GENOME RELATIONSHIPS AMONG *Lotus* SPECIES BASED ON RANDOM AMPLIFIED POLYMORPHIC DNA (RAPD)

by

Lázara Pereira Campos

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfilment of the requirements for the degree of Master of Science

August 1992

Genome relationship among Lotus species based on random amplified polymorphic DNA (RAPD)

SHORT TITLE:

Genome relationship among Lotus species based on RAPD

ABSTRACT

The usefulness of RAPDs (Random Amplified Polymorphic DNA) to distinguish among different taxa of *Lotus* was evaluated. The following species were included: *L. corniculatus*, *L. tenuis*, *L. alpinus*, *L. japonicus*, and *L. uliginosus*. Several accessions for each species were studied. Following DNA extraction, amplification reactions were performed in a Hybaid DNA Thermal Cycler, and the product visualized according to a standard procedure. Twenty primers were used for each species/accession. Clear bands and several polymorphisms were obtained for all primers. A phenogram was drawn based on the genetic distance among the species. *L. alpinus* appears as the most distant species from *L. corniculatus*, followed by *L. uliginosus*, *L. tenuis*, and *L. japonicus*. With the exception of *L. alpinus*, these findings are in agreement with previous experimental studies in the *L. corniculatus* group. The use of a greater number of primers and increased number of species may provide a greater resolution of the systematics of these taxa.

RÉSUMÉ

L'utilité des RAPDs (Random Amplified Polymorphic DNA) pour distinguer différents taxons de *Lotus* à été évaluée. Les espèces suivantes furent étudiées: *L. corniculatus*, *L. tenuis*, *L. alpinus*, *L. japonicus* et *L. uliginosus*. Pour chaque espèce, plusieurs accessions ont été testés. A la suite de l'extraction de l'ADN, des réactions d'amplification en chaîne ont été entreprises utilisant un cycleur thermique Hybaid. Le produit fut par la suite visualisé, selon une procédure standard. Vingt amorces ont été utilisées pour chaque espèce/accession. Des bandes ainsi que plusieurs polymorphismes ont été obtenus avec toutes les amorces. Un phénogramme fut dessiné à partir des distances génétiques entre les espèces. *Lotus alpinus* semble être l'espèce la plus distante de *L. corniculatus*, suivie de *L. uliginosus*, *L. tenuis* et *L. japonicus*. A l'exception de *L. alpinus*, ces résultats sont en accord avec d'autres études expérimentales effectuées précédemment dans le groupe de *L. corniculatus*. L'utilisation d'un plus grand nombre d'amorces et d'espèces pourrait ainsi procurer un degré plus élevé de résolution qu'en à la systématique de ces taxons.

Acknowledgements

The author extends her sincere gratitude to Professor William Grant, my thesis supervisor, for his continued guidance, support, advice, encouragement and assistance during the course of tnis study.

The author gratefully acknowledges Dr. John Raelson for helpful suggestions and encouragement in using RAPDs for the analyses, and Dr. M. Fortin for his suggestions and lab facilities support.

The author is indebted to Dr. Hoan Lee, Dr. Alfredo Rivas, and Dr. A. Ahmad, for their encouraging discussions and the use of their laboratory facilities.

The help from Anne-Marie l'Heureux in preparing the French summary is also highly appreciated. Nick Tinker was very helpful in introducing the programs used for the phylogenetic analyses.

A sincere word of appreciation is dedicated to the late Professor Mamdouh A. Fanous and to his wife Estelle, whose support, friendship, and entertaining moments provided since the day my husband and I arrived in Canada will remain ingrained in our memory forever as sincere proof of affection.

Special thanks are extended to my friends Shana, Suyanee, Anne-Marie, Yolanda, Mazé, Marjorie, and Michelle, for their valuable friendship that created

a wonderful atmosphere essential to overcome the most difficult obstacles.

My deepest gratitude is extended to my husband Paulo and my son Daniel for sharing some of the happiest moments of my life, and for giving me the strength and the motivation to pursue my goals.

Claim to originality

To the author's knowledge, the present work is the first to examine the genome relationships among *Lotus* species using Random Amplified Polymorphic DNA markers.

Table of Contents

List of Tables

List of Figures

1. INTRODUCTION

The genus *Lotus* (Leguminosae) is a large polymorphic group, comprising approximately 200 annual and perennial species (Grant, 1965). Birdsfoot Trefoil (*Lotus corniculatus* L.) is a tetraploid ($2n = 24$) forage legume which has distinct advantages for forage production on wet, acid, or shallow soils. The agronomic advantages of Birdsfoot Trefoil have been well reviewed (Seaney and Henson, 1970; Grant and Marten, 1985).

Various authors have proposed different species as the progenitors of the tetraploid cultivated species *Lotus corniculatus*. The arguments favoring one hypothesis or the other have been based upon the inheritance of the few genetic markers available or upon meiotic analyses of hybrids. Several diploid ancestors have been proposed including *L. alpinus* Schleich., *L. japonicus* (Regel) Larsen, *L. tenuis* Waldst. et Kit, and *L. uliginosus* Schkuhr.

In recent years, the availability of methods based on DNA markers has given origin to a new discipline called molecular systematics, in which species are classified according to DNA polymorphisms. The polymorphisms observed are visualized in acrylamide and agarose gels, providing unique information that is complementary to other traditional approaches of classification (for example, see Song et al., 1988, Dally and Second, 1990, Wang et al., 1992). A variety of computer programs have also become available for generating phylogenetic

trees based on DNA information from different individuals, making the task of studying genome relationships more efficient and accurate.

Despite all the advances achieved in molecular genetic techniques and their application to the study of the systematics of plant species, to the author's knowledge none of the molecular approaches have been devoted to study genome relationships in the genus *Lotus*. Therefore, the purpose of this thesis is to make use of DNA polymorphisms to study genome relationships among different *Lotus* species in the *Lotus corniculatus* group and to use this information to aid in resolving the origin of the tetraploid species, *Lotus corniculatus*.

2. LITERATURE REVIEW

1. Genome relationships among *Lotus* species

Several studies have been made to help in elucidating the origin of the tetraploid species *L. corniculatus*. It has been proposed that *L. corniculatus* is an autotetraploid of *L. tenuis* (Dawson, 1941), or an autotetraploid of *L. alpinus* (Larsen, 1954). The analysis of artificial autotetraploids of these species have been carried out by Tome and Johnson (1945); Raelson and Grant (1988) showed that these artificial tetraploids do not resemble *L. corniculatus* in morphology or fertility. Evidence suggests that *L. corniculatus* is indeed an allotetraploid involving at least two species. Stebbins (1950) proposed that *L. corniculatus* is a segmental allotetraploid. Somaroo and Grant (1972) proposed that *L. japonicus* and *L. alpinus* could be the ancestral species, based on the fact that the artificial amphidiploid (*L. japonicus* x *L. alpinus*)^2 can be easily crossed with *L. corniculatus*, producing progeny with high fertility and meiotic regularity. Ross and Jones (1985) have proposed that either *L. alpinus* or *L. tenuis* could be the maternal parent of *L. corniculatus*, since the hybrid matched both species for rhizobium specificity, a character that is inherited maternally. Raelson and Grant (1988) carried out an isoenzyme survey for different accessions of four diploid *Lotus* species, *L. alpinus*, *L. japonicus*, *L. tenuis* and *L. uliginosus*, and the tetraploid *L. corniculatus*, to determine whether isoenzyme

information could provide enough evidence to elucidate the origin of *L. corniculatus*; they found that the hybrids *L. japonicus* x *L. alpinus* and *L. tenuis* x *L. alpinus*, or their reciprocal cross, contained all the alleles found in *L. corniculatus*, whereas *L. uliginosus* was distinct from all the other species for several isoenzyme alleles at several loci.

Although these studies have shed some light on the question of the origin of the tetraploid, it is evident that further studies are necessary to solve some of the controversies. It is anticipated that the use of molecular techniques will help in elucidating some of these discrepancies.

2. Methods based on Restriction Fragment Length Polymorphism (RFLP)

In recent years, much progress has been achieved in using methods based on DNA markers for genetic studies in several animal and plant species. The most common method used for the detection of DNA polymorphisms has been the restriction fragment length polymorphisms (RFLPs) (Bishop and Skolnick 1980, Botstein et al. 1980, Soller and Beckmann 1983). RFLPs are a product of changes in the bases within a restriction enzyme target site, deletion or insertion within a restriction fragment, or rearrangement of DNA (Welsh et al. 1991). These polymorphisms are DNA segments which are originated from only one of the parents, and have been proven to be inherited in a Mendelian fashion. The major strength of RFLPs is that they have the

potential to identify an almost unlimited number of polymorphisms (Kan and Dozy, 1978; Botstein et al., 1980). Despite their recognized power, analyses based upon RFLPs are normally very expensive and time consuming, and difficult to be used in species with large and complex genomes (Weining and Langridge, 1991). Furthermore, the need to construct specific probes that hybridize to multiple tandem-repetitive sequences is another complicating factor in this method.

3. Methods based on the Polymerase Chain Reaction

The introduction of the polymerase chain reaction (PCR) has dramatically modified the procedures and results in molecular genetics. It started with its introduction by Saiki et al. (1985), followed by Mullis and Faloona (1987). An extensive review of the technique is provided in Erlich et al. (1991) and the main features are summarized here. The polymerase chain reaction involves *in vitro* enzymatic synthesis of several millions of copies of a specific DNA segment, based on the annealing and extension of two oligonucleotide primers that flank the target region in duplex DNA. A PCR cycle involves three steps: denaturation of the DNA, primer binding, and DNA synthesis. Each one of these steps is performed at a different temperature (typically 94° to 98°C, 37° to 65°C, and 72°C). After the denaturation of the DNA, each primer will hybridize to one of the two strands in a way that the extension from each 3' hydroxyl end will be

directed toward the other. Following annealing, the primers are extended on the template strand with a DNA polymerase. Each one of the newly synthesized DNA segments will become a template for the next round, resulting in the exponential accumulation of a discrete fragment with the terminus consisting of the 5' end of the primer. In about 20-30 cycles, it is possible to amplify the original sequence in the order of millions of copies (Van Brunt, 1990).

The discovery of a thermostable DNA polymerase (Taq) from the thermophilic bacterium *Thermus aquaticus* by Saiki et al. (1988) has significantly contributed to the proliferation of studies based on PCR, since it enables multiple rounds of DNA synthesis, involving successive heating and cooling, without the need to add fresh enzymes each time. This was a significant advance compared to the initial studies that used the Klenow fragment of *Escherichia coli* DNA polymerase I, which is inactivated by high temperatures used for strand separation. Furthermore, amplification with Taq DNA polymerase significantly increased the specificity of the reaction, allowing the detection of many amplifications of the PCR product as a single ethidium bromide-stained band on an electrophoretic gel (Saiki et al., 1988).

One of the main advantages of using methods based on PCR is that minute quantities of DNA can be used to obtain useful information, compared to the need for large amounts of high molecular size DNA that are required by methods based on Southern blotting (Southern, 1975).

Considerable advances have been made in studies of genome relationships using this powerful technique. The range of PCR applications is vast, including introduction of new nucleotide information via PCR primers, *in vitro* analysis of protein-DNA interactions (footprinting), mapping and sequencing in the Human Genome Project, gene expression, and molecular evolutionary studies in plants and animals (Erlich et al. 1991).

In plants, some new PCR applications have been reported in the literature. For example, Martin et al. (1991) isolated DNA sequences that are linked to important plant genes by PCR, in which synthetic primers were used to amplify random sequences from genomic DNA. The feasibility of the method is demonstrated by analyzing a pair of tomato near-isogenic lines that differ for a region on chromosome 5 that contains a gene *(Pto)* conferring resistance to *Pseudomonas syringae*. Of four amplified products that were further investigated, three were confirmed by segregation analysis to be tightly linked to the *Pto* gene. The linked sequences identified by this method can be very useful for detecting the presence of the target gene in plant populations.

Guilluy et al. (1991) took advantage of PCR capability to isolate sequences associated with stigma self-incompatibility in *Brassica oleracea*. They used PCR amplifications with primers synthesized to the *S6* cDNA sequence, to demonstrate the presence of mRNA homologous to the stigma *S*-locus gene (*SLG*) in anthers during early microsporogenesis.

Weining and Langridge (1991) used PCR to reveal and map polymorphisms in *Triticum aestivum*, *Hordeum vulgare*, and *Secale cereale*.

Rieseberg et al. (1992) used PCR-amplified chloroplast DNA fragments and subsequent cladistic and phenetic analyses to reconstruct the evolutionary history of *Datiscaceae*.

One of the constraints of the methods based purely on PCR (Amplified Sequence Polymorphism, ASPs) is that they require specific DNA target information to design the amplification primers. Several strategies have been used to overcome this limitation. Welsh and McClelland (1990) and Williams et al. (1990) have described simple methods for assessing genomic variability, and for construction of genomic maps, based on amplification of genomic DNA with single primers of arbitrary nucleotide sequence. These primers have been shown to detect polymorphisms in the absence of specific nucleotide sequence information in DNA from bacterial, fungal (*Neurospora crassa*), human, and plant origins. The terms RAPD (random amplification of polymorphic DNA) (Williams et al., 1990) and AP-PCR (arbitrary primed polymerase chain reaction) fingerprinting (Welsh and McClelland 1990) have been applied to identify these techniques. RAPD and AP-PCR differ only in the method of primer choice used; in RAPD, the oligonucleotides are randomly produced, whereas in AP-PCR the oligonucleotides are chosen arbitrarily from among specifically designed primers. Since RAPD does not require any DNA target information, it becomes

much less time and cost consuming than those solely based on PCR (Williams et al., 1990). The observed fingerprints for a given DNA sample will depend on the length and sequence of the primer, as well as on the optimization of reaction conditions such as reagent concentrations and annealing temperature. Since the likelihood of a primer binding to genomic DNA is $(1/4)^n$, where n is the length in bases of the primer, oligonucleotides with shorter sequences of base pairs, such as used in RAPD, will provide a higher probability for priming events. Anollés et al. (1991) have reported DNA amplification with arbitrary primers as short as 5 nucleotides. They were able to produce detailed and relatively complex DNA profiles and detect DNA polymorphisms from a wide variety of organisms.

In the past two years several analyses have been performed using RAPD-PCR, to solve genome relationships among different individuals. Lankhorst et al. (1991) have adapted a RAPD assay for tomato, using a set of 11 oligonucleotide decamer primers. The potential of the original RAPD assay to generate polymorphic DNA markers with a given set of primers was further increased by combining two primers in a single PCR. Comparisons of fingerprints of *Lycopersicon esculentum*, *L. pennellii*, and the *L. esculentum* chromosome 6 substitution line LA 1641, which carries chromosome 6 from *L. pennellii*, three chromosome 6-specific RAPD markers could be directly identified among the set of amplified DNA fragments. One of the RAPD markers

was found to be linked to the nematode resistance gene *Mi*, which was considered valuable for further studies in menatode resistance.

Crowhurst et al. (1991) have used RAPDs to assess genome variability between 21 isolates from *Fusarium solani f. sp.* cucurbitae races 1 and 2. Based on RAPD marker patterns the isolates fell into two distinct groups, corresponding to mating populations MPI and MPV. Four isolates that could not be assigned to one or the other mating population by traditional means were distinguished by RAPD patterns, showing the value of this method.

Welsh et al. (1991) have determined the usefulness of AP-PCR in varietal identification and parentage determination of six maize inbreds and five hybrids derived from these inbreds. Results suggest that not only were they successful in their approach, but also that single-gel experiments could be carried out to identify the different varieties.

Yao and Sink (1991) report progress towards mapping the asparagus genome using RFLP and RAPD markers. They have identified nine random primers screened in PCR which show genotype specific polymorphism. These polymorphic DNA segments were amplified and visualized by electrophoresis, excised from the gel, and reamplified using the same primer. Following, reamplification, the DNAs were labeled with ^{32}P and used as probes in RFLP. Two of the random primers evaluated showed specific polymorphisms between *Asparagus officinalis* and *Asparagus densiflorus* that may be useful as selection

markers for somatic fusion products between these two species.

Hu and Quiros (1991) used RAPD markers to identify broccoli and cauliflower cultivars from different seed merchants. They found greater differences for each crop between cultivars from different seed companies than from the cultivars sold within the same company.

Quiros et al. (1991) reported the application of RAPD markers for genetic studies in *Brassica*. Of a total of 65 markers, 16 were determined to be in the A genome, 37 in the B genome, and 12 in the C genome. Of the 37 B-genome-specific markers, 11 were mapped on four independent chromosomes of *B. nigra* with the aid of a *B. napus-nigra* disomic alien addition line.

Weeden (1991) evaluated the potential of RAPD markers to augment studies on genetic mechanisms in plants. In his survey he examined species of the following genera: *Pisum, Lens, Cicer, Lathyrus, Phaseolus, Lupinus, Cucumis, Brassica, Silene, Vitis,* and *Malus*. The frequency of observed polymorphisms correlated well with other measures of genetic diversity. However, he found a significant noise level in RAPD data that interfered with the quantification of the diversity.

No molecular systematic studies have been carried out in the genus *Lotus* and a preliminary exploration is attempted in this thesis.

3. MATERIALS AND METHODS

The source of material for the present study was obtained from the world *Lotus* collection maintained by W. F. Grant on the Macdonald Campus of McGill University. A complete list of the species and accession numbers are presented in Table 1. The molecular analyses were performed according to the following procedures:

DNA EXTRACTION - The method described in Appels and Dvorak (1982), modified by Doyle and Beachy (1985), was used for DNA extraction. In this procedure, 0.19 g of fresh leaf tissue is ground in liquid nitrogen in a mortar and pestle. Then 1 ml of extraction buffer (0.1 *M* NaCl, 0.1 *M* EDTA, 50 m*M* Tris-HCl pH 7.0) is added. The slurry is extracted with two volumes of phenol-chloroform (1:1) and precipitated with ethanol. RNA is eliminated from the precipitate by treatment with DNAse-free pancreatic RNAse. Ethanol precipitation is used to collect RNA-free DNA. In soybean, this procedure yielded from 50 to 100 mg of DNA per 0.19 g of fresh leaf tissue (Doyle and Beachy, 1985).

Taq AND DNA OPTIMUM CONCENTRATIONS - A series of assays were carried out to evaluate both Taq polymerase and DNA concentration. For each species/accession, several assays were carried out to determine DNA concentrations in order to obtain between 20 and 25 ng of genomic DNA. This

was accomplished by comparing the dilutions to a known genomic DNA standard after separation on a 1.2% agarose gel and staining with ethidium bromide. After the optimum DNA concentrations were established for each species/accession, the effect of the following Taq polymerase concentrations in *Lotus* were evaluated: 1.0 unit, 1.3 units, 1.5 units, and 2.3 units.

PRIMERS - The arbitrary decamer oligonucleotides were obtained from Operon Technologies, Inc. (Alameda, CA, USA). The primers were used for the amplification of random DNA sequences. The nucleotide sequence of each primer is shown in Table 2. The criteria used to select these primers was the CG content which was between 50 and 70%.

AMPLIFICATION CONDITIONS - The amplification reactions were performed under conditions similar to those used by Williams et al. (1990). A volume of 25 µl was used, containing 10 m*M* Tris-HCl, pH 8.3, 50 m*M* Kcl, 2 m*M* $MgCl_2$, 100 µ*M* each of dATp, dCTp, dGTp, and genomic DNA, and 1.0 unit of Taq DNA polymerase (GIBCO BRL). The amplifications were carried out in a Hybaid DNA Thermal Cycler, programmed for 45 cycles of 1 minute at 94°C to denature, 1 minute at 35°C for the annealing primer, and 2 minutes at 72°C for extension. Tubes containing all the reaction components except for the genomic DNA, were included as a control when a new primer was used. After the cycling was completed, 15 µl of each sample was run in 1% agarose gel and visualized in the presence of ethidium bromide. The molecular

Table 1 - List of species and accessions that were analysed.

Species	Accession number	Chromos. number	Origin
L. tenuis Waldst. et Kit.	131	12	USDA, Ames, Iowa. Origin: Turkey.
	222	12	Plant Introduction Station, Geneva, N.Y. P.I. N° 246731. Origin: Spain.
	296	12	P.I.S., Geneva, N.Y. P.I. N° 229569. Origin: Greece.
	298	12	P.I.S., Geneva, N.Y. P.I. N° 251148. Origin: Yugoslavia.
L. uliginosus Schkuhr	865	12	Museum of Natural History, Paris, France, cultivated. Origin: St. Côme d'Olt, Aveyson, France.
	854	12	Origin: Niedersachsen, Germany.
	201	12	Hortus Botanicus, Coimbra, Portugal.
	858	12	Origin: Hessen, Germany.
L. corniculatus L.	554	24	P.I.S., Geneva, N.Y. Origin: Samsun, Turkey.
	279	24	P.I.S., Geneva, N.Y. P.I. N° 161878, Argentina.
	247	24	Origin: Loir-et-Cher, Herbault, France.
	811	24	P.I.S., Geneva, N.Y.. P.I. N° 380896. Origin: Iran.

Table 1, continued

Table 1, continued

Species	Accession number	Chromos. number	Origin
L. japonicus (Regel) Larsen	129	12	Origin: Gifu,Japan Collector: I. Hirayoshi.
	557	12	Origin: Iwaya, Awaji, Hyogo Prefecture, Japan. Collector: Y. Huziwara.
	581	12	Origin: Jima Islands, Japan. N. Satomi.
L. alpinus Schleich.	77	12	Origin: Valley of Emaney, Swiss Alps. Collector: C. Favarger.
L. corniculatus cv. 'Leo'	B764	24	Emile Lods Research Station, Macdonald Campus, McGill University.

standard used was the *lambda* DNA digested by *EcoRI/Hind III*. The gels were photographed under UV light with Polaroid film 655 or 667.

DATA ANALYSES - The occurrence of polymorphisms detected by the analyses were scored as 1 and absences were scored as 0. Therefore, a sequence of 0's and 1's was generated for each primer/species. The number of polymorphisms among species was expressed in terms of genetic distance. The programs **Gendist, Neighbour,** and **Drawgram**, written by J. Felsenstein of the University of Washington, were used to construct a phenogram. **Gendist** calculates the genetic distance between individuals tested according to the method proposed by Nei (1972). This method makes the following assumptions: 1) the differences between populations arise from genetic drift; 2) all loci have the same rate of neutral mutation; 3) the genetic variability in the population is initially at equilibrium between mutation and genetic drift. The genetic distance D is given by the following expression:

$$D = -\ln\left(\frac{\sum_m \sum_i p_{1mi}\, p_{2mi}}{\left[\sqrt{\sum_m \sum_i (p_{1mi})^2} \right] \left[\sqrt{\sum_m \sum_i (p_{2mi})^2} \right]} \right)$$

where m is summed over loci, i over alleles at the m-th locus, and p_{1mi} is the frequency of the i-th allele at the m-th locus in population 1. According to the above specified assumptions, for a sample of many equivalent loci, Nei's genetic distance is expected to rise linearly with time.

The output of **Gendist** is used as an input to **Neighbor.** This last program constructs a tree by successive clustering of lineages, setting branch lengths as the lineages join. The tree does not assume an evolutionary clock, and therefore it is an unrooted tree. The average linkage method was chosen for clustering. The output consists of a tree file which is used as input into the program **Drawgram** to create the phenogram as shown in Figure 1 of the results.

Table 2. List of primers used and respective oligonucleotide sequence.

PRIMER	OLIGONUCLEOTIDE SEQUENCE 5' to 3'
H-03	A G A C G T C C A C
H-04	G G A A G T C G C C
H-05	A G T C G T C C C C
H-07	C T G C A T C G T G
H-12	A C G C G C A T G T
H-13	G A C G C C A C A C
H-14	A C C A G G T T G G
H-19	C T G A C C A G C C
G-02	G G C A C G G A G G
G-08	T C A C G T C C A C
G-09	C T G A C G T C A C
G-10	A G G G C C G T C T
G-12	C A G C T C A C G A
G-13	C T C T C C G C C A
G-14	G G A T G A G A C C
G-16	A G C G T C C T C C
G-17	A C G A C C A G C A
G-18	G G C T C A T G T G
G-19	G T C A G G G C A A
FCP4	T C C T A A G C G G

4. RESULTS AND DISCUSSION

4.1. Optimum DNA and Taq concentration

Figure 1 illustrates the visualization of the ideal DNA concentration tentatively obtained for each species. Approximately 120 assays were carried out to obtain the optimum concentration for each species. Figure 2 shows the results of Taq concentration for the species *L. corniculatus* accession 811. It can be observed that the concentrations of 1.3, 2.3, and 1.0 units give clear bands. This result was generally consistent for all the species/accessions, and therefore the concentration of 1.0 unit was chosen arbitrarily for the analyses.

4.2. Polymorphic bands

RAPD polymorphisms were evaluated using 20, 10-base primers for each species/accession (see Table 2). None of these primers had been previously selected to be used in *Lotus*. In order to detect if the amplification products were reliable polymorphics, three replications were performed for each primer. Several polymorphic bands were observed for each primer. Most primers produced multibanded 'fingerprints' with sizes ranging from 300 bp to 3000 bp. Figures 3, 4, and 5 illustrate results obtained for primers H-03, H-04, and H-05, respectively. Primers H-03 (Figure 3) and H-05 (Figure 5) reproduced bands for all accessions analyzed, whereas primer H-04 did not reproduce any band for

the accession 854 of *L. uliginosus*. The same results were obtained for three replications performed for each primer. The negative control (see lane 17) did not produce any bands in Figures 3 and 5, but did exhibit a very weak band in Figure 4. However, the latter band did not prevent the interpretation of the polymorphisms of the species. One clear difference can be detected, between the accessions for the species *L. uliginosus* (lanes 13-16) and the other species. These results are in agreement with the observations of the isoenzyme study of Raelson and Grant (1988), who excluded this species from the ancestry of *L. corniculatus*. As pointed out by those authors, such an exclusion would lead to some conflicts, since this is the only species that contains tannin (Ross and Jones, 1985) and the phenolic compound delphinidin (Harney and Grant, 1965), components that are also present in *L. corniculatus*. Further evidence for differences between these species was found by Crompton (1982) who conducted an electron micrographic survey of pollen morphology among species of the Loteae. Through principal component analysis of several morphological pollen characters for a number of taxa, he concluded that *L. uliginosus* is dissimilar to *L. tenuis* and *L. corniculatus*.

Polymorphisms were also detected within the species. For example, on lane 16, the accession 865 of *L. uliginosus* appears distinct from the other three accessions for this species on both Figures 3 and 5. Similarly, on Figure 3 and 5 lane 7 indicates a clear polymorphism for the accession 557 of *L. japonicus*,

which differs from the other two accessions for this species. It would be risky to infer that such accessions are wrongly classified based on the analyses of just a few primers. A more general analyses based on the performance of several primers combined is presented in the next section.

A series of assays including all primers for each species/accession was also carried out. Figure 6 illustrates the results obtained for *L. corniculatus* cv. 'Leo' using twenty different primers. It can be observed that primers G-02, G-09, and G-20 failed to produce bands. However, analyses performed for all species/accessions did not show any primer consistently yielding a smaller or larger number of bands.

4.3. Genetic distance analysis

The phenogram obtained for all primers combined is presented in Figure 7. The phylogenetic tree clusters species that are closely related and the topology of the tree portrays the relationships among the species/accessions. *Lotus alpinus* and *L. uliginosus* appear as the most distant species from *L. corniculatus*, followed by *L. tenuis* and *L. japonicus*. Accessions within the same species generally appear very closely related, which is good evidence for the quality of the experimental data. Apart from *L. alpinus*, the relationship among the species is similar to the arrangement of the species as determined by Raelson and Grant (1988). These results suggest that RAPDs could be more

sensitive than isoenzymes for phylogenetic studies in the genus *Lotus*. On the other hand, it should be pointed that only one accession of *L. alpinus* was included in the present study. Since variability within species cannot be ruled out, it would be advisable to include additional accessions of this species in future studies in order to confirm the present findings.

Dally and Second (1990) made a complete study of genome relationships in rice, using chloroplast DNA restriction patterns to study genetic distances which they compared to a cladistic-mutation analyses of the different accessions. They found that in general the results from both methods had a close congruence with each other.

The results presented here are not intended to be a final analyses of the genomic relationships among the *Lotus* species which have been studied in this thesis. This study represents a first step in using DNA markers as a tool to implement studies of molecular systematics in this large genus. The inclusion of several other species in the *L. corniculatus* group not considered in the present study could help in elucidating the origin of the tetraploid *L. corniculatus*. Overall the power of RAPDs to reveal polymorphisms among *Lotus* species is clearly revealed.

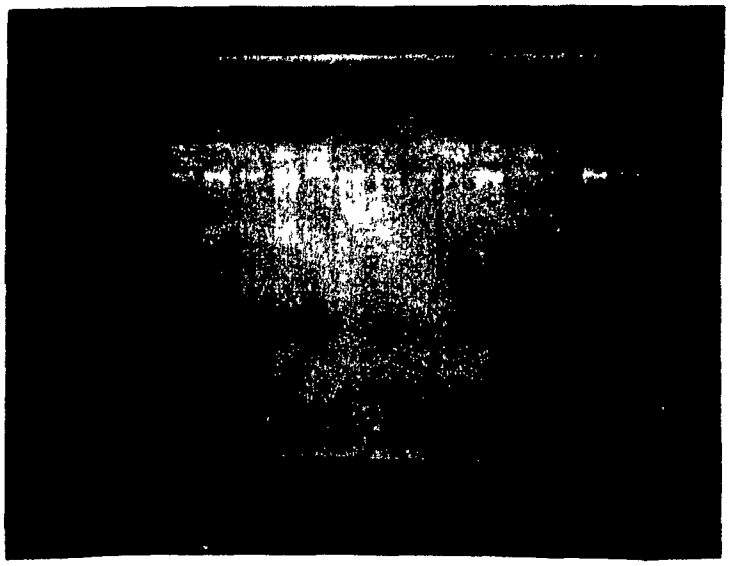


Figure 1. Visualization of optimum DNA concentration in 1.2% agarose gel. Note the similar response for all species/accessions. Labels for each lane are as follows: a, b= markers; 1 = *Lotus alpinus*, 2-5 = *Lotus corniculatus*, 6-8= *Lotus japonicus*, 9-12 = *Lotus tenuis*, 13-16 = *Lotus uliginosus.*

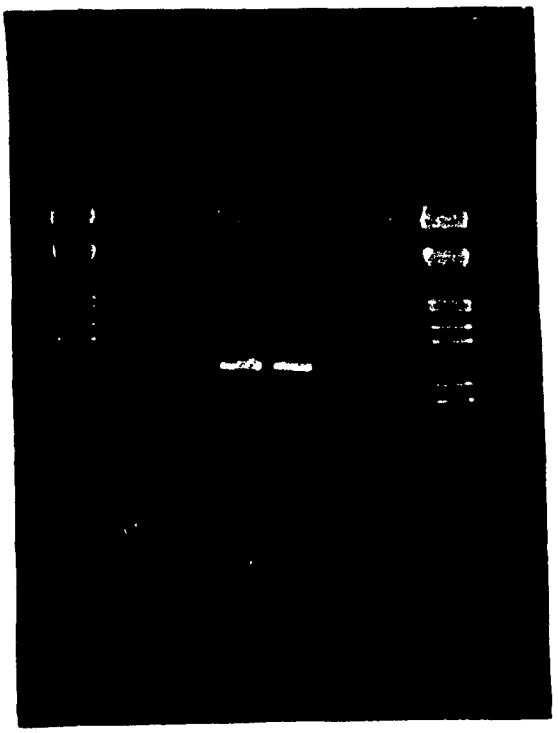


Figure 2. Visualization of optimum Taq concentration in 1.2% agarose gel. Lanes 1 and 8: *lambda* DNA, lanes 2 and 7: empty, lane 3: 1.3 units, lane 4: 2.3 units, lane 5: 1.0 unit, lane 5: 1.5 units.

Figure 3. PCR fingerprint of several accessions of *Lotus*, using the primer H-03. Labels for each lane are as follows: a, b= markers; 1 = *Lotus alpinus*, 2-5 = *Lotus corniculatus*, 6-8=*Lotus japonicus*, 9-12 = *Lotus tenuis*, 13-16 = *Lotus uliginosus*, 17 = negative control. Additional information is given in Table 1.

Figure 4. PCR fingerprint of several accessions of *Lotus* using the primer H-04. Labels for each lane are as follows: a, b= markers; 1 = *Lotus alpinus*, 2-5 = *Lotus corniculatus*, 6-8= *Lotus japonicus*, 9-12 = *Lotus tenuis*, 13-16 = *Lotus uliginosus*, 17 = negative control. Additional information is given in Table 1.

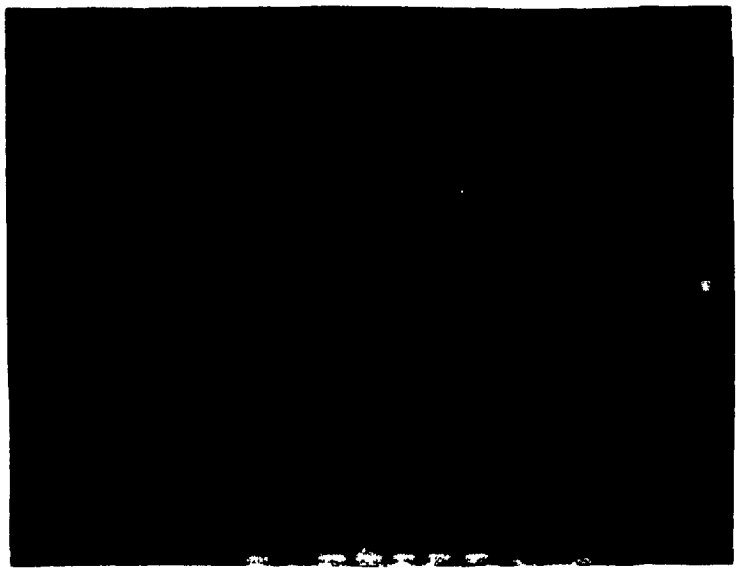


Figure 5. PCR fingerprint of several accessions of *Lotus* using the primer H-04. Labels for each lane are as follows: a, b= markers; 1 = *Lotus alpinus*, 2-5 = *Lotus corniculatus*, 6-8= *Lotus japonicus*, 9-12 = *Lotus tenuis*, 13-16 = *Lotus uliginosus*, 17 = negative control. Additional information is given in Table 1.

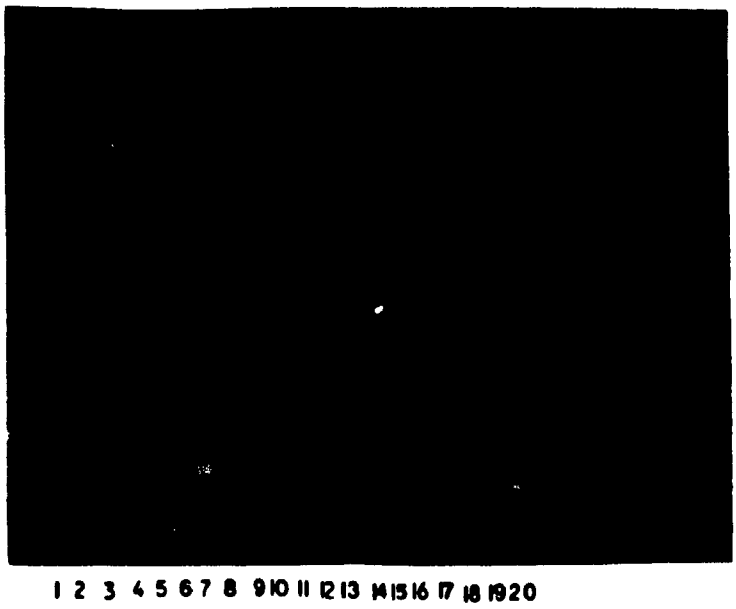


Figure 6. PCR fingerprint of *Lotus corniculatus* cv. 'Leo' using twenty different primers, as follows: lane 1 = G-02, lane 2 = FCP-4, lane 3 = G-08, lane 4 = G-09, lane 5 = G-10, lane 6 = G-11, lane 7 = G-12, lane 8 = G-14, lane 9 = G-17, lane 10 = G-18, lane 11 = G-19, lane 12 = G-20, lane 13 = G-13, lane 14 = H-03, lane 15 = H-05, lane 16 = H-15, lane 17 = H-14, lane 18 = H-16, lane 19 = H-04, lane 20 = H-13.

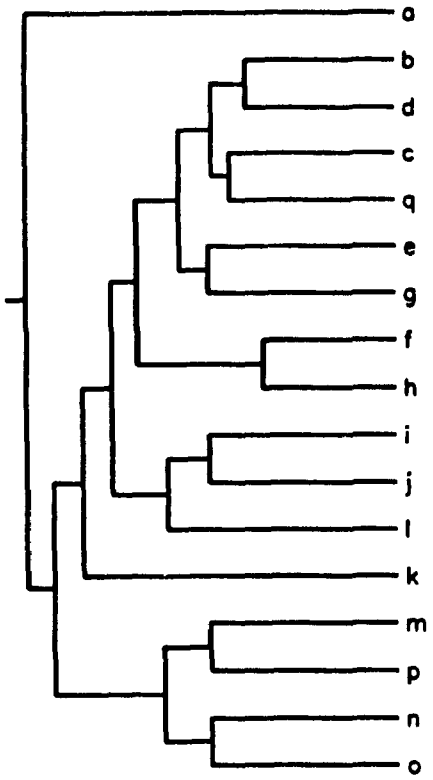


Figure 7. Phenogram showing the genetic distances among several species/accessions of *Lotus*. * Key to the letters: a = *Lotus alpinus*, b = *Lotus corniculatus* 247, c = *Lotus corniculatus* 279, d = *Lotus corniculatus* 554, e = *Lotus corniculatus* 811, f = *Lotus japonicus* 129, g = *Lotus japonicus* 557, h = *Lotus japonicus* 581, i = *Lotus tenuis* 131, j = *Lotus tenuis* 222, k = *Lotus tenuis* 296, l = *Lotus tenuis* 298, m = *Lotus uliginosus*, n = *Lotus uliginosus*, o = *Lotus uliginosus*, p = *Lotus uliginosus*, q = *Lotus corniculatus* cv. 'Leo'.

5. CONCLUSIONS

The use of RAPDs to study genome relationships among *Lotus* species was examined. Clear bands were obtained for all the accessions evaluated and several polymorphisms were detected. The species *L. uliginosus* showed sufficient differences to be excluded as an ancestral form of *L. corniculatus*, which confirms previous results from isoenzyme studies.

In this study *L. alpinus* is considered to be the most distantly related. While other studies have shown *L. alpinus* to be more closely related to *L. corniculatus*, this study would suggest that further analyses of *L. alpinus* are required and that the morphological features of these closely related taxa should be investigated in light of the various experimental evidence.

The study of additional accessions of these species and the use of an increased number of primers may provide a greater resolution of the relationships of these taxa. Additional species in the *L. corniculatus* group, not considered in this study should be investigated using RAPD markers.

It is considered that the use of RAPDs is an excellent technique for delimiting taxa in the genus *Lotus*.

6. REFERENCES

Anollés, G. C., Bassam, B. J., and Gresshoff, P. M. (1991). DNA amplification fingerprinting using very short arbitrary oligonucleotide primers. **Biotech. 9**:553-557.

Appels, R. and Dvorak, J. (1982). Relative rates of divergence of spacer and gene sequences within the rRNA region of species in the *Triticale*: Implications for the maintenance of homogeneity of a repeated gene family. **Theor. Appl. Genet. 63**:361-365.

Bishop, D. and Skolnick, M. H. (1980). Numerical considerations for linkage studies using polymorphic DNA markers in humans. Banbury Rep. **4**: Cancer Incidence in Defined Populations. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp 421-433.

Botstein, D., White, R. L., Skonic, M., and David, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphism. **Am. J. Hum. Genet. 32**:314-331.

Crompton, C. W. (1982). Pollen morphology of the tribe Loteae (Leguminosae) by light and scanning electron microscopy. M.Sc. Thesis, McGill University, Montreal, Canada.

Crowhurst, R. N., Hawthorne, B. T., Rikkeriuk, E. H. A., and Templeton, M. D. (1991). Differentiation of *Fusarium solani* f. sp. *cucurbitae* races 1 and 2 by

ramdom amplification of polymorphic DNA. **Curr. Genet. 20**:391-396.

Dally, A. M. and Second, G. (1990). Chloroplast DNA diversity in wild and cultivated species of rice (Genus *Oriza*, section *Oriza*). Cladistic-mutation and genetic-distance analysis. **Theor. Appl. Genet. 80**:209-222.

Dawson, C. D. R. (1941). Tetrasonic inheritance in *Lotus corniculatus* L. **J. Genet. 42**:49-72.

Doyle, J. J. and Beachy, R. N. (1985). Ribosomal gene variation in soybean (*Glycine*) and its relatives. **Theor. Appl. Genet. 70**:369-376.

Erlich, H. A., Gelfand, D., and Sninsky, J. J. (1991). Recent advances in the polymerase chain reaction. **Science, 252:**1643-1650.

Grant, W. F. (1965). A chromosome atlas and interspecific hybridization index for the genus *Lotus* (Leguminosae). **Can. J. Genet. Cytol. 7**:457-471.

Grant, W. F. and Marten, G. C. (1985). Birdsfoot trefoil. **In: Forages: Science of Grassland Agriculture**, 4th. ed., M. E. Heath, D. D. Metcalfe and R. F. Barnes, eds., Iowa State Univ. Press, Ames, Iowa. pp. 98-108.

Guilluy, C.-M, Trick, M., Heizmann, P., and Dumas, C. (1991). PCR detection of transcripts homologous to the self-incompatibility gene in anthers of *Brassica*. **Theor. Appl. Genet. 82**:466-472.

Harney, P. M. and Grant, W. F. (1965). A polygonal presentation of chromatographic investigation on the phenolic content of certain species of *Lotus*. **Can. J. Genet. Cytol. 7**:40-51.

Hu, J., and Quiros, C. F. (1991). Identification of broccoli and cauliflower cultivars with RAPD markers. **Plant Cell Rep. 10**:505-511.

Kan, Y., and Dozy, A. (1978). Antenatenal diagnosis of sickle-cell anaemia by DNA analysis of amniotic-fluid cells. **Lancet, 2**: 910-912.

Lankhorst, R. M. K., Vermunt, A., Weide, R. Liharska, T., and Zabel, P. (1991). Isolation of molecular markers for tomato (*L. esculentum*) using random amplified polymorphic DNA (RAPD). **Theor. Appl. Genet. 83**:108-114.

Larsen, K. (1954). Cytotaxonomical studies in *Lotus*. I. *Lotus corniculatus* L. sens. lat. **Bot. Tidsskr 51**: 205-211

Martin, G. B., Williams, J. G. K., and Tanksley, S. D. (1991). Rapid identification of markers linked to a *Pseudomonas* resistance gene in tomato by using random primers and near-isogenic lines. **Proc. Natl. Acad. Sci. 88**:2336-2340.

Mullis, K.B., and Faloona, F. (1987). Specific synthesis of DNA *in vitro* via a polymerase catalysed chain reaction. **Methods Enzimol. 155**: 335-350.

Nei, M. (1972). Genetic distance between populations. **Am. Nat. 106**:283-292.

Quiros, C. F., This, J. H. P., Chevre, A. M., and Delseny, M. (1991). Development and chromosomal locatization of genome-specific markers by polymerase chain reaction in *Brassica*. **Theor. Appl. Genet. 82**:627-632.

Raelson, J.V. and Grant, W.F. (1988). Evaluation of hypotheses concerning the origin of *Lotus corniculatus* (Fabaceae) using isoenzyme data. **Theor. Appl.**

Genet. 76: 267-276.

Riesberg, L. H., Hanson, M. A., and Philbrick, C. T. (1992). Androdioecy is derived from dioecy in *Datiscaceae*: Evidence from restriction site mapping of PCR-amplified chloroplast DNA fragments. **Syst. Bot. 17**(2):324- 336.

Ross, M.D. and Jones, W.T. (1985). The origin of *Lotus corniculatus*. **Theor. Appl. Genet. 71**:284-288.

Saiki, R.K., Scarf, S., Faloona, F., Mullks, K.B., Horn, G.T., Erlich, H.A., and Arnheim, N. (1985). Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle-cell anaemia. **Science, 230**:1350-1354.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G., Mullis, K. B., and Erlich, H. A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. **Science 239**:487-491.

Seaney, R.R. and Henson, P.R. (1970). Birdsfoot trefoil. **Adv. Agron. 22**:119-157.

Soller, M. and Beckmann, J. S. (1983). Genetic polymorphism in varietal identification and genetic improvement. **Theor. Appl. Genet. 67**:25-33.

Somaroo, B. H. and Grant, W. F. (1972). Crossing relationships between synthetic *Lotus* amphidiploids and *Lotus corniculatus*. **Crop Sci. 12**:103-105.

Song, K. M., Osborn, T. C., Williams, P. H. (1988). Brassica taxonomy on nuclear restriction fragment length polymorphisms (RFLPs). 1. Genome

evolution of diploid and amphidiploid species. **Theor. Appl. Genet. 75**:784-794.

Southern, E. M. (1975). Detection of specific sequence among DNA fragments separated by gel electrophoresis. **J. Mol. Biol. 98**:503-517.

Stebbins, Jr., G. L. (1950). Variation and evolution in plants. Columbia Univ. Press, New York and London. 643 pp.

Tome, G. A. and Johnson, I. J. (1945). Self- and cross-fertility relationships in *Lotus corniculatus* L. and *Lotus tenuis* Wald. et Kit.. **J. Am. Soc. Agron.** 37: 1011-1022.

Van Brunt, J. (1990). Amplifying genes: PCR and its alternatives. **Biotech. 8**:29-31.

Wang, Z. Y., Second, G., and Tanksley, S. D. (1992). Polymorphisms and phylogenetic relationships among species in the genus *Oriza* as determined by analysis of nuclear RFLPs. **Theor. Appl. Genet. 83**:565-581.

Weeden, N. (1991). Applications and potential for RAPD markers in plant genetics. **Am. J. Bot. 78**(6):90.

Weining, S., and Langridge, P. (1991). Identification and mapping of polymorphism in cereals based on the polymerase chain reaction. **Theor. Appl. Genet. 82**:209-216.

Welsh, J. and McClelland, M. (1990). Fingerprinting genomes using PCR with arbitrary primers. **Nucleic Acids Res. 18**:7213-7218.

Welsh, J., Honeycutt, R. J., McClelland, M., and Sobral, B. W. S. (1991). Parentage determination in maize hybrids using the arbitrarily primed polymerase chain reaction (AP-PCR). **Theor. Appl. Genet. 82**:473-476.

Williams, J.G.K., Kubelik, A., Livak, J., Rafalski, J.A., and Tingey, S.V. (1990). DNA polymorphism amplified by arbitrary primers are useful as genetic markers. **Nucleic Acids Res. 22**:6531-6535.

Yao, K. S., and Sink, K. C. (1991). Progress in mapping the asparagus genome by means of RFLP and RAPD markers. **Hortscience 26**:98.

APPENDIX

Following, is a more detailed description of the programs used in the phylogenetic analyses.

The program **Gendist** version 3.4 (copyright 1989-1991 by the University of Washington and by Joseph Felsenstein) computes genetic distances from gene frequencies. This program can be used to compute three different measures of genetic distance from a set of gene frequencies in different populations or species: Nei's genetic distance, Cavalli-Sforza's chord measure, and Reynolds, Weir, and Cockerham's genetic distance. In the present analysis Nei's genetic distance was chosen. The input data are presented in Table 3. It consists of the number of species evaluated, the number of loci, a line containing the numbers of alleles at each of the loci, and the gene frequences expressed by the absence or presence of polymorphysms expressed as sequences of 0's and 1's. The 20 primers are scored in sequence for each species/accession. The following settings were chosen within **Gendist**:

A Input file contains all alleles at each locus? One omitted at each locus

N Use Nei genetic distance? Yes

C Use Cavalli-Sforza chord measure? No

R Use Reynolds genetic distance? No

L Form of distance matrix? Square

Table 3. Input format of the data to be used in Gendist.

---

17 144*

2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2

2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2

2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2**

a*** 0 0 0 1 1 0 1 0 1 1 1 0 0 0 1 0 1 0 1 0 0 0 0 0 1 0 1 0 0 1 1 0 0 0 1 0 0 1 0 0 0 0 0 1 1

1 0 1 0 0 0 0 0 0 1 0 1 0 0 1 1 0 0 0 0 0 1 0 1 1 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 1 0 1 1 1 1 0 0 0

0 1 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 1 1 0 1 1 0 1 0 0 0 1 1 1 1 0 1 1 0 0 0 0 0 0 0 0 0 0 0

b 0 0 0 1 0 0 1 0 1 1 1 0 0 0 1 0 0 0 1 0 0 0 0 0 1 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1

1 0 0 0 0 0 0 0 1 1 0 0 0 0 1 0 1 1 0 0 0 0 1 1 0 1 0 1 0 0 0 0 0 0 0 1 0 1 0 1 0 1 0 1 0 0 0 1 0 0

0 0 1 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 1 0 1 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0

c 0 0 0 0 0 0 1 0 1 1 1 0 0 0 1 0 0 1 1 0 0 0 0 0 1 0 0 0 1 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0

0 0 0 0 0 0 0 1 1 0 0 0 0 1 1 1 0 0 0 1 0 1 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 0 1 0

0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0

d 0 0 0 1 0 0 1 0 1 1 1 1 0 0 1 0 1 0 1 0 0 0 0 0 1 0 0 0 1 1 0 0 0 0 0 1 0 1 0 0 0 0 0 1 0 1

1 0 0 0 0 0 0 0 0 1 0 0 0 0 1 0 1 0 0 0 0 0 0 1 0 0 0 1 0 0 0 1 0 0 0 0 0 1 0 0 0 1 0 0 0 0 0 1 1 0

0 0 1 1 0 1 0 0 0 0 0 0 1 1 0 0 0 0 1 0 0 1 0 1 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0

e 1 0 0 0 0 0 1 0 1 1 0 1 0 0 1 0 0 0 1 0 0 0 0 0 1 0 0 0 1 1 0 0 0 0 0 0 1 1 0 0 0 0 0 1 0 1

0 0 0 0 0 0 0 1 1 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0 1 1 0 0 1 0 1 0 0 0 0 0 1 0 0 0 1 0 0 0 0 0 1 0 1

0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

f 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 1 0 0 0 0 1 1 1 0 1 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0 0 1 1

0 0 0 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0 1 0 1 0 0 0 0 0 0 0 0 0 1 0 0 1 0 0 0 0 0 0 0 1 0 1 0

0 0 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 1 0 1 0 0 0 0 0 1 0 0 1 1 1 1 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0

g 1 0 0 0 0 0 1 0 1 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 1 0 0 0 0 1 0 0 1 0 0 0 0 0 0 0 0

0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 1 0 0 1 0 0 1 0 0 0 1 0 0 0 0 0 1 0 0

1000000000001000000101010000101111000010000000000

h 0000000010100010011100011011000100001010000001

1000000011000001000000100101000110100100100100101

0000000100001000000111000001001111000000000000000

i 1100001110100000001101001010000000000000100011

0000000110000101000000101010001001001011110100010100

0000000000010000001110000000011110010000000001000

j 1000011001101000011101011011000000000000100011

0000000010000101001000000000011000001000100000100000

0001000000010000001110000000011110011000011100000

k 1000010000000011000011010011100000000000010010

0000000010010000001000000000000001011010100000100000

0000010000000010000001000000001111000000001100010

l 1000001001101000000101000010000000000000100010

1000000100000000100100010101001000000011110000000000

0000000010010000001010001100001110000010000000001

m 1000000000000000001000000111010000000100000000

0101101001000000000001000000000000100000000000000

0000000000000000100100100010001111000000100000001

n 0010001000100000001000000111010000000000100000

0001100001100010000000000101100001010000001001000

0000000010101101100100100010000111000000000000000

o 0010001000000101000000101111010010000101000001

0001110001000000000000001001100001010010001000000

0010010000011101100000000000011111000000000000000

p 0000000000000010000000000101100000000000000000

0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 1 0 0 0 1 0 0 0 0 0 0 0

0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 1 0 1 0 0 0 0 0 0 1 0 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

q 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0

1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0

0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 1 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0

---

* number of species = 17, number of loci = 144

** number of alleles at each loci = 2

*** Key to letters: a = *Lotus alpinus*, b-e = *Lotus corniculatus*, f-h = *Lotus japonicus*, i-l = *Lotus tenuis*, m-p = *Lotus uliginosus*, q = *Lotus corniculatus* cv. 'Leo'.

Table 4. Square matrix of genetic distances generated by Gendist.

17

a 0.0000 0.2693 0.3351 0.2970 0.3448 0.4590 0.3448 0.3848
0.3351 0.3747 0.4700 0.4480 0.3951 0.4480 0.5155 0.4372 0.3848

b 0.2693 0.0000 0.1740 0.1576 0.2693 0.2785 0.2336 0.2513
0.1907 0.2970 0.3448 0.2877 0.3951 0.3646 0.4265 0.3351 0.1823

c 0.3351 0.1740 0.0000 0.1991 0.2076 0.3064 0.2076 0.2785
0.2693 0.3254 0.3747 0.3351 0.3064 0.3547 0.3951 0.3064 0.1740

d 0.2970 0.1576 0.1991 0.0000 0.2249 0.3646 0.2424 0.3547
0.2877 0.3254 0.3951 0.3951 0.4055 0.4159 0.3951 0.3848 0.2249

e 0.3448 0.2693 0.2076 0.2249 0.0000 0.2424 0.1991 0.2693
0.2785 0.3351 0.3254 0.3448 0.3159 0.4055 0.4055 0.2970 0.2513

f 0.4590 0.2785 0.3064 0.3646 0.2424 0.0000 0.2424 0.1415
0.2693 0.3254 0.3159 0.3351 0.3064 0.3747 0.4159 0.2513 0.2076

g 0.3448 0.2336 0.2076 0.2424 0.1991 0.2424 0.0000 0.2336
0.2249 0.2785 0.2877 0.3064 0.2603 0.3448 0.4055 0.2603 0.1991

h 0.3848 0.2513 0.2785 0.3547 0.2693 0.1415 0.2336 0.0000
0.1907 0.2785 0.3448 0.3646 0.3547 0.4055 0.4700 0.3351 0.2513

i 0.3351 0.1907 0.2693 0.2877 0.2785 0.2693 0.2249 0.1907
0.0000 0.1991 0.3351 0.2249 0.3646 0.3951 0.4812 0.3848 0.2785

j 0.3747 0.2970 0.3254 0.3254 0.3351 0.3254 0.2785 0.2785
0.1991 0.0000 0.2603 0.2603 0.3646 0.4159 0.4812 0.3646 0.3351

k 0.4700 0.3448 0.3747 0.3951 0.3254 0.3159 0.2877 0.3448
0.3351 0.2603 0.0000 0.3448 0.2970 0.4480 0.4265 0.2785 0.3064

l 0.4480 0.2877 0.3351 0.3951 0.3448 0.3351 0.3064 0.3646

	0.2249	0.2603	0.3448	0.0000	0.3159	0.3646	0.4480	0.2970	0.3064
m		0.3951	0.3951	0.3064	0.4055	0.3159	0.3064	0.2603	0.3547
	0.3646	0.3646	0.2970	0.3159	0.0000	0.2076	0.2785	0.1991	0.2424
n		0.4480	0.3646	0.3547	0.4159	0.4055	0.3747	0.3448	0.4055
	0.3951	0.4159	0.4480	0.3646	0.2076	0.0000	0.1991	0.2249	0.3254
o		0.5155	0.4265	0.3951	0.3951	0.4055	0.4159	0.4055	0.4700
	0.4812	0.4812	0.4265	0.4480	0.2785	0.1991	0.0000	0.2785	0.3254
p		0.4372	0.3351	0.3064	0.3848	0.2970	0.2513	0.2603	0.3351
	0.3848	0.3646	0.2785	0.2970	0.1991	0.2249	0.2785	0.0000	0.2076
q		0.3848	0.1823	0.1740	0.2249	0.2513	0.2076	0.1991	0.2513
	0.2785	0.3351	0.3064	0.3064	0.2424	0.3254	0.3254	0.2076	0.0000

---

* Key as in Table 3

M Analyse multiple data sets? No

O Terminal type (IBM PC, VT52, ANSI)? IBM PC

1 Print indications of progress of run? Yes

Using these settings **Gendist** will output a square matrix of genetic distances as shown in Table 4. The first line contains the number of species, and each subsequent line contains the genetic distances for each species. The output is ready to be used by **Neighbor**.

The **Neighbor** program (copyright 1991 by the University of Washington and by Joseph Felsenstein, written by Joseph Felsenstein) implements the Neighbor-Joining method and the UPGMA method of clustering. It constructs a tree by successive clustering of lineages, setting branch lengths as the lineages join. The tree is not rearranged thereafter. Since an evolutionary clock is not assumed, the generated tree is in fact unrooted. The following settings were chosen from the Neighbor menu:

N Neighbor-joining or UPGMA tree? UPGMA

O Outgroup root? No

L Lower-triangular data matrix? No

R Upper-triangular data matrix? No

S Subreplicates? No

J Randomize input order of sequences? No. Use input order

M Analyze multiple data sets? No

O Terminal type (IBM PC, ANSI, VT52)? IBM PC

1 Print out the data at start of run? No

2 Print indications of progress of run? Yes

3 Print out tree? Yes

4 Write out trees onto tree file? Yes

The UPGMA option constructs a tree by successive (aglomerative) clustering using an average-linkage method of clustering. The output consists of a file containing a tree and the length of the interior segments, as shown in Table 5.

The program **Drawgram** (copyright 1990-1991 by Joseph Felsenstein, written by Joseph Felsenstein) uses the output file from **Neighbor** to interactively plot a cladogram or phenogram-like rooted tree diagram. Several options available in this program allow the construction of the plot according to the particular needs of the user. In particular, adjustments can be made for the orientation of the tree and branches, style of tree, label sizes and angles, tree depth, margin sizes, stem lengths, and placement of nodes in the tree.

The following options from the main menu were used to generate the trees shown in Figure 5:

(1) Tree grows: Horizontally

(2) Style of tree: Phenogram

(3) Use branch lengths: Yes

(4) Angle of labels: 90.0

(5) Horizontal margins: 1.73 cm

(6) Vertical margins: 2.24 cm

(7) Depth/Breadth of tree: 0.53

(8) Stem-length/tree depth: 0.05

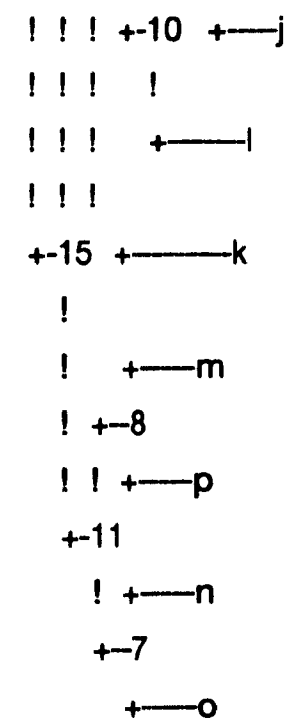
(9) Character ht / tip space: 0.3333

(10) Ancestral nodes: Intermediate

In the generated phenogram, nodes are connected to other nodes and to tips by a horizontal and then a vertical line, giving a very good idea of horizontal levels. Option (3) refers to whether the tree has branch lengths that are being used in the diagram. If the tree that was produced had a full set of branch lengths, it will be assumed as a default that the branch lengths are to be used in the diagram, unless the contrary is specified. Option (7) specifies the ratio between the depth and the breadth of the tree. It is initially set near 0.5, to approximate a v-shaped tree. Depth and breadth are described as if the tree grew vertically, so that depth is always measured from the root to the tip without including the labels. Option (10) controls the horizontal positions of the nodes and can greatly affect the appearance of the tree. It is not possible to control the vertical positions insofar as the use or non-use of branch lengths are specified. The intermediate option within (10) horizontally places the node halfway between its immediate descendants.

Table 5. Tree file generated using Neighbor and options specified in the text.

```
 17 Populations
Neighbor-Joining/UPGMA method version 3.42
 UPGMA method
 Negative branch lengths allowed
  +--------a*
  !
  !                 +---b
  !               +-2
  !               ! +---d
  !             +-4
  !             ! ! +-----c
  !             ! +-3
  !           +-9     +-----q
  !           ! !
  !           ! ! +-----e
  !         +-12 +-5
 -16         ! !   +-----g
  !        ! !
  !        ! ! +---f
  !     +-13 +-1
  !     ! !    +---h
  !     ! !
  !     ! !    +---i
  ! +-14 ! +-6
```



Between	And	Length
16	a	0.19510
16	15	0.01466
15	14	0.01521
14	13	0.01595
13	12	0.01258
12	9	0.02195
9	4	0.01720
4	2	0.01874
2	b	0.07880
2	d	0.07880
4	3	0.01054
3	c	0.08700
3	q	0.08700

9	5	0.01519
5	e	0.09955
5	g	0.09955
12	1	0.06594
1	f	0.07075
1	h	0.07075
13	10	0.02798
10	6	0.02175
6	i	0.09955
6	j	0.09955
10	l	0.12130
14	k	0.16523
15	11	0.05675
11	8	0.02414
8	m	0.09955
8	p	0.09955
11	7	0.02414
7	n	0.09955
7	o	0.09955

* Key as in Table 3