitable for primer design in the *psbI*-end of the inversion, in the gene *psbM* bordering to *psbI* outside the inversion, and in the gene *psbK* next to the *trnC*-end of the inversion according to the gene-order in *S. uncinata*. The gene order should be the following in species where rearrangement hasn't occurred: *psbI* next to *psbK* and *trnC* next to *psbM*. Whereas in *H. lucidula*, *trnC* is bordered by *petN*, which in *S. uncinata* seems to have been transposed and is located more than 36 kb downstream from *psbM*. The gene *ycf66* which is located between *petN* and *psbM* in *H. lucidula*, is not present in *S. uncinata* at all. It is therefore difficult to know which gene should be adjacent to *psbM* in *Selaginella* species with no inversion.

DNA-sequences amplified by the primers psbIR and psbKF and verified in the nucleotide databases, were all of similar length (~ 800 bp) and matched the same sites in *S. uncinata*, namely a non-coding region outside *psbK* extending toward *trnC* in *S. uncinata*. This might seem contradictory, since in a non-inversion scenario *psbI* is supposed to be adjacent to *psbK*. But because this is not the gene order in *S. uncinata* it is safe to propose that the region matching the PCR-products, should be the region between *psbI-psbK* in the species where the inversion is

absent. Further, if we look at the length of the genes *psbI* – *psbK* in *H. lucidula*, we find that it is 828 bp long, just about as long as the DNA-sequences from the *Selaginella* species. It is also important to note that the species where a PCR-product was received for the primers *psbIR*+*psbKF*, the amplification did not result in any PCR-product when using the primers *psbIR*-*psbMF*, which represents a presence of the inversion. All of the above indicate the absence of the inversion in all species with only one exception. In the case of *S. plana* the situation is the opposite and the PCR-products show that *S. plana* has the inversion just like *S. uncinata*.

4.4 Phylogenetic implications of a distinct subgroup and sources of uncertainty

We were able to find sites for amplification at both ends of the inversion region. Unfortunately this did not give as clear results at the *psbM*-end region of the inversion as we wished for. More than half of the DNA-fragments could not be verified, either because there wasn't enough DNA present in the samples to be sent for sequencing (*S. selaginoides*, *S. kraussiana*, *S. diffusa* and *S. pulcherrima*) or because the sequence received was corrupt (*S. arizonica*), or simply because there was no match in the nucleotide databases (*S. martensii*). The failure to get PCR-products of good quality could depend on the quality of the primers, especially *trnC*R. The failures could also be due to the uncertainty of which genes could be present, their gene order and of course the length between *psbM* and *trnC* in *Selaginella* species. The region in *H. lucidula* is about 2 kb long, but we do not know how long it should be in *Selaginella*.

The successful PCR-amplifications on the other hand gave consistent results. Even though they varied somewhat in length (~750-1300 bp), the DNA-sequences matched the same sites in *S. uncinata* and in other land plants in the nucleotide databases. The corresponding sites in *S. uncinata* were part of the *psbM*-gene and a couple of hundred bp long non-coding region outside this gene, extending toward *psbI*. A partially non-coding region including the gene *petN* seems to be present as well. The seemingly well-conserved gene is confirmed to be present in all DNA-fragments sequenced, as it matched in all cases for *petN* in about 99 land plants. This is further indication that the inversion indeed is absent. It also shows that in inversion-free species, the *petN* is located in the same position as in *H. lucidula*, next to *psbM*. Based on this information, we can conclude that the inversion is clearly not an overall feature for members of Selaginellaceae. It seems to have been acquired later in the evolution of the family. Petra Korall and Niklas Wikström (not published) did an analysis including all previous species in her studies adding *S. uncinata*, to investigate its phylogenetic position. She obtained a phylogeny, where *S. uncinata* was resolved in a sister clade of *S. plana*, *S. pervillei* and *S. willdenovii* (Figure 1). Well into the laboratory phase, we included *S. denticulata* (classified in a sister clade to the *S. plana* group) as well as *S. willdenovii*. The later was not included in the results of this study, but it was screened for the presence of the inversion. According to the PCR-products the inversion is present in both *S. willdenovii* as in *S. denticulata*. This indicates that the inversion is present in the clade *S. plana* – *S. denticulata*. In this study, there were no species examined in the clade closest to the *S. plana* / *S. denticulata*. Therefore, it is not clear where the inversion could have arisen. According to the results in this study, it is evident that the inversion is present in the *S. uncinata* clade, but it cannot be concluded whether the inversion arose at the separation of this clade.

4.5 Expect the unexpected - Conclusions

The aim of the study was to shed light upon whether or not the transposition and the inversion in *S. uncinata* could be regarded as shared by all taxa of the family Selaginellaceae, as was proposed by the team behind the sequencing of the *S. uncinata* chloroplast genome. If this would have been the case, it would mean that the rearrangements are basal and a novelty for the family.

However, we can conclude that these rearrangements do not appear to be basal and it would be safe to assume that they have arisen during the evolution of the Selaginellaceae family.

It is fair to conclude that the presence of the transposition in most of the sampled species, points to the emergence of the transposition first in the “rhizophoric” clade, as drawn up by Korall (2003). The information obtained about the inversion, implicates that the clade including *S. plana*, *S. uncinata* and *S. denticulata* belong to a monophyletic group. They all seem to carry the inversion, as opposed to the rest of the sampled species. It is important to understand that this is only an inkling, and that there is need for validation and further investigation to better comprehend where the inversion arose in the Selaginellaceae family.

4.6 Future considerations

Naturally, as in many studies, there is always room for broader sampling. The evaluation of the transposition in *S. selaginoides* would be of great importance, as well as screening for the transposition in the other member of the subgroup *Selaginella*. In the case of the inversion’s evolutionary origin, a broader sampling is needed, especially in the clade closest to the *S. plana*-*S. denticulata* group. Design of better primers would also be helpful. Additionally, it would be interesting to try to unravel the gene order around the genes *trnC* and *psbM* and the position of the gene *ycf66* in other *Selaginella* species.

4.7 Up-to-date findings

After the finishing of the practical part of this thesis and during the final touch-ups of the text, it came to my attention that the genome of a new *Selaginella* species has been sequenced. *Selaginella moellendorffii*’s full genome was sequenced by a group of scientists in 2011 (Banks *et al.*, 2011). Unfortunately there is no possibility here to include the findings of the sequencing, of *Selaginella moellendorffii*.

5 Acknowledgements - “I never put off till tomorrow what I can possibly do the day after” – Oscar Wilde

Due to procrastination, I think I have rewritten this part at least ten times. Hopefully this will be the last.

First of all I would like to thank Niklas Wikström, who made me understand the importance of systematics and helped me through the theoretical and practical parts with great patience. My gratitude goes out to Petra Korall, without whom none of this would have been possible and who helped me immensely with sharing her experience, time, advice and her literature to facilitate my job. In addition to thanking, I also owe both Niklas and Petra an apology for taking so long in completing the thesis (and I do hope that the phrase “better late than never” applies...). I also would like to thank Didrick Vanhoecker for the *Isoetes* samples.

Now to the VIPs. Thank you, Johnny Newell for pushing me and for giving me opportunity to complete this dusty work. Thank you, sister Judit Simon-Karlsson for being an inspiration to me and for always giving me good and honest advice. Mom and Dad, what would we girls be without you? Thank you for always giving me your love and support. Thank you Margaret Newell for always making me feel welcome and for the help in taking care of Patrick during the last phase of the work.

I’d like to thank my friends who have given me much comfort before, during and after the actual thesis was done (and whom I’m not sure I deserve): Julia, Jennie, Julieta, Eric, Rosana, Linda Götulf, Eszter, Eugen, Steve, Peffelito, Fran, Linda Romero, Lina, Vera and Christian Steger.

6 Literature cited

- Banks, J.A., Nishiyama, T., Hasebe, M., Bowman, J.L., Gribskov, M., dePamphilis, C., Albert, V.A., Aono, N., Aoyama, T., Ambrose, B.A., Ashton, N.W., Axtell, M.J., Barker, E., Barker, M.S., Bennetzen, J.L., Bonawitz, N.D., Chapple, C., Cheng, C., Correa, L.G., Dacre, M., DeBarry, J., Dreyer, I., Elias, M., Engstrom, E.M., Estelle, M., Feng, L., Finet, C., Floyd, S.K., Frommer, W.B., Fujita, T., Gramzow, L., Gutensohn, M., Harholt, J., Hattori, M., Heyl, A., Hirai, T., Hiwatashi, Y., Ishikawa, M., Iwata, M., Karol, K.G., Koehler, B., Kolukisaoglu, U., Kubo, M., Kurata, T., Lalonde, S., Li, K., Li, Y., Litt, A., Lyons, E., Manning, G., Maruyama, T., Michael, T.P., Mikami, K., Miyazaki, S., Morinaga, S., Murata, T., Mueller-Roeber, B., Nelson, D.R., Obara, M., Oguri, Y., Olmstead, R.G., Onodera, N., Petersen, B.L., Pils, B., Prigge, M., Rensing, S.A., Riano-Pachon, D.M., Roberts, A.W., Sato, Y., Scheller, H.V., Schulz, B., Schulz, C., Shakirov, E.V., Shibagaki, N., Shinohara, N., Shippen, D.E., Sorensen, I., Sotooka, R., Sugimoto, N., Sugita, M., Sumikawa, N., Tanurdzic, M., Theissen, G., Ulvskov, P., Wakazuki, S., Weng, J.K., Willats, W.W., Wipf, D., Wolf, P.G., Yang, L., Zimmer, A.D., Zhu, Q., Mitros, T., Hellsten, U., Loque, D., Otilar, R., Salamov, A., Schmutz, J., Shapiro, H., Lindquist, E., Lucas, S., Rokhsar, D. and Grigoriev, I.V. 2011. The *Selaginella* Genome Identifies Genetic Changes Associated with the Evolution of Vascular Plants. *Science*. 332 (6032): 960-963
- Engels, B. 2004. *Amplify 3.1* wrengels@facstaff.wisc.edu
- Héban, C. and Lec, D. W. 1984. Ultrastructural basis and development control of blue iridescence in *Selaginella* leaves. *American Journal of Botany*. 71 (2): 216-219
- Jermy, A.C. 1986. Subgeneric names in *Selaginella*. *Fern Gazette*, 13 (2): 117-118
- Jermy, A. C. 1990. in K. U. Kramer et al. *The families and genera of plants: Pteridophytes and gymnosperms*. Springer, Berlin: p. 39-45
- Kenrick, P. and Crane, R. 1997. The origin and early evolution of plants on land. *Nature*. 389 (6646); 33-38
- Korall, P. Kenrick, P. and Therrien, J. P. 1999. Phylogeny of Selaginellaceae: evaluation of generic/subgeneric relationships based on *rbcL* gene sequences. *International Journal of Plant Sciences*. 160 (3): 585-594
- Korall, P. and Kenrick, P. 2002. Phylogenetic relationships in Selaginellaceae based on *rbcL* sequences. *American Journal of Botany*. 89 (3): 506-517
- Korall, P. and Kenrick, P. 2004. The phylogenetic history of Selaginellaceae based on DNA sequences from the plastid and nucleus: extreme substitution rates and rate heterogeneity. *Molecular Phylogenetics and Evolution*. 31 (3): 852-864
- Korall, P. 2003. Phylogeny of Selaginellaceae. Doctoral Thesis, Department of Botany, Stockholm University, Stockholm, Sweden
- Mossberg, B. and Stenberg, L. 2003. *Den nya nordiska floran*. Wahlström & Widstrand, Sweden: p. 38
- Ohyama, K., Fukuzawa, H., Kohchi, T., Sano, T., Sano, S., Shirai, H., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S., Inokuchi, H. and Ozeki, H. 1988. Structure and organization of *Marchantia polymorpha* chloroplast genome. 203 (2): 281-198
- “xerophyte”. A *Dictionary of Biology*. Oxford University Press. Great Clarendon Street, Oxford OX2 6DP. Fifth edition. 2004. Market House Books Ltd., 1985, 1990, 1996, 2000, 2004. ISBN-13: 978-0-19-860917-9. Printed in Great Britain by Clays Ltd. St. Ives plc.
- Rambaut, A. 2002. *Se-Al v2.0a11 Carbon*. University of Oxford, UK.
- Raven, P. H., Evert, R. F. and Eichhorn, S. E. 2003. *Biology of Plants, 6th edition*. W.H. Freeman and Company, New York, New York, USA: p. 425-442
- Rydin, C. & Wikström, N. (2002) Phylogeny of isoetes (lycopsida): Resolving basal relationships using rbcL sequences. *Taxon*, **51**, 83-89
- Staden package v1.6.0. 2005. <http://staden.sourceforge.net>
- Sugiura, C., Kobayashi, Y., Aoki, S., Sugita, C. and Sugita, M. 2003. Complete chloroplast sequence of the moss *Physcomitrella patens*: evidence for the loss and relocation of *rpoA* from the chloroplast to the nucleus. *Nucleic acid research*. 31 (18): 5324-5331
- Thomas, B. A. 1997. Upper Carboniferous herbaceous lycopsids. *Review of Palaeobotany and Palynology* 95: 129-153
- Tsuji, S., Ueda, K., Nishiyama, T., Hasebe, M., Yoshikawa, S., Konagaya, A., Nishiuchi, T. and Yamaguchi, K. 2007. The chloroplast genome from a lycophyte (microphylophyte) *Selaginella uncinata*, has a unique inversion, transposition and many gene losses. *Journal of Plant Research*. 120 (2): 281-290
- Wikström, N., Kenrick, P. 1997. Phylogeny of Lycopodiaceae (Lycopsida) and the relationships of *Phylloglossum drummondii* Kunze based on *rbcL* sequences. *International Journal of Plant Science*. 158(6): 862-871

Wolf, P. G., Rowe, C. A., Sinclair, R. B. and Hasebe, M. 2003. Complete nucleotide sequence of the chloroplast genome from a leptosporangiate fern, *Adiantum capillus-veneris* L. *DNA Research*. 10 (2): 59-65

Wolf, P. G., Karol, K. G., Mandoli, D. F., Kuehl, J., Arumuganathan, K., Ellis, M. W., Mishler, B. D., Kelch D. G., Olmstead, R. G. and Boore G. L. 2005. The first complete chloroplast genome of a lycophyte, *Huperzia lucidula* (Lycopodiaceae). *Gene*. 350 (2): 117-128