Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

Faculty of Landscape Architecture,
Horticulture and Crop Production Science

Developing species-diagnostic microsatellite markers in *Lepidium* species

Nathan Zivi

Degree Project 30 hec
Alnarp 2016

Developing species-diagnostic microsatellite markers in *Lepidium* species

Utveckling av mikrosatellitmarkörer för artbestämning inom släktet *Lepidium*

Nathan Zivi

Supervisor:	Mulatu Geleta Dida, SLU, Department of Plant Breeding
Co-Supervisors:	Cecilia Gustafsson, SLU, Department of Plant Breeding
	Stefan Andersson, Lund University, Department of Biology
Examiner:	Rodomiro Octavio Ortiz Rios, SLU, Department of Plant Breeding

Credits: 30 hec
Project level: A2E
Course title: Master's Thesis Project in Biology
Course code: EX0800
Subject: Biology

Place of publication: Alnarp
Year of publication: 2016

Publication: http://stud.epsilon.slu.se

Keywords: *Lepidium*, *Lepidium campestre*, microsatellite, SSR, SNP, domestication, breeding, molecular markers

SLU, Swedish University of Agricultural Sciences
Faculty of Landscape Architecture, Horticulture and Crop Production Science
Department of Plant Breeding

ACKNOWLEDGEMENTS

I would like to thank all the people that have supported and helped me complete this thesis. My supervisor, Mulatu Geleta, for his guidance and patience for my questions. My co-supervisor, Cecilia Gustafsson, for her help with the DNA extraction machines and when Mulatu was not available. Helle Turesson for her help with the fragment analyzer. And all the wonderful people in the lab for their help and company. I want to thank my friends, Catja Selga and Salvatore Caruso, for allowing me to vent my frustrations to them. I also want to thank my friend Elina Nordlund for being a listening ear and a helpful hand when I felt overwhelmed. And importantly, I want to thank my parents and family for their unwavering support in more ways than I can count. Without them, I would have never had the opportunity to pursue this degree.

I am so grateful to Lund University and the Swedish University of Agricultural Sciences (Sveriges Lantbruksuniversitet) for the opportunities and facilities provided. This thesis project was completed at SLU as part of a Master's degree in Plant Science at Lund University.

SUMMARY

Lepidium campestre is a potential oil and catch crop that is currently undergoing multiple studies in order to domesticate the species. Traits from four closely related species have been identified as desirable by researchers at the SLU, but more information is needed before the closely related species can be used to their fullest potential. 62 microsatellite loci were studied to develop species-diagnostic markers to facilitate introgression of the desirable traits into *L. campestre* and to better describe the relationships of the five species of *Lepidium*. Seven loci were shown to be informative between *L. campestre* and at least one of the four closely related species. Eight loci contained one or more informative SNPs/indels. Genetic diversity within *L. campestre* is estimated to be high. Diversity does not appear to group by geography. More thorough studies need to be completed to confirm this. Studies have been conducted on the relationships of some species examined in this thesis, but they are not in agreement. No conclusive data regarding these relationships was found in this thesis.

Table of contents

INTRODUCTION

The *Lepidium* genus is represented by about 230 species worldwide (Al-Shehbaz et al., 2006). It belongs to the Brassicaceae family and therefore shares common heritage with many important crop cultivars for agriculture, horticulture, and industry including *Brassica napus, Brassica oleracea, Lunaria annua,* as well as the important model organism: *Arabidopsis thaliana*. This genus is characterized by most often growing in a rosette, raceme flowering pattern with angustiseptate fruits and often mucilaginous seeds, and a chromosome basic number of x = 8 (Al-Shehbaz et al., 2006; Anonymous, n.d. a). Al-Shehbaz (1986) described several adaptations that promote autogamy including reduced flower size, scentless flowers, and anther dehiscence before bud opening, to be widespread in the genus. Species of *Lepidium* can be found on every continent except Antarctica (Al-Shehbaz et al., 2006; Anonymous, n.d. a). They often grow in fields, roadsides, and disturbed areas and at least a few *Lepidium* species are often seen as weeds in agricultural environments with *L. draba* showing allelopathic effects towards cereal crops (Haragan, 1991; Qasem, 1994; Larson et al., 2000; Anonymous, n.d. a). In this thesis, five species of *Lepidium* are examined as part of the *L. campestre* domestication and breeding program at SLU-Alnarp. The species are *L. campestre* (L.) R. Br., *L. heterophyllum* Benth., *L. hirtum* (L.) Sm., *L. draba* L. and *L. graminifolium* L. Three sub-species of *L. hirtum*, namely *L. hirtum* ssp. *atlanticum* (Ball) Maire, *L. hirtum* ssp. *calycotrichum* (Kunze) Thell., and *L. hirtum* ssp. *nebrodense* (Raf.) Thell. were included in this study.

Currently, there is a domestication and breeding project being performed at Sveriges Lantbruksuniversitet in Alnarp, Sweden. *L. campestre* was identified to be a potential oil and catch crop and was evaluated by Merker and Nilsson (1995). Desirable traits already inherent in *L. campestre* made it a good target for domestication. Some other traits were deemed deficient in *L. campestre*, but were found in four closely related *Lepidium* species. The domestication project is now working to identify microsatellite loci that will facilitate the introgression of the desirable traits from the four closely related *Lepidium* species into *L. campestre*.

OBJECTIVES

- To identify and develop species-diagnostic molecular markers for five *Lepidium* species that are being used in interspecific hybridization as part of domestication and breeding
- To further illuminate the relatedness between *Lepidium* species

BACKGROUND

Lepidium species

L. campestre is a diploid biennial plant native to Europe and Asia and has a chromosome count of 2n = 16 (Lee et al., 2002; L. campestre). It has a number of desirable traits that led it to be chosen for domestication in Sweden as an oil and catch crop. These include winter hardiness, high seed yield, growth pattern favoring undersowing, and a flower morphology that promotes self-fertilization (Merker and Nilsson, 1995; Al-Shehbaz, 1986; Eriksson, 2009; Geleta et al., 2014). *L. campestre* does have drawbacks as well that will need to be addressed before the plant is ready for agricultural production. Having a perennial character is seen as a favorable trait for *L. campestre*, so the trait needs to be found in the wild gene pool of *L. campestre* or bred into it from another species. Other traits to be improved include oil content, oil quality, seed yield, and pod shatter resistance (Eriksson, 2009; Geleta et al., 2014). The oil content and quality of *L. campestre* needs to be improved in order for this species to become economically viable to harvest and separate oils (Murphy, 1996; Nilsson et. al., 1998; Stymne, 2005). In order to use *Lepidium* as a source of food oil, anti-nutritional glucosinolates and their derivatives need to be removed during the purification process or new cultivars with acceptable levels of glucosinolates and euric acid should be developed through breeding (Andersson et al., 1999; Ivarson et al., 2016).

There are four closely related species have been identified as potential sources of desirable traits that could be introgressed to *L. campestre* through application of breeding techniques. *L. heterophyllum* is a diploid perennial plant native to Europe and has a chromosome count of 2n = 16 (Lee et al. 2002). Its perennial nature and the relative ease of crossing with *L. campestre* made it a target for the breeding program. The hybridization of *L. campestre* with *L. heterophyllum* has been successful and some hybrids have been found to be high yielding and/or perennial. The hybrids have varying seed size, which provides opportunities for developing varieties with larger seeds (Mulatu Geleta, personal communication). *L. hirtum* is a perennial plant found around the Mediterranean Sea and has a chromosome count of 2n = 16 (Anonymous, n.d. f). Perenniality is the main target trait in this species to be used in *Lepidium* domestication and breeding programs. An effort to develop viable hybrids of *L. campestre* and *L. hirtum* is underway. *L. draba*, previously known as *Cardaria draba*, is a perennial polyploid plant native to southeastern Europe

and central Asia and has a chromosome count of either 2n = 32 or 2n = 64 (Mulligan and Frankton, 1962; Bon et al., 2005; Anonymous, n.d. c). *L. draba* has traits worth integrating into *L. campestre* and traits that would best be avoided. The special interest in this species is the traits for indehiscence. This will stop seed loss, through broken seed pods, during harvest (Al-Shehbaz, 1986). Its susceptibility to frost is one trait to be avoided (Larson, 2000). *L. graminifolium* has relatively high oil content and low erucic acid content as compared to the other four *Lepidium* species included in this study (Mulatu Geleta, personal communication), therefore it can be used to increase oil content and reduce erucic acid content in *L. campestre*. It is polyploid and has a chromosome count of 2n = 48 and is native to the Mediterranean region (Brandes, 1995; Mummenhoff et al. 2001; Lee et al. 2002). Since *L. draba* and *L. graminifolium* are polyploids and are not as closely related to *L. campestre* as *L. heterophyllum* and *L. hirtum*, obtaining viable *L. campestre* x *L. draba* and *L. campestre* x *L. graminifolium* hybrids requires advanced techniques, for example embryo rescue.

History of *Lepidium* research at SLU

At the end of the 1980s, the search for a perennial oilseed crop was started at SLU. Samples of wild species were examined and eventually species from the *Lepidium* and *Barbarea* genera were chosen (Eriksson, 2009). This was followed by field trials on the selected species: *L. campestre*, *B. stricta*, *B. verna* and *B. vulgaris* examining seed yield and size, seed dormancy, winter hardiness, plant height, flowering and self-fertility (Merker and Nilsson, 1995). After this study, *B. stricta* was removed from consideration for domestication. Further analysis of oil content was done on the remaining species (Nilsson et al., 1998). The study revealed that *L. campestre* had about 20% oil by dry weight while *B. verna* and *B. vulgaris* had about 30% oil by dry weight. The oil content of *L. campestre* consisted mostly of C18:3 (linolenic acid) at 34-39% and C22:1 (erucic acid) at 22-25%, but also contained significant levels of C18:1 (oleic acid) at 12-16% and C18:2 (linoleic acid) at 8-11%. The oil content of *B. verna* was C22:1 at about 50%, C18:2 at 14%, and C18:1 at 13%. The oil content of *B. vulgaris* was 26%-32% C22:1 at 25-32%, C18:2 at about 23%, and C18:1 (oleic acid) at about 23%.

Andersson et al. (1999) evaluated the chemical composition for technical and nutritional application of these three species. They found that the oil and protein contents were lower in all

species when compared to rapeseed, and must be increased before they are agriculturally useful. They found that the amino acid composition is acceptable for human consumption but expressed concern for the high erucic acid content and presence of glucosinolates. Research on the species continued with a study on their effectiveness as catch crops undersown in barley (Börjesdotter, 1999). The plants were grown in Uppsala, Kristianstad, and Svalöv. They were all able to cope with the winter and produce seeds. Other traits studied were how seed yield and quality were affected by plant density and nitrogen supply. Börjesdotter (1999) also discovered a high propensity towards pod shattering in all three species during harvest. In the end, *L. campestre* was chosen for domestication. There are many reasons for this decision, some of which include: it has a good agronomic structure with upright flower stems branching in the upper part; it is a biennial with closely related perennials; it is more winter hardy than many other Brassicaceae species; it is resistant to the *Brassica* pest, pollen beetle, due to its small flower size; it has high seed yield and acceptable seed size; and it is self-fertile and diploid which makes the development of pure bred lines more simple (Eriksson, 2009; Börjesdotter, 1999). This does not even take into account the oil composition of *L. campestre*, which, even in its pre-domestication state, shows promise for industry, nor its qualities as a catch crop.

Eriksson (2009) described four important problems that need to be solved in order for *L. campestre* to be suitable for wide agricultural use. They are as follows: reducing or removing the tendency for pod shattering, increasing oil content, improving oil quality, and making it perennial. Recent work by several researchers on *L. campestre* has contributed to address these problems using both traditional breeding techniques and genetic modification techniques (Börjesdotter, 2000; Aronsson, 2000; Eriksson, 2009; Ivarson et al., 2013; Geleta et al., 2014; Ivarson et al., 2016).

Potential uses of *L. campestre*

The main objectives of domesticating *L. campestre* are for use as an oilseed crop and catch crop. These require detailed study on various desirable traits and their subsequent improvement through different plant breeding techniques.

Seed oil

Increasing demand for oil, growing concern over climate change, and pollution of the environment

have led many researchers to look for ways to increase vegetable oil production globally for applications in number a of areas including food, fuel and industry. Carlsson et al. (2011) hope that 40% of fossil oils will be replaced by plant oils within 20 years. Currently it is not economically feasible to replace fossil oils in many applications. This is partly due to the fact that extracting and separating oils after harvest is expensive (Princen, 1983; Stymne, 2005; Eriksson, 2009). It may be better to have plant species with seeds that are already high in oil content and desirable fatty acids to reduce costs (Murphy, 1996). This goal can be achieved with traditional breeding practices, with GM techniques, or with a combination of both. According to Carlsson et al. (2011), over-expression or silencing of genes would be an effective way to get a desirable oil content and qualities. This has been shown to be the case in other species, including cottonseed and soybean, by different research teams (Liu et. al., 2002; Pham et al., 2010). Recent research by Ivarson et al. (2016), has shown that this is also possible in *L. campestre*. Oleic acid content can be increased while erucic acid content can be decreased using RNAi silencing of genes. They were able to increase oleic acid from 11% in the wild type used to about 80% in the third generation while erucic acid content was decreased from 20.3% in the wild type to 0.1% in the third generation.

Most oilseed crops contain saturated palmitic (C16:0) and stearic (C18:0) acids, monounsaturated oleic acid (C18:1) and polyunsaturated linoleic (C18:2) and linolenic (C18:3) acids (Singh et al., 2005). *Brassica* species often contain erucic acid (C22:1) in high levels, which is unhealthy for human consumption (Mikolajczak et al., 1961). *L. campestre* also contains other unhealthy fatty acids like eicosenoic acid (C20:1), as well as glucosinolates (Nilsson et al., 1998; Andersson et al., 1999). The European Food Safety Authority (2008) recommends that the amount of glucosinolates in animal feeds should be limited to 1 – 1.5 mmol per Kg feed for monogastric animals with lower concentrations for young animals. To have a successful crop plant, there should be varieties that cater to the needs of the potential customers both by increasing the desirable fatty acids in the oil and eliminating the undesirable fatty acids and other compounds. Although erucic acid is not fit for consumption, it is used in the production of lubricants, plasticizers, slip agents, and foam suppressants (Princen, 1983). Varieties with low erucic acid and high oleic acid contents can be used for biofuel or food (Eriksson, 2009). The techniques used by Ivarson et al. (2016) can be applied to the manipulation of genes for other fatty acids as well. According to Carlsson et al.

(2011) the keys to having a plant useful to the oil industry are generally high oil content, tailored plants to have a high purity of a single fatty acid, and specialty oils not found in other oil crops. This approach will lead to creating a general and niche market that will help offset the costs of domestication of *L. campestre*.

Catch crop

Along with reduction of fossil fuel use, pollution of the environment has been a topic of discussion for decades. If our water systems and farm lands continue to be polluted, our ability to sustain the quality of life we enjoy now will be diminished. There are a couple of farming practices, especially in traditional cereal farming in temperate regions, that lead to large amounts of pollution that not only strips the land of nutrients and pollutes the water, but also costs farmers more time and money (Bergström & Brink, 1986). Leaving the soil bare for the winter and tilling of the soil leads to nitrogen runoff through mineralization of soil organic matter (Macdonald et al., 1989; Jensen, 1991; Goss et al., 1993; Hoffmann & Johnsson, 1998). When tilling occurs, more fuel needs to be used, which leads to more leaching of nitrogen, and the leaching of nitrogen causes more fertilizer to be needed to make up for the loss of nutrients (Hoffmann & Johnsson, 1998; Merker et al., 2010). A catch crop can help reduce leaching of nitrogen through a number of ways. Depending on the catch crop system used, annual sowing and tilling will be reduced which causes less fuel to be used and nitrogen lost (Merker et al., 2010). Lal et al. (2007) believe that no-till farming can help address: climate change, soil degradation, desertification, biodiversity decline, food insecurity, increase carbon storage in soil, and can help achieve better nutrient balance in the soil. There are differences in nitrogen leaching in spring vs. fall ploughing of soil when catch crops are used. Ploughing the catch crop into the soil in the spring is best for reducing nitrogen leaching (Aronsson, 2000). In the long run, ploughing tends to cause more damage by leaving the soil susceptible to crusting, compaction, and erosion (Lal et al 2007). It may be tempting to sow a completely new crop in the fall to keep the ground covered through the winter, however, spring sown catch crops do better to prevent nitrogen leaching in this case. Most likely because they have had time to establish themselves and there is no need for another round of soil cultivation (Goss et al., 1993; Aronsson, 2000).

Catch crops are established by undersowing them in a main crop. This can be done when the main

crop is being sown or at a different time. One drawback of using a catch crop is that it might compete with the main crop over nutrients. It is desirable for the catch crop to mainly take up nutrients in the fall and winter so that there will be little or no effect on the main crop (Jensen, 1991; Aronsson, 2000). Competition for nitrogen can be exacerbated by the time at which the catch crop is planted. What is best for the environment is not always the best for the farmer. Sowing both crops at the same time will reduce cost to the farmer and pollution of the environment (Känkänen & Eriksson, 2007). However there is evidence that when the crops are planted simultaneously, the main crop is at risk of poor growth (Kvist, 1992; Ohlander et al., 1996). There are two ways that the catch crop can hurt the main crop. Either by outright competition or by not releasing nitrogen back into the soil quick enough after being ploughed into the soil in the spring (Aronsson, 2000).

The best catch crop will grow quickly after the main crop harvest, but will have little to no effect on the main crop. In some studies, it has been found that catch crops can increase yield of the main crop, have no effect, or decrease yield of the main crop. Aronsson (2000) found that undersowing with perennial ryegrass had no effect on the yield of the main crop. Andersen and Olsen (1993) found that undersowing with perennial ryegrass had a small negative effect on barley yields but a considerable negative effect was caused by Italian rye. Munkholm and Hansen (2012) observed negligible negative effects by fodder radish while both ryegrass and dyer's woad showed negative effects on the main barley crop. Fortunately for *L. campestre*, it had a positive effect on the yield of barley, and also had a comparable yield to winter oilseed rape when it was undersown as a catch crop (Merker et al., 2010).

Simple Sequence Repeats (SSR; Microsatellites)

Simple Sequence Repeats (SSR; microsatellites) are a type of molecular markers based on polymerase chain reaction (PCR) that has been widely used for many different genetic studies. For example, SSRs have been used to create linkage maps for animals and plants, identification of individuals or cultivars, and population genetics studies, among others (Kalia et al., 2011). SSRs consist of two to six base-pair repeat units, e.g. (AC), (CTC), (GACG), (CGAGT) and (CGATGT). These units can be repeated a number of times because their structure makes them more prone to errors when DNA is replicated (Tautz & Renz, 1984; Mason, 2014)

The properties of SSRs, such as being highly polymorphic, abundant, located throughout the genome, and ease and reliability of amplification has made them useful in genetic studies (Cregan & Quigley, 1997; Mason, 2014). In dicots, SSRs are less likely to be in coding regions than in monocots, but in both, they are more abundant in non-coding regions (Morgante et al., 2002; Shi et al., 2013). Mutation rates on dinucleotide repeats have been studied in *Arabidopsis thaliana* and Marriage et al. (2009) discovered that as the length of the SSR increased, it was more prone to mutation, most mutations caused a gain or loss of one repeat, and the average mutation rate was 8.87×10^{-4}. Some prior knowledge of the target genome is needed in order to develop SSR primers, which is a drawback when compared to markers that do not need prior knowledge of the target genome, such as Inter Simple Sequence Repeats (ISSR). Primers are used to amplify the sequences between the primer annealing sites using PCR and then the amplified sequences must be visualized. This could be done by sequencing, gel electrophoresis and capillary electrophoresis (Cregan & Quigley, 1997; Mason, 2014). Sequence data can be especially useful in proper identification of SSR alleles as well as finding Simple Nucleotide Polymorphism (SNPs) and insertions/deletions (indels) within the SSR containing PCR products. An SSR that has been developed for one species can often be used in other closely related species. This is especially useful to understand relatedness between species. SSRs can also be used to reduce breeding time via marker assisted selection that involves inter-specific hybridization (Patel et al., 2015).

MATERIAL AND METHODS

Material

In this study, the number of accessions that represent *L. campe*stre, *L. heterophyllum*, *L. hirtum*, *L. draba*, and *L. graminifolium* were ten, three, four, two and two, respectively (Table 1). These accessions originated from wild collected plants from around Sweden and from accessions obtained from botanical gardens and seed banks around the world. Uneven sampling of species is due to the higher interest in *L. campestre*, and the desire for a quick look at the other four species in order to guide future work. Of the 10 samples *of L. campestre* used in this study, six were from Sweden, two from Greece, one from France, and the origin of one sample is unknown. Among the three *L. heterophyllum* genotypes, one was from Sweden, one sample from Spain was received from the USDA, and the third one is of unknown origin. The four samples of *L. hirtum* used in this study represent three subspecies. These subspecies are *L. hirtum* ssp. *atlanticum*, L. *hirtum* ssp. *calycotrichum* and *L. hirtum* ssp. *nebrodense*. There are two samples of *L. hirtum* ssp. *atlanticum* and one sample of the two other subspecies. All samples were received from the USDA. The *L. hirtum* ssp. *atlanticum* samples originated from Morocco, the *L. hirtum* ssp. *calycotrichum* sample originated from Spain, and the *L. hirtum* ssp. *nebrodense* sample originated from Italy. The two *L. draba* samples are from Gotland in Sweden. Information about the origin of the *L. graminifolium* received from a Denmark Botanical garden is not provided with the samples, whereas the second sample received from the USDA, was originally from Spain.

Table 1: List of samples representing the five *Lepidium* species used for sequencing and fragment analysis of the SSR loci.

Sample code	**Taxon**	**Received/collected from**	**Country of origin**	**Used for**
C-19A	*L. campestre*	Arild, Skåne	Sweden	Fragment analysis
C-52	*L. campestre*	USDA-ARS	France	Fragment analysis
C-89	*L. campestre*	IPK, Germany	Greece	Fragment analysis
C-92	*L. campestre*	IPK, Germany	Greece	Sequencing
C-124	*L. campestre*	IPK, Germany	France	Fragment analysis
Alb-6	*L. campestre*	Albrunna, Öland	Sweden	Fragment analysis
Gäv-2	*L. campestre*	Gävle, Gastrikland	Sweden	Sequencing
Mör-6	*L. campestre*	Mörbylånga, Öland	Sweden	Sequencing
Spjut-2	*L. campestre*	Spjutstorp, Skåne	Sweden	Fragment analysis
Vik-2	*L. campestre*	Viken, Skåne	Sweden	Fragment analysis
Häst-3	*L. heterophyllum*	Hästvada, Skåne	Sweden	Sequencing
H-56a	*L. heterophyllum*	USDA-ARS	Spain	Sequencing
H-56b	*L. heterophyllum*	USDA-ARS	Spain	Fragment analysis
H-98/48a	*L. heterophyllum*	Marburg Bot. garden, Germany	?	Sequencing
H-98/48b	*L. heterophyllum*	Marburg Bot. garden, Germany	?	Fragment analysis
Hi-53a	*L. hirtum* ssp. *nebrodense*	USDA-ARS	Italy	Sequencing
Hi-53b	*L. hirtum* ssp. *nebrodense*	USDA-ARS	Italy	Fragment analysis
Hi-57a	*L. hirtum* ssp. *atlanticum*	USDA-ARS	Morocco	Sequencing
Hi-57b	*L. hirtum* ssp. *atlanticum*	USDA-ARS	Morocco	Fragment analysis
Hi-58a	*L. hirtum* ssp. *calycotrichum*	USDA-ARS	Spain	Sequencing
Hi-58b	*L. hirtum* ssp. *calycotrichum*	USDA-ARS	Spain	Fragment analysis
Hi-87a	*L. hirtum* ssp. *atlanticum*	USDA-ARS	Morocco	Sequencing
Hi-87b	*L. hirtum* ssp. *atlanticum*	USDA-ARS	Morocco	Fragment analysis
Draba-1	*L. draba*	Gotland	Sweden	Sequencing
Draba-2	*L. draba*	Gotland	Sweden	Sequencing
G-27a	*L. graminifolium*	Denmark Botanic garden	?	Sequencing
G-27b	*L. graminifolium*	Denmark Botanic garden	?	Fragment analysis
G-55a	*L. graminifolium*	USDA-ARS	Spain	Sequencing
G-55b	*L. graminifolium*	USDA-ARS	Spain	Fragment analysis

Growing plants and DNA extraction

Seeds representing each sample were planted on plastic trays. Plants were sampled when they reached 2-3 weeks old for DNA extraction. Clean, disease free, young leaves were placed into 2 ml eppendorf tubes containing two glass beads and immediately placed in liquid nitrogen. The tubes were then kept at -80°C until the DNA was extracted. Throughout this study, DNA, lab equipment, and chemicals were handled with care to avoid cross contamination. Good laboratory practices were followed to reduce the likelihood of mistakes and accidents.

Machine extraction: The Qiacube robot was used for DNA extraction. All components of the DNA extraction buffer (0.1 M Tris, 20 mM EDTA, 1.4 M NaCl, and CTAB) were mixed and then held at 57°C for 30 minutes. After cooling down to 20°C, the pH was adjusted to 7.5 and the volume was adjusted to the final volume (depending on how much buffer was needed) using MilliQ-H_2O, NaOH, and HCl and again held at 57°C to await use. Frozen tubes containing sampled leaf tissue were placed into pre-cooled Retsch tube holder. The sample containing tube-holder was immersed in liquid nitrogen and kept there until it stopped fizzing. The leaf tissue was grinded for 90 seconds at 30 Hz using a Retsch MM400 shaker. The samples were refrozen in liquid nitrogen and ground again. One milliliter of pre-heated CTAB was added to each sample. The samples were incubated at 52°C for one hour. The samples were inverted a few times during incubation to ensure mixing of the samples with buffer. After incubation, the samples were centrifuged at 13200 rpm for 15 minutes. 200 µl of supernatant was loaded onto the Qiagen Robot sample plate, and robotic DNA extraction was conducted. The extracted DNA was allowed to solubilize overnight at 4°C. One microliter of RNase A (Qiagen 10 mg/ml) was added per 100 µl sample, and incubated and gently shaken at 37°C for one hour. DNA purity and concentration was measured on NanoDrop and test ran on a 1% agarose gel (1X TAE) containing GelRed.

Manual extraction: Manual extraction was done on DNA samples that failed or were of poor quality when the above Qiacube based procedure was followed. Extraction buffer was made in the same was as previously stated. A solution of chloroform and isoamylalcohol was prepared at a 24:1 ratio. A washing buffer was made (76% ethanol, 10mM NH_4Ac), and adjusted to the final volume using MilliQ-H_2O. Leaf tissue was grinded in a Retsch tube holder and incubated in CTAB as described above. 500 µl of supernatant was transferred into a new tube. 500 µl of

chloroform/isoamylalcohol solution was added and spun for 3 minutes at 12000 rpm. 400 µl of the top phase was transferred into a new tube containing 400 µl of ice cold isopropanol, inverted several times, and spun for three minutes at 12000 rpm. The supernatant was poured off and 500 µl of washing buffer was added. It was left to sit for two minutes before inverting and spinning for 3 minutes at 12000 rpm. The supernatant was discarded again and the tube re-spun. The remaining liquid was pipetted away. The pellet was dried and then re-suspended in 50 µl of TE buffer. The DNA was allowed to solubilize overnight at 4°C. One microliter of RNase A (Qiagen 10 mg/ml) was added per 100 µl sample, and incubated and gently shaken at 37°C for one hour. DNA purity and concentration was measured using NanoDrop (ND-1000) spectrophotometer and also ran on a 1% agarose gel (1X TAE) containing GelRed. One microliter of DNA sample was loaded onto the measurement pedestal of NanoDrop to determine the DNA quality and concentration. DNA samples with a 260/280 value below 1.5 were discarded and DNA was re-extracted from a fresh leaf tissue sampled from the same plant.

New SSR Primer-pairs (unpublished) developed based on Restriction site Associated DNA Sequencing (RAD-sequencing) within the ongoing Mistra-Biotech research program (http://www.slu.se/mistrabiotech) were used for genotyping (Table 2). In total, 62 new SSR primer-pairs (Table 2) developed based on DNA sequences were screened for selective amplification of target SSR loci (without amplification of unintended targets) across all samples of *L. campestre*.

Table 2: List of 62 SSR loci initially tested, their repeat motifs (based on a consensus sequence), and expected fragment size. **Bold** loci were chosen for further study.

SSR loci	Repeat Motif	Expected fragment size	SSR loci	Repeat Motif	Expected fragment size
Lep-021	**(AT)9**	**234**	Lep-052	(AT)6	200
Lep-022	**(AGTG)4**	**176**	**Lep-053**	**(AG)7**	**293**
Lep-023	(AT)6	171	Lep-054	(TA)9	237
Lep-024	(ATTT)4	384	**Lep-055**	**(GTT)6**	**282**
Lep-025	(TAAT)4	187	Lep-056	(TGTAA)5	398
Lep-026	(AT)6	306	Lep-057	(TA)8	301
Lep-027	**(GAGATG)4**	**212**	**Lep-058**	**(CTT)8**	**199**
Lep-028	**(TTTC)4**	**309**	Lep-059	(AAAG)4	187
Lep-029	(CT)6	376	Lep-060	(GA)8	247
Lep-030	(TATGT)6	109	Lep-061	(TTC)8	316
Lep-031	**(AG)10**	**320**	Lep-062	(AT)8	222
Lep-032	(AT)6	127	Lep-063	(TA)8	252
Lep-033	**(AT)7**	**164**	Lep-064	(AT)6	180
Lep-034	(AG)10	164	Lep-065	(CAA)6	232
Lep-035	(TA)8	189	Lep-066	(TCT)6	338
Lep-036	(ATC)6	327	Lep-067	(TA)9	222
Lep-037	(AT)6	138	Lep-068	(ATATAG)6	124
Lep-038	**(GT)8**	**228**	**Lep-069**	**(AT)7**	**346**
Lep-039	(CTT)6	143	Lep-070	(TATT)4	175
Lep-040	**(AAAG)6**	**311**	**Lep-071**	**(AT)7**	**331**
Lep-041	(AAG)8	296	Lep-072	(TATTT)4	211
Lep-042	**(AT)6**	**252**	**Lep-073**	**TATTT**[a]	**307**
Lep-043	**(AT)7**	**177**	Lep-074	(GA)7	294
Lep-044	**(AG)7**	**182**	Lep-075	(GA)7	210
Lep-045	**(AT)8**	**383**	**Lep-076**	**(TC)13**	**257**
Lep-046	(GAA)6	115	**Lep-077**	**(TC)6**	**284**
Lep-047	**(AAAAC)4**	**243**	Lep-078	(AT)6	169
Lep-048	**(GT)7**	**119**	**Lep-079**	**(CT)8**	**373**
Lep-049	**(GAAAGA)6**	**151**	Lep-080	(TC)8	164
Lep-050	**(TA)9**	**197**	**Lep-081**	**(TA)7**	**376**
Lep-051	(AT)6	207	Lep-082	(AAATA)4	296

Note: [a] = No consensus sequence found for *Lep-73*, therefore number of repeats is unknown.

SSR PCR

Out of these 62 primers, 26 (42%) were deemed acceptable for analysis using samples from all five species included in this study (Table 2). SSR data based on the 26 SSR loci was generated by sequencing the PCR products and by performing capillary electrophoresis to determine fragment sizes. Sequencing of the amplified products was conducted to confirm that they match the original sequence used to design the SSR primer-pairs. 29 samples were used in total between the two methods; 14 samples for sequencing and 15 samples for fragment analysis. The 14 samples used for sequencing were three *L. campestre*, three *L. heterophyllum*, four *L. hirtum*, two *L. draba*, and two *L. graminifolium*. The 15 samples used for fragment analysis were seven *L. campestre*, two *L. heterophyllum*, four *L. hirtum*, and two *L. graminifolium*. There was no overlap in samples used following the two techniques. A compiled list of samples can be seen in Table 1 and the primer-pairs used for final analysis of the five species are given in bold font in Table 2.

A master mix was prepared on ice. The master mix consisted of 16.5 µl of ddH_2O, 2.5 µl of 10x PCR buffer, 1.5 µl of 25 mM $MgCl_2$, 0.3 µl of 25 mM dNTPs, 0.75 µl of 10 µM forward primer, 0.75 µl of 10 µM reverse primer, and 0.25 U/µl Taq polymerase. This comes to 22.5 µl per sample. All components were pipetted into 1.5 ml eppendorf tubes. Then, 2.5 µl of DNA (10ng/µl) and 22.5 µl master mix were pipetted separately into the same PCR tube. PCR was completed for primers with a melting temperature above 62°C as follows: initial denaturation at 95°C for 3 minutes, 95°C for 30 seconds to denature the DNA, 57°C for 30 seconds so the primers anneal to the DNA, 72°C for 45 seconds for the Taq polymerase to elongate the DNA (the last three steps were repeated 40 times in total). This was followed with the samples being kept at 72°C for 20 minutes to allow the complete extension of shorter fragments, and then down to 4°C until the samples were withdrawn from the machine. For primers with a melting temperature near or below 59°C, the same temperature profile was used except that the annealing temperature was 53°C. PCR products were stored at 4°C until gel electrophoresis.

M13 primer SSR PCR

A master mix was prepared on ice. The master mix consisted of 16.3 µl of ddH_2O, 2.5 µl of 10x PCR buffer, 1.5 µl of 25 mM $MgCl_2$, 0.3 µl of 25 mM dNTPs, 0.2 µl of 10 µM forward primer, 0.75 µl of 10 µM reverse primer, 0.75 µl of 10 µM universal fluorescent-labeled M13 primer, and

0.2 5U/µl Taq polymerase. This comes to 22.5 µl per sample. All components were pipetted into 1.5 ml eppendorf tubes. 2.5 µl of DNA (10ng/µl) sample and 22.5 µl master mix were pipetted separately into the same PCR tube. PCR was performed as follows for primers with a melting temperature above 62°C: initial denaturation at 95°C for 15 minutes, 94°C for 30 seconds to denature the DNA, 57°C for 90 seconds so the primers anneal to the DNA, 72°C for 60 seconds for the Taq polymerase to elongate the DNA (the last three steps were repeated 35 times in total), then 8 times of 94°C for 30 seconds, 53°C for 45 seconds, 72°C for 45 seconds, then extension at 60°C for 20 minutes, and then down to 4°C until the samples were withdrawn from the machine. For primers with a melting temperature near or below 59°C, initial denaturation at 95°C for 15 minutes, 94°C for 30 seconds to denature the DNA, 53°C for 90 seconds so the primers anneal to the DNA, 72°C for 60 seconds for the Taq polymerase to elongate the DNA (the last three steps were repeated 35 times in total), then 8 times of 94°C for 30 seconds, 53°C for 45 seconds, 72°C for 45 seconds, then extension at 60°C for 20 minutes, and then down to 4°C until the samples were withdrawn from the machine. PCR product was stored at 4°C until it was run on a gel.

Gel electrophoresis

Extracted DNA and SSR PCR products were run on 1% and 1.5% agarose gels (1X TAE), respectively, containing GelRed. Gels were loaded with PCR products containing 1x loading buffer. Three microliters of 50 bp DNA ladder was used to estimate sizes of PCR products and a 100 bp ladder was used for genomic DNA. Gels were visualized under UV light.

PCR product purification

The QIAquick® PCR Purification Kit was used to purify PCR products used for DNA sequencing according to the instruction provided with the kit. Five volumes of DNA binding buffer (buffer PB) were added to one volume of PCR product and mixed. If the color of the mixture was orange or violet, 10 µl of 3 M sodium acetate, at pH 5.0, was added and mixed so that the mixture turned yellow. The solution was pipetted into a QIAquick column (provided) that sits on a 2 ml collection tube (provided) and was centrifuged for 60 seconds at 13200 rpm to bind the DNA to the membrane. The flow-through was discarded and the column was placed back onto the collection tube. 750 µl of washing buffer (buffer PE) was added to the column and spun for 60 seconds at 13200 rpm to wash the DNA. The flow-through was discarded and the column was placed back

onto the collection tube. The column was centrifuged for a final time before discarding the collection tube and placing the column in a new 1.5 ml tube. 50 µl of elusion buffer (buffer EB) was added to the column and was spun for 60 seconds at 13200 rpm. The column was discarded and the now purified PCR product was kept at -20°C. Quality of the purified product was confirmed via a 1.5% agarose gel electrophoresis.

DNA sequencing

DNA sequencing was performed by Eurofins Scientific in Germany. Purified PCR product was prepared using the submission guidelines for Eurofins (Eurofins). The concentration and volume per sample was about 5 ng/µl and 15 µl, respectively.

Capillary electrophoresis

Capillary electrophoresis was done using a 3500 Series Genetic Analyzerfrom ThermoFisher at SLU, Department of Plant Breeding. Samples were prepared for capillary electrophoresis following 96-well plate format. Before the PCR products were transferred to PCR plates for capillary electrophoresis, about 5 µl of each PCR product were run on agarose gels to rate the concentration based on intensity of the band using scales 1 to 5. This was done so that similar and optimum concentrations of PCR products were used for the analysis. Those PCR products that were rated 4 or 5 were diluted 12.5x whereas those rated 3 were diluted 7x. Similarly, those that were rated 1 or 2 were diluted 5x using ddH_2O water.

Data generated through capillary electrophoresis was analyzed using GeneMarker V2.4.0 software (SoftGenetics, State College, Pennsylvania, USA), which include peak identification and determination of the size of fragments. The size of the fragments was determined based on GeneScan 600 internal size standard. The SSR loci with lots of missing data or low quality data were excluded from further analysis. Data imputation was not performed because it is beyond the purpose of this thesis and because of the study's small sample size.

Data analysis

Sequencing data received from Eurofins was preliminarily aligned using ClustalX2 version 2.0 (Larkin et al., 2007). Then the data were further organized using BIOEDIT's sequence alignment

editor version 7.2.5 (Hall 1999). Sequences that did not match the consensus sequence were removed. Only the sequences that contained the full SSR were used to identify informative loci or for cluster analysis. SNPs/indels were searched for using BIOEDIT's sequence alignment editor, but not all could be identified due to time constraints. Cluster analysis was done using the program FreeTree version 0.9.1.50 downloadable online (Pavlicek et al., 1999). Dendrograms were generated using Unweighted Pair Group Method with Arithmetic Mean (UPGMA), the Jaccard similarity coefficient, and bootstrapping with 10,000 resamples. The reference dendrograms with bootstrap values that were generated were visualized using TreeView X version 0.5.0 downloadable online (Page, 2005).

RESULTS

Sequence data

Out of a total of 293 purified PCR products sent to Eurofins for sequencing, 206 of them contained the full length of their target microsatellites, and a sequence amplified by *Lep-21* primer-pairs contained two microsatellites. Others gave partial sequences of the target PCR products that either did not contain a full length of the microsatellite sequences or completely missed the microsatellites. The microsatellites that were not captured by sequencing were located close to the forward primers' annealing sites, and consequently trimmed off after sequencing due to low sequence quality. Of the 26 primer-pairs, three (*Lep-42*, *Lep-58*, *Lep-69*) failed to amplify their targets, five (*Lep-31*, *Lep-43*, *Lep-47*, *Lep-49*, and *Lep-73*) were removed due to missing or partial sequence data for some samples whereas two microsatellites were found adjacent to each other in sequences amplified by *Lep-21*. This leaves a total of 19 loci to examine. Further, three primer-pairs (*Lep-55*, *Lep-73*, and *Lep-76*) did not amplify the expected SSR. The matching *L. campestre* consensus SSR sequence was found on record for *Lep-55* and *Lep-76*, but not for *Lep-73*, and all three were included in this study. Within *L. campestre*, data was missing for one locus, and 39% of the remaining loci (7 out of 18 loci) were polymorphic. *L. heterophyllum* had missing data at six loci for at least one sample. 54% of the remaining loci (7 out of 13 loci) were polymorphic among the three samples. Of the loci that had data for at least two samples in *L. hirtum*, 94% (16 out of 17 loci) were polymorphic. Interestingly, 57% of the loci that have complete data (8 out of 14) were polymorphic among the two samples representing *L. hirtum* ssp. *atlanticum*. Among the eight loci that had data for both *L. draba* samples, only two (25%) were polymorphic. In the eight loci that had data for both *L. graminifolium* samples, only one (12.5%) was polymorphic (Table 3). Transferability of loci within sequence data between *L. campestre* and the other species was examined. The analysis suggests that all *L. campestre* SSR loci are transferable to *L. heterophyllum*. In *L. hirtum*, only one SSR locus (*Lep-28*) appears to be absent, therefore, 95% of *L. campestre* SSR loci are transferable to *L. hirtum*. Only 53% of the loci were amplified in *L. draba* suggesting medium level transferability of SSR loci between *L. campestre* and *L. draba*. The transferability of the SSR loci to *L. graminifolium* was 50% suggesting medium level transferability. These percentages are for sequence data only.

There were 3 primer-pairs (*Lep-40*, *Lep-44*, and *Lep-77*) that gave data for all 14 *L. campestre* samples. Comparisons between species using complete data could be made across 12 loci between *L. campestre* and *L. heterophyllum*, across 16 loci between *L. campestre* and *L. hirtum*, across 8 loci between *L. campestre* and *L. draba*, across 7 loci between *L. campestre* and *L. graminifolium*, and across 11 loci between *L. campestre, L. heterophyllum*, and *L. hirtum*.

Table 3: List of the 19 SSR loci, length of their repeat motifs in the original consensus Rad-sequences and across the 14 *Lepidium* samples sequenced.

		L. campestre			*L. heterophyllum*			*L. hirtum*				*L. draba*		*L. graminifolium*	
SSR Loci	Repeat motifs	C-92	Gäv-2	Mör-6	H-56	H-98/48a	Häst-3	Hi-53a	Hi-57a	Hi-58a	Hi-87a	Dra-1	Dra-2	G-27a	G-55a
Lep-21a	(AT)9	10	10	10	6	6	7	*	16	*	*	-	-	-	-
Lep-21b	(GT)4	4	4	4	8	4	4	3	5	3	3	-	-	-	-
Lep-22	(AGTG)4	4	4	4	2	2	2	4	5	4	*	-	-	-	-
Lep-27	(GAGATG)4	2	2	3	*	7	7	1	3	*	2	-[a]	-[a]	-	-
Lep-28	(TTTC)4	4	4	4	*	4	4	-	-	-	-	-	5	-	-
Lep-33	(AT)7	7	7	7	7	6	6	10	*	7	7	-[a]	-[a]	-	-
Lep-38	(GT)8	8	8	-	7	8	8	7	7	8	7	-[a]	-[a]	8	8
Lep-40	(AAAG)6	6	6	6	5	5	5	5	6	6	6	3	3	3	3
Lep-44	(AG)7	7	7	7	8	9	9	6	22	10	6	4	4	7	7
Lep-45	(AT)8	8	8	8	8	8	*	7	8	8	8	8	8	8	8
Lep-48	(GT)7	5	5	5	7	*	7	5	9	5	6	5	5	5	5
Lep-50	(TA)9	9	9	9	*	5	5	6	6	6	6	-[a]	-[a]	6	6
Lep-53	(AG)7	7	7	7	7	7	7	8	9	8	11	-[b]	-[b]	-[a]	-[a]
Lep-55	(GTT)6	6	5	5	*	5	*	6	5	5	5	5	6	5	*
Lep-71	(AT)7	8	8	8	7	7	8	10	11	5	10	10	2	-	-
Lep-76	(TC)13	9	*	*	12	*	9	-[a]	*	10	14	-[a]	-[a]	-[a]	*
Lep-77	(TC)6	6	6	6	6	6	6	6	6	6	5	6	6	6	6
Lep-79	(CT)8	8	8	8	8	8	8	9	11	8	8	9	9	*	4
Lep-81	(TA)7	8	8	7	5	5	5	9	8	8	8	5	*	5	8
% polymorphic loci		39			54			94				25		12.5	

[a] = gels showed good banding pattern, but sequencing and capillary electrophoresis failed or (regarding *L. draba*) was not done

[b] = Two bands, not sent for sequencing

* = either sequencing failed, but capillary electrophoresis succeeded, or only partial microsatellite sequence was obtained and hence the full length could not be determined

- = the SSR locus provided no alleles and is considered absent

Sequences containing SNPs and indels

The sequencing data also provided some informative SNPs and indels (Table 4). Most of these are species-diagnostic while some are only diagnostic to individuals, but could be important to note during breeding when individuals with such mutations are used.

Table 4: List of potential species-diagnostic SNPs/INDELs identified in the sequences amplified by using SSR primer-pairs.

Primer	Marker type	Location[a]	C	H	Hi-neb	Hi-alt	Hi-cal	D	G
Lep-31	SNP	+56	C	A*	*	N/A*	N/A	A	A*
	SNP	+57	T	A*	*	N/A*	N/A	C	C*
	INDEL	+136	G	-*	*	*	N/A	-	-*
Lep-43	SNP	+78	G	*	C	C	C	C	N/A*
Lep-44	SNP	-65	A	A	A	A	A	G	A
	SNP	-59	A	A	A	A	A	T	A
	SNP/INDEL	-38	C	G	C	C	-	-	C
	INDEL	-57	--	--	--	--	--	CG	--
	INDEL[b]	-53	+45	+45	+45	+45	-45	-45	--
	INDEL	-8	--------	AGGTGAAT	AGGTGAAT	AGGTGAAT	AGGTGAAT	AGGTGAAT	--------
	INDEL	+14	---	GAG	---	---	---	GAG	---
Lep-49	SNP	+100	A	G	A	A*	A	N/A	A
Lep-50	SNP	+9	A	A*	C	C	C	N/A	C
	SNP	-11	T	A	A	A*	A	N/A	A
	INDEL	-1	--	TG*	--	--	--	N/A	--
	INDEL	+38	-	A*	-	-	-	N/A	-
Lep-71	SNP	-68	G	T	G	G	G	G	N/A
	SNP	-2	T	A	A	A	A	A	N/A
Lep-73	SNP	+36[c]	G	A	N/A	*	*	*	*
Lep-79	SNP	-205	G	A	G	G*	G	G	*
	SNP	-191	T	A	A	A	A	A	*
	SNP	-63	C	C	C	C	C	T	C*
	SNP	-12	T	C	C	C	C	C	C*
	INDEL	+52	-	T	-	-	-	N/A	-*

[a] = "+" refers to downstream and "–" refers to upstream from the first nucleotide (5'-end) position of the *L. campestre* consensus sequence microsatellite. * = missing data. N/A = No amplification. C is *L. campestre*, H is *L. heterophyllum*, Hi-neb is *L. hirtum* ssp. *nebrodense*, Hi-atl is *L. hirtum* ssp. *atlanticum*, Hi-cal is *L. hirtum* ssp. *calycotrichum*, D is *L. draba*, G is *L. graminifolium*.

[b] = "+45" represents "AGAGAGAGAAATTAACTAAGGAGAGGTAAGGAGGAATAGGTGAAT and "-45" is absence of this sequence.

[c] = There is no *L. campestre* consensus sequence found for *Lep-73*

Fragment size data

Only 15 *Lepidium* samples with seven, two, two, and four samples representing *L. campestre*, *L. graminifolium*, *L. heterophyllum*, and *L. hirtum*, respectively, were used for fragment analysis. Fragment size data was generated by capillary electrophoresis. Of the 26 loci, amplification of one (Lep-49) failed, and seven (*Lep-22*, *Lep-27*, *Lep-38*, *Lep-40*, *Lep-45*, and *Lep-48*, *Lep-73*) were removed due to missing or low quality data. Good quality data was generated for a total of 18 SSR loci. Of these loci, *L. campestre* samples had data missing in four loci (including *Lep-71* which had amplified previously for sequencing, but failed in the case of capillary electrophoresis) and it was polymorphic in 71% of the remaining loci (10 out of 14 loci). *L. heterophyllum* samples had data missing in six loci and it was polymorphic in 67% of the remaining loci (8 out of 12 loci). *L. hirtum* samples had data missing in eight loci and it was polymorphic in all remaining loci. Like in the sequence data, *L. hirtum* ssp. *atlanticum* showed a high level of polymorphism. Six of the seven loci that had complete data showed polymorphism between the samples, while in the last sample, both samples are identically heterozygous which indicate the presence of more than one alleles in all seen loci of *L. hirtum* ssp. *atlanticum*. *L. graminifolium* samples had data missing in 12 loci while one locus only had one sample, but with two alleles. It was polymorphic in 67% of the remaining loci (4 out of 6 loci). Transferability of loci within fragment size data between *L. campestre* and the other species was examined. This analysis suggests a relatively high transferability (86%) between *L. campestre* and *L. heterophyllum*. The transferability of the SSR loci between *L. campestre* and *L. hirtum* is relatively high at 80%. In *L. graminifolium*, the transferability of *L. campestre* SSR loci is 33% suggesting a low level of transferability. Analysis was also done using partially complete data. Using this data, comparisons could be done using 12 loci within *L. campestre*, 11 loci between *L. campestre* and *L. heterophyllum*, in 8 loci between *L. campestre* and *L. hirtum*, in 5 loci between *L. campestre* and *L. graminifolium*, and in 7 loci between *L. campestre*, *L. heterophyllum*, and *L. hirtum*. Fragment size data was organized by fragment length as seen in Table 5 below.

Table 5: Fragment size variation among 15 *Lepidium* samples, with seven, two, two and four samples representing *L. campestre*, *L. hirtum*, *L. heterophyllum*, and *L. graminifolium* from left to right in that order.

		L. campestre							*L. heterophyllum*		*L. hirtum*				*L. graminifolium*	
Primer	**Repeat Motif**	**C-19A**	**C-52**	**C-89**	**C-124**	**Alb-6**	**Spjut-2**	**Vik-2**	**H-56b**	**H-98/48b**	**Hi-53b**	**Hi-57b**	**Hi-58b**	**Hi-87b**	**G-27b**	**G-55b**
Lep-21	(AT)9	252	252	*	252	252	252	*	252	244	238	264/6	238	240	-	-
Lep-28	(TTTC)4	325	325	325	325	325	325	325	325	327	-[a]	-[a]	-[a]	-[a]	-[a]	-[a]
Lep-31	(AG)10	330	334	330	329	330	330	337	330	321/31	321	*	321/30	320	321/30	321/30
Lep-33	(AT)7	188	188	188	188	188	188	188	188	185	184/90	192/9	183	183	-	-
Lep-42	(AT)6	260	260	261/65	265/9	262/9	-[a]	-[a]	267	265	-[a]	-[a]	-[a]	-[a]	-[a]	-[a]
Lep-43	(AT)7	204	204	193	193	204	204	193	-	-	-	-	-	-	-	-
Lep-44	(AG)7	195	195	195/7	195/7	197	197	197	187	189	*	*	*	186	*	*
Lep-47	(AAAAC)4	256	251	256	256	256	256	256	251	251	251	251	251	-[a]	-[a]	-[a]
Lep-50	(TA)9	196; 213	196; 213	196; 213	196; 213	196; 213	196; 213	196; 213	*	*	*	*	*	*	196; 213	196; 213
Lep-53	(AG)7	307	307	292	307	307	307	292	292	292	*	*	*	*	-[a]	-[a]
Lep-55	(GTT)6	300	300	300	300	300	300	*	*	*	300	297	*	*	*	*
Lep-58	(CTT)8	205	205	205	205	205/14	205/14	205/14	-[a]	-[a]	-[a]	-[a]	-[a]	209	205	205
Lep-69	(AT)7	362	362	376	362	-[a]	-[a]	-[a]	354/9	-[a]	350	360	350	-[a]	-	-
Lep-71	(AT)7	+	+	+	+	+	+	+	350	350	356	358	323/57	356/58	-	-
Lep-76	(TC)13	265	265	265	265/73	265/73	265/73	265	266	274	-[a]	247/57	242	247/57	-[a]	247/57
Lep-77	(TC)6	300	300	300	300	300	300	300	*	*	297	300	297	297	*	*
Lep-79	(CT)8	389	389	389	389	389	389	*	388/91	388	*	395	389	389	383/6	383/6
Lep-81	(TA)7	393	393	393	393	393	393	393	389	389	397/7	395	405	397	386	386
% polymorphic loci		71							67		100				60	

[a] = gels showed good banding pattern, but both sequencing and capillary electrophoresis failed

* = capillary electrophoresis failed, but sequencing was successful

+ = *L. campestre* failed to amplify PCR product for capillary electrophoresis, but was found present in the sequencing data

- = the SSR locus provided no alleles and is considered absent

Cluster analysis

Both sequence data and fragment size data were used for cluster analysis. Cluster analysis was conducted based on Jaccard similarity coefficient, and bootstrapping was done with 10,000 resamples. The reference dendrograms, supported by bootstrap values, are presented below (Figures 1, 2, 3, 4). Cluster analysis based on the sequence data was completed for comparison between *L. campestre* and *L. heterophyllum*, *L. campestre* and *L. hirtum*, *L. campestre* and *L. draba*, *L. campestre* and *L. graminifolium*, and *L. campestre*, *L. heterophyllum*, and *L. hirtum*. For fragment size data, similar comparisons were made, except for *L. draba* which was not used. Dendrograms were generated within *L. campestre*, between *L. campestre* and *L. heterophyllum*, between *L. campestre* and *L. hirtum*, and between *L. campestre*, *L. heterophyllum*, and *L. hirtum* are given below.

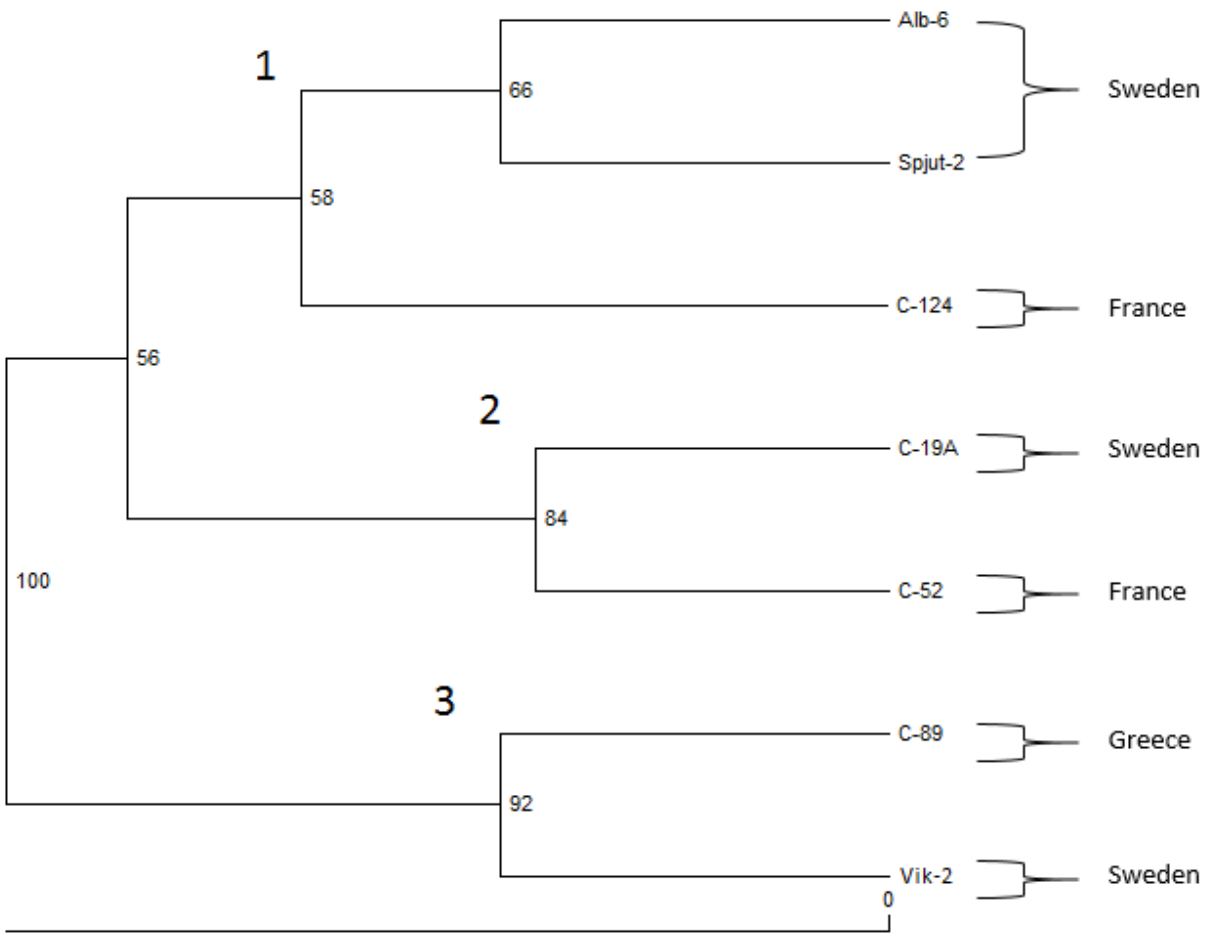

Figure 1: UPGMA (Unweighted Pair Group Method with Arithmetic Mean) dendrogram depicting relationship between seven samples of *L. campestre* based on fragment size data. There are three main groups. In group 1, there are two samples from Sweden and one from France. In group 2, there is one sample from both Sweden and France. In group 3, there is one sample from both Greece and Sweden.

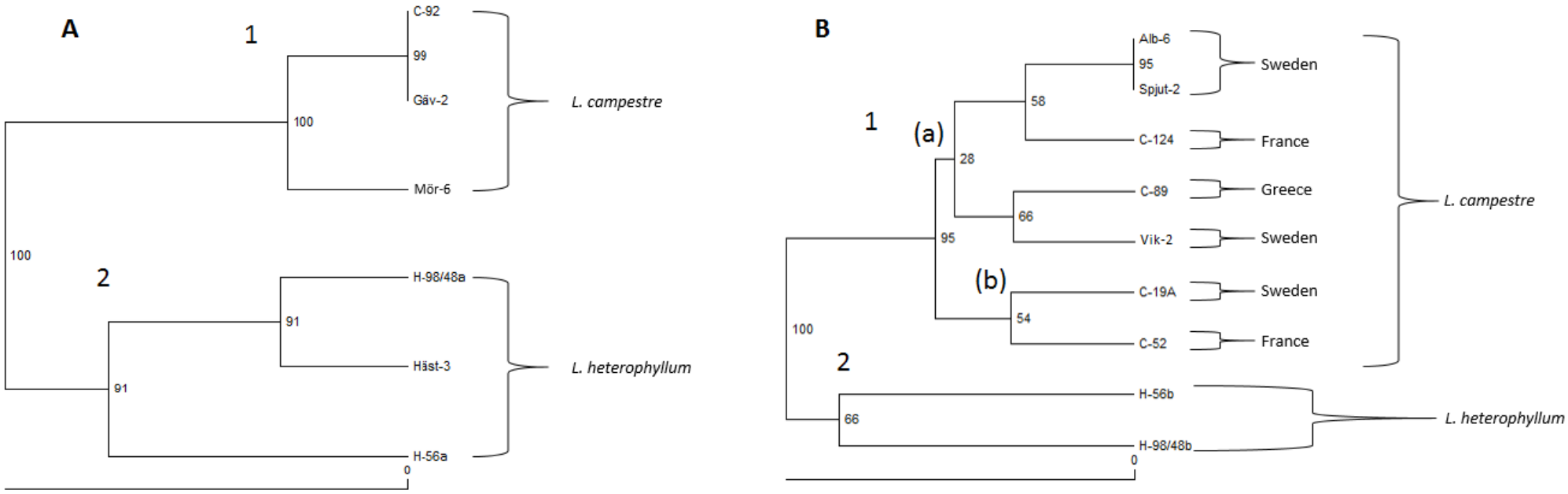

Figure 2: UPGMA dendrograms depicting relationship between *L. campestre* and *L. heterophyllum* based on (A) sequence data and (B) fragment size data. In dendrogram A, there are two main groups. They are separated by species. Group 1 is *L. campestre* and group 2 is *l. heterophyllum*. In dendrogram B there two main groups, with group 1 having two sub-groups. Group 1 contains all *L. campestre* and group 2 contains all *L. heterophyllum*. In the sub-groups of group 1, (a) contains five samples. Three from Sweden, one from France, and one from Greece. Sub-group (b) contains two samples, one from both Sweden and France.

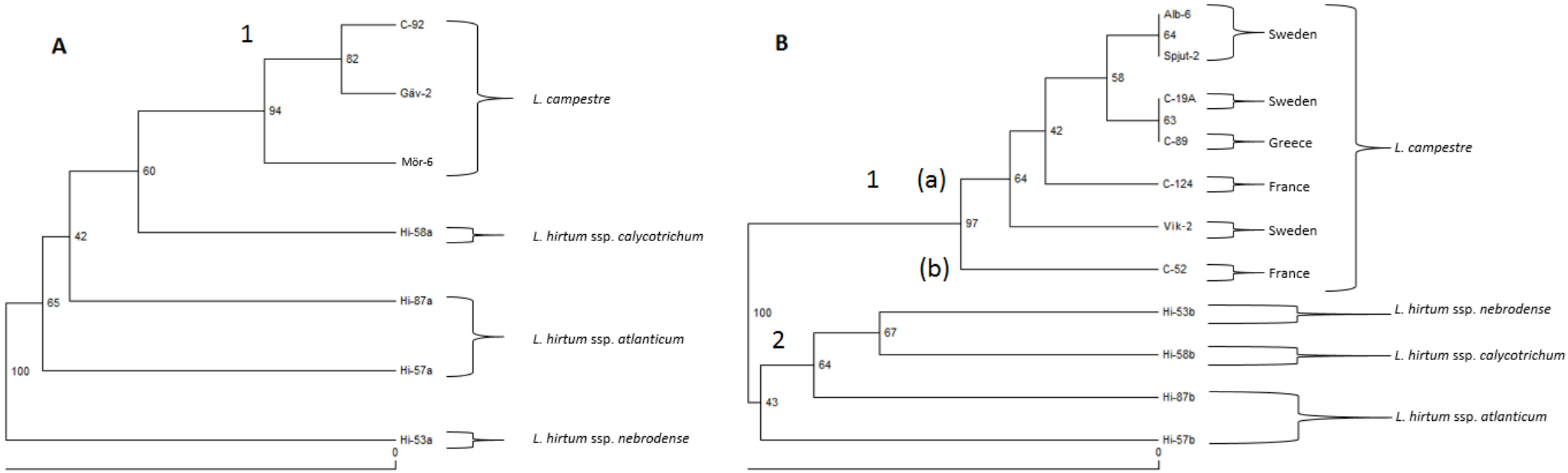

Figure 3: UPGMA dendrograms depicting relationship between *L. campestre* and *L. hirtum* based on (A) sequence data and (B) fragment size data. In dendrogram A, there is one group separating *L. campestre* from *L. hirtum*, but there is no grouping with regards to *L. hirtum* subspecies. In dendrogram B, there are two main groups. Group 1 contains *L. campestre* and group 2 contains *L. hirtum*. Sub-grouping in *L. campestre* shows 2 groups. Sub-group (a) contains six samples, four from Sweden, one from Greece, and one from France. Sub-group (b) only contains one sample from France.

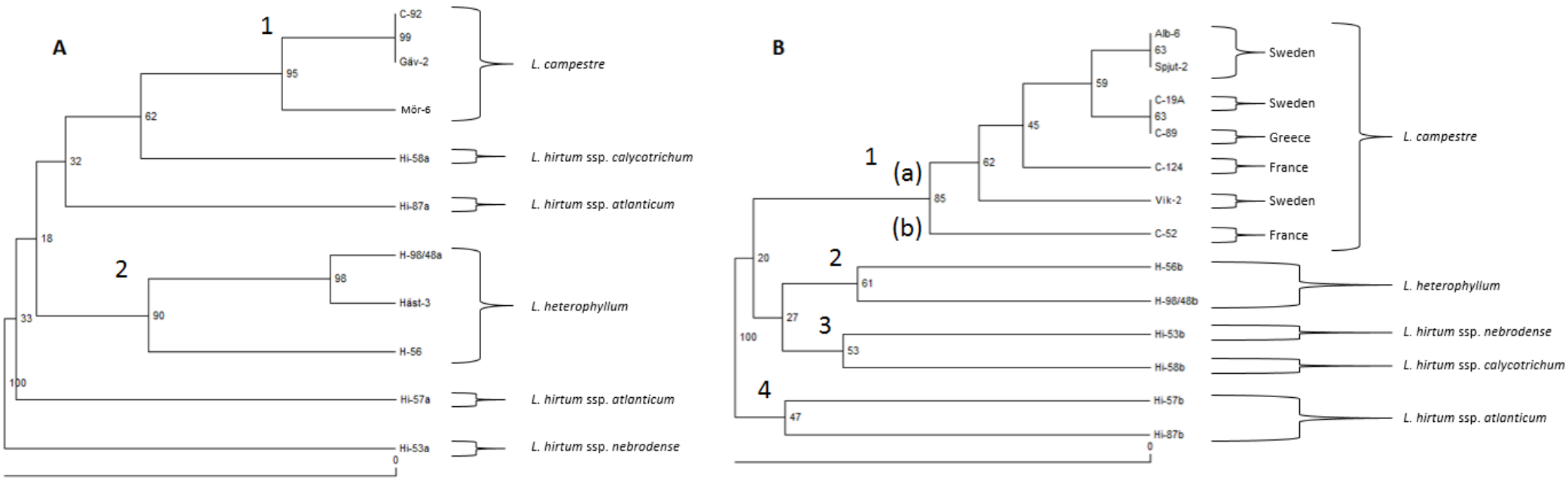

Figure 4: UPGMA dendrograms depicting relationship between *L. campestre*, *L. heterophyllum* and *L. hirtum* based on (A) sequence data and (B) fragment size data. Dendrogram A contains two main groups. Group 1 contains *L. campestre* and group 2 contains *L. heterophyllum*. *L. hirtum* samples are not strongly grouped with any sample, except *L. hirtum* ssp. *nebrodense* which is separated from all other samples. Dendrogram B has four groups. Group 1 contains *L. campestre* and two sub-groups. Sub-group (a) contains six samples, four from Sweden, one from Greece, and one from France, while sub-group (b) only contains one sample from France. Group 2 contains *L. heterophyllum*. Group 3 contains *L. hirtum* ssp. *nebrodense* and *L. hirtum* ssp. *calycotrichum*. Group 4 contains both samples of *L. hirtum* ssp. *atlanticum*.

DISCUSSION

Informative loci

In the DNA sequence based SSR analysis, there were seven loci that are potentially informative for breeding when comparing *L. campestre* to other species (Table 6). Loci that contain useful genetic variation between species while being monomorphic within a species are informative for breeding. Table 6 shows the loci in which *L. campestre* had monomorphic SSR alleles (within the consensus sequence and samples examined) while the other *Lepidium* species examined had SSR alleles that were differently monomorphic or polymorphic with no overlap. Three of the loci (*Lep-22*, *Lep-40*, and *Lep-44*) can be used for *L. heterophyllum*, four of the loci can be used for *L. hirtum* (*Lep-21b*, *Lep-44*, *Lep-50*, and *Lep-53*), three can be used for *L. draba* (*Lep-40*, *Lep-44*, *Lep-79*), and two for *L. graminifolium* (*Lep-40* and *Lep-50*). The SSR alleles in Table 6 are useful to identify successful hybrids during interspecific hybridization between *L. campestre* and other *Lepidium* species. The use of these markers in genome wide association studies (GWAS) or quantitative trait loci (QTL) analysis may reveal association/linkage with desirable traits. These types of analyses, and others including pedigree and population studies, has been extensive and are becoming more popular Kalia et al. 2011). For example, if a locus shows association with perenniality, they can be used in marker assisted selection that involves interspecific hybridization between *L. campestre* and perennial *Lepidium* species. This will speed up the breeding process by signaling to researchers that a specific plant most likely has the desired trait. Plants that most likely do not have the desired trait can be removed. The cost of the breeding program is reduced, and allows researchers to focus on the plants that are most promising. With *Lepidium*, the selection of plants can happen within the first couple of weeks of the life of new plants. This is because the plants grow quickly from seeds and only a small amount of plant tissue is needed to extract DNA to make an assessment. Qucik identification of which plants most likely have the desired trait allows money and time to be saved.

This information can also be used to identify different species or cultivars (Rongwen et al. 1995; Szewc-Mcfadden et al. 1996; Hokanson et al. 1998; Lamboy, 1998). Identification can be done for unknown wild collected plants and incorrectly categorized or unknown plants that have already been collected. Not every plant can easily be identified in the field, especially since these species

have overlapping ranges and live in similar environments (Anonymous, n.d. a-f). Being able to quickly distinguish closely related species can help with the breeding process that involves hybridization between species. Fewer mistakes based on assumed or misidentified species allow plants to be quickly verified before too much time and money have been spent on the wrong plants. Verifying that each plant is the expected species adds a layer of protection against potentially costly mistakes due to mistaken identity or low stock of valuable seeds.

Table 6: List of potentially informative SSR loci that can be used to differentiate *L. campestre* from one or more of the four other *Lepidium* species included in this study. Values in the cells refer to the number of repeats of the SSR motifs as determined through sequencing. Values in **bold** show the species that can potentially be differentiated from *L. campestre* at the corresponding locus.

		L. campestre			***L. heterophyllum***			***L. hirtum***				***L. draba***		***L. graminifolium***	
Locus	**Repeat motifs**	**C-92**	**Gäv-2**	**Mör-6**	**H-56a**	**H-98/48a**	**Häst-3**	**Hi-53a**	**Hi-57a**	**Hi-58a**	**Hi-87a**	**Dra-1**	**Dra-2**	**G-27a**	**G-55a**
Lep-21b	(GT)4	4	4	4	8	4	4	**3**	**5**	**3**	**3**	**-**	**-**	**-**	**-**
Lep-22	(AGTG)4	4	4	4	**2**	**2**	**2**	4	5	4	*	**-**	**-**	**-**	**-**
Lep-40	(AAAG)6	6	6	6	**5**	**5**	**5**	5	6	6	6	**3**	**3**	**3**	**3**
Lep-44	(AG)7	7	7	7	**8**	**9**	**9**	**6**	**22**	**10**	**6**	**4**	**4**	7	7
Lep-50	(TA)9	9	9	9	*	5	5	**6**	**6**	**6**	**6**	*	*	**6**	**6**
Lep-53	(AG)7	7	7	7	7	7	7	**8**	**9**	**8**	**11**	**-**	**-**	*	*
Lep-79	(CT)8	8	8	8	8	8	8	9	11	8	8	**9**	**9**	*	4

- = the locus is absent

* = sequencing failed (missing)

Genetic polymorphism

Microsatellites

When looking at the data of sequenced microsatellites, some interesting conclusions can be made. The primer-pairs were developed based on Rad-sequences generated using *L. campestre* samples, and they are expected to work best in this species. However, most of these SSR loci are transferable to other *Lepidium* species as well. Transferability has been studied in other species of plants including important crop plants. Kuleung et al. (2003) studied SSR transferability in wheat, rye, and triticale to varying success. Transferability studies between taxa on both agriculture plants and wild species were summarized by Ellis and Burke (2007). The summarized studies covered traditional SSRs and Expressed Sequence Tag SSRs (EST-SSRs). EST-SSRs are a cheaper alternative and more likely to be transferable due to the amount of genomic data that is publically available (Ellis & Burke, 2007). Most of the studies they summarized found high levels of transferability. This study found medium to high transferability between *Lepidium* species (Table 7). The lowest total transferability was in *L.* draba at 53%, while the highest was *L. heterophyllum* at 96%. *L. hirtum* is also very high with 92%, and *L. graminifolium* shows medium transferability at 54% of loci transferable. This is a positive sign for future population genetic studies and the current breeding program since there are many loci that can be used.

Table 7: The transferability of *L. campestre* loci to the other species from both sets of data brought into one table. Some loci worked for only one set of data while others worked for both. This table allows a look at total transferability shown in all data.

SSR Loci	*L. campestre*	*L. heterophyllum*	*L. hirtum*	*L. draba*	*L. graminifolium*
Lep-21a	+	+	+	-	-
Lep-21b	+	+	+	-	-
Lep-22	+	+	+	-	-
Lep-27	+	+	+	-	-
Lep-28	+	+	-	+	-
Lep-31	+	+	+	N/A	+
Lep-33	+	+	+	-	-
Lep-38	+	+	+	-	+
Lep-40	+	+	+	+	+
Lep-42	+	+	-	N/A	-
Lep-44	+	+	+	+	+
Lep-45	+	+	+	+	+
Lep-47	+	+	+	N/A	-
Lep-48	+	+	+	+	+
Lep-50	+	+	+	-	+
Lep-53	+	+	+	-	-
Lep-55	+	+	+	+	+
Lep-58	+	-	+	N/A	+
Lep-69	+	+	+	N/A	-
Lep-71	+	+	+	+	-
Lep-76	+	+	+	-	+
Lep-77	+	+	+	+	+
Lep-79	+	+	+	+	+
Lep-81	+	+	+	+	+
% transferable	N/A	96	92	53	54

+ = Locus present
- = Locus absent
N/A = Not Applicable

The selection of the 26 primer-pairs studied was predicated on the uniformity of bands on an agarose gel. However, 39% of the sequenced loci and 71% of the fragment loci analyzed were polymorphic within *L. campestre*. Sequencing can show which specific bases are different and can give insight on whether it is the length of the SSRs that is causing differences in allele sizes or if there are indels. On the other hand, capillary electrophoresis will only show the length of fragments. This can cause a discrepancy in alleles due to two assumptions. The first is that variation between alleles is due to size differences in the microsatellite. Indels outside the SSR would completely change the allele in fragment size data, but could possibly be undetected in the sequencing data. The second is that heterozygosity was not considered during sequencing. Because heterozygosity was not considered, heterozygotes would only be discovered by capillary electrophoresis in this study. Some of the *L. campestre* samples showed heterozygosity or their loci showed polymorphism after capillary electrophoresis, while only one allele was seen when sequenced (*Lep-44* and *Lep-50*; *Lep-53*). One locus (*Lep-81*) was the other way around with the fragment size data being homozygous and monomorphic, while the sequence data was polymorphic. It is possible that only one allele would show up in the sequencing data because only three individuals were sequenced. Heterozygosity is also important with regards to both *L. draba*, and *L. graminifolium*. They are both polyploids and their type of ploidy could affect the likelihood that they are heterozygous, maybe even with more than two alleles. This study would only pick this up within the fragment size data for *L. graminifolium*. No heterozygosity with more than two alleles was found.

Polymorphisms within and between species are useful in genetic diversity studies or to better understand relationships between taxa (Chase et al., 1996; Gupta et al., 2003; Tang et al., 2007). However, having homozygous, monomorphic loci is important when it is necessary to predict the outcome of a cross. In order to avoid undesirable loci, some steps need to be taken. It would be best to first run capillary electrophoresis on PCR products and, after learning which locus is homozygous, sending those to be sequenced. Those are the loci that should be used initially for mapping the genome and then, hopefully, also be useful to develop inbred lines for breeding.

There are a couple of ideas that can be discussed with the data pertaining to polymorphism. The first is that the rate of polymorphism within a species could be overstated. This may be because

only two samples are needed to say the species are polymorphic, while all samples need to be present and have the same allele to be monomorphic. The second is the increase in polymorphism between sequence and fragment size data (Table 8). In the sequence data, 24 loci provided a total of 77 alleles (3.21 alleles/locus), while in the fragment size data, 23 loci provided 85 alleles (3.7 alleles/locus). This can be attributed to the fact that when analyzing the sequence data, the only alleles that count are within the microsatellite, while fragment size data shows an allele for any change in DNA length. This could be indels that are outside of the microsatellite, or within the microsatellite. This can create more alleles while the actual microsatellite remains unchanged.

This analysis revealed that higher genetic variation exists in *L. hirtum* (2.22 alleles/locus in sequence data and 2.06 alleles/locus in fragment size data) than the other species included in this study, including *L. campestre* (1.42 alleles/locus in sequence data and 1.75 alleles/locus in fragment size data). The high rate can be explained by there being three different subspecies studied. These subspecies came from different regions along the Mediterranean Sea (Table 1). In the dendrograms (Figures 3 and 4) where *L. hirtum* samples are included, the two samples of *L. hirtum* ssp. *atlanticum* were distantly related to each other. In order to get a clearer picture of the relations between subspecies, more genetic studies need to be done with more samples and loci than used in this study.

Table 8: Number of samples successfully used (NS), number of loci analyzed (NL), number of alleles detected (NA) and average number of alleles per locus (A/L) for DNA sequence and fragment size based analysis of *Lepidium* species.

		L. campestre	***L. heterophyllum***	***L. hirtum***	***L. draba***	***L. graminifolium***	**Total/ Average**
Sequence data	NS	4*	3	4	2	2	15
	NL	24	20	18	10	11	24
	NA	34	29	40	12	12	77
	A/L	1.42	1.45	2.22	1.20	1.09	3.21
Fragment size data	NS	7	2	4	NU	2	15
	NL	20	6	18	NU	16	23
	NA	35	10	37	NU	24	85
	A/L	1.75	1.67	2.06	NU	1.50	3.70

*including *L. campestre* consensus sequence. NU = Not used

SNPs/indels

SNPs/indels data can be used in quicken the breeding process. They have been used effectively as molecular markers in *Brassica napus*, rice, wheat, apple, and maize (Chagné et al. 2007; Septiningsih et al. 2009; Asif et al. 2011; Harper et al. 2012; Naidoo et al. 2012). If the allele(s) within a species are different from allele(s) in the other species, it can be used in the same way an SSR is used. In Table 4, the SNPs and indels that are informative from all sequence data are presented. Individual SNPS and indels could be used as an identification tool for specific individuals and heritage as has been shown in recent studies in forensics and animal husbandry (Heaton et al. 2002; Homer et al. 2008). Both of these studies found that SNPs were a useful tool in identification of individuals. Their methods can easily be used in plants as well. SNPs/indels can also be used to help differentiate species. This has been done with other plant species, including *Capsicum* (Jeong et al. 2010).

In this data, *Lep*-44 is a good example of how individual identification and species differentiation can be completed using SNPs/indels. Depending on the results of genome wide association studies (GWAS) and quantitative trait loci (QTL) analysis, they could also be useful for marker assisted selection. In Table 4, at locus *Lep-44* there was a large deletion in Hi-58a. This deletion was unique to this sample. If it could be shown that the deletion is near an important gene and associated with favorable allele, then the deletion could be used to confirm that the allele it is associated with was most likely passed down. In this situation, there was also a difference in SSR length which makes the use of the deletion as a marker less relevant, however this might not always be the case. It is easy to imagine an SNP or indels in a single individual of a species that would be informative. In the *Lepidium* domestication process, the breeding project is not only looking to breed traits into *L. campestre*, but also to try to find the traits within *L. campestre* in order for an easier breeding process. Since the success of crosses between species are not high (Mulatu Geleta, personal communication), for example, a single *L. campestre* individual might be found to have a pod less prone to shattering would be extremely valuable The SSR associated with the relevant genes might be the same between species. In this case, SNPs or indels data can be useful in order to be able to see if the desired traits are inherited.

The SSR locus that produced the most interesting SNPs and indels was *Lep-44* (Figure 5). Here, *L. campestre*, *L. heterophyllum*, *L. hirtum*, and *L. draba* can be identified either by length of SSR or by using SNPs and deletions. The "G/C" SNP in the sequence differentiates all *L. heterophyllum* samples from samples of the other species. *L. draba* samples have two unique SNPs, and two unique indels. *L. hirtum* is a bit trickier, but the presence of a "C" where *L. heterophyllum* has a "G" rules that species out, the deletion of the sequence "AGGTGAAT" rules out *L. campestre* and *L. graminifolium*, and the presence of an "A" where *L. draba* has a "G" or a "T" or the lack of a "GA" rules *L. draba* out, leaving *L. hirtum* as the only choice. There is also the ability to identify certain individuals of these species by using SSR length.

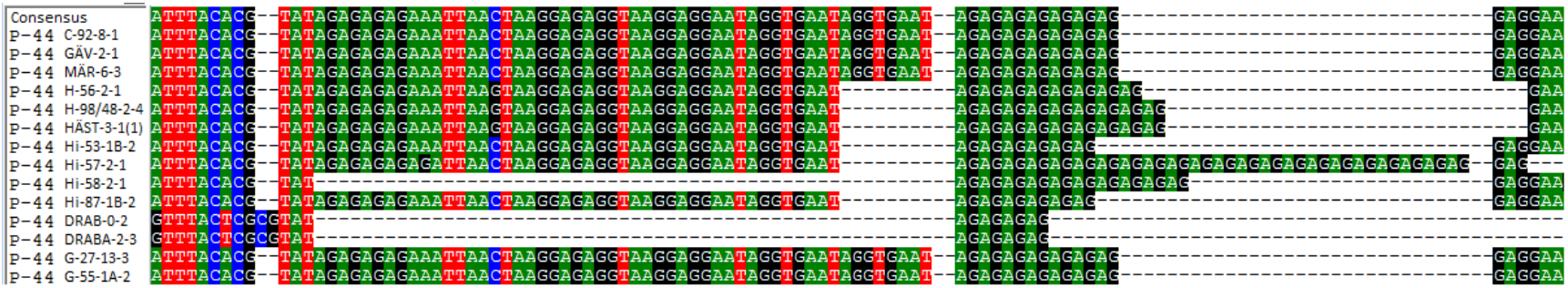

Figure 5: Partial sequence of PCR products of 14 samples representing five *Lepidium* species amplified using *Lep-44* SSR primer-pair.

Cluster analysis

By using the sequence and fragment size data of the different *Lepidium* species, cluster analysis can be made. This data should be added to the already known phylogenetics of these five species. The phylogenetic relationships between some *Lepidium* species have previously been presented by Mummenhoff et al. (2001; 2009) and Lee et al. (2002). Within the breeding program, they can be used to identify the species more closely related. This may have an effect on which species will continue to be used in or introduced into the program, and which will be removed in the future. Crosses between *L. campestre* and the other species have been attempted. Some of these crosses have failed to produce viable hybrids. So far, species (*L. heterophyllum* and *L. hirtum*) that are more closely related have more readily crossed with *L. campestre* than the species that are more distantly related. All attempted crosses of *L. campestre* with *L. draba* and *L. graminifolium* have not resulted in viable hybrids. The cause of the failure could be the difference in ploidy level or the distance of the relationships. The success rate of crosses between *L. campestre* and *L. hirtum* is low. The most successful crosses were those between *L. campestre* and *L. heterophyllum* (Mulatu Geleta, personal communication) suggesting a close phylogenetic relationship between these two species.

One particularly interesting dendrogram is the one comparing *L. campestre* samples based on the fragment size data. There were three samples (Spjut-2, C-19A, and Vik-2) collected from around Skåne, one sample (Alb-6) from Öland, and three samples from around Europe (C-52 and C-124 from France and C-89 from Greece). This showed no clear grouping of the individuals according to their geographic origin and the branches were supported by moderate to high bootstrap values (Figure 1). This pattern of not grouping according to geography was also observed in the other dendrograms (Figures 2, 3, and 4). So far, no genetic diversity studies have been conducted on *L. campestre*, but a possible explanation why adding samples from other countries might not add new traits is that the genetic diversity is fairly evenly spread throughout its range. Although there are not many samples in this study, what is seen suggests that genetic diversity within *L. campestre* is not linked to geography. If genetic diversity studies that involve large number of populations/accessions each represented by ten or more samples are conducted and find this to be true, priority can be given to more local populations of *L. campestre* thereby reducing costs for the breeding program without a big risk of losing desirable genetic variants.

The dendrograms (Figure 2) comparing *L. campestre* and *L. heterophyllum* show the expected groupings between the two species. The relationship between *L. hirtum* subspecies deserves a closer look. Each of the dendrograms comparing *L. hirtum* to other species shows a different relationship between its subspecies. In Figure 3, both dendrograms A and B, have a close relationship between the Hi-57 and Hi-87 individuals**.** This is expected because they are of the same subspecies (*L. hirtum* ssp. *atlanticum*). *L. campestre, L. heterophyllum,* and *L. hirtum* are compared in Figure 4. In dendrogram A, two of the four *L. hirtum* individuals (Hi-58a, *L. hirtum* ssp. *calycotrichum,* and Hi-87a, *L. hirtum* ssp. *atlanticum*) are more closely related to *L. campestre* than the *L. heterophyllum* individuals are. This adds credence to Lee et al. (2002) who found *L. hirtum* ssp. *calycotrichum* more closely related to *L. campestre* than *L. heterophyllum*, but runs counter to Mummenhoff et al.'s (2001; 2009) data that found *L. hirtum* subspecies, including *calycotrichum* and *nebrodense,* more distantly related to *L. campestre* than *L. heterophyllum*. However, Hi-57b and Hi-87b are both *L. hirtum* ssp. *atlanticum* and follow the expected pattern of being more distantly related to *L. campestre* in dendrogram B. Therefore, more data that includes relatively slow evolving gene sequences may be needed in order to establish more robust phylogenetic relationships between these species.

REFERENCES

Al-Shehbaz, I. A. (1986). The genera of Lepidieae (Cruciferae: Brassicaceae) in the southeastern United States. *Journal of the Arnold Arboretum, 67*(3), 265-311.

Al-Shehbaz, I. A., Beilstein, M. A., & Kellogg, E. A. (2006). Systematics and phylogeny of the Brassicaceae (Cruciferae): an overview. *Plant Systematics and Evolution, 259*(2-4), 89-120.

Andersen, A., & Olsen, C. C. (1993). Rye grass as a catch crop in spring barley. *Acta Agriculturae Scandinavica B-Plant Soil Sciences, 43*(4), 218-230.

Andersson, A. A., Merker, A., Nilsson, P., Sørensen, H., & Åman, P. (1999). Chemical composition of the potential new oilseed crops *Barbarea vulgaris*, *Barbarea verna* and *Lepidium campestre*. *Journal of the Science of Food and Agriculture, 79*(2), 179-186.

Anonymous. (n.d.). *Lepidium* in Flora of North America. Retrieved September 05, 2016 from http://www.efloras.org/florataxon.aspx?flora_id=1&taxon_id=118034.

Anonymous. (n.d.). *Lepidium campestre* in Flora of North America. Retrieved September 05, 2016 from http://www.efloras.org/florataxon.aspx?flora_id=1&taxon_id=200009577.

Anonymous. (n.d.). *Lepidium draba* in Flora of North America. Retrieved September 05, 2016 from http://www.efloras.org/florataxon.aspx?flora_id=1&taxon_id=242328876.

Anonymous. (n.d.). *Lepidium heterophyllum* in Flora of North America Retrieved September 05, 2016 from http://www.efloras.org/florataxon.aspx?flora_id=1&taxon_id=250095029.

Anonymous. (n.d.). Tropicos | Name - *Lepidium graminifolium*. Retrieved September 05, 2016 from http://www.tropicos.org/Name/4100536.

Anonymous. (n.d.). Tropicos | Name - *Lepidium hirtum* (L.) Sm. Retrieved September 05, 2016 from http://www.tropicos.org/Name/4100202.

Aronsson, H. (2000). Nitrogen turnover and leaching in cropping systems with ryegrass catch crops. *Acta Universitatis Agriculturae Sueciae: Agraria*, (214).

Asif, M., Shaheen, T., Tabbasam, N., Zafar, Y., & Paterson, A. H. (2011). Marker-assisted breeding in higher plants. *Alternative Farming Systems, Biotechnology, Drought Stress and Ecological Fertilisation* (pp. 39-76): Springer.

Bergström, L., & Brink, N. (1986). Effects of differentiated applications of fertilizer N on leaching losses and distribution of inorganic N in the soil. *Plant and Soil, 93*(3), 333-345.

Bon, M. C., Hurard, C., Gaskin, J., & Risterucci, A. M. (2005). Polymorphic microsatellite markers in polyploid *Lepidium draba* L. ssp. *draba* (Brassicaceae) and cross-species amplification in closely related taxa. *Molecular Ecology Notes, 5*(1), 68-70.

Börjesdotter, D. (1999). Potential oil crops. Cultivation of *Barbarea verna*, *Barbarea vulgaris* and *Lepidium campestre*. *Acta Universitatis Agriculturae Sueciae: Agraria,* (193).

Börjesdotter, D. (2000). Oljehaltiga fånggrödor – vårgyllen, sommargyllen och fältkrassing. *FAKTA Jordbruk, 9*.

Brandes, D. (1995). The flora of old town centres in Europe. In H. Sukopp, M. Numata, & A. Huber (Eds.), *Urban Ecology as the Basis of Urban Planning* (pp. 49-58): SPB Academic Publishing..

Carlsson, A. S., Yilmaz, J. L., Green, A. G., Stymne, S., & Hofvander, P. (2011). Replacing fossil oil with fresh oil–with what and for what? *European Journal of Lipid Science and Technology, 113*(7), 812-831.

Chagné, D., Carlisle, C. M., Blond, C., Volz, R. K., Whitworth, C. J., Oraguzie, N. C., Crowhurst, R. N., et al. (2007). Mapping a candidate gene (*MdMYB10*) for red flesh and foliage colour in apple. *Bmc Genomics, 8*(1), 1.

Chase, M., Kesseli, R., & Bawa, K. (1996). Microsatellite markers for population and conservation genetics of tropical trees. *American Journal of Botany*, 51-57.

Cregan, P., & Quigley, C. (1996). Simple sequence repeat DNA marker analysis. In G. Caetano-Anolles & P. M. Gresshoff (Eds.), DNA Markers: Protocols, Applications, and Overviews (pp. 173-185). New York: J. Wiley and Sons.

EFSA [European Food Safety Authority]. (2008). Glucosinolates as undesirable substances in animal feed - Scientific Opinion of the Panel on Contaminants in the Food Chain. *EFSA Journal* (590), 1-76.

Ellis, J. R., & Burke, J. M. (2007). EST-SSRs as a resource for population genetic analyses. *Heredity, 99*(2), 125-132.

Eriksson, D. (2009). Towards the domestication of *Lepidium campestre* as an undersown oilseed crop. *Acta Universitatis Agriculturae Sueciae: Agraria, 2009*(65).

Eurofins. Mix2Seq - Smart • Easy • Quick. Retrieved November 08, 2016 from http://www.eurofinsgenomics.eu/en/custom-dna-sequencing/tube-sequencing-service/mix2seq.aspx.

Geleta, M., Zhu, L. H., Stymne, S., Lehrman, A., & Hansson, S. O. (2014). *Domestication of Lepidium campestre as part of Mistra Biotech, a research programme focused on agro-biotechnology for sustainable food*. In C. Batello, L. Wade, S. Cox, N. Pogna, A. Bozzini, & J. Choptiany (Eds.), *Perennial Crops for Food Security. Proceedings of the FAO Expert Workshop*, Rome, Italy, 28-30 August 2013 (pp. 141-147). FAO, Italy. ISBN 978-92-5-107998-0 (print). E-ISBN 978-92-5-107999-7 (PDF).

Goss, M., Howse, K., Lane, P., Christian, D., & Harris, G. (1993). Losses of nitrate-nitrogen in water draining from under autumn-sown crops established by direct drilling or mouldboard ploughing. *Journal of Soil Science, 44*(1), 35-48.

Gupta, P. K., Rustgi, S., Sharma, S., Singh, R., Kumar, N., & Balyan, H. S. (2003). Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. *Molecular Genetics and Genomics, 270*(4), 315-323.

Hall, T. A. (1999). *BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series, 41*, 95-98.Haragan, P. D. (2015). *Weeds of Kentucky and Adjacent States: A Field Guide*. Lexington: University Press of Kentucky.

Harper, A. L., Trick, M., Higgins, J., Fraser, F., Clissold, L., Wells, R., Hattori, C., et al. (2012). Associative transcriptomics of traits in the polyploid crop species *Brassica napus*. *Nature Biotechnology, 30*(8), 798-802.

Heaton, M. P., Harhay, G. P., Bennett, G. L., Stone, R. T., Grosse, W. M., Casas, E., Keele, J. W., et al. (2002). Selection and use of SNP markers for animal identification and paternity analysis in US beef cattle. *Mammalian Genome, 13*(5), 272-281.

Hoffmann, M., & Johnsson, H. (2000). Nitrogen leaching from agricultural land in Sweden: Model calculated effects of measures to reduce leaching loads. *AMBIO: A Journal of the Human Environment, 29*(2), 67-73.

Hokanson, S., Szewc-McFadden, A., Lamboy, W., & McFerson, J. (1998). Microsatellite (SSR) markers reveal genetic identities, genetic diversity and relationships in a *Malus*× *domestica* Borkh. core subset collection. *Theoretical and Applied Genetics, 97*(5-6), 671-683.

Homer, N., Szelinger, S., Redman, M., Duggan, D., Tembe, W., Muehling, J., Pearson, J. V., et al. (2008). Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. *PLoS Genet, 4*(8), e1000167.

Ivarson, E., Ahlman, A., Lager, I., & Zhu, L.-H. (2016). Significant increase of oleic acid level in the wild species *Lepidium campestre* through direct gene silencing. *Plant Cell Reports, 35*(10), 2055-2063.

Ivarson, E., Ahlman, A., Li, X., & Zhu, L.-H. (2013). Development of an efficient regeneration and transformation method for the new potential oilseed crop *Lepidium campestre*. *BMC Plant Biology, 13*(1), 1.

Jensen, E. S. (1991). Nitrogen accumulation and residual effects of nitrogen catch crops. *Acta Agriculturae Scandinavica, 41*(4), 333-344.

Jeong, H.-J., Jo, Y. D., Park, S.-W., & Kang, B.-C. (2010). Identification of *Capsicum* species using SNP markers based on high resolution melting analysis. *Genome, 53*(12), 1029-1040.

Kalia, R. K., Rai, M. K., Kalia, S., Singh, R., & Dhawan, A. K. (2011). Microsatellite markers: an overview of the recent progress in plants. *Euphytica, 177*(3), 309-334.

Känkänen, H., & Eriksson, C. (2007). Effects of undersown crops on soil mineral N and grain yield of spring barley. *European Journal of Agronomy, 27*(1), 25-34.

Kuleung, C., Baenziger, P. S., & Dweikat, I. (2004). Transferability of SSR markers among wheat, rye, and triticale. *Theoretical and Applied Genetics, 108*(6), 1147-1150.

Kvist, M. (1992). Catch crops undersown in spring barley-competitive effects and cropping methods. *Crop Production Science, 15.* Department of Crop Production Science, Swedish University of Agricultural Sciences.

Lal, R., Reicosky, D. C., & Hanson, J. D. (2007). Evolution of the plow over 10,000 years and the rationale for no-till farming. *Soil and Tillage Research, 93*(1), 1-12.

Lamboy, W. F. (1998). Using simple sequence repeats (SSRs) for DNA fingerprinting germplasm accessions of grape (*Vitis* L.) species. *Journal of the American Society for Horticultural Science, 123*(2), 182-188.

Larkin, M. A., Blackshields, G., Brown, N., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics, 23*(21), 2947-2948.

Larson, L., Kiemnec, G., & Smergut, T. (2000). Hoary cress reproduction in a sagebrush ecosystem. *Journal of Range Management*, 556-559.

Lee, J.-Y., Mummenhoff, K., & Bowman, J. L. (2002). Allopolyploidization and evolution of species with reduced floral structures in *Lepidium* L. (Brassicaceae). *Proceedings of the National Academy of Sciences (USA), 99*(26), 16835-16840.

Liu, Q., Singh, S. P., & Green, A. G. (2002). High-stearic and high-oleic cottonseed oils produced by hairpin RNA-mediated post-transcriptional gene silencing. *Plant Physiology, 129*(4), 1732-1743.

Macdonald, A. J., Powlson, D. S., Poulton, P. R., & Jenkinson, D. S. (1989). Unused fertiliser nitrogen in arable soils—its contribution to nitrate leaching. *Journal of the Science of Food and Agriculture, 46*(4), 407-419.

Marriage, T. N., Hudman, S., Mort, M. E., Orive, M. E., Shaw, R. G., & Kelly, J. K. (2009). Direct estimation of the mutation rate at dinucleotide microsatellite loci in *Arabidopsis thaliana* (Brassicaceae). *Heredity, 103*(4), 310-317.

Mason, A. S. (2015). SSR genotyping. In J. Batley (Ed.), Plant Genotyping: Methods and Protocols (pp. 77-89): Humana Press.

Merker, A., Eriksson, D., & Bertholdsson, N.-O. (2010). Barley yield increases with undersown *Lepidium campestre*. *Acta Agriculturae Scandinavica Section B–Soil and Plant Science, 60*(3), 269-273.

Merker, A., & Nilsson, P. (1995). Some oil crop properties in wild Barbarea and Lepidium species. *Swedish Journal of Agricultural Research, 25*, 173-178.

Mikolajczak, K., Miwa, T., Earle, F., Wolff, I., & Jones, Q. (1961). Search for new industrial oils. V. Oils of Cruciferae. *Journal of the American Oil Chemists Society, 38*(12), 678-681.

Morgante, M., Hanafey, M., & Powell, W. (2002). Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nature Genetics, 30*(2), 194-200.

Mulligan, G. A., & Frankton, C. (1962). Taxonomy of the genus *Cardaria* with particular reference to the species introduced into North America. *Canadian Journal of Botany, 40*(11), 1411-1425.

Mummenhoff, K., Brüggemann, H., & Bowman, J. L. (2001). Chloroplast DNA phylogeny and biogeography of *Lepidium* (Brassicaceae). *American Journal of Botany, 88*(11), 2051-2063.

Mummenhoff, K., Polster, A., Mühlhausen, A., & Theißen, G. (2009). *Lepidium* as a model system for studying the evolution of fruit development in Brassicaceae. *Journal of Experimental Botany, 60*(5), 1503-1513.

Munkholm, L. J., & Hansen, E. M. (2012). Catch crop biomass production, nitrogen uptake and root development under different tillage systems. *Soil Use and Management, 28*(4), 517-529.

Murphy, D. J. (1996). Engineering oil production in rapeseed and other oil crops. *Trends in Biotechnology, 14*(6), 206-213.

Naidoo, R., Watson, G., Derera, J., Tongoona, P., & Laing, M. (2012). Marker-assisted selection for low phytic acid (lpa1-1) with single nucleotide polymorphism marker and amplified fragment length polymorphisms for background selection in a maize backcross breeding programme. *Molecular Breeding, 30*(2), 1207-1217.

Nilsson, P., Johansson, S. Å., & Merker, A. (1998). Variation in seed oil composition of species from the genera *Barbarea* and *Lepidium*. *Acta Agriculturae Scandinavica B-Plant Soil Sciences, 48*(3), 159-164.

Ohlander, L., Bergkvist, G., Stendahl, F., & Kvist, M. (1996). Yield of catch crops and spring barley as affected by time of undersowing. *Acta Agriculturae Scandinavica B-Plant Soil Sciences, 46*(3), 161-168.

Page, R. (2005). TreeView X v0.5.0. Available online at: http://darwin.zoology.gla.ac.uk/~rpage/treeviewx/index.html [Verified November 2016]

Patel, D. A., Zander, M., Dalton-Morgan, J., & Batley, J. (2015). Advances in Plant Genotyping: Where the Future Will Take Us. *Plant Genotyping: Methods and Protocols*, 1-11.

Pavlicek, A., Hrda, S., & Flegr, J. (1998). Free-Tree--freeware program for construction of phylogenetic trees on the basis of distance data and bootstrap/jackknife analysis of the tree robustness. Application in the RAPD analysis of genus *Frenkelia*. *Folia biologica, 45*(3), 97-99.

Pham, A.-T., Lee, J.-D., Shannon, J. G., & Bilyeu, K. D. (2010). Mutant alleles of FAD2-1A and FAD2-1B combine to produce soybeans with the high oleic acid seed oil trait. *BMC Plant Biology, 10*(1), 1.

Princen, L. (1983). New oilseed crops on the horizon. *Economic Botany, 37*(4), 478-492.

Qasem, J. (1994). Allelopathic effect of white top (Lepidium draba) on wheat and barley. *Allelopathy Journal, 1*(1), 29-40.

Rongwen, J., Akkaya, M., Bhagwat, A., Lavi, U., & Cregan, P. (1995). The use of microsatellite DNA markers for soybean genotype identification. *Theoretical and Applied Genetics, 90*(1), 43-48.

Septiningsih, E. M., Pamplona, A. M., Sanchez, D. L., Neeraja, C. N., Vergara, G. V., Heuer, S., Ismail, A. M., et al. (2009). Development of submergence-tolerant rice cultivars: the *Sub1* locus and beyond. *Annals of Botany, 103*(2), 151-160.

Shi, J., Huang, S., Fu, D., Yu, J., Wang, X., Hua, W., Liu, S., et al. (2013). Evolutionary dynamics of microsatellite distribution in plants: insight from the comparison of sequenced brassica, Arabidopsis and other angiosperm species. *PLoS One, 8*(3), e59988.

Singh, S. P., Zhou, X.-R., Liu, Q., Stymne, S., & Green, A. G. (2005). Metabolic engineering of new fatty acids in plants. *Current Opinion in Plant Biology, 8*(2), 197-203.

SoftGenetics. (2016). *Software PowerTools for Genetic Analysis*. Retrieved November 09, 2016 from http://www.softgenetics.com/.

Stymne, S. (2005). Nya vegetabiliska oljor för tekniskt bruk med genteknik. *Frö och Oljeväxtodlarna*, Försöks- och forskningsrapport.

Szewc-McFadden, A., Kresovich, S., Bliek, S., Mitchell, S., & McFerson, J. (1996). Identification of polymorphic, conserved simple sequence repeats (SSRs) in cultivated *Brassica* species. *Theoretical and Applied Genetics, 93*(4), 534-538.

Tang, R., Gao, G., He, L., Han, Z., Shan, S., Zhong, R., Zhou, C., et al. (2007). Genetic diversity in cultivated groundnut based on SSR markers. *Journal of Genetics and Genomics, 34*(5), 449-459.

Tautz, D., & Renz, M. (1984). Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Research, 12*(10), 4127-4138.