Characterisation of Genetic Diversity and Molecular Dissection of Seed Yield and Persistence in Swiss Mattenklee (*Trifolium pratense* L.)

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Doris Herrmann

von

Langnau BE

Promotionskomitee

Prof. Dr. Beat Keller (Vorsitz)

PD Dr. Catherine Feuillet

Dr. Roland Kölliker (Leitung der Dissertation)

Zürich, 2006

Table of contents

Summary

Red clover (*Trifolium pratense* L.) is an important component of permanent pastures and meadows as well as of grass-clover leys in temperate regions. Mattenklee landraces, i.e. persistent and locally adapted Swiss red clover landraces, and wild red clover populations may present particularly valuable genetic resources. However, landraces and wild populations are poorly characterised and the genetic origin of Mattenklee is largely unknown. Although through the last decades red clover cultivars improved for traits such as forage yield or disease resistance have been developed, cultivars often suffer from low seed yield and persistence. These two complex traits are influenced by many components, may be negatively correlated and have therefore proven to be difficult to improve. Molecular markers may help to rapidly characterise genetic resources. In addition, the development of markers closely linked to seed yield and persistence may help to complement and speed up traditional breeding procedures.

The first objective of this thesis was to determine genetic diversity and distinctness of Mattenklee landraces and wild clover populations and to investigate the potential ancestry of Mattenklee. The second objective was to identify quantitative trait loci (QTLs) for seed yield and persistence for the future development of molecular markers to apply in marker assisted breeding and to detect alternative traits linked to the target traits, which are easier to score, thus allowing for improved selection of seed yield and persistence.

In the first part of the project, eight Mattenklee landraces, eight Mattenklee cultivars and three field clover cultivars were characterised using 24 individual plants and amplified fragment length polymorphism (AFLP) markers. Furthermore, an optimised bulking strategy using AFLP markers and two bulked samples per population consisting of twenty plants each was applied to characterise 120 populations of six different red clover groups (Mattenklee landraces, Mattenklee cultivars, Swiss wild clover populations, field clover cultivars, Dutch wild clover populations and Dutch landraces). The results of these two studies revealed Mattenklee landraces as a distinct group showing a high genetic diversity, which was clearly higher than diversity of Mattenklee and field clover cultivars. The investigations based on bulked samples showed that Swiss wild clover populations are a highly diverse and distinct group, which was clearly separated from all other red clover groups. This separation emphasises that the ancestry of red clover landraces is primarily found in introduced germplasm rather than in natural wild clover populations. Mattenklee landraces and Swiss wild clover populations represent distinct and diverse genetic resources for the improvement of red clover cultivars and for conservation and restoration of diversity. In addition, the optimised bulking strategy may help to improve

diversity analyses not only in red clover populations, but also in other outcrossing grassland species.

In the second part of the project, a F_1 red clover population segregating for seed yield and persistence was established. Persistence, eight seed yield components (seed yield per plant, number of seeds per plant, seed yield per head, number of seeds per head, head number per plant, thousand-seed weight, percent seed set, time of flowering) and four additional morphological traits (number of stems, length of stem, width of medial leaflet, length of medial leaflet) were investigated in the field. A genetic linkage map using 42 simple sequence repeat (SSR) markers and 216 AFLP markers was constructed and QTL analyses for the above mentioned traits were performed. The genetic linkage map covered a total length of 444.2 cM and an average distance between loci of 1.7 cM was observed. A total of 57 QTLs were identified for the thirteen investigated traits including three QTLs for seed yield explaining together 33.8 % of the total variation and one QTL for persistence explaining 11.0 % of the variation. While there was no negative correlation between seed yield and persistence, some traits showed high positive correlations with the two target traits. Length of stem and thousand seed weight were both highly correlated with persistence. In addition, head number per plant and length of stem were highly correlated with seed yield and QTLs for these traits were often detected in the same genome region. These genome regions rich in QTLs represent valuable candidate regions for further characterisation and the development of marker closely linked to seed yield and persistence. The two key components head number per plant and length of stem were identified, which may facilitate further improvement of seed yield and persistence.

The findings of this thesis may contribute to the efficient management of genetic resources not only in red clover but also in other outcrossing species. A basis for the implementation of marker assisted breeding strategies and therefore for the efficient and progressive improvement of red clover breeding was provided. Moreover, this thesis forms a first step towards the systematic exploration of the red clover genome and may help to identify genes of economic interest.

Zusammenfassung

Rotklee (*Trifolium pratense* L.) bildet einen wichtigen Bestandteil von Natur- und Ansaatwiesen in gemässigten Zonen. Insbesondere die ausdauernde und lokal angepassten Schweizer Landsorten, so genannte Mattenklee-Hofsorten, sowie Wildklee Populationen von Rotklee könnten ein wichtiges Reservoir genetischer Diversität darstellen. Diese zwei Gruppen von Rotklee sind aber schlecht charakterisiert und auch die Herkunft der Mattenklee-Hofsorten ist weitgehend unbekannt. Obwohl gezielte Züchtung in den letzten Jahrzehnten zu verbesserten Rotklee Zuchtsorten bezüglich Ertrag und Krankheitsresistenz geführt hat, weisen diese Sorten oft noch Mängel bezüglich Samenertrag und Ausdauer auf. Diese zwei komplexen Merkmale sind von vielen Faktoren beeinflusst und eine negative Korrelation scheint zwischen diesen Merkmalen zu existieren. Samenertrag und Ausdauer sind deshalb zwei schwierig zu verbessernde Merkmale in der Rotklee Züchtung. Molekulare Marker, die mit diesen zwei Merkmalen gekoppelt sind, stellen eine viel versprechende Möglichkeit dar, traditionelle Züchtungsmethoden zu unterstützen. Ausserdem bilden molekulare Marker auch eine effiziente Methode zur Charakterisierung von genetischen Ressourcen.

Ziel des ersten Teils dieser Dissertation war es, Mattenklee-Hofsorten und Schweizer Wildklee Populationen genetisch zu charakterisieren und die Herkunft des Schweizer Mattenklees abzuklären. Ziel des zweiten Teils war die Identifikation von Genen, die die quantitativen Merkmale Samenertrag und Ausdauer kontrollieren (so genannte QTL, quantitative trait loci) sowie von korrelierten Merkmalen, die eine vereinfachte indirekte Selektion von Samenertrag und Ausdauer ermöglichen.

Im ersten Teil dieser Arbeit wurden je 24 Pflanzen von acht Mattenklee-Hofsorten, acht Mattenklee-Zuchtsorten and drei Ackerklee-Zuchtsorten mit Hilfe von AFLP (amplified fragment length polymorphism) Markern charakterisiert. Ausserdem wurde eine Methode zur Untersuchung von Mischproben mit Hilfe von AFLP Markern optimiert. Mit der optimierten Methode wurden zwei Mischproben mit je 20 Pflanzen pro Population verwendet um 120 Populationen von sechs verschiedenen Rotklee Gruppen zu charakterisieren (Mattenklee-Hofsorten, Mattenklee-Zuchtsorten, Schweizer Wildklee Populationen, Ackerklee-Zuchtsorten, Holländische Wildklee Populationen und Holländische Landsorten). Beide Diversitätsstudien zeigten, dass die Mattenklee-Hofsorten eine eigenständige Rotklee Gruppe bilden, die eine höhere Diversität aufweist als Mattenklee- und Ackerklee-Zuchtsorten. In der Diversitätsstudie, in welcher Mischproben untersucht wurden, konnte zusätzlich gezeigt werden, dass Schweizer Wildklee Populationen eine hohe Diversität aufweisen und sich am deutlichsten von den übrigen fünf Gruppen abgrenzen. Diese deutliche Abgrenzung weist darauf hin, dass der

Ursprung von Mattenklee nicht im Schweizer Wildklee zu finden ist. Mattenklee-Hofsorten sowie Schweizer Wildklee Populationen stellen eigenständige, vielfältige genetische Ressourcen für die Futterpflanzenzüchtung und für die Erhaltung der Diversität dar. Ausserdem kann die optimierte Methode von Mischproben zu einer Vereinfachung der Charakterisierung von fremdbefruchtenden Pflanzenarten beitragen.

Im zweiten Teil dieser Arbeit wurde eine F_1 Population, die für Samenertrag und Ausdauer aufspaltet, etabliert. Ausdauer, acht Samenertragskomponenten (Samenertrag pro Pflanze, Anzahl Samen pro Pflanze, Samenertrag pro Blütenstand, Anzahl Samen pro Blütenstand, Anzahl Blütenstände pro Pflanze, Tausendkorngewicht, Prozent Samenansatz und Blühzeitpunkt) und vier zusätzliche morphologische Merkmale (Anzahl Triebe, Trieblänge, Breite des mittleren Fiederblattes und Länge des mittleren Fiederblattes) wurden in einem Feldversuch untersucht. Eine genetische Karte wurde mit Hilfe von 42 SSR (simple sequence repeat) Markern und 216 AFLP Markern berechnet. Die genetische Karte wies total eine Länge von 444.2 cM auf und hatte einen durchschnittlichen Abstand zwischen zwei Markern von 1.7 cM. Für die 13 untersuchten Merkmale konnten insgesamt 57 QTLs lokalisiert werden. Dabei wurden für den Samenertrag drei QTLs, die insgesamt 33.8 % der Gesamtvarianz erklärten, und für Ausdauer ein QTL, der 11.0 % der Gesamtvarianz erklärte, nachgewiesen. Während keine negative Korrelation zwischen Samenertrag und Ausdauer aufgezeigt werden konnte, wiesen diese zwei Merkmale hohe positive Korrelationen zu anderen Merkmalen auf. Für Ausdauer zeigte sich eine hohe Korrelation zu Trieblänge und Tausendkorngewicht. Für Samenertrag konnte eine hohe Korrelation zu Anzahl Blütenstände pro Pflanze und Trieblänge nachgewiesen werden. Ausserdem wurden oft QTLs von diesen Merkmalen in den gleichen Regionen des Genoms lokalisiert. Diese Regionen, in denen viele QTLs nachgewiesen werden konnten, bilden eine gute Basis für die Entwicklung von Markern, die eine enge Kopplung zu Samenertrag und Ausdauer aufweisen. Trieblänge und Anzahl Blütenstände pro Pflanze konnten als zwei einfach zu untersuchende Merkmale bestimmt werden, die eine indirekte Selektion für Samenertrag und Ausdauer erlauben.

Die Ergebnisse dieser Dissertation können zu einer effizienten Erhaltung von genetischen Ressourcen von Rotklee und von anderen Fremdbefruchtern beitragen. Eine gute Basis für die markergestützte Züchtung und dadurch für eine effiziente und fortschrittliche Rotkleezüchtung wurde geschaffen. Ausserdem bildet diese Dissertation einen wichtigen ersten Schritt in Richtung systematischer Untersuchung des Rotklee Genoms und kann zur Identifizierung von wichtigen Genen beitragen.

1 General Introduction

1.1 Red clover - an important forage crop

Grasslands, including savanna, shrubland, temperate grassland and tundra, cover about 40 % of the global land area (White et al. 2000). Besides their importance for producing forage for domestic livestock, grasslands are complex and highly diverse ecosystems, which form habitats for many plant and animal species. At least fifteen percent of the centres of plant diversity are found in grassland areas including natural habitats of ancestors of major cereals such as wheat (*Triticum aestivum*) and rice (*Oryza sativa*), which harbour important genetic resources for improving cultivated crop species (White et al. 2000). Red clover (*Trifolium pratense* L.) is an important component of temperate grasslands, which cover about eight percent of the global land area. In Switzerland approximately 40 % of the total surface is covered with permanent pastures and meadows as well as with grass-clover leys (Bundesamt für Statistik 2004).

1.1.1 Distribution and performance

Red clover is abundant in most temperate regions in the world. It is found in Europe from the Mediterranean region to North Scandinavia, in most of North America, in South America north from Chile and Argentina to higher elevations in Peru and Mexico and occurs in parts of New Zealand, Australia, China and Japan (Taylor and Quesenberry 1996). Red clover is adapted to a wide range of soil types and environmental conditions, e.g. it tolerates annual rainfall ranging from 310 to 1910 mm (Smith et al. 1985). It is of high ecological value due to its ability to fix atmospheric nitrogen and to furnish fixed nitrogen to other plants. N-fixation of red clover may reach up to 373 kg ha^{-1} in grass-clover mixtures (Boller and Nösberger 1987). Furthermore, red clover generally increases total dry matter yield of pastures and meadows by up to 2 t ha^{-1} and may improve forage crude protein content by 1 to 3 % in mixture with grasses when compared to pure grass stands (Taylor and Quesenberry 1996).

1.1.2 Systematics and genetics

The roughly 20’000 species of the family of Fabaceae are divided into three subfamilies: Mimosoideae, Caesalpinioideae, and the numerically and economically dominant Faboideae (Papilionoideae). Red clover and several other important crops such as soybean (*Glycine max*) or pea (*Pisum sativum*) as well as the model legume *Lotus japonicus* all belong to the same subfamily Faboideae, but to different tribes. The tribe Trifoliaea includes the genus *Trifolium* and the genus *Medicago*, which comprises for example the forage crop alfalfa (*Medicago sativa*) and the model legume *Medicago truncatula* (Fig. 1.1; Choi et al. 2004).

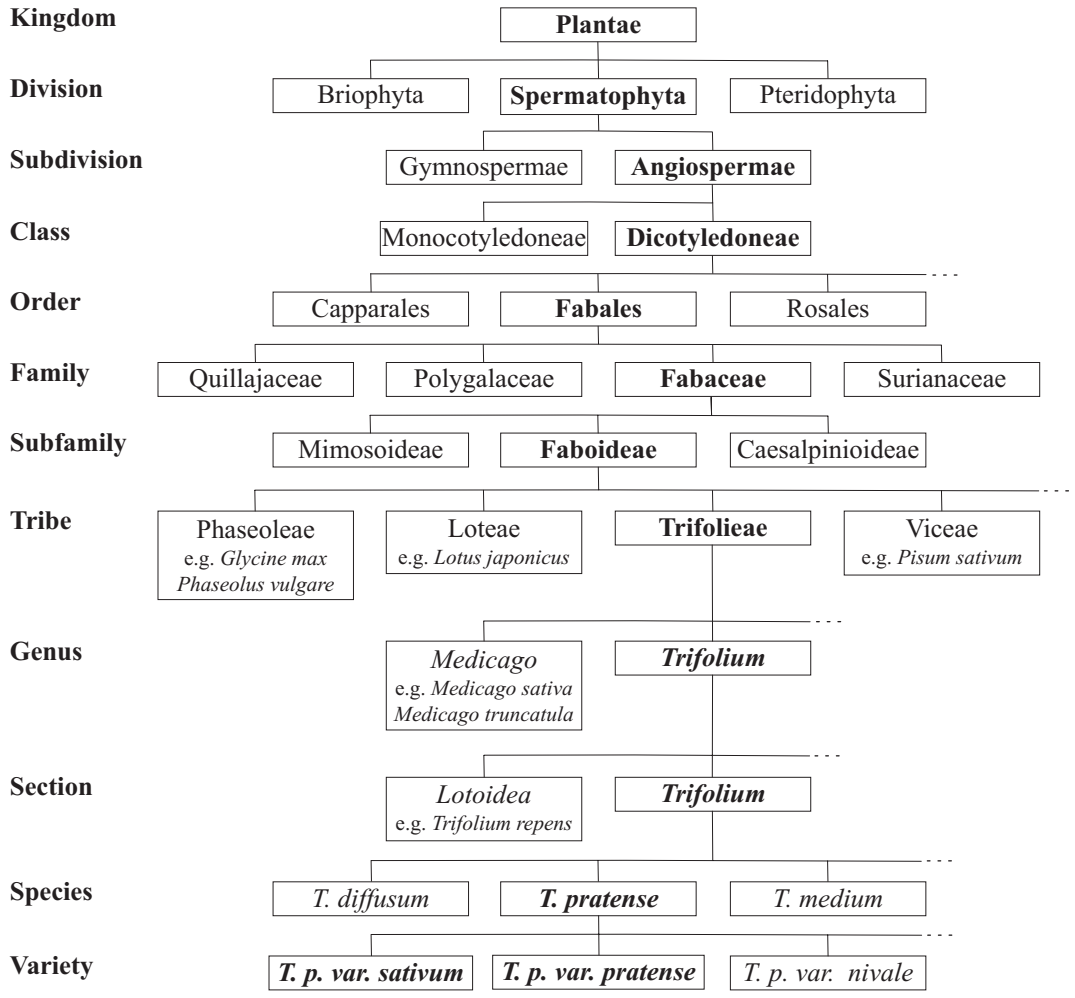


Fig. 1.1 Taxonomic relationship of the family of Fabaceae. Key taxa and species mentioned in the thesis are listed (Baltisberger 1995; Choi et al. 2004; Zohary and Heller 1984)

The genus *Trifolium* comprises approximately 250 annual and perennial species including red clover and other species of agricultural importance, such as white clover (*Trifolium repens*) or alsike clover (*Trifolium hybridum*). The genus is divided into eight sections: red clover *(Trifolium pratense)* belongs to the section *Trifolium* together with species such as diffuse clover (*Trifolium diffusum*) and zigzag clover (*Trifolium medium*), whereas white clover belongs to the section *Lotoidea* (Fig. 1.1; Zohary and Heller 1984).

Although several varieties of the species *Trifolium pratense* (*T. p.*) have been identified, the taxonomy of the species is complex and remains unclear (Hess et al. 1970; Zohary and Heller 1984). Cultivated red clover is generally referred to as the variety *T. p.* var. *sativum*, whereas in Switzerland two varieties of wild clover are distinguished: *T. p*. var. *nivale* with pink or white flowers in the alpine or subalpine area and *T. p.* var. *pratense* found in permanent pastures and

meadows of Swiss lowlands with pink to red flowers (Fig. 1.1; Hess et al. 1970; Zohary and Heller 1984).

Red clover is a perennial, diploid (2n = 2x = 14) and outcrossing species with a gametophytic self incompatibility system determined by one locus (Taylor and Quesenberry 1996). This S locus is characterised by up to 200 different alleles, which is higher than the number of S-alleles in most other self incompatible species (Lawrence 1996). One of these alleles, the S_f allele, induces self-fertility and is dominant over the other S-alleles (S_n). When heterozygous, self fertile plants (S_f S_n) are crossed with unrelated self-incompatible plants (S_n S_n), progenies segregate approximately 1:1 for self-fertility and self-incompatibility as expected (Taylor and Quesenberry 1996). Although stocks of the S_f allele have been maintained through selection for several generations and are available for research, severe inbreeding depression minimises their application in red clover breeding (Taylor and Quesenberry 1996).

1.1.3 History of red clover cultivation and breeding

The centre of diversity of red clover is located in the eastern Mediterranean region from where wild red clover spread across most of western Europe (Taylor and Quesenberry 1996). A large number of wild regional accessions adapted to a wide variety of environmental conditions developed over time. After a long history as a wild plant, cultivation of red clover began in the third century AD in southern Europe, i.e. in Spain. From there it was introduced into northern Europe (e.g. Flanders 1566 or England 1645), into Russia (1776) and with colonists into the United States, Australia and most other temperate regions of the world. The value of red clover for replacing the fallow in crop rotation and therefore for furnishing nitrogen to subsequent crops and for producing feed for cattle was first realised by Flemish farmers, who soon became an example for entire Europe (Taylor and Quesenberry 1996).

For nearly 300 years, there were no targeted breeding efforts and red clover populations were only slightly improved for resistance and persistence by cultivation practices. Progress in plant physiology, phytopathology, and crop science made targeted breeding possible. At the beginning of the twentieth century, and most importantly between the two world wars, improved red clover cultivars were obtained by targeted breeding for resistance to several diseases and pests such as anthracnose (*Colletotrichum trifolii*) or nematodes (*Ditylenchus dipsaci*). However, as no reliable means for seed production were available, these improved cultivars were not available to farmers until about 1960, when seed of good quality became available in sufficient quantities (Taylor and Quesenberry 1996).

The development of red clover cultivation in Switzerland was similar to the development in the world. Soon after its introduction in northern Europe, red clover became a crucial component of

Switzerland's grassland agriculture. Traditionally, red clover was propagated by harvesting seeds on farm and sowing them again the following year. Because it was easier to harvest seeds in the third year than in the second year, when plants still have many leaves, there was a certain selection for persistence. This on farm seed production led to the development of many landraces with superior persistence, which are highly adapted to local climates and are commonly called 'Mattenklee landraces' (Boller 2000b). In the late 20th century, seed of improved red clover cultivars became easily available to farmers and there was no further need to produce seeds on farm. Consequently, Mattenklee landraces began to disappear (Boller 2000b). A collection was made on farms in 1971/1972 preserving around 100 Mattenklee landraces (Nüesch 1976), which may be a valuable source of germplasm for red clover breeding. However, no targeted and comprehensive characterisation of this collection was undertaken (Boller 2000b). In addition, the origin of Mattenklee landraces is unclear (Fig. 1.2). Knowledge on the origin of germplasm is important in order to avoid redundancy in genetic resources and to optimally manage and exploit genetic diversity. In phenotypic evaluations, Mattenklee have been compared to Swiss wild clover populations as well as to cultivated clovers of different European countries. Some traits such as time of flowering and number of internodes suggested no close relationship with Swiss wild clover populations, but a close relationship with cultivated clover (Nüesch 1960). Cultivated clover was imported into Switzerland in the eighteenth century from Brabant and Flanders, two regions in Belgium and the Netherlands (Fig. 1.2). Therefore, this represents a likely source of starting material from which Mattenklee landraces may have developed. However, inter breeding with wild clover populations may also have occurred during landrace establishment (Fig. 1.2; Nüesch 1960).

Breeding of forage crops began in Switzerland about one hundred years ago. However, only since 1950 different Mattenklee landrace collections have been used to develop Mattenklee cultivars (Fig. 1.2). The diploid cultivar Changins (1964) was developed, which was the basis for the tetraploid cultivar Vanessa (1988). Changins is based on Mattenklee landraces collected around 1950 by Samuel Badoux (Boller 2000b). Furthermore, a few landraces of the preserved Mattenklee landraces collected in 1971/1972 were used to develop the cultivar Formica (1993), which is characterised by a low oestrogen content. However, most of the Swiss cultivars are based on the diploid cultivar Renova (1964), which was selected from Mattenklee landraces collected around 1950 by Bruno Nüesch (Boller 2000b). Duplication of the chromosome set of Renova led to the still widely used tetraploid cultivar Temara (1972) and a cross between Renova and the Mattenklee landrace Rütti to the diploid cultivar Rüttinova (1984). Further selection for resistance to crown rot (*Sclerotinia trifolium*) or nematodes resulted in the diploid cultivars Milvus (1993), Corvus (1998), Pica (1998), Pavo (2002) and Merula (2002) as well as

the tetraploid cultivars Larus (1998) and Astur (1998). Some of these cultivars are not only cultivated in Switzerland but also in different European countries (e.g. Milvus in France or Temara in France, Germany and Austria; Boller 2000b). Swiss Mattenklee cultivars are characterised by an early time of flowering, high resistance to anthracnose or crown rot and an improved persistence resulting particularly in a high forage yield in the second production year. However, Mattenklee cultivars show low seed yield and often a lower forage yield in the first production year, particularly for the first cut, when compared to European field clover cultivars (Bundessortenamt 2003; Deneufbourg 2004; Lehmann and Briner 1998).

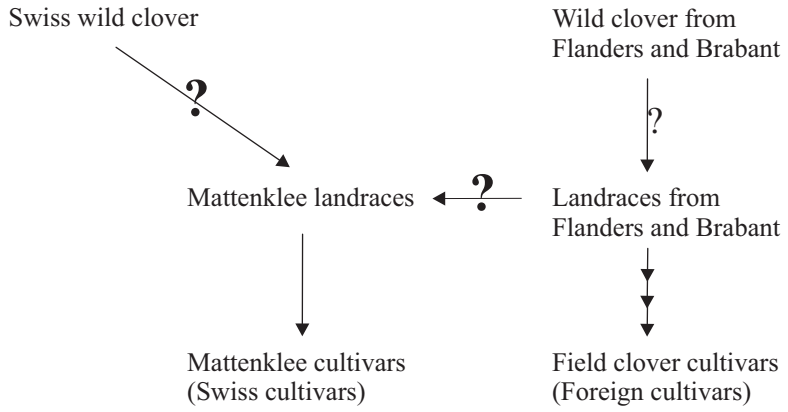


Fig. 1.2 Relationship among the six red clover groups Mattenklee landraces, Mattenklee cultivars, Swiss wild clover populations, field clover cultivars and wild clover populations and landraces from Flanders and Brabant. Question marks indicate that relationship between these two groups is unclear

1.1.4 Objectives of red clover breeding

High forage yield of good quality is the main objective in red clover breeding. Forage of good quality includes a high content of crude protein as well as a good digestibility contributing to a well-balanced and high-class feed of cattle. Forage yield is influenced by many factors such as pest and disease resistance, persistence, high competition or adaptation to environments and therefore these factors play an important role in red clover breeding. However, as forage yield and forage quality are influenced by various factors, which often are negatively correlated, the major difficulty in breeding programs is to find a compromise between all these factors (Taylor and Quesenberry 1996).

1.1.4.1 Persistence and seed yield - two traits of major agronomic importance

Red clover does usually not sustain longer than three years and one of the major deficiencies of forage stands is the lack of temporally stable proportion of legumes. Legumes enhance the quality of forage and furnish fixed nitrogen to other plant species (Taylor and Quesenberry 1996). This is particularly important in organic farming systems, where nitrogen supply is often limiting. As relevance of organic farming or sustainable agriculture in general increased in the last decades, selection for persistence in red clover became more important. Persistence may be defined as the ability to produce constant forage yields over three or more growing seasons. Therefore, it does not only include the survival but also the performance of individual plants for three or more years. Persistence is the result of adaptation of the crop to its stress load. Stress may be defined as any factor, whether physiogenic (e.g. management system) or pathogenic (e.g. *Fusarium* resistance), that contributes to less than luxuriant growth (Taylor and Quesenberry 1996). Therefore, poor management practices such as insufficient lime or fertiliser applications may lead to inadequate stand maintenance. It was also reported that flowering in the year of sowing (Choo 1984) or frequent cutting (Wiersma et al. 1998) have a negative influence on persistence. However, for permanent enhancement of persistence, targeted breeding is necessary. Breeding for single factors influencing persistence such as *Fusarium* resistance, has not proven to be very successful. The most efficient way to breed for persistence is to expose plants to their environment and to select for survival (Taylor and Quesenberry 1996). Such an approach is very complex and very time consuming and may take twelve years before superior genotypes can be identified (Christie and Martin 1999).

Another trait of agronomic importance in red clover breeding is seed yield. An unsatisfactory seed yield leads to high seed production costs and limits the success of the cultivar in the marketplace. Improved cultivars often show a dissatisfactory level of seed yield, which was often only considered a by-product of forage production and was therefore neglected in breeding programs. Several studies investigated possibilities to improve seed yield by management practices such as irrigation (Oliva et al. 1994) or selection of soil types specifically favourable for seed production (Belzile 1991). However, for a steady improvement of seed yield, targeted selection in breeding programs is a promising option, but is complicated by several factors influencing this complex trait. Seed yield is quantitatively inherited and affected by many components (Montardo et al. 2003; Oliva et al. 1994). It can usually only be measured in the field in the second year of cultivation and measurement is time consuming. Moreover, seed yield was reported to be negatively correlated with important agronomic traits such as forage yield (Steiner et al. 1997). This negative correlation seems to be particularly pronounced for

persistence, as Swiss Mattenklee cultivars, continuously selected for persistence, in particular lack a good seed yield potential (Deneufbourg 2004; Lehmann and Briner 1998).
Over all, persistence and seed yield are two important traits in breeding, but are difficult and time consuming to improve. Molecular markers for persistence and seed yield would therefore represent a very helpful tool to complement traditional breeding procedures. Another prerequisite for improvement of red clover cultivars is to characterise the association of different seed yield components as well as the association of seed yield with other important agronomic traits such as persistence.

1.2 Diversity - a basis for breeding and ecosystem stability

1.2.1 Definition and relevance

The most prevalent usage of the term 'biodiversity' or 'diversity' is a synonym for the 'variety of life', or in some more details a definition given by the US congress Office of Technology Assessment (OTA) refers to diversity as: 'Biologic diversity refers to the variety and variability among living organisms and the ecological complexes in which they occur…' (Gaston 1996). Diversity can be measured at different levels, i.e. from individuals over populations to species, and with different methods, i.e. differences of genotypes or differences of phenotypes. Principally one can differentiate between inter-species diversity, i.e. diversity among species, and intra-species diversity, i.e. diversity within species (Gaston 1996).
Preservation of diversity is crucial for several reasons. Species added to an ecosystem may increase ecosystem stability, but the increase in function may increase more slowly as more species are included in the system (rivet hypothesis; Gaston 1996). Therefore, inter-species diversity may contribute to ecosystem stability and promote resistance to disturbance. In grasslands, primary productivity of diverse plant communities has been shown to be more resistant to and to recover better from drought stress than the productivity of less complex plant communities (Tilman and Downing 1994). Intra-species diversity, i.e. diversity within and among populations of a species, is reported in many studies to be correlated with fitness (Fischer and Matthies 1998; Menges 1991). In a study comparing 34 data sets of different plants, invertebrates and vertebrates the mean weighted correlation between genetic diversity and population fitness was 0.432. This correlation was highly significant and explained 19 % of the variation of fitness (Reed and Frankham 2003). Higher fitness may contribute to a consistent proportion of a species in a plant community and therefore intra-species diversity may also substantially influence ecosystem stability. Urbanisation and agriculture are strongly

transforming and affecting inter-species as well as intra-species diversity (White et al. 2000). For example, more intensive management practices in agriculture, such as more frequent defoliation or increased fertilisation, were reported to result in lower diversity of meadow fescue (Kölliker et al. 1998). Fragmentation of grasslands by structures such as roads can lead to genetically isolated and reduced population sizes with a decreased diversity and fitness (White et al. 2000).

Apart from the important role of diversity for ecosystem stability, intra-species diversity within and among populations may also play an important role in plant breeding. Genetic diversity of breeding material used for cultivar development may substantially influence the success of a resulting cultivar through various mechanisms such as heterosis, general combining ability, inbreeding depression or self incompatibility (Kölliker et al. 2005a). For example, in perennial ryegrass (*Lolium perenne*) two polycrosses of six parental plants with contrasting levels of genetic diversity were composed. It has been reported that progenies of the parental plants with high diversity had a 3.8 % higher forage yield compared to the progenies of the parental plants with a low diversity (Kölliker et al. 2005a). Therefore, maintaining or improving diversity of a cultivar in general may be important. A decrease of diversity by human selection, i.e. a decrease of diversity of cultivars compared to landraces or wild accessions, has been reported for species such as pepper (Hernandez-Verdugo et al. 2001) or meadow fescue (Kölliker et al. 1998). For the improvement of diversity in a breeding program specific plants within the breeding pool may be selected or diversity may be broadened by including landraces or wild populations. In addition, wild populations as well as landraces may be a source for traits such as resistance to biotic or abiotic stresses, forage quality or male sterility and can be integrated in breeding programs for the targeted improvement of a specific trait (Rao et al. 2003).

1.2.2 Characterisation - a prerequisite for preservation and utilisation

In order to maintain and conserve diversity, there are primarily two different methods used, i.e. ex-situ or in-situ conservation. Ex-situ conservation involves the conservation of diversity of plants away from their area of origin or development. This is often performed in seed gene banks, i.e. seeds in cold storage, but it includes also field gene banks as used for maintenance of apple varieties in orchards. In the last decades many large gene banks were developed (Becker 1993). However, there are two major restraints to an efficient use of the genetic resources. On the one hand, the conserved material is often insufficiently characterised and often redundant. On the other hand, the maintenance of gene banks is expensive and laborious, as seeds have a limited viability in cold storage and must be regenerated periodically (Becker 1993). In-situ conservation involves the management of diversity of agro- and natural ecosystems in the area

of origin or development. In-situ conservation allows for the continued evolution of accessions while conserving adaptive traits in a broad sense. In addition, the system can also increase the control of local communities over their genetic resources. As diverse natural ecosystems but also agro-ecosystems are often endangered, it is very important that valuable and diverse regions or territories are selected and protected allowing for an efficient and optimal management of in-situ conservation (Jarvis et al. 2000).

For the identification of diverse and valuable territories for in-situ conservation characterisation of diversity is a decisive prerequisite. Characterisation is also crucial for an efficient ex-situ conservation of diversity in gene banks. It enables the estimation of diversity within a population as well as of distinctness and relationship of populations and different groups such as wild populations or cultivars. Thus, redundancy in gene banks can be reduced, which will allow to minimise the number of samples and the cost of conservation. Furthermore, characterisation of diversity in gene banks and of in-situ conserved resources will allow the targeted utilisation of genetic resources in breeding. An efficient tool to characterise diversity in particular within a species but also among species are genetic markers. Genetic markers are specific chromosomal fragments within a genome, which presence and absence can be monitored, respectively. They allow for a rapid assessment of genetic diversity on the genome level and have been widely used to characterise genetic resources in various plant species (Beebe et al. 2001; Kubik et al. 1999).

1.3 Marker assisted selection – a concept to improve cultivars

Beside the efficient and reliable characterisation of diversity, markers may be a very helpful tool to complement traditional breeding procedures and may allow for a more efficient improvement of cultivars. The concept of marker assisted selection (MAS) is not to directly select for a target trait, but for a marker tightly linked with the target trait. The idea is to simplify and increase efficiency of selection as the marker is simpler and less laborious to identify or selection can be performed before the trait can be scored visually. The lower the distance between the marker locus and the target trait, the lower is the probability for recombination between these two loci and the more reliable the marker is. Therefore, in the ideal case the locus of the marker is identical with the locus of the gene controlling the target trait (Jain et al. 2002). The potential value of markers and their association with agronomic traits has been known for more than 80 years. The usefulness of MAS was recognised as early as 1923 when Sax demonstrated the association of seed size with seed coat pigmentations in beans (Sax 1923). However, for a long time, MAS was restricted by the limited number of morphological markers available. Molecular markers or DNA markers are numerous and their development represents a milestone in

genetics as they provide the capacity for complete coverage of crop genomes (Jain et al. 2002). Therefore, molecular markers present the basic prerequisite to identify tightly linked markers to a target trait. Further steps towards MAS are to align these markers in a high density linkage map and to map the trait relative to the molecular markers on this linkage map (Jain et al. 2002). Some notable examples for the successful use of MAS for qualitatively inherited traits were reported in major crops such as rice, where MAS was applied to pyramid resistance genes against bacterial blight (*Xanthomonas oryzeae* pv. *oryzeae*), blast (*Pyricularia grisea*) or gall midge (*Orseolia oryzae*, Jain et al. 2002). MAS also provides a new and promising opportunity to improve complex quantitatively inherited traits such as seed yield or persistence, which are difficult to improve through traditional breeding procedures. However, forage crop in general and red clover in particular lagged behind this development and basic as well as more detailed investigations are needed before MAS can be applied in breeding programs.

1.4 Tools for diversity analysis and marker assisted selection

1.4.1 Molecular markers

Markers are an efficient tool to characterise genetic diversity of species and are the basic prerequisite for the application of MAS. Basically, a marker is a specific chromosomal fragment within a genome representing either coding or non-coding DNA and whose inheritance can be followed (Staub et al. 1996). An ideal marker should be co-dominant, highly polymorphic, reliable, abundant, neutral and independent of the environment. There are different levels within an organism where markers can be detected. Phenotypic markers are detected on the morphological or physiological level, isozymes markers on the protein level and DNA markers, often called molecular markers, at the DNA level. Molecular markers have three major advantages when compared to phenotypic and isozyme markers. They are unaffected by the environment, detectable at all stages of plant growth and often highly polymorphic. The first type of molecuar markers in the form of restriction fragment length polymorphism (RFLP) had become available in 1980 (Botstein et al. 1980). The development of the polymerase chain reaction (PCR; Mullis and Faloona 1987) opened vast possibilities to develop new marker techniques. Molecular marker techniques can be divided in techniques which are either based on amplification of undefined or anonymous elements (anonymous markers) and therefore are independent of prior sequence knowledge, or in sequenced based techniques needing prior knowledge of sequences for the development (sequence based markers; Paterson 1998).

1.4.1.1 Anonymous marker techniques

Randomly amplified polymorphic DNA (RAPD)

RAPD markers are generated by PCR amplification of genomic DNA segments using single random primers. They have been widely used in diversity studies of many species (Campos-De-Quiroz and Ortega-Klose 2001; Gustafson et al. 1999) and also occasionally together with other marker systems to construct genetic linkage maps (Warnke et al. 2004). RAPD markers are rather inexpensive, because no prior sequence information is needed and a large number of markers can be generated using readily available primers. However, RAPD markers are dominant and suffer from severe reproducibility problems (Jones et al. 1997a), which ultimately limited their use in the last years.

Amplified fragment length polymorphism (AFLP)

In 1995, Vos et al. described AFLP as a new technique for DNA fingerprinting. DNA is first digested with a rare cutting enzyme (e.g. *Eco*RI) with a recognition site of 6 bp, and a frequent cutting enzyme (e.g. *Mse*I) with a recognition site of 4 bp. Then, two oligonucleotides, called adapters, with a core sequence and overhangs complementary to the cut ends are ligated to each end of the fragments. Specific primers to both adapter sequences are used for subsequent PCR amplifications. A first amplification, called preamplification, is used to enrich for fragments which have been restricted by both enzymes. In complex genomes, the primers used in the preamplification have one additional specific nucleotide, which reduces the number of fragments amplified. A second amplification, called selective amplification, is performed using diluted preamplification products and primers with typically one to three selective nucleotides. At last amplified fragments are separated and visualised using fluorescent tags, radioactive isotopes or silver staining. Polymorphisms between plants arise through point mutations, inversions, deletions or translocations, which lead to the loss or gain of restriction sites or a change of fragment length between restriction sites (Vos et al. 1995).

The AFLP technique has many advantages. It requires no prior sequence information and allows for the use of standard primers. The technique is highly reproducible and a large number of markers are screened in a single assay. However, AFLP markers are mainly dominant and the analysis is technically demanding (Jones et al. 1997a).

AFLP markers are particularly useful for diversity studies in species where little sequence information is available. They allowed for the rapid estimation of diversity within and among population of many species (Beebe et al. 2001; Kölliker et al. 2001; Sawkins et al. 2001). The dominant nature limits the application in construction of linkage maps. However, AFLP markers are very useful to fill up gaps in genetic maps, and are therefore optimal markers to construct maps in combination with other marker systems for instance in white clover in

combination with SSR markers (Jones et al. 2003) or in ryegrass in combination with RFLP, isozymes, and EST markers (Jones et al. 2002b).

1.4.1.2 Sequence based marker techniques

Restriction fragment length polymorphism (RFLP)

For the analysis of RFLP markers, DNA is first restricted, then fragments are separated and hybridised with labelled sequence specific probes, such as random genomic clones or cDNA clones. The advantages of RFLP are the codominant nature of the markers, their reliability and specificity. Furthermore, cDNA-based RFLP have the potential to detect homologous sequences in distantly related genomes, which makes them especially useful to construct consensus maps (Isobe et al. 2003). The major drawbacks of RFLP markers are that probes need to be developed, the technique is labour and time consuming and large amounts of high quality DNA are needed (Staub et al. 1996). As RFLP markers were the first molecular markers available, they have been important in genetic mapping and have been used to generate saturated genetic maps in several crops such as tomato (*Solanum lycopersicum*), rice, barley (*Hordeum vulgare*) or wheat (Jain et al. 2002). In the last years they have been still occasionally used to construct linkage maps, either as the only markers system available, e.g. in red clover (Isobe et al. 2003), or in combination with other marker systems (Jones et al. 2002b).

Simple sequence repeats (SSR)

SSR markers, also called microsatellites, are short nucleotide core elements, i.e. 1-6 nucleotides, which are tandemly repeated. SSR markers are amplified using primers designed to regions flanking the repetitive element and separated by gel electrophoresis. Polymorphisms arises due to changes in the number of tandemly repeated core elements (Brown et al. 1996). SSR markers are codominant, highly polymorphic, highly reproducible and allow for a large sample throughput. They have the disadvantage that prior sequence information is needed to design the primers, and therefore the development of SSR markers is very expensive and time consuming (Brown et al. 1996).

SSR markers are useful for diversity studies where sufficient sequence information is available (Kubik et al. 1999; Mengoni et al. 2000). But first of all, they are a good choice for linkage mapping as they are highly specific, codominant and also show certain transferability among species. SSR markers were developed and widely used for the construction of maps for example in soybean or wheat (Cregan et al. 1999a; Liu et al. 2005) as well as in forage crops such as perennial ryegrass (Jones et al. 2002a), white clover (Barrett et al. 2004) or red clover (Sato et al., submitted).

The name SSR marker is mainly used for markers, which were developed using genomic DNA. A special kind of SSR markers, which were developed using expressed sequence tags (EST),

are called EST-SSR markers. An EST represents a cDNA clone derived from a mRNA or a part thereof, generally 150 – 400 bp in length. An efficient method to design primer pairs for EST-SSR markers is to apply the information of an EST database. Large EST databases are available for species such as soybean, rice or the model legume *Medicago truncatula*. As they are based on the conserved coding region of the genome, EST-SSR markers have a greater interspecific transferability when compared to genomic SSR markers, but they are likely to be less polymorphic (Jain et al. 2002). The association with coding regions makes them especially useful for application in marker assisted selection, as there is an increased probability that the EST-SSR marker linked with the target trait coincides with the gene controlling the target trait. EST-SSR markers have been used in white clover to construct a linkage map (Barrett et al. 2004) or for a diversity study in tall fescue (Mian et al. 2005).

Sequence-tagged site (STS)

STS is a short unique sequence that identifies one or more specific loci, which can be amplified through PCR. Each STS is characterised by a pair of PCR primers, which are designed by sequencing or using information of DNA databases. Basically, all sequence based marker techniques are STS, but in particular this name is used for sequences of cloned DNA fragments of RFLP, AFLP or RAPD markers. Polymorphisms are usually the result of mutations in the primer binding site, resulting in one allele failing to amplify (Jain et al. 2002). This conversion of anonymous markers into STS markers results in easyly scorable markers and is of special interest when the marker is closely linked to a target trait. An example for conversion of two RFLP markers closely linked to resistance genes against bacterial blight in STS markers is given in rice, where these STS markers have been used to simplify pyramiding these resistance genes in breeding lines (Huang et al. 1997a).

Single nucleotide polymorphism (SNP)

SNP markers detect single base mutations between homologous DNA fragments. In the human genome, SNP have been shown to be the most abundant polymorphisms with an estimated average frequency of one SNP per thousand base pairs. Although SNP markers suffer from some disadvantages, such as being tetra-allelic and not multi-allelic as for example SSR markers, their abundance makes them very attractive. There is also some evidence that the stability and therefore the reliability of SNP markers are higher than for other marker systems such as SSR or AFLP markers. The availability of EST databases and to some extent genome-wide sequences as well as new techniques such as pyrosequencing facilitated the development and genotyping of SNP markers, which probably will result in a wide application in near future also in plant species (Jain et al. 2002). Although several examples of development and application of SNP markers in plants species were already reported for soybean (Zhu et al.

2003), wheat (Tyrka et al. 2004), or rice (Feltus et al. 2004), they require substantial sequence information and are not suitable for less investigated crops such as forage legumes.

Comparative anchor tagged sequence (CATS)

CATS offer an efficient approach to generate anchor markers, allowing for rapid information transfer from model plants to crop species. The approach is based on differences in the rate of evolutionary DNA changes in a genome. Primer pairs are developed in regions, which are conserved in different species and genera. These primer pairs flank regions, which are not conserved and have therefore different length in different species and genera. Whereas the conserved regions are mainly expressed sequences, the regions of different length are intron sequences (Schauser et al. 2005). Therefore, CATS markers have a high interspecific transferability and offer a very promising approach for comparative genetics in crops such as legumes. However, this marker technique has one major drawback. For the detection of conserved regions and primer development extensive sequence knowledge of at least two species for the family under study is needed. Nevertheless, for legumes first CATS were developed using information of the two model plants *Medicago truncatula* and *Lotus japonicus* as well as of soybean (Schauser et al. 2005).

1.4.2 Bulked samples - an efficient approach for diversity studies

Molecular markers represent an efficient tool for the investigation of genetic diversity within and among plant populations. For a reliable characterisation of diversity in outcrossing species the following plays an important role. A single population of an outcrossing species, i.e. a cultivar, landrace or plants of a species of a defined meadow, consists of many different genotypes. Therefore, the greater the number of plants per population included in a diversity study, the higher the probability to include rare alleles, i.e. the better the population is represented (Crossa 1989). Genotypes or alleles occurring at frequencies of at least 10 % are likely to be detectable in a study analysing 40 plants, whereas 100 plants have to be investigated to detect all alleles occurring at frequencies of at least 5 % (Crossa 1989). In order to determine diversity within and among populations of outcrossing species such as red clover, often around 25 individuals are analysed (Gaudeul et al. 2000; Semerikov et al. 2002), resulting in a probability higher than 90 % to include an allele occurring at a frequency of 10 % in a population (Kraft and Sall 1999). On the other hand, if only 10 plants are included, this probability decreases to 65 %. Therefore, if less than 20 plants are included in a study the population may not be represented adequately, which may result in a misleading estimation of relationship among populations or in a incapability to distinguish populations from each other (Crossa 1989).

In large diversity studies, analysis of 20 or more individuals per population leads to a very high sample number. In order to reduce the number of samples processed, DNA or plant material of several individual plants may be combined into a single bulked sample (Michelmore et al. 1991). Thereby, a certain loss of information of rare alleles has to be taken into account (Kölliker et al. 2001; Gilbert et al. 1999), but on the other hand, bulked samples accumulate population specific markers (Michelmore et al 1991). Characterisation of diversity using bulked samples have been reported in several crops such as white clover (Kölliker et al. 2001), rescuegrass (*Bromus catharticus*; Puecher et al. 2001) or alfalfa (Segovia-Lerma et al. 2003). This method may be an efficient approach to characterise diversity of outcrossing species particularly when a high number of populations are investigated. However, there is little experimental evidence as to how this approach influences the estimates of relationship among populations, i.e. how the number of plants combined in a bulked sample or the number of bulked sample investigated per population influences this relationship.

1.4.3 Linkage mapping

The construction of a reliable and robust linkage map is a basic prerequisite for MAS as well as a first step towards identification of genes of economic interest. A linkage map shows the relative genetic distances between markers, genes or other DNA sequences and their assignment to linkage groups on the basis of the recombination frequencies from all their pairwise combinations. The concept of linkage maps was first reported by Thomas H. Morgan who postulated that Mendelian genetic factors (genes), which are assigned to the same chromosome are usually co-transmitted from parent to progeny and that chromosomes are the basis for the existence of linkage groups (Morgan 1911). Therefore, linkage map distances are usually reported in units of Morgan or centi Morgan (cM). One cM corresponds to a recombination frequency of 1 %, i.e. 1 % of the tested progenies show a recombination or crossing over between the two markers investigated. Pairs of markers located on the same linkage group show a recombination frequency of less than 50 %, whereas pairs of independent markers located on separate linkage groups show a frequency of 50 % (Jones et al. 1997b).

Calculations of recombination frequencies may become very complex when large numbers of individuals and markers are investigated. To handle such calculations, computer programs such as JoinMap (Van Ooijen and Voorrips 2001) or MAPMAKER (Lander et al. 1987) have been developed. Basically, JoinMAP runs through the following steps, which are analogically performed by other computer programs. The first step is to calculate all pairwise recombination frequencies based on maximum likelihood estimators as well as the corresponding LOD score. The LOD score is an indicator for the likelihood of linkage, i.e. the probability that two loci are

linked with a given recombination value over the probability that the two are not linked. In a second step, the linkage groups are established based on the LOD value of the pairwise estimation and in the third step, the markers are aligned in the linkage groups by adding loci one after the other (Van Ooijen and Voorrips 2001).

One of the advantages of JoinMap compared to other mapping computer programs is that it can handle F_1 populations based on heterozygous parents of outcrossing plant species, which is far more complicated than to handle progenies of homozygous parents. In a F_2 population of a cross between two fully homozygous parents (aaxbb), only two alleles segregate at one locus and only three genotypes codes are possible for progenies: homozygote as the one parent (aa), homozygote as the other parent (bb) and heterozygote as the F_1 (ab). In a F_1 populations based on an outcrossing species up to four alleles segregate at one locus. Therefore, at one locus nine different combinations of genotypes of the two parents are possible (abxcd, abxac, abxab, abxaa, abxcc, aaxab, ccxab, aaxbb, aaxaa). Of these nine types, five fundamentally different marker types can be defined for the construction of a linkage map: loci in the parents are heterozygous for both parents with four (abxcd), three (abxac) or two (abxab) different alleles, or loci are heterozygous in one parent and homozygous in the other parent (abxaa; aaxab; Van Ooijen and Voorrips 2001).

For the calculation of a linkage map of an outcrossing species, the marker type abxcd is more informative compared to the two marker types abxaa and aaxab. While recombination frequency can be calculated between type abxcd and all other marker types, recombination frequency can not be estimated between marker types abxaa and aaxab. Thus, a higher number of recombination frequencies can be calculated for the marker type abxcd resulting in a more precise estimation of location when compared to the marker types abxaa and aaxab (Maliepaard et al. 1997). Such valuable markers of the type abxcd are only available in codominant markers system such as SSR, whereas in dominant marker system such as AFLP mainly markers of the type abxaa and aaxab can be scored. However, only a very limited number of SSR markers was available in red clover until very recently (Sato et al., submitted). As accuracy of estimation is also lower for higher recombination frequency, i.e. for higher distance between two markers, the reliability of a map depends also on the number of markers investigated (Maliepaard et al. 1997). Therefore, AFLP markers, where a large number of polymorphic bands are produced in a single assay, can nevertheless be very valuable to fill up gaps in a linkage map and to contribute to the construction of a reliable map.

Besides the marker type, the number of investigated progenies determines substantially the accuracy of the estimation of recombination frequencies, the power of detecting linkage and therefore the reliability of a map (Maliepaard et al. 1997). The adequate consideration of these

factors, i.e. number of markers, number of progenies and type of markers, allows for the construction of a reliable and stable linkage map, which is a prerequisite to interpret and apply the map in further analyses.

1.4.4 Analysis of quantitatively inherited traits

In plant breeding, many traits of economic interest such as seed yield, forage yield or persistence are quantitatively inherited. At the beginning of the 20th century, the Swedish geneticist and plant breeder Herman Nilsson-Ehle, was the first to propose the multiple-factor hypothesis of quantitatively inherited traits. Quantitative traits are controlled by a large number of genes (polygenes) each with a small additive effect on the character. Each polygene is inherited according to the laws proposed by Mendel, i.e. law of uniformity, of segregation and of independent assortment (Nilsson-Ehle 1909). In the last decades, the term quantitative trait locus (QTL) became the accepted term for an individual and independent polygene contributing to continuous variation. Using controlled crosses between parents of dissimilar phenotypes, scientists could for example estimate heritability of a quantitative trait and the number of relevant loci involved in genetic control. These studies form the conceptual framework of our understanding of quantitative inheritance, but they rarely analyse the roles of individual QTL, as with statistical genetics details about any one locus are difficult to discern (Young 1996).

In 1923 the first time linkage between a marker and a quantitative trait was demonstrated in bean, where seed size differences were associated with seed coat colour (Sax 1923). This concept was further elaborated by Thoday (1961), who suggested that if segregation of simply inherited monogenes (markers) could be used to detect linked QTLs, then it should be possible to map and characterise all the QTLs involved in a complex trait. However, only with the discovery of molecular markers it became feasible to assign QTLs to specific regions of a linkage group and to determine the individual contribution of each QTL to the phenotypic variation. Paterson et al. (1988) were the first to resolve quantitative traits into discrete Mendelian factors using a RFLP linkage map of tomato and to identify QTLs controlling fruit mass. Since then, QTLs have been mapped for several traits in many crop species such as wheat (Keller et al. 1999), soybean (Mansur et al. 1996), alfalfa (Brouwer and Osborn 1999) or perennial ryegrass (Cogan et al. 2005). However, to the authors knowledge no QTL analyses in red clover have been reported. Such analyses form an important first step towards MAS of traits of economic interest such as seed yield or persistence

There are several approaches to calculate the probability of the location of a QTL. The simplest ones are single marker analyses, which are performed on each marker separately and where the linkage map is used only for sorting the markers. Such approaches use analysis of variance

(ANOVA) or nonparametric analyses like the rank sum test of Kruskal-Wallis. In this test all individuals are ranked according to the quantitative trait and classified according to their marker genotype. A QTL closely linked to the tested marker will result in large differences in average rank of the marker genotypes classes. A test statistic based on the ranks in the genotype classes is then calculated (Van Ooijen 2004). Single marker analysis is a good first step with the disadvantage that the power of detecting a QTL decreases with increasing distance between the marker and the QTL.

An improved approach is interval mapping (IM) also called simple interval mapping (SIM), as it considers several markers for the calculation of a QTL. For each position in the genome (e.g. every one centi Morgan) the likelihood for the presence of a segregating QTL is calculated. For the calculation of this probability, the genetic information from markers surrounding this position is used. If the two flanking markers are fully informative, i.e. codominant markers of the type abxcd, only these two markers are integrated in the calculation as additional markers do not provide additional information. If, however, one or both of the flanking makers are not fully informative, i.e. dominant markers of for example the type abxaa, extra information is gained by employing neighbouring markers. Therefore, QTL calculation with not fully informative markers is possible, but fully informative markers provide more accurate probabilities and faster computing. The resulting probability is tested against the probability that no QTL is present at this position (Van Ooijen 2004).

Further development of interval mapping led to multiple QTL mapping (MQM; Jansen and Stam 1994) or composite interval mapping (CIM; Zeng 1994). It extends interval mapping by applying cofactors. When testing for a QTL in a particular interval, for each other QTL of the trait outside of this interval the next flanking marker is used as cofactor. Effects of QTLs outside of the interval are included in the calculation and the resulting residual variation can be reduced. Therefore, the power of QTL analysis is increased and the confidence interval of a QTL can be narrowed. This improves the resolution of two neighbouring QTLs as well as the probability to detect minor QTLs (Van Ooijen 2004).

In principle, interval mapping can be performed using a maximum likelihood algorithm (MapQTL; Van Ooijen 2004) or multiple regression (PlabQTL; Utz and Melchinger 2003). Studies comparing these two methods find very little differences between the two methods with regard to the significance of QTLs detected. The regression model offers a great advantage over the maximum likelihood method in terms of computing speed, but estimation of proportion of phenotypic variance explained tends to be biased (Xu 1995). However, up to date only the maximum likelihood algorithm of MapQTL is capable to handle F_1 progenies of outcrossing species.

Molecular markers using bulked samples form an efficient and reliable tool to characterise genetic resources of outcrossing species such as red clover. Molecular markers are also a basic prerequisite for the construction of a stable linkage map, for QTL analyses and therefore for marker assisted selection and to elucidate the genetic control for complex traits such as seed yield and persistence.

1.5 Objectives

In order to evaluate the value of the Mattenklee landraces and wild clover populations as genetic resources for diversity and breeding and to elucidate the genetic control of seed yield and persistence, two important traits in red clover breeding, the objectives of this study were twofold:

1. Estimation of the over all value of the six red clover groups Mattenklee landraces, Mattenklee cultivars, Swiss wild clover populations, Field clover cultivars, Dutch landraces and Dutch wild clover populations as genetic resources in order to optimally manage preserved germplasm and to identify potential populations to include in breeding programs. The objectives of this part of the project were to:
   - provide a strategy for large scale analyses of outcrossing species such as red clover based on AFLP markers and bulked samples
   - determine genetic diversity particularly within and among Mattenklee landraces and wild red clover populations
   - investigate the relationship of these populations in order to assess the distinctness of the individual red clover groups and to determine the ancestry of Mattenklee
2. Establishment of a basis for the development of molecular markers to assist breeding for seed yield and persistence in red clover. For this purpose, in the second part of the project a F_1 population segregating for seed yield components and persistence was established. Eight seed yield components and persistence were investigated in a field experiment. Moreover, a linkage map of this segregating F_1 population was constructed and QTL analyses were performed for these two traits, in order to:
   - characterise the association among seed yield components as well as between seed yield components and persistence
   - identify components which are easy to score, thus allowing for improved selection for these two traits

- identify genome regions containing QTLs for seed yield components and persistence for the future development of molecular markers to assist improvement of these two traits in red clover

2 Swiss Mattenklee landraces, a distinct and diverse genetic resource of red clover (*Trifolium pratense* L.)

R. Kölliker, D. Herrmann, B. Boller, F. Widmer

Theoretical and Applied Genetics (2003) 107:306-315

Abstract

Genetic variability within and among 19 landraces and cultivars of red clover (*Trifolium pratense* L.) was investigated by means of amplified fragment length polymorphism (AFLP) analysis in order to assess the potential value of Swiss Mattenklee landraces as genetic resources for plant breeding and the preservation of biodiversity. Populations were classified into three groups according to their origin and agronomic features: Mattenklee landraces (8), Mattenklee cultivars (8) and field clover cultivars (3). Analysis of molecular variance based on 276 polymorphic AFLP markers revealed 80% of total variability to be due to variability within populations while 12% were attributed to variability among groups. Stepwise discriminant analysis identified a subset of 126 AFLP markers which best separated individual plants into the three respective groups. Genetic distances between populations were considerably larger among groups than among populations within the same group, providing further evidence for the genetic distinction between Mattenklee landraces, Mattenklee cultivars and field clover cultivars. AFLP markers identified two landrace clusters, containing three and four populations respectively, which, together with one additional landrace, may sufficiently represent the genetic variability of all eight landraces investigated. The results of this study strongly suggest that Swiss Mattenklee landraces form a genetically distinct group of red clover. The data obtained provide criteria on how to efficiently manage, preserve and exploit Mattenklee germplasm.

Keywords: Mattenklee, *Trifolium pratense* L., landraces, genetic variability, AFLP

2.1 Introduction

To efficiently preserve, manage and exploit genetic resources, detailed knowledge on the genetic variability within a germplasm collection is indispensable. Such information may assist plant breeders in deciding which germplasm to include in breeding programs and may also allow the identification of accessions that substantially contribute to the overall diversity of the species (Grenier et al. 2000).

Landraces, also known as local populations, traditional cultivars, or farmers varieties (Zeven 1998), provide a valuable resource for plant breeding as well as for the preservation of genetic diversity. In various crops such as wheat (Skovmand et al. 2001), barley (Lakew et al. 1997), rice (Yang et al. 1994), maize (Zeven 2000) and pearl millet (Ouendeba et al. 1995), numerous landraces have been collected, characterised and exploited for several purposes. Landraces may be used as starting populations for cultivar development (Lakew et al. 1997) or as sources for

the introgression of genes and QTLs conferring resistance to biotic (Huang et al. 1997b) and abiotic stresses (Forster et al. 2000). In some forage legumes (e.g. red clover, *Trifolium pratense* L.), landraces may be of particular value since modern cultivars are genetically not as far advanced as compared to other crops such as grain cereals (Hill et al. 1988; Woodfield and Caradus 1994). Although a number of landraces have been described and utilised in forage crops such as alfalfa (*Medicago sativa* L.; Julier 1996), white clover (*Trifolium repens* L.; Annicchiarico and Piano 1997) and red clover (*T. pratense* L.; Kouame and Quesenberry 1993), there is little information available on the genetic variability of such populations and their relationship to modern cultivars.

Molecular markers allow for a rapid assessment of genetic diversity directly at the genome level and have been extensively used to characterise genetic resources in various plant species (Crouch et al. 2000; Fahima et al. 1999; Grenier et al. 2000; Semagn et al. 2000). Amplified fragment length polymorphism (AFLP; Vos et al. 1995) markers are particularly useful for diversity studies. The technique is based on generic PCR primers and allows the detection of a large number of loci in a single assay (Powell et al. 1996).

Red clover, *Trifolium pratense* L., is one of the most important forage legumes of temperate climates. Due to its ability to fix atmospheric nitrogen and its high nutritive value (Taylor and Quesenberry 1996), it is widely used for grass-clover leys in crop rotation and also is an important component of permanent pastures and meadows. Red clover is an extremely polymorphic, diploid species (2n = 14). Zohary and Heller (1984) identified more than 40 forms or varieties, but no detailed taxonomy of this species has yet been established (Taylor and Quesenberry 1996). However, this extensive variability is only partly exploited. Of 4233 classified red clover accessions, only 42% are advanced cultivars, 12% are landraces and the remaining 45% consist of wild ecotypes (Taylor and Quesenberry 1996). Landraces and ecotypes of red clover not only form a valuable source for breeding, they substantially contribute to the biodiversity of natural and managed grasslands. Changes in land-use patterns and the increased use of elite cultivars threatens the genetic diversity of native and naturalised ecotypes and emphasises the need for an efficient management of genetic resources (Morris and Greene 2001).

Red clover cultivation in Europe dates back to the third century and reached economic importance in Flanders around 1600. From there, cultivation spread rapidly to most temperate regions of the world (Taylor and Quesenberry 1996). In Switzerland, farmers probably started red clover seed production in the 18th century based on Flemish plant material (Koblet and Nüesch 1960; Merkenschlager 1934). Seed was traditionally harvested in the third year after sowing or later. Thus, a certain selection for persistent genotypes was practiced. Over time, a

Swiss form of red clover, known as Mattenklee, was developed, which is characterised by increased persistency and early flowering (Koblet and Nüesch 1960; Nüesch 1976). Mattenklee landraces specifically adapted to local climates and conditions were developed and maintained on individual farms. Germplasm from Mattenklee landraces was also integrated in the Swiss clover breeding program and today a broad range of Mattenklee cultivars is available and widely used throughout Europe (Boller 2000b). Mattenklee cultivars and landraces are distinguished from other forms of red clover mainly by their improved persistency which makes them particularly useful for ley farming systems. Less persistent forms of red clover are generally used in shorter crop rotation and are, in this paper, referred to as field clover.

Seed from improved red clover cultivars became widely available in the late 20th century. Thus, the use of Mattenklee landraces declined dramatically. Today only small seed lots of approximately 100 old landraces are still available, which were collected on Swiss farms in 1971/72 (Nüesch 1976). These landraces may provide a valuable genetic resource for the further advancement of red clover cultivars as well as for the preservation of genetic diversity in less intensively managed pastures and meadows. However, there is no information available on their genetic compositions or structures. Although Isozyme and RAPD markers have been used to characterise cultivars from North America, Europe, Japan and Chile (Campos-De-Quiroz and Ortega-Klose 2001; Kongkiatngam et al. 1996; Yu et al. 2001), there is a general lack of studies on the molecular characterisation of ecotypes and landraces of red clover.

In the present study we used AFLP markers to assess the potential value of landraces of red clover as genetic resources. The objectives were (1) to characterise genetic variability within and among Mattenklee landraces and cultivars; (2) to determine whether Mattenklee landraces form a distinct group of red clover; and (3) to compare genetic diversity between Mattenklee populations and field clover cultivars.

2.2 Materials and Methods

Plant material

The 19 populations of red clover (*Trifolium pratense* L.) analysed in this study represented three groups according to their origin and their agronomic features (Table 2.1). Mattenklee landraces (group I) consisted of old landraces collected in 1971/72 from Swiss farms where they were maintained for many decades (Nüesch 1976). Mattenklee cultivars (group II) are Swiss cultivars of red clover which are distinguished from field clover cultivars (group III) mainly through their improved persistence. In addition, the white clover (*T. repens*) cultivar Bombus was included in the analysis as a reference for the comparison of genetic distances. Plant material from

24 randomly selected individual plants of each landrace and each cultivar was collected for AFLP analysis. Individual plants will be referred to as genotypes, while the term population will be used for landraces and cultivars and the term group refers to the three types of populations, i.e. Mattenklee landraces, Mattenklee cultivars and field clover cultivars.

Table 2.1 Red clover (*T. pratense* L.) landraces and cultivars and white clover (*T. repens* L.) cultivar used for investigation

Name	Origin[a]	Elevation (m asl)	Longitude (°E)	Latitude (°N)	Last year of selection
Group I: Mattenklee landraces (ML)					
LR8	Bubikon	510	8° 49’	47° 16’	
LR127	Zäziwil	700	7° 40’	46° 54’	
LR189	Sumiswald	700	7° 45’	47° 02’	
LR239	Köniz	680	7° 25’	46° 56’	
LR292	Lanzenhäusern	810	7° 21’	46° 51’	
LR300	Burgistein	830	7° 30’	46° 47’	
LR325	Affoltern i.E.	800	7° 44’	47° 04’	
LRDet	Dettenbühl	508	7° 38’	47° 15’	
Group II: Mattenklee cultivars (MC)					
Corvus	Zurich				1991
Formica	Zurich				1989
Merula	Zurich				1994
Milvus	Zurich				1990
Pavo	Zurich				1995
Pica	Zurich				1991
Renova	Zurich				1964
Rüttinova	Zurich				1980
Group III: Field clover cultivars (FC)					
Lucrum	Germany				n.a.[b]
Merviot	Belgium				n.a.
Mont Calme	Changins				1970
White clover					
Bombus	Zurich				1994

[a]Location of collection (landraces) or cultivar development (cultivars). Unless otherwise mentioned, all localities are situated in Switzerland

[b]Information not available

AFLP analysis

DNA was extracted from fresh plant material using the DNeasy 96 Plant Kit (Qiagen, Hilden, Germany) and quantified using PicoGreen (Molecular Probes, Eugene, Ore., USA) and a fluorimeter (LS-30; Perkin Elmer Instruments, Shelton, Conn., USA) as well as by visual inspection on a 1% (w/v) agarose gel.

AFLP analysis was performed following the method of Vos et al. (1995). AFLP templates were prepared by restriction digestion and adaptor ligation of 1 µg genomic DNA. Adaptors and primers (see Table 2.2) were synthesised by Microsynth, Balgach, Switzerland. *Eco*RI + C and *Mse*I + A primers were used to amplify double-digested, adaptor ligated DNA (20 ng) in a 20-µl reaction containing 1 x PCR buffer, 5 pmol of each primer, 1.5 mM $MgCl_2$, 0.2 mM dNTP and 0.5 U *Taq* DNA polymerase (Invitrogen, Carlsbad, Calif., USA). PCR was performed in a MJ PT-200 (MJ Research, Waltham, Mass., USA) thermocycler using an initial denaturation step of 2 min at 94°C, 26 cycles of 1 min at 94°C, 1 min at 56°C and 1 min at 72°C, followed by a final extension of 5 min at 72°C. Amplification products were diluted by adding 100 µl H_2O. For subsequent selective amplification, *Eco*RI and *Mse*I primers with three additional nucleotides each were used (see Table 2.2). Selective amplification reactions contained 1 µl of diluted pre-amplified AFLP template, 5 pmol 6-FAM, HEX or TET labelled *Eco*RI + CNN primer, 6 pmol *Mse*I + CNN primer, 1.5 mM $MgCl_2$, 0.2 mM dNTP and 0.4 U *Taq* DNA polymerase in a total volume of 20 µl. PCR was performed using a touchdown PCR protocol with an initial denaturation of 2 min at 94°C, 12 cycles of 1 min at 94°C, 30 s at 65°C (-0.7°C per cycle) and 1 min at 72°C, followed by 23 cycles of 30 s at 94°C, 30 s at 56°C and 1 min at 72°C with a final extension of 5 min at 72°C. AFLP fragments were analysed on an ABI PRISM 310 Genetic Analyser using POP 4 polymer and a 47 cm x 50 µm capillary (Applied Biosystems, Foster City, Calif, USA).

Data analyses

AFLP patterns were analysed using Genescan 3.1 and Genotyper 3.7 software (Applied Biosystems). AFLP markers were visually scored for presence (1) or absence (0) and entered into a binary matrix containing the AFLP profile of each genotype. Each AFLP pattern was independently scored by two different persons and only polymorphic bands which could be scored unequivocally were included in the analysis. Five genotypes were excluded from further analysis due to repeatedly unscoreable AFLP patterns.

The binary data matrix with no missing values was subjected to multivariate analysis procedures as described by Semagn et al. (2000) using the SAS v. 8.0 statistical package (SAS Institute, Cary, N.C., USA). (1) Principle component analysis was carried out using the SAS procedure PROC PRINCOMP and the entire set of polymorphic markers obtained through AFLP analysis;

(2) PROC STEPDISC was employed to identify a subset of AFLP markers which were the best discriminating factors among the 19 red clover populations ($P = 0.15$ for adding and retaining variables); and (3) canonical discriminant functions were calculated using the previously identified subset of AFLP markers and the SAS procedure PROC CANDISC.

Analysis of molecular variance (AMOVA; Excoffier et al. 1992) was used to calculate variance components within and among groups and populations. Pairwise genetic distances between populations were calculated using the coancestry coefficient of Reynolds et al. (1983). Significance of variance components and differentiation between populations was tested using a nonparametric procedure based on 1000 permutations according to Excoffier et al. (1992). AMOVA and coancestry coefficients were computed using the Arlequin 2.0 software (Schneider et al. 2000). Cluster analysis was performed on coancestry coefficients using the unweighted pair-group procedure with an arithmetic mean (UPGMA) of the NTSYS-pc 2.1 software (Rohlf 2000). Reliability of the clustering was tested by computing Mantel test statistics for the correlation of the distance matrix and the cophenetic matrix. Bootstrap analysis was performed on marker frequencies per population using the software package PHYLIP (Felsenstein 2002) with 10'000 replicated datasets and Reynold's genetic distance. Bootstrap values were transferred to the dendrogram obtained by cluster analysis of coancestry coefficients which showed identical topology as the dendrogram obtained through bootstrap analysis. Genetic variability within populations was estimated by calculating average pairwise Euclidean distance (Excoffier et al. 1992; Huff et al. 1993) among genotypes.

A model-based approach as described by Pritchard et al. (2000) was used to infer population structure on the data set and to assign individuals to a pre-defined number of populations. This method is suitable to demonstrate the presence of population structure and to identify a meaningful number of populations present in the data. Calculations were performed using the Structure 2.0 software (http://pritch.bsd.uchicago.edu; Pritchard and Wen 2002). Independent runs for K (the number of populations) between 1 and 20 were performed with two different models (no admixture and admixture model) based on 100'000 iterations.

2.3 Results

Characteristics of red clover AFLP markers analysed

The six AFLP primer combinations generated a total of 276 polymorphic markers, ranging in size from 60 to 345 base pairs (Table 2.2). The number of polymorphic markers detected with each primer combination ranged from 36 to 58 with an average of 46. Each of the 451 genotypes was characterised through a unique AFLP phenotype (data not shown). The number

of polymorphic markers within each red clover group ranged from 215 for field clover cultivars to 267 for Mattenklee landraces. Within individual populations, the lowest number of polymorphic markers was found in the Mattenklee cultivar Pica with 151 markers while the Mattenklee landrace LR239 showed with 229 the highest number of polymorphic markers (data not shown). Although marker frequencies varied considerably among groups and populations, no group- or population specific-markers were identified.

Table 2.2 AFLP primer combinations used for selective amplification and number of polymorphic markers detected within each group of 19 *T. pratense* L. landraces and cultivars

Primer pairs (5'→3')	Number of polymorphic markers across			
	Mattenklee landraces[a]	Mattenklee cultivars[b]	Field clover cultivars[c]	All populations
E[d]+ACT / M[e]+CAC	41	36	32	41
E+ACA / M+CAC	40	40	43	43
E+ACT / M+CTA	50	47	36	51
E+AGA / M+CTA	54	54	39	58
E+AGG / M+CAC	35	33	27	36
E+AGT / M+CTA	47	44	38	47
Total	267	254	215	276

[a]Eight landraces, 24 individuals each

[b]Eight cultivars, 24 individuals each

[c]Three cultivars, 24 individuals each

[d]Primer core specific for *Eco*RI site (Vos et al. 1995)

[e]Primer core specific for *Mse*I site (Vos et al. 1995)

Distinction of groups

Principle component analysis based on all 276 polymorphic markers resulted in a moderate separation of the 451 genotypes into the three respective groups (Fig. 2.1). While the first three principle components (PCs) explained 17% of the total variation among samples, each of the following PCs explained less than 2%. The relatively poor separation of the three groups is also illustrated by the low level of genetic variation attributed to variation among groups (12%) as revealed by analysis of molecular variance (AMOVA; Table 2.3A).

Stepwise discriminant analysis using the 19 populations as class variables identified a set of 126 AFLP markers which discriminated best among the 19 populations analysed. The first three discriminant functions (Can) of a canonical discriminant analysis based on these 126 AFLP

markers explained 81% of the total variation observed among genotypes. A plot of Can 1 (53%), Can 2 (18%) and Can 3 (10%) revealed a clear separation of the three red-clover groups (Fig. 2.2). Can 1 mainly separated Mattenklee landraces (group mean = -8.7; Fig. 2.2, black symbols) from Mattenklee cultivars (group mean = 6.9; Fig. 2.2, grey symbols) and field clover cultivars (group mean = 6.1; Fig. 2.2, white symbols), while Can 3 was the most powerful discriminant between Mattenklee cultivars (group mean = 2.4) and field clover cultivars (group mean = -1.4). The better discrimination of the three groups achieved with the set of 126 AFLP markers was also reflected in AMOVA where the proportion of genetic variability among groups was increased to 18% of the total variability detected (Table 2.3B). Stepwise discriminant analysis using the three groups as class variables identified 97 AFLP markers as best discriminating factors and resulted in a similar separation of the three groups where Can 1 explained 84% and Can 2, 16% of the variation (data not shown). In order to optimally separate populations, only the set of 126 markers as determined with stepwise discriminant analysis and populations as class variables was considered for further analysis.

Table 2.3 Analysis of molecular variance (AMOVA) for eight landraces and 11 cultivars of *T. pratense* L. using 276 polymorphic AFLP markers (**A**) and a subset of 126 markers which are the best differentiating factors among the 19 populations as determined using stepwise discriminant selection (**B**)

Source of variation	*df*	Sum of squares	Variance component[a]	% Total variance
A				
Variance among groups[b]	2	1282.0	4.0	12.0
Variance among landraces and cultivars	16	1451.0	2.7	8.1
Variance within landraces and cultivars	431	11454.7	26.6	79.9
B				
Variance among groups	2	987.2	3.1	17.9
Variance among landraces and cultivars	16	934.8	1.9	11.1
Variance within landraces and cultivars	431	5359.3	12.4	71.0

[a]Components were significant at $P < 0.001$; the probability of obtaining a more extreme random value computed from non-parametric procedures (1000 data permutations)

[b]Three groups consisting of eight Mattenklee landraces, eight Mattenklee cultivars and three field clover cultivars, respectively (Table 2.1)

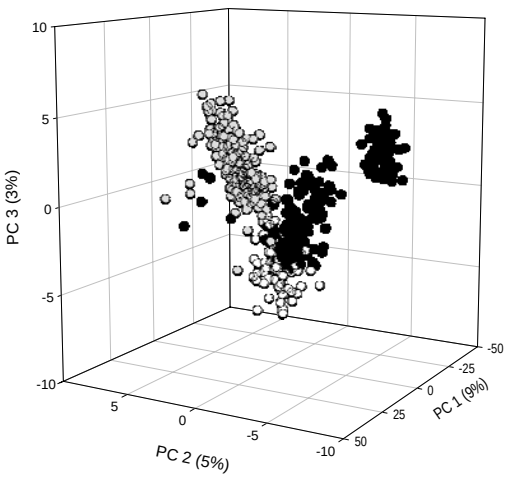


Fig. 2.1 Principle component analysis of 451 *T. pratense* L. plants from 19 populations and three groups using 276 AFLP markers (*black* = Mattenklee landraces; *grey* = Mattenklee cultivars; *white* = field clover cultivars)

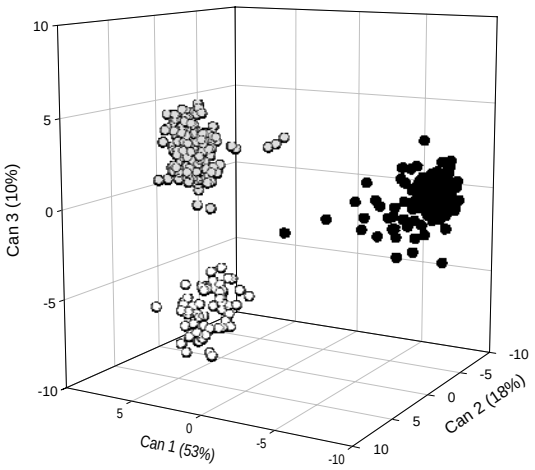


Fig. 2.2 Canonical discriminant analysis of 451 *T. pratense* L. plants from 19 populations and three groups using 126 AFLP markers which differentiated best among the 19 populations as determined by stepwise discriminant selection (*black* = Mattenklee landraces; *grey* = Mattenklee cultivars; *white* = field clover cultivars)

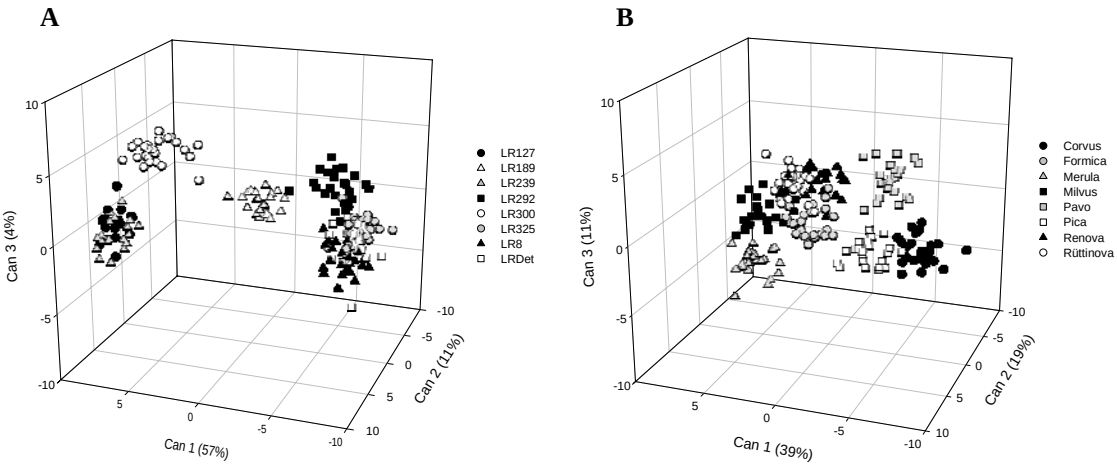


Fig. 2.3 Canonical discriminant analysis of eight Mattenklee landraces (**A**) and eight Mattenklee cultivars (**B**) using 126 AFLP markers which differentiated best among the 19 populations as determined by stepwise discriminant selection

Table 2.4 Genetic distance for eight landraces and 11 cultivars of *T. pratense* L. and one cultivar of *T. repens* L. based on 126 AFLP markers. Coancestry coefficient (Reynold et al. 1983) calculated using 24 individuals per population (below diagonal) and significance of F_{st} values (above diagonal; + = significant difference ($P < 0.05$) determined using a non-parametric permutation approach and 1000 permutations (Excoffier et al. 1992), T = no significant difference). For description of populations refer to Table 1

Type	Mattenklee cultivars								Mattenklee landraces								Field clover cultivars			White clover
	REN	RÜT	MIL	FOR	COR	PIC	MER	PAV	239	127	300	189	325	DET	8	292	MON	MER	LUC	BOM
REN		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
RÜT	0.03		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
MIL	0.10	0.04		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
FOR	0.08	0.06	0.06		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
COR	0.10	0.15	0.25	0.20		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PIC	0.08	0.11	0.23	0.19	0.07		+	+	+	+	+	+	+	+	+	+	+	+	+	+
MER	0.11	0.11	0.17	0.19	0.17	0.11		+	+	+	+	+	+	+	+	+	+	+	+	+
PAV	0.06	0.09	0.20	0.16	0.09	0.04	0.09		+	+	+	+	+	+	+	+	+	+	+	+
239	0.37	0.40	0.54	0.51	0.43	0.43	0.49	0.35		T	+	+	+	+	+	+	+	+	+	+
127	0.42	0.46	0.61	0.59	0.50	0.50	0.55	0.40	0.01		+	+	+	+	+	+	+	+	+	+
300	0.34	0.36	0.48	0.45	0.39	0.40	0.44	0.33	0.03	0.05		+	+	+	+	+	+	+	+	+
189	0.26	0.28	0.40	0.36	0.34	0.33	0.38	0.29	0.20	0.25	0.13		+	+	+	+	+	+	+	+
325	0.18	0.22	0.37	0.35	0.25	0.23	0.32	0.22	0.25	0.31	0.25	0.17		T	+	+	+	+	+	+
DET	0.22	0.25	0.41	0.39	0.29	0.26	0.36	0.24	0.24	0.28	0.24	0.20	0.00		+	+	+	+	+	+
8	0.22	0.24	0.41	0.40	0.30	0.27	0.35	0.24	0.29	0.34	0.28	0.21	0.03	0.02		+	+	+	+	+
292	0.22	0.28	0.42	0.41	0.31	0.29	0.36	0.26	0.24	0.29	0.21	0.17	0.03	0.03	0.03		+	+	+	+
MON	0.23	0.26	0.40	0.38	0.32	0.29	0.33	0.24	0.34	0.39	0.32	0.22	0.16	0.17	0.16	0.17		+	+	+
MER	0.26	0.30	0.45	0.44	0.33	0.30	0.37	0.27	0.41	0.48	0.42	0.32	0.21	0.22	0.22	0.26	0.10		+	+
LUC	0.28	0.34	0.46	0.47	0.37	0.33	0.39	0.28	0.40	0.48	0.41	0.33	0.25	0.25	0.27	0.29	0.16	0.09		+
BOM	0.86	0.97	1.14	1.16	1.09	1.08	1.13	0.98	0.92	1.07	1.02	0.97	0.85	0.89	0.97	0.99	1.00	1.03	1.01	

Comparison of landraces and cultivars

In order to elucidate population relationships based on individual genotypes, canonical discriminant analysis was carried out separately for Mattenklee landraces and Mattenklee cultivars (Fig. 2.3). Due to the small sample size of only three populations no meaningful result was obtained with canonical discriminant analysis for field clover cultivars and these cultivars were therefore excluded from this analysis. For Mattenklee landraces, the first three canonical discriminant functions explained 72% of the variation and allowed to clearly separate all individuals of LR300 and LR189 (Fig. 2.3A). The landraces LR127 and LR239 formed a distinct cluster. The same was true for LR8, LRDet, and LR325 which formed a cluster close to LR292. The Mattenklee cultivars Corvus, Pavo, Pica and Merula were separated by the first three canonical discriminant functions which explained 69% of the variation in this data set (Fig. 2.3B). The remaining four Mattenklee cultivars formed one cluster with Milvus and Formica being slightly separated from the rest. Discrimination of populations based on individual plants was confirmed by plotting class means of the first three canonical discriminant functions for each population (data not shown).

Genetic distances between red clover populations, expressed as coancestry coefficients, ranged from 0 (LRDet/LR325) to 0.61 (Milvus/LR127; Table 2.4). Comparisons of red clover populations with the white clover cultivar Bombus resulted in coancestry coefficients which ranged from 0.86 to 1.13 (Table 2.4). All F_{st} values were significant at $P < 0.05$ except for the LRDet/LR325 and the LR127/LR239 comparison. The average coancestry coefficient within groups was 0.12 for Mattenklee cultivars and field clover cultivars and 0.17 for Mattenklee landraces. The average coancestry coefficient between groups was 0.36 for Mattenklee landraces and Mattenklee cultivars, 0.30 for Mattenklee landraces and field clover cultivars, and 0.34 for Mattenklee cultivars and field clover cultivars. UPGMA clustering of genetic distances (coancestry coefficients) resulted in two major clusters, one comprising all Mattenklee cultivars, the other including all Mattenklee landraces as well as the field clover cultivars (Fig. 2.4). Cluster analysis including white clover as an outgroup (UPGMA as well as Neighbour joining) did result in the same topology of the red clover cluster (data not shown). Mattenklee landraces LR239, LR127 and LR300 formed a tight cluster (Fig. 2.4, ML I) and were clearly separated from the Mattenklee landrace LR189, which formed a separate cluster (ML II), the landraces LR325, LRDet, LR292 (ML III) and the field clover cultivars (FC I). The Mattenklee cultivars also formed a distinct cluster and were subdivided into two additional clusters (MC I + II; Fig. 2.4). The correlation coefficient between cophenetic values derived from the dendrogram and the coancestry coefficients was $r = 0.81$ ($P < 0.0001$), indicating a good fit of the clustering with the original distance matrix. UPGMA clustering based on individual plants and Euclidean

squared distance was largely congruent with the clustering based on coancestry coefficients of populations. Eighty nine percent of all individual plants were correctly assigned to one of the four major clusters (ML I-III, MC I+II, FC I; data not shown).

The proportions of individual populations assigned to inferred clusters were used to identify meaningful groups based on the model-based clustering method of Pritchard et al. (2000). Although proportions assigned to individual clusters were smaller for the admixture model when compared to the non-admixture model, the population structure revealed by both models was largely congruent (data not shown). Since the use of the non-admixture model is recommended for dominant markers (Pritchard and Wen 2002) and yielded more pronounced clusters when compared to the admixture model, only results obtained with the non-admixture model are presented. For each number of populations (*K*), meaningful clusters were identified which contained at least one population represented with a proportion of 50% or more. For $K = 3$ and $K = 4$, all populations were assigned to the respective number of meaningful clusters. The obtained clusters reflected the topology of the dendrogram obtained through cluster analysis at coancestry coefficient values of 0.3 and 0.2 respectively (Fig. 2.4). For larger values of *K*, the number of clusters containing at least one population represented with a proportion larger than 50% increased to a value of nine for $K = 12$, but remained constant for any larger value of *K* tested. The clusters obtained with values of *K* ranging from 12 to 20 were congruent with the topology of the UPGMA dendrogram at coancestry coefficient values of 0.1 (Fig. 2.4).

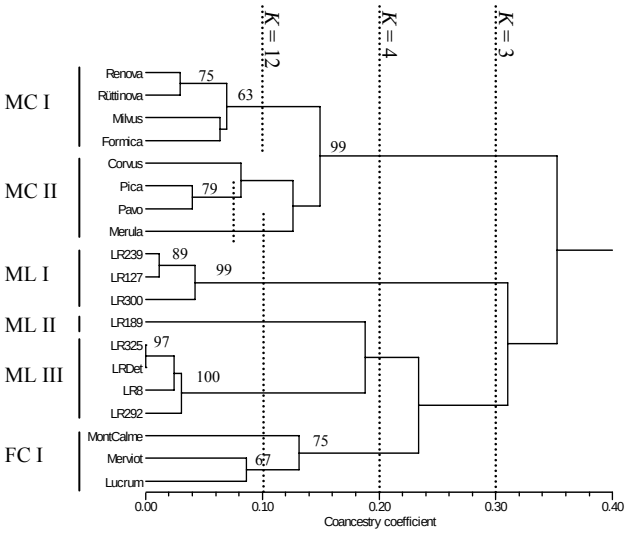


Fig. 2.4 UPGMA clustering of 19 *T. pratense* L. populations from three groups (*MC* = Mattenklee cultivars; *ML* = Mattenklee landraces; *FC* = field clover cultivars) based on coancestry coefficients (Reynolds et al. 1983) derived from 126 AFLP markers and 24 individual plants per population. *Numbers* above branches indicate bootstrap values derived from 10000 re-sampling cycles (only values above 60% are shown). *Dotted lines* indicate the tree topology which corresponds to the grouping obtained with a model-based clustering method (Pritchard et al. 2000) assuming three different numbers of populations (*K*)

Table 2.5 Genetic diversity within populations and groups for eight landraces and 11 cultivars of *T. pratense* L. based on 126 AFLP markers

Item	Percentage of polymorphic markers[a]	Euclidean squared distance[b]
Within populations[c]		
LR8	63.5	24.6
LR127	72.2	23.4
LR189	69.0	24.1
LR239	81.0	28.5
LR292	64.3	23.9
LR300	65.1	25.5
LR325	77.0	28.0
LRDet	65.9	26.4
Corvus	56.3	22.2
Formica	60.3	23.0
Merula	61.1	23.2
Milvus	61.1	22.6
Pavo	71.4	26.1
Pica	58.7	21.5
Renova	80.1	31.1
Rüttinova	72.2	28.7
Lucrum	57.9	21.2
Merviot	62.7	21.6
Mont Calme	72.2	25.3
Average within groups		
Mattenklee landraces	69.8	25.8
Mattenklee cultivars	65.2	24.8
Field clover cultivars	64.3	22.7

[a]A total of 126 markers which are the best differentiating factors among the 19 populations as determined by stepwise discriminant selection was scored

[b]Average distances for pairwise comparisons of genotypes

[c]See Table 2.1 for description of populations

Variability within landraces and cultivars

The total extent of genetic variability detected with the set of 126 selected AFLP markers (see above) was primarily due to variation between genotypes within populations (71%; Table 2.3). AMOVA performed on each individual group separately revealed the variance component for

within-population variation to be higher for Mattenklee landraces (12.8%) when compared to Mattenklee cultivars (12.4%) and field clover cultivars. (11.4%, data not shown). Within individual populations, genetic diversity expressed as Euclidean squared distance ranged from 21.2 for the field clover cultivar Lucrum to 31.1 for the Mattenklee cultivar Renova (Table 2.5). Average Euclidean distance across all populations was 24.8. While the Euclidean distance was higher than the average within five Mattenklee landraces, the same was true for only three Mattenklee and one field clover cultivar. The percentage of polymorphic markers observed within individual populations ranged from 56.3 for the Mattenklee cultivar Corvus to 81.0 for the Mattenklee landrace LR239, with an average of 66.9. A comparison of the three red clover groups revealed the largest genetic diversity within Mattenklee landraces, followed by Mattenklee cultivars and field clover cultivars which showed the least diversity (Table 2.5).

2.4 Discussion

Genetic variability within germplasm collections may strongly affect their value as genetic resources for breeding. AFLP analysis of 19 red clover populations from different groups (Mattenklee landraces, Mattenklee cultivars and field clover cultivars) revealed a substantial amount of genetic variability within this germplasm collection. The largest proportion of the total variability detected was attributed to variability between genotypes within populations (Table 2.3). Red clover is an outcrossing species with a high degree of gametophytic self-incompatibility (Taylor and Quesenberry 1996) and populations are therefore composed of heterogenous individuals. Consequently, high levels of within-population variability are expected. The amount of genetic variability detected within Mattenklee landraces was slightly higher than the variability within Mattenklee cultivars or field clover cultivars, and comparable to values previously reported for red clover (Campos-De-Quiroz and Ortega-Klose 2001) and other forage species such as white clover (*T. repens*; Kölliker et al. 2001) and meadow fescue (*Festuca pratensis*; Kölliker et al. 1998).

Principle component analysis (PCA) based on all 276 polymorphic AFLP markers moderately separated individual red clover plants into the respective groups (Fig. 2.1). The main reason for the incomplete separation apparently is the high variability observed within populations. However, PCA, where no class information is used in order to calculate principle components, already indicates a clear distinction between the three groups. The separation was greatly improved using canonical discriminant analysis and a subset of 126 AFLP markers which best discriminated among populations. Discriminant analysis proved powerful to sort individual plants of strongly heterogenous populations into biologically meaningful groups. Similarly,

discriminant analysis was successfully employed to separate lowland and central-highland ecotypes of *Phytolacca dodecandra* which were previously not separated by principle component analysis (Semagn et al. 2000).

Despite the clear separation of the three groups, no clear-cut distinction between Mattenklee and field clover was possible based on AFLP analysis. While discriminant analysis revealed a slightly closer relationship between Mattenklee cultivars and field clover cultivars when compared to Mattenklee landraces (Fig. 2.2), cluster analysis based on coancestry coefficients clearly separated Mattenklee landraces from Mattenklee cultivars but placed field clover cultivars into a subcluster within Mattenklee landraces (Fig. 2.4). While bootstrap values for the cluster containing all Mattenklee cultivars were higher than 60% and therefore considered relevant, only lower values were observed for the cluster containing Mattenklee landraces and field clover cultivars. However, using a model-based clustering method based on three inferred populations revealed the same close relationship between Mattenklee landraces and field clover cultivars (Fig. 2.4). This, together with the fact that the latter method does not rely on prior population information when compared to discriminant analysis, is a clear indication for a closer relationship of Mattenklee landraces to field clover cultivars when compared to Mattenklee cultivars.

A similar situation was observed for Indian wheat genotypes where *Triticum durum* landraces formed a cluster with *T. dicoccum* cultivars and were clearly separated from *T. durum* cultivars (Pujar et al. 1999). The authors speculated a limited number of domestication events during the evolution of *T. dicoccum* cultivars to be one reason for the observed clustering. This is unlikely to be the case for red clover, where a similar overall selection intensity can be assumed for Mattenklee cultivars and field clover cultivars. However, it can not be excluded that field clover cultivars have been used to improve pastures and meadows where Mattenklee landraces were maintained, or, that Swiss red clover landraces have been used in the development of field clover cultivars as it is known for Mont Calme (Boller 2000b). Another possible explanation for the clear separation of Mattenklee cultivars could be a strong selection targeting mainly one trait, i.e. persistence. Nevertheless, Mattenklee landraces, Mattenklee cultivars and field clover cultivars form three genetically distinct groups. To further elucidate genetic relationships in the *T. pratense* complex, more detailed studies involving wild clover populations as well as a larger number of field clover cultivars are necessary.

Genetic distances (coancestry coefficients) between populations were considerably larger among groups than among populations within the same group, providing further evidence for the genetic distinction between Mattenklee landraces, Mattenklee cultivars and field clover cultivars (Table 2.4). Coancestry coefficients between red clover populations reached up to

50 % of the values obtained from comparisons between red clover populations and the white clover cultivar Bombus. White clover (*T. repens*) is not a close relative of red clover and hybrids between both species have only been obtained by means of embryo rescue (Taylor and Quesenberry 1996). Hence, the variability observed among Mattenklee and field clover populations is quite remarkable.

Due to high levels of intra-population variation, separation of closely related cultivars of outbreeding species can be difficult (Guthridge et al. 2001; Kongkiatngam et al. 1995; Yu and Pauls 1993). However, we observed almost complete separation of five out of eight Mattenklee cultivars using discriminant analysis on individual plants (Fig. 2.3B). This separation was also largely congruent to cluster analysis based on coancestry coefficients and partially reflected the breeding history of the cultivars. Rüttinova was directly selected from Renova while Pica and Pavo were developed using Mattenklee germplasm as well as wild clover ecotypes collected in Croatia (Boller 2000a). Cluster analysis also separated older cultivars (last selection before 1991) from newer cultivars (Table 2.1, Fig. 2.4). In general, there was a substantial amount of variability observed among Mattenklee cultivars with coancestry coefficients comparable to values observed among cultivars of alfalfa (Mengoni et al. 2000) and white clover (Gustine et al. 2002). For red clover, two studies report similarly high levels of variability among cultivars (Campos-De-Quiroz and Ortega-Klose 2001; Kongkiatngam et al. 1996) while relatively little genetic diversity was found among 34 North American red clover cultivars by Yu et al. (2001).

One major concern in maintaining and exploiting germplasm collections is the identification of populations which truly contribute to the genetic diversity of a collection and are not simply duplicates of populations already represented (Gilbert et al. 1999). The Mattenklee landraces analysed in this study were not only distinct from Mattenklee cultivars and field clover cultivars, but most landraces were also clearly separated from each other. Although variability among landraces was larger than variability among populations of the two other groups, there were two pairs of landraces where no significant population differentiation was observed (Table 2.4, Fig. 2.4). The Swiss Mattenklee landrace collection at FAL-Reckenholz, Zurich consists of populations collected from farms where landraces were traditionally maintained for many decades. However, the exact origin of some landraces is difficult to determine (Nüesch 1976) and it can not be excluded that some landraces were supplemented with external germplasm in times of seed scarcity. This could be one possible explanation for the high genetic similarity of LR239/LR127 and LR325/LRDet. For the latter pair this is particularly likely since LRDet was officially recommended for cultivation for many years (Badoux et al. 1967). Based on genetic distances, it appears appropriate to summarise the eight Mattenklee landraces in two major germplasm pools (MR I, MR III; Fig. 2.4). The Mattenklee landraces analysed in this study may

be sufficiently represented through one population of each pool and the population LR189, which displayed high genetic distance to all landraces. However, considering the low genetic variability among some of the morphologically distinct Mattenklee cultivars, closely related landraces such as LR8 and LR292 may still represent valuable genetic resources.

The clustering of Mattenklee landraces based on coancestry coefficients did not reflect the geographic location where the populations were sampled from. For example, the genetic distance between LR8 and LR292 was quite low although the two populations were located 155 km apart, while LR189 and LR325 showed considerable genetic distance but originated from locations only 5 km apart. No significant correlation between geographic separation and genetic distance was also observed in wild emmer wheat (*Triticum dicoccoides*; Fahima et al. 2002) and white clover (Kölliker et al. 2001). Such a lack of correlation may be explained by a sharp local differentiation as opposed to a gradual change in allele frequencies across the geographic range of a species (Fahima et al. 2002). In addition to local climatic conditions, varying sources of initial red clover germplasm as well as differences in selection targets and intensities, may be plausible explanations for a strong local differentiation of Mattenklee landraces.

The results of this study strongly suggest that Swiss Mattenklee landraces form a distinct genetic resource of red clover and are genetically different from Mattenklee cultivars and field clover cultivars. Due to the high genetic variability, Mattenklee landraces form a valuable gene pool for red clover breeding as well as for the preservation of biodiversity. AFLP analysis combined with AMOVA and canonical discriminant analysis proved highly effective for identifying putative duplicates and for determination of genetic variability within and among Mattenklee landraces. Such information complements morpho-physiological evaluations and allows for an efficient management and exploitation of Mattenklee germplasm collections.

Acknowledgments

The authors thank Yvonne Häfele for excellent technical assistance and Beat Reidy (SHL Zollikofen, Switzerland) for careful revision of the manuscript.

3 Optimisation of bulked AFLP analysis and its application for exploring diversity of natural and cultivated populations of red clover

D. Herrmann, B. Boller, F. Widmer, R. Kölliker

Genome (2005) 48:474-486

Abstract

Landraces and wild populations of red clover (*Trifolium pratense* L.) may represent a significant yet poorly characterised genetic resource of temperate grasslands. A bulking strategy with amplified fragment length polymorphism (AFLP) markers was optimised to characterise 120 red clover populations in 6 different groups: Swiss wild clover populations, Mattenklee landraces, Mattenklee cultivars, field clover cultivars, Dutch wild clover populations, and Dutch landraces. Analysis of 2 bulked samples/population consisting of 20 plants each with 12 AFLP primer combinations was found optimal for determining genetic diversity and relationships within and among red clover populations and groups. Swiss wild clover populations were clearly separated from all other red clover groups and variability within and among populations was shown to be particularly high in wild clover populations and Mattenklee landraces, emphasising their value as genetic resources for improvement of red clover cultivars as well as for conservation and restoration of biodiversity. This study shows that the ancestry of red clover landraces is primarily found in introduced cultivars rather than in natural wild clover populations. In addition, the methodological considerations presented here may help to improve diversity analyses using bulked samples.

Keywords: AFLP, bulked samples, *Trifolium pratense* L., cultivars, landraces, wild populations

3.1 Introduction

Molecular markers such as amplified fragment length polymorphism (AFLP; Vos et al. 1995) allow for a rapid assessment of plant genetic diversity and are suitable for the investigation of ancestry of cultivated species, as well as for the analysis of relationship among natural and cultivated populations (Cresswell et al. 2001; Hernandez-Verdugo et al. 2001; Prashanth et al. 2002). To optimally manage genetic resources for the improvement of cultivars, as well as to maintain and restoring biodiversity, knowledge of genetic diversity within species is indispensable. However, detailed analysis of germplasm diversity requires a large number of accessions to be analysed. DNA or plant material of several individual plants may be combined into a single bulked sample in order to reduce the number of samples processed and to accumulate population-specific markers (Michelmore et al. 1991). Although bulked AFLP analysis is frequently used (Kölliker et al. 2001; Puecher et al. 2001; Segovia-Lerma et al. 2003), there is little experimental evidence as to how the bulking method (i.e. the number of plants combined in a sample and the number of bulked samples analysed) influences estimates of relationship among populations.

Red clover (*Trifolium pratense* L.) is an important component of both ley grass farming and of permanent pastures and meadows in temperate regions. Although several varieties have been identified, the taxonomy of this outcrossing and polymorphic species is complex and remains unclear (Hess et al. 1970; Zohary and Heller 1984). At present, wild red clover populations, initially introduced from the southeastern Mediterranean region into western Europe (Taylor and Quesenberry 1996), are an important component of natural grasslands and may provide valuable genetic resources for breeding, as well as for the conservation and restoration of biodiversity in agricultural ecosystems. After a long history as a wild plant, farmers started to propagate and maintain red clover seed in the 16th century (Taylor and Quesenberry 1996). Thus, unique and locally adapted landraces developed under moderate human interference (Boller 2000b; Taylor and Quesenberry 1996). The value of such landraces has been shown for various crops (Skovmand et al. 2001; Tosti and Negri 2002; Yang et al. 1994) but very few red clover landrace collections have been established and maintained (Bingefors and Akerberg 1961; Nüesch 1976). Swiss red clover landraces, generally referred to as Mattenklee landraces, are believed to originate from Flanders and Brabant (Volkart 1929), regions considered as the early centre of domestication of red clover (Taylor and Quesenberry 1996).

Although the genetic diversity of natural and cultivated populations of red clover was analysed in several studies (Campos-De-Quiroz and Ortega-Klose 2001; Greene et al. 2004; Kongkiatngam et al. 1996; Kölliker et al. 2003; Mosjidis et al. 2004; Semerikov et al. 2002; Ulloa et al. 2003; Yu et al. 2001) and some Mattenklee landraces have been used to develop cultivars that are broadly used in Switzerland and various other countries (Boller 2000b), the vast gene pool available in Mattenklee landraces is not yet completely characterised and barely exploited. In addition, there is a lack of knowledge as to the ancestry of red clover landraces or of the interactions between cultivated and wild populations. Therefore, the objectives of this study were (*i*) to provide a strategy for large scale analyses of outcrossing species such as red clover based on AFLP markers and bulked samples, (*ii*) to determine genetic diversity in particularly within and among Mattenklee landraces and wild red clover populations, and (*iii*) to investigate the relationship of these populations to red clover cultivars and populations from the Netherlands, which represent the early centre of red clover domestication.

3.2 Materials and Methods

Plant material

The 120 populations analysed in this study represented 6 groups of red clover (*Trifolium pratense* L.; Table 3.1). Mattenklee landraces (group 1) were collected from Swiss farms in

1971 and 1972 (Nüesch 1976) and stored as seed in a germplasm collection. Plants of each landrace were grown for seed regeneration and multiplication from 1999 to 2003. Mattenklee cultivars (group 2) were bred in Switzerland and are distinguished from foreign field clover cultivars (group 3) by increased persistence. Swiss wild clover populations (group 4) were collected as seed from at least 50 individual plants in natural grasslands in Switzerland. Seed samples of 2 landraces (group 5) and 4 wild clover populations (group 6) from the Flanders and Brabant regions were kindly supplied by the Centre for Genetic Resources the Netherlands (CGN, Wageningen). *Trifolium repens* L. 'Bombus', a white clover, was included as a reference. Plants of each population were grown from seed in the greenhouse (16 h light (17°C) : 8 h dark (13 °C)). An individual trifoliate leaf and (or) individual plants combined into a single sample consisting of 1 trifoliate leaf from different plants (bulked sample) were collected for DNA extractions.

Table 3.1 Red clover (*Trifolium pratense* L.) populations of 6 red clover groups and 1 white clover (*Trifolium repens* L.) population used in this study

		Geographic location of collection sites[b]		
Name	Origin[a]	Elevation (m asl)	Longitude (°E)	Latitude (°N)
Mattenklee cultivars				
'Changins'	Agroscope RAC (CH)			
'Corvus'	Agroscope FAL (CH)			
'Formica'	Agroscope FAL (CH)			
'Leisi'	Unterwagenburg (CH)			
'Merula'	Agroscope FAL (CH)			
'Milvus'	Agroscope FAL (CH)			
Field clover cultivars				
'Diper'	Agri obtentions (F)			
'Lucrum'	Saatzucht Steinach (D)			
'Merviot'	CLO-DvP (B)			
'Mont Calme'	Agroscope RAC (CH)			
'Suez'	Agrogen (CZ)			
'Violetta'	CLO-DvP (B)			
Swiss wild clover populations				
Appenzell	Brülisau (CH)	950	9°27'18"	47°17'52"
Boppelser Waid	Boppelsen (CH)	580	8°25'08"	47°28'09"
Chaischbu	Trub (CH)	980	7°50'54"	46°56'58"
Dettenbühl L.i.	Wiedlisbach (CH)	580	7°38'02"	47°15'29"
Dettenbühl TW	Wiedlisbach (CH)	580	7°37'58"	47°15'28"
Hemmental	Hemmental (CH)	578	8°36'01"	47°45'46"
Hilterfingen	Hilterfingen (CH)	645	7°40'23"	46°43'55"
Längenschwand	Eggiwil (CH)	999	7°45'48"	46°53'06"
Langnau	n.a.[c] (CH)	n.a.	n.a.	n.a.
Lenggenwil	Lenggenwil (CH)	567	9°08'37"	47°27'42"
Oberehrendingen	Oberehrendingen (CH)	590	8°21'19"	47°29'15"
Sihlfeld	Zurich (CH)	415	8°30'23"	47°22'32"
Unterwagenburg	Unterwagenburg (CH)	575	8°38'31"	47°29'25"
Dutch wild clover populations				
CGN07319	Roosteren-Stein, Limburg (NL)	30	5°48'	51°05'
CGN07330	Hoogcruts, Limburg (NL)	150	5°50'	50°47'
CGN07346	Diessen, Noord Brabant (NL)	40	5°11'	51°28'
CGN07351	Boerehol-Breskens, Zeeland (NL)	3	3°34'	51°23'

Table 3.1 (*continued*)

Name	Origin[a]	Geographic location of collection sites[b]		
		Elevation (m asl)	Longitude (°E)	Latitude (°N)
Dutch landraces				
CGN20064 (Rode Maasklaver)	n.a. (NL)	n.a.	n.a.	n.a.
CGN20065 (Rosendaalse Rode Klaver)	n.a. (NL)	n.a.	n.a.	n.a.
Mattenklee landraces				
LR1	Hirzel (CH)	700	9°35'54"	48°14'08"
LR2	Uttigen (CH)	580	7°33'45"	46°47'52"
LR3	Brütten (CH)	613	8°39'36"	47°29'00"
LR4	Eriswil (CH)	880	7°51'12"	47°04'00"
LR5	Schmitten (CH)	640	7°15'12"	46°50'52"
LR6	Affoltern i.E. (CH)	800	7°43'10"	47°04'52"
LR8	Bubikon (CH)	510	8°49'40"	47°15'58"
LR9	Eriswil (CH)	880	7°51'58"	47°03'46"
LR21	Grossdietwil (CH)	570	7°53'08"	47°10'21"
LR50	Huttwil (CH)	800	7°50'56"	47°05'38"
LR51	Huttwil (CH)	800	7°50'56"	47°05'38"
LR52	Ufhusen (CH)	705	7°54'10"	47°06'06"
LR60	Huttwil (CH)	700	7°51'00"	47°06'22"
LR62	Ursenbach (CH)	650	7°46'00"	47°08'28"
LR71	Huttwil (CH)	700	7°51'33"	47°06'19"
LR75	Niederwangen (CH)	600	7°22'36"	46°55'44"
LR76	Bern (CH)	560	7°21'45"	46°56'50"
LR78	Bern (CH)	560	7°22'10"	46°57'18"
LR86	Frauenkappeln (CH)	600	7°21'33"	46°57'24"
LR88	Riedbach (CH)	570	7°21'00"	46°56'43"
LR119	Bowil (CH)	820	7°42'42"	46°53'28"
LR121	Oberthal (CH)	950	7°41'20"	46°55'12"
LR125	Zäziwil (CH)	790	7°40'12"	46°53'40"
LR127	Zäziwil (CH)	700	7°40'18"	46°54'22"
LR128	Bowil (CH)	923	7°43'24"	46°53'10"
LR138	Lauperswil (CH)	900	7°42'30"	46°57'06"
LR140	Signau (CH)	985	7°45'15"	46°53'45"
LR149	Lauperswil (CH)	690	7°44'12"	46°56'44"
LR154	Signau (CH)	733	7°45'42"	46°55'08"
LR157	Schüpbach (CH)	730	7°44'06"	46°55'40"
LR160	Rüegsau (CH)	570	7°40'30"	47°01'32"
LR161	Rüegsau (CH)	570	7°39'42"	47°01'40"
LR163	Bigenthal (CH)	785	7°39'15"	46°58'44"
LR167	Goldbach i.E. (CH)	620	7°40'06"	47°00'04"
LR176	Krauchthal (CH)	700	7°34'45"	47°00'54"
LR186	Affoltern i.E. (CH)	800	7°43'20"	47°03'32"
LR189	Sumiswald (CH)	700	7°45'40"	47°02'06"
LR198	Sumiswald (CH)	735	7°45'12"	47°02'54"
LR199	Wasen i.E. (CH)	850	7°47'36"	47°02'13"
LR210	Sumiswald (CH)	850	7°44'42"	47°01'28"
LR222	Ostermundigen (CH)	550	7°29'27"	46°56'47"
LR225	Belpberg (CH)	800	7°31'36"	46°52'42"
LR226	Belpberg (CH)	792	7°31'12"	46°52'56"
LR229	Belpberg (CH)	792	7°31'12"	46°52'56"
LR231	Köniz (CH)	700	7°26'00"	46°55'00"
LR239	Köniz (CH)	680	7°24'44"	46°54'30"
LR242	Köniz (CH)	650	7°25'10"	46°54'50"
LR247	Köniz (CH)	800	7°25'38"	46°53'00"
LR249	Englisberg (CH)	840	7°27'16"	46°53'47"
LR260	Köniz (CH)	645	7°23'10"	46°54'00"
LR261	Gasel (CH)	640	7°23'18"	46°54'00"
LR262	Niederwangen (CH)	600	7°23'06"	46°55'36"
LR264	Köniz (CH)	480	7°22'08"	46°54'20"
LR272	Niederscherli (CH)	650	7°23'40"	46°53'30"
LR279	Köniz (CH)	700	7°25'20"	46°54'00"
LR280	Oberuzwil (CH)	625	9°07'40"	47°25'46"

Table 3.1 (*concluded*)

		Geographic location of collection sites[b]		
Name	Origin[a]	Elevation (m asl)	Longitude (°E)	Latitude (°N)
LR282	Schwarzenburg (CH)	770	7°19'39"	46°49'34"
LR288	Schwarzenburg (CH)	830	7°21'12"	46°49'40"
LR291	Lanzenhäusern (CH)	764	7°22'10"	46°51'00"
LR292	Lanzenhäusern (CH)	810	7°21'24"	46°50'48"
LR294	Ueberstorf (CH)	650	7°19'42"	46°52'22"
LR300	Burgistein (CH)	830	7°28'55"	46°47'50"
LR307	Riggisberg (CH)	760	7°28'40"	46°48'40"
LR309	Riggisberg (CH)	760	7°28'40"	46°48'40"
LR310	Riggisberg (CH)	760	7°28'40"	46°48'40"
LR311	Riggisberg (CH)	800	7°28'40"	46°48'40"
LR313	Riggisberg (CH)	805	7°28'10"	46°48'35"
LR314	Rüti (CH)	900	7°28'30"	46°47'15"
LR315	Rüeggisberg (CH)	850	7°25'20"	46°48'40"
LR318	Riggisberg (CH)	780	7°28'40"	46°48'40"
LR321	Oberönz (CH)	470	7°41'42"	47°10'50"
LR322	Courrendlin (CH)	490	7°21'30"	47°20'00"
LR323	Kaufdorf (CH)	538	7°30'10"	46°50'25"
LR324	Courrendlin (CH)	490	7°21'30"	47°20'00"
LR325	Affoltern i.E. (CH)	800	7°44'06"	47°03'56"
LR326	Affoltern (CH)	800	7°43'40"	47°04'14"
LR327	Weier i.E. (CH)	765	7°45'33"	47°03'25"
LR328	Affoltern i.E. (CH)	748	7°45'20"	47°03'56"
LR329	Oeschenbach (CH)	730	7°44'54"	47°06'08"
LR330	Oeschenbach (CH)	650	7°45'22"	47°07'16"
LR332	Rüedisbach (CH)	760	7°42'54"	47°06'30"
LR333	Häusernmoos (CH)	790	7°44'24"	47°05'10"
LR336	Schmidigen (CH)	775	7°43'48"	47°05'00"
LR340	Neuenegg (CH)	520	7°16'15"	46°54'20"
LR343	Kriechenwil (CH)	530	7°13'43"	46°54'46"
LR346	Ueberstorf (CH)	650	7°18'42"	46°52'00"
LR351	Arni b. Biglen (CH)	890	7°39'08"	46°55'30"
LRDet	Wiedlisbach (CH)	508	7°38'08"	47°15'18"
LRRüt	Rütti Zollikofen (CH)	550	7°27'50"	46°59'25"
White clover cultivar				
'Bombus'	Agroscope FAL (CH)			

[a] Breeding company (cultivars) or name of collection sites (landraces and wild clover populations)

[b] For Mattenklee landraces the geographic locations of the farms rather than actual collection sites are given

[c] Information not available

AFLP Analyses

Leaf tissue was ground in liquid nitrogen and DNA was extracted using the DNeasy 96 plant kit (Qiagen, Hilden, Germany). AFLP analysis was performed following the method of Vos et al. (1995). From each sample, 2 AFLP templates were prepared by digesting 500 ng DNA with the restriction enzyme combinations *Eco*RI-*Mse*I or *Pst*I-*Mse*I and by ligating the corresponding oligonucleotide adaptors (Muluvi et al. 1999; Vos et al. 1995) in a total volume of 60 µL. Preselective PCR amplification with primers corresponding to the adaptor core sequences (E, P and M; Table 3.2) extended by 1 selective nucleotide (E + A or P + A and M + C) was performed in a 20-µL reaction containing 2 µL of AFLP template. PCR contained 1x PCR buffer, 0.25 µmol/L of each primer, 1.5 mmol of $MgCl_2$/L, 0.2 mmol/L of each dNTP and 0.5 U

Taq DNA polymerase (Invitrogen, Carlsbad, Calif.) and was performed using a Biorad iCycler (Biorad, Hercules, Calif.). PCR conditions consisted of 2 min at 72 °C, 2 min at 94 °C, 26 cycles of 1 min at 94 °C, 1 min at 56 °C and 1 min at 72°C, followed by a final extension of 5 min at 72 °C. Amplification products were diluted in 100 µL deionised H_2O and 1 µL was used for selective amplification in a total volume of 20 µL containing 0.25 µmol/L of fluorescently labelled E- or P-specific primer extended by 3 selective nucleotides (Table 3.2), 0.3 µmol/L of M-specific primer extended by 3 selective nucleotides (Table 3.2), 1.5 mmol of $MgCl_2$/L, 0.2 mmol/L of each dNTP and 0.4 U of *Taq* DNA polymerase (Invitrogen). PCR was performed using a touchdown protocol with an initial denaturation of 2 min at 94°C, 12 cycles with 30 s at 94 °C, 30 s at 65 °C (decreasing the temperature by 0.7 °C after each cycle), and 1 min at 72 °C, followed by 23 cycles of 30 s at 94 °C, 30 s at 56 °C and 1 min at 72 °C with a final extension of 2 min at 72 °C. AFLPs were analysed on an ABI Prism 3100 genetic analyser using POP-4 polymer and 36-cm capillaries (Applied Biosystems, Foster City, Calif.).

Table 3.2 AFLP primer combinations used during optimisation of bulking strategy and to detect polymorphic loci among 120 red clover (*Trifolium pratense* L.) and 1 white clover (*Trifolium repens* L.) population

	Primer pairs (5' → 3')	No. of loci		Primer pairs (5' → 3')	No. of loci
1	E[a]+AGA/M[b]+CTA	n.a.[d]	15	P[c]+AAC/M+CAT	n.a.
2	E+AGT/M+CTA	n.a.	16	P+ACA/M+CAC	n.a.
3	E+ACT/M+CAC	n.a.	17	P+ACT/M+CAC	n.a.
4	E+ACT/M+CTA	n.a.	18	P+AGA/M+CAC	n.a.
5	E+AGG/M+CAC	n.a.	19	P+AGG/M+CTA	n.a.
6	E+AAC/M+CAT	n.a.	20	P+AGT/M+CAT	n.a.
7	E+ACA/M+CTA	n.a.	21	P+AAC/M+CAC	32
8	E+AGT/M+CAT	n.a.	22	P+ACA/M+CTA	38
9	E+AAC/M+CAC	39	23	P+ACT/M+CTA	29
10	E+AGA/M+CAT	23	24	P+AGA/M+CAT	33
11	E+AGA/M+CAC	34	25	P+AGG/M+CAC	44
12	E+ACA/M+CAC	35	26	P+AGT/M+CAC	32
13	E+AGT/M+CAC	35			
14	E+AGG/M+CTA	36			
	Subtotal	**202**		**Subtotal**	**208**

[a] Primer core of *Eco*RI adaptor 5'-GACTGCGTACCAATTC-3'

[b] Primer core of *Mse*I adaptor 5'-GATGAGTCCTGAGTAA-3'

[c] Primer core of *Pst*I adaptor 5'-GACTGCGTACATGCAG-3'

[d] Not available; primer combination was only used for optimisation of the bulking strategy

Data analyses

AFLP data was analysed using the Genotyper 3.6. software (Applied Biosystems) where relative migration units (RMU) reflected the size and relative fluorescence units (RFU) reflected the quantity of amplicons detected. AFLP loci were characterised by the primer combination used for selective amplification and the size of the amplicon in RMU. The term AFLP marker was used to describe the positive allele (i.e., amplicon) of an AFLP locus. AFLP markers were visually scored for presence (1) or absence (0) and entered into a binary data matrix.

Analysis of molecular variance (AMOVA) was used to calculate variance components within and among groups and populations. Significance of variance components was tested using a non-parametric procedure based on 1000 permutations (Excoffier et al. 1992). F_{st} values were used as a measure of genetic distance between pairs of red clover groups. Significance of F_{st} values was tested in order to elucidate the significance of differentiation between groups using a non-parametric procedure based on 1000 data permutations (Excoffier et al. 1992). Pairwise genetic distances between groups were calculated using coancestry coefficients derived from F_{st} values (Reynolds et al. 1983). AMOVA and coancestry coefficients, as well as significance of variance components and F_{st} values, were computed using the Arlequin 2.0 software (Schneider et al. 2000). Principal coordinate analysis (PCoA) based on coancestry coefficients was performed by double centering the distance matrix, i.e., transforming to a matrix where the diagonal elements can be considered to be the squared distance from each population to the centroid of the space, using NTSYSpc 2.1 software (Rohlf 2000).

Cluster analysis based on marker frequencies per group, in which each population consisted of 2 bulked samples, and genetic distances (Nei 1972) was calculated using the unweighted pair-group method with an arithmetic mean (UPGMA) of the NTSYSpc 2.1 software (Rohlf 2000). Reliability of the clustering was tested with Mantel test statistics for correlation of distance matrices and cophenetic matrices obtained by transforming the hierarchical system of clusters into a symmetrical matrix (Rohlf 2000). Bootstrap analysis was performed with 1000 data permutations on marker frequencies per group and Nei's genetic distances using the PHYLIP software package (Felsenstein 2002).

To estimate the relationship among Swiss wild clover populations and correlations of Swiss wild clover populations to environmental factors, redundancy analyses (RDA) were calculated. Monte Carlo permutation tests based on all canonical axes were performed in order to test significance of environmental factors on the genetic structure of Swiss wild clover populations. RDA and permutation tests were computed using the Canoco 4.5 software (ter Braak and Smilauer 2002).

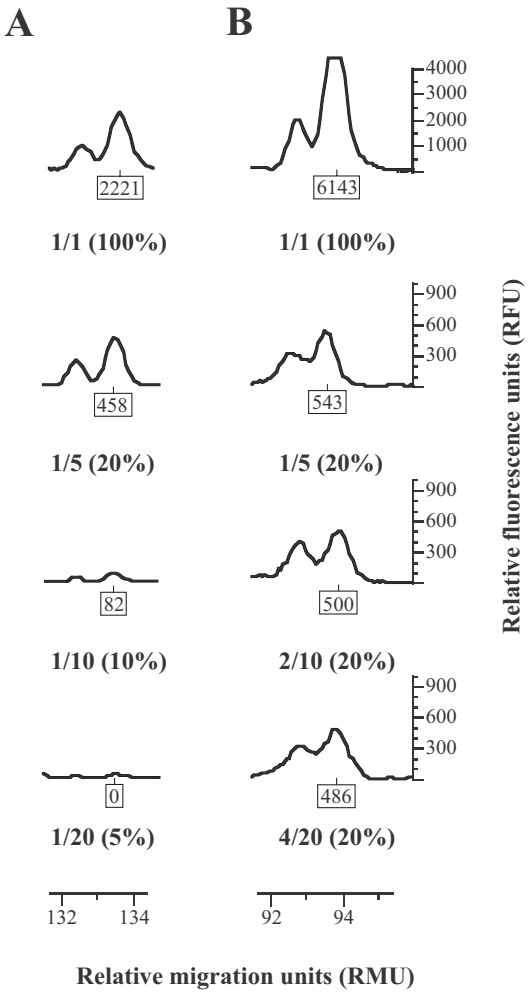


Fig. 3.1 Two series of electropherograms showing detection limits of AFLP markers occurring in only one plant in different bulked samples (**A**) or at a constant frequency of 20% in different bulked samples (**B**). Ratios indicate the number of plants containing the respective marker and total number of plants in the sample. The percentage of plants in the sample containing the marker is indicated in parentheses. The relative fluorescence units (RFU) are indicated in boxes

3.3 Results

Bulking strategy for large-scale analysis of red clover populations

For the efficient genetic distinction of a large number of red clover populations, an AFLP approach on bulked leaf samples was established. For this purpose, 4 steps of optimisation were performed to determine the number of plants needed in each bulked sample (steps 1 and 2), the number of bulked samples required for each population (step 3) and the number of AFLP analyses needed for each sample for a reliable differentiation (step 4).

In step 1, the detection limit for markers present in only one single plant was determined in different bulked samples containing variable numbers of plants not carrying the respective marker. Bulked DNA samples of 5, 10, 15 and 20 individual plants as well as single DNA samples of the same 20 individual plants of the Mattenklee landrace LR300 (Table 3.1) were analysed using 6 AFLP primer combinations (Nos. 1-5 and 12 in Table 3.2). Markers present in

only one plant were still unambiguously scorable (i.e., detected at intensities larger than 300 RFU) in bulked samples consisting of 5 plants (marker frequency = 20%) but not in samples consisting of 10 (10%) or 20 (5%) plants (Fig. 3.1 A). In bulked samples of 10 plants, markers had to be present in at least 2-3 plants (20%-30%) to be scorable, and in samples consisting of 20 plants 4-5 plants (20%-25%) were required (Fig. 3.1 B). Thus, markers present at frequencies of at least 20%-30% were scorable in bulked samples regardless of the number of plants combined in the sample.

In step 2, the effect of the number of plants combined in a bulked sample on the number of polymorphic loci detected between pairs of populations was assessed. Two bulked DNA samples consisting of 10 plants and 1 bulked DNA sample consisting of 15 or 20 plants were prepared from 4 populations (LR8, LR189, LR292, LRDet; Table 3.1) and compared using 6 AFLP primer combinations (1-5 and 12 in Table 3.2). On average, 19.3 polymorphic loci were observed between pairs of populations with samples of 10 plants, 19.8 polymorphic loci with samples of 15 plants and 21.2 polymorphic loci with samples of 20 plants (Table 3.3).

Table 3.3 Number of polymorphic loci between pairs of populations represented by bulked samples consisting of 10, 15 and 20 plants from 4 Mattenklee landraces

	10 plants/sample[a]			15 plants/sample[b]			20 plants/sample[b]		
	LR8[c]	LR189	LR292	LR8	LR189	LR292	LR8	LR189	LR292
LR189	20			23			23		
LR292	12	22		16	23		19	22	
LRDet	20	21	21	18	19	20	23	22	18

[a] Number of polymorphic loci with 2 bulked samples/population

[b] Number of polymorphic loci with 1 bulked sample/population

[c] For description of populations see Table 3.1

In step 3, the possibility of increasing the total number of markers by analysing 2 bulked samples per population was tested. A total of 45 polymorphic loci were scored across 3 populations (LR8, LR292, LR300; Table 3.1) with 2 bulked samples consisting of leaves of 20 or 15 plants each using 6 AFLP primer combinations (1-5 and 12, Table 3.2). Although 10-16 markers were detected in individual bulked samples, the 2 bulked samples of the same population consisting of the same number of different plants yielded a total of 14-22 markers (Table 3.4). Thus, the use of 2 bulked samples per population resulted in 2-7 additional markers.

With a total of 45 loci, this corresponds to an increase of up to 16% when compared with analysis based on a single bulked sample per population.

Table 3.4 Number of markers detected in 2 bulked samples (A and B) consisting of 15 or 20 plants from 3 Mattenklee landraces

	15 plants/sample			20 plants/sample		
	LR8[a]	LR292	LR300	LR8	LR292	LR300
Sample A[b]	12	14	15	12	12	15
Sample B[b]	10	12	15	12	12	16
Common[c]	8	7	8	10	7	11
Total[d]	14	19	22	14	17	20

[a] For description of populations see Table 3.1

[b] Numbers represent markers detected in each bulked sample

[c] Numbers represent markers common in both samples

[d] Total number of markers detected in both samples

In step 4, the minimal number of AFLP primer combinations required to obtain robust estimates of genetic diversity was evaluated. Two bulked samples consisting of leaves of 20 plants were analysed for 12 populations (3 Mattenklee landraces, 3 Mattenklee cultivars, 3 field clover cultivars and 3 Swiss wild clover populations). UPGMA dendrograms based on Euclidean squared distances of these 12 populations were compared for different combinations of 24 AFLP primer combinations (3-26, Table 3.2). The correlations between pairs of cophenetic matrices obtained from dendrograms based on 6 different primer combinations each ranged from 0.41 to 0.91. The corresponding dendrograms showed substantial differences for the primer combinations used (data not shown). Comparing cophenetic matrices based on 12 different primer combinations resulted in correlations ranging from 0.88 to 0.96 and dendrograms revealed largely congruent topologies (data not shown). The correlation between cophenetic matrices derived from the 12 AFLP primer combinations based on *Pst*I-*Mse*I and the 12 AFLP primer combinations based on *Eco*RI-*Mse*I was 0.92.

For all experiments described above, differences between bulked samples of the same population were mostly smaller than differences between bulked samples of different populations, regardless of the number of plants per bulked sample and the bulking method (leaf material or DNA; data not shown).

Based on the results from these optimisations, all further analyses were performed using 2 bulked leaf samples of 20 plants/population and 12 AFLP primer combinations (9-14 and 21-26 in Table 3.2).

Distinction and genetic diversity of red clover groups

The 12 AFLP primer combinations generated a total of 410 polymorphic loci across all 242 samples analysed. Of these, 212 loci were polymorphic within and among red clover populations, while the remaining 198 loci only differentiated the red clover populations from the white clover population. The dataset was complete with no missing values across all 242 samples. Genetic distances, expressed as coancestry coefficients, between red clover populations ranged from 0 (e.g. LR1/LR75) to 1.1 ('Merula'/Hilterfingen) while comparisons of red clover populations with the white clover population resulted in genetic distances from 2.6 (LR272) to 3.3 ('Corvus'). The number of loci detected with the different primer combinations ranged from 23 (E + AGA/M + CAT) to 44 (P + AGG/M + CAC) with an average of 34 loci (Table 3.2). The 6 primer combinations based on *Eco*RI-*Mse*I (202 loci) and those detected with the 6 primer combinations based on *Pst*I-*Mse*I (208 loci) yielded almost the same number of loci. Although marker frequencies varied considerably among red clover groups and populations, no group-specific markers were identified. However, one population specific marker (P + AGG/M + CAC, 237 RMU) was exclusively detected in the Mattenklee cultivar 'Corvus'. The reproducibility of AFLP analysis was demonstrated by including 40 replicated samples in the cluster analysis. Although minor differences of less than 10 out of the 212 polymorphic loci were observed between some replicates, cluster analysis always grouped replicated samples closest to each other (data not shown). Replicated samples, in addition to the white clover population, were excluded from further analyses.

Analysis of molecular variance (AMOVA) showed that the majority of the observed genetic variability was due to variation between the 2 bulked samples within populations (Table 3.5). However, a significant proportion of the total variation was attributed to variation among the 6 red clover groups.

Genetic distances between red clover groups, expressed as coancestry coefficients, ranged from 0.06 (Dutch landraces versus field clover cultivars) to 0.25 (Swiss wild clover populations versus Dutch landraces) with an average of 0.14 (Table 3.6). Swiss wild clover populations showed the largest distances to the other red clover groups. The distance between Mattenklee landraces and Mattenklee cultivars was small (0.08) and comparable to the distance between field clover cultivars and Dutch landraces (0.06), Dutch landraces and Dutch wild clover (0.07) as well as field clover cultivars and Mattenklee cultivars (0.07). On the other hand, the distance

was larger for comparisons of Mattenklee landraces and the remaining red clover populations (0.12–0.16). F_{st} values of 14 pairwise comparisons of the different red clover groups were significant at $P < 0.05$ (Table 3.6), indicating significant genetic differences between groups. However, there was no significant distance between field clover cultivars and Dutch landraces. This was congruent with the results obtained by a pairwise Monte Carlo permutation test based on canonical axes derived from redundancy analysis (data not shown). Average genetic distance, expressed as coancestry coefficients, of the field clover 'Suez' to all other field clover cultivar was 0.51, while average genetic distance among the 5 remaining field clover cultivars was 0.36. In addition, PCoA separated 'Suez' from all other field clover cultivars (population 1 in Fig. 3.2).

Table 3.5 Analysis of molecular variance (AMOVA) for the 6 red clover groups based on 410 AFLP loci and 120 populations represented by 2 bulked samples each

Source of variation	d.f.	Sum of squares	Variance components[a]	Percentage of variation
Variance among groups	5	293.86	1.98	12.01
Variance among populations within groups	114	2022.91	3.25	19.71
Variance within populations	120	1350.00	11.25	68.28

[a] Components were significant at $P < 0.001$, the probability of obtaining a more extreme random value computed from non-parametric procedures (1000 data permutations)

Table 3.6 Genetic distances between the 6 red clover groups based on 410 AFLP loci and 120 populations with 2 bulked samples each (below diagonal: coancestry coefficients (Reynold et al., 1983); above diagonal: significance of F_{st} values determined using a non-parametric permutation approach with 1000 permutations, * = $P < 0.05$, ns = no significant difference)

	Field clover cultivars	Mattenklee landraces	Dutch landraces	Mattenklee cultivars	Swiss wild clover	Dutch wild clover
Field clover cultivars		*	ns	*	*	*
Mattenklee landraces	0.12		*	*	*	*
Dutch landraces	0.06	0.16		*	*	*
Mattenklee cultivars	0.07	0.08	0.10		*	*
Swiss wild clover	0.25	0.16	0.25	0.19		*
Dutch wild clover	0.12	0.13	0.07	0.10	0.18	

The first 3 axes of PCoA explained 73.2% of the total variation observed among samples (Fig. 3.2). While the first principal coordinate accounted for 32.9% of the variance and mainly separated individual populations within groups, the second principal coordinate accounted for 24.3% of the variance and clearly separated Swiss wild clover populations (group mean = -0.26) from the remaining 5 red clover groups (group means between –0.01 and 0.18). The third principal coordinate explained 16.0% of the variance and separated Mattenklee landraces (group mean = -0.04) from field clover cultivars, Dutch landraces and Dutch wild clover (group means 0.11 to 0.16). PCoA grouped the 4 Mattenklee cultivars 'Changins', 'Corvus', 'Formica' and 'Leisi' close to the Mattenklee landraces, whereas the remaining 2 Mattenklee cultivars 'Merula' and 'Milvus', were grouped among field clover cultivars and Dutch populations. The field clover 'Mont Calme' (population 2 in Fig. 3.2) also grouped close to the Mattenklee landraces. On the other hand, some Mattenklee landraces such as LR222 were grouped close to field clover cultivars and Dutch landraces, while several other Mattenklee landraces, such as LR62, were grouped close to some of the Swiss wild clover populations (populations 3 and 4 in Fig. 3.2).

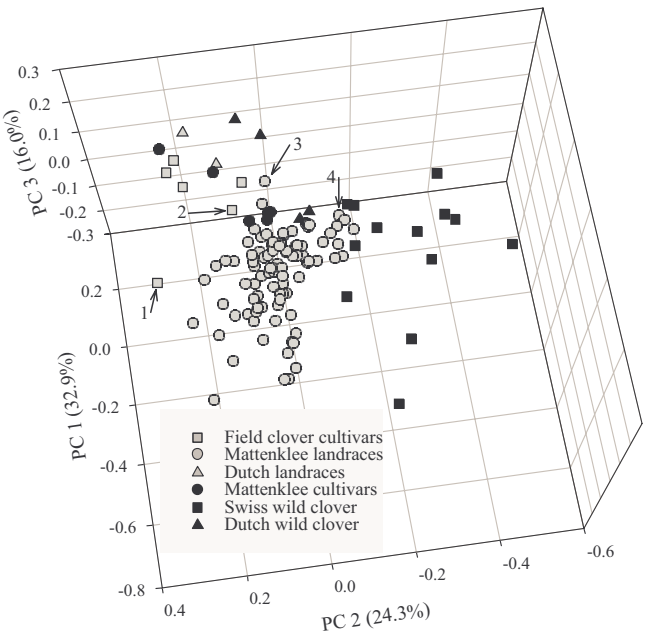


Fig. 3.2 Principal coordinate analysis where data points represent 120 populations of the 6 red clover groups (Mattenklee landraces, Mattenklee cultivars, field clover cultivars, Swiss wild clover, Dutch wild clover and Dutch landraces). Analysis based on coancestry coefficients (Reynolds et al., 1983) using 410 AFLP loci from 120 populations represented by 2 bulked samples each. Numbers indicate samples of particular interest described in the text: 1, 'Suez'; 2, 'Mont Calme'; 3, LR222; 4, LR62

UPGMA clustering based on Nei's genetic distances derived from marker frequencies per group clearly separated Swiss wild clover populations from the remaining 5 red clover groups (Fig. 3.3). Mattenklee landraces and Mattenklee cultivars formed a separate cluster and were

clearly separated from field clover cultivars. Dutch wild clover and Dutch landraces were grouped between the cluster containing Mattenklee landraces, Mattenklee cultivars, and field clover cultivars and the Swiss wild clover branch. The correlation coefficient between cophenetic values derived from the dendrogram and coancestry coefficients was r = 0.78, indicating a good fit of the data from the cluster analysis with the original distance matrix. The reliability of the dendrogram was supported by bootstrap values higher than 85% for most nodes. A lower value of 53% was only observed for the node between the Dutch landraces and Dutch wild clover populations (Fig. 3.3).

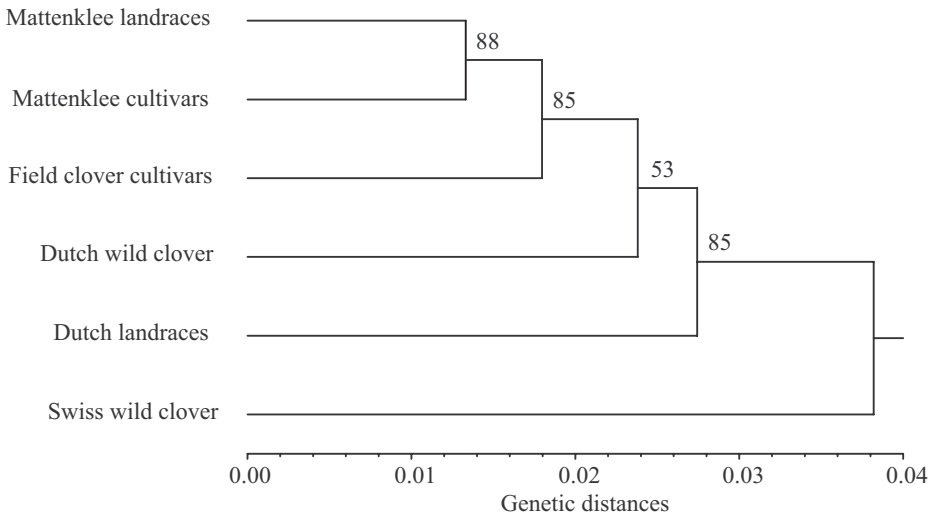


Fig. 3.3 UPGMA clustering of the 6 red clover groups (Mattenklee landraces, Mattenklee cultivars, field clover cultivars, Swiss wild clover, Dutch wild clover and Dutch landraces). Analysis was based on genetic distances (Nei, 1972) derived from marker frequencies per group, in which each population consisted of 2 bulked samples. Analysis was performed using 410 AFLP loci from 120 populations represented by 2 bulked samples each. Numbers above branches indicate the percentage of bootstrap values derived from 1000 permutations

Euclidean squared distances and mean sum of squares between the 2 bulked samples of each population within a group were highest for Swiss wild clover populations followed by Mattenklee landraces (Table 3.7). Between populations of each individual group, average Euclidean squared distances and mean sum of squares were also clearly higher for the Swiss wild clover populations when compared to the other groups (Table 3.7).

Table 3.7 Genetic differences between populations and bulked samples within the 6 red clover groups based on 410 AFLP loci and 120 populations with 2 bulked samples each

	Euclidean squared distances[a]	Mean sum of squares[b]
Mattenklee landraces (89 populations)		
Between populations	30.10	16.68**
Between bulked samples	22.64	11.32
Swiss wild clover (13 populations)		
Between populations	36.50	22.37**
Between bulked samples	25.00	12.50
Mattenklee cultivars (6 populations)		
Between populations	31.90	21.05**
Between bulked samples	19.17	9.58
Field clover cultivars (6 populations)		
Between populations	30.90	20.90**
Between bulked samples	20.50	10.25
Dutch wild clover (4 populations)		
Between populations	31.30	20.08**
Between bulked samples	22.00	11.00
Dutch landraces (2 populations)		
Between populations	29.00	16.50n.s.
Between bulked samples	17.00	8.50

[a]Average distances for pairwise comparisons of populations and bulked samples, respectively

[b] Derived from analysis of molecular variance

** variance component significant at $P < 0.01$; n.s variance component not significant

Diversity and geographic location of Swiss wild clover populations

The 4 canonical axes of a redundancy analysis (RDA) of the 12 Swiss wild clover populations, based on 410 AFLP loci and the 4 geographic variables altitude, longitude, latitude and the distance of each Swiss wild clover population to the closest collection site of Mattenklee landraces, explained 29 % of the total variation (Fig. 3.4). While the first axis was mainly correlated with the environmental variable distance (64 %), the second axis was mainly explained by longitude and latitude with a product-moment correlation coefficient of 75 % and 74 %, respectively. RDA moderately separated the populations based on distance to the closest

Mattenklee landrace collection site (triangles versus circles in Fig. 3.4) and based on longitude of wild clover collection sites (black versus grey in Fig. 3.4).

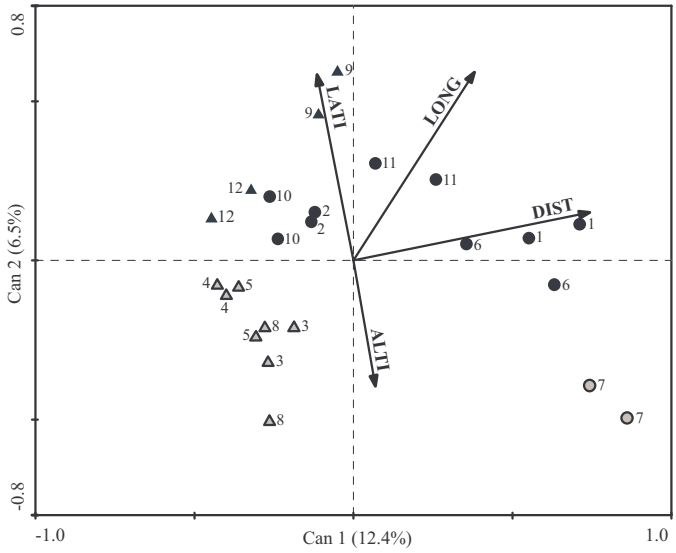


Fig. 3.4 Redundancy analysis of twelve Swiss wild red clover populations based on 410 AFLP loci and 4 geographic variables : ALTI, altitude (m above sea level); LATI, latitude (°N); LONG, longitude (°E); DIST, distance (km) to the closest Mattenklee landrace collection site. Numbers indicate the 12 Swiss red clover populations represented by 2 bulked samples each: 1, Appenzell; 2, Boppelser Waid; 3, Chaischbu; 4, Dettenbühl L.i.; 5, Dettenbühl TW; 6, Hemmental; 7, Hilterfingen; 8, Längenschwand; 9, Lenggenwil; 10, Oberehrendingen; 11, Sihlfeld; 12, Unterwagenburg (for details of populations see Table 1). Grey symbols mark populations collected between 7°37'E and 7°51'E and black symbols indicate populations collected between 8°21'E and 9°28'E. Triangles indicate populations collected closer than 11 km from the closest Mattenklee landrace collection site and circles indicate populations collected further than 11 km from the closest Mattenklee landrace collection site

The correlation coefficient between altitude and the 2 variables distance and longitude was 1 % and 5 %, respectively, while the correlation coefficient among all other pairs of environmental variables ranged from 44 to 60 %. Monte Carlo permutation testing on all canonical axes was significant ($P < 0.05$) for the combination of all 4 environmental variables indicating significant influence of the environmental variables on the genetic structure of Swiss wild clover populations as detected with AFLP analysis. Individual analysis of each environmental variable revealed significant effects of distance ($P = 0.0015$.) and longitude ($P = 0.0249$), explaining 9 % and 7 % of the total variation, respectively.

3.4 Discussion

AFLP analysis on bulked leaf samples allowed us to distinguish the 6 different red clover groups and to investigate the genetic diversity within and among red clover populations. Two bulked samples per population, each consisting of 20 individual plants, were selected for the analysis, based on information from the literature and on the results of our optimisation. The reasons of doing so were three fold. First, the greater the number of plants included in a study the higher the probability of rare markers also being included (Crossa 1989), i.e., the population is better represented. To describe distinctness, uniformity and stability of a variety, UPOV (the International Union for the Protection of New Varieties of Plants) calls for analysis of 60 plants/population. Second, increasing the number of plants within bulked samples did not influence the relative detection limit of less frequent markers (Fig. 3.1) and markers present at a frequency of 20%-30% were readily detected in bulked samples. The detection limit for specific markers may be substantially influenced by the level of heterozygosity at individual loci. In addition, Williams et al. (1993) have shown the number of positive alleles required to give rise to a RAPD amplicon in bulked samples varies substantially among loci. However, the detection limit of 20%-30% obtained in the present study corresponded to limits found in other species such as lupin (Gilbert et al. 1999) and white clover (Kölliker et al. 2001). Finally, analysing 2 bulked samples/population resulted in up to 16% more markers when compared with analysis based on a single bulked sample per population (Table 3.4). Although the analysis of several bulked samples per population has been previously suggested (Gilbert et al. 1999), many larger scale studies have been performed by analysis of just 1 bulked sample/population (Kölliker et al. 2001; Sawkins et al. 2001; Segovia-Lerma et al. 2003). However, the results of the optimisation experiments presented in this study and the high variation observed between bulked samples of the same population (Table 3.5) demonstrate the advantage of analysing 2 or possibly more bulked samples in outcrossing species.

The use of 12 AFLP primer combinations resulted in robust dendrograms with correlations of 0.88 – 0.96 between pairs of cophenetic matrices, while using only 6 AFLP primer combinations yielded correlations as low as 0.41. This is in congruence with the findings of Hollingsworth and Ennos (2004) who showed that the number of loci examined substantially affects the topology of dendrograms. In contrast to *Eco*RI, AFLP markers generated using *Pst*I as the rare cutting enzyme tend not to form clusters on linkage maps (Castiglioni et al. 1998; Keim et al. 1997; Young et al. 1999b). Thus, the choice of restriction enzymes may also influence results obtained from diversity analyses. However, in the present study correlations between cophenetic matrices derived from primer combinations based on *Eco*RI-*Mse*I and *Pst*I-*Mse*I, respectively, were as high as 0.92 and the choice of restriction enzymes did not influence

the number of polymorphic loci detected across the entire dataset (Table 3.2). Although the optimal number of loci to be analysed and the optimal restriction enzymes to be used may vary among species and genepools (Pejic et al. 1998; Powell et al. 1997; Young et al. 1999b; Zhu et al. 1998), the methodological considerations presented here help to improve diversity analyses using bulked samples.

The optimised AFLP analysis using bulked samples was applied to provide information on genetic diversity and potential ancestry within and among different red clover groups. As expected for an outcrossing species (Hamrick et al. 1979), AFLP analysis of 120 red clover populations revealed the majority of the diversity was due to diversity within populations, i.e., variation between bulked samples of the same population (Table 3.5). Although the estimation of diversity within populations was only based on 2 bulked samples and a loss of rare markers has certainly to be taken into account, the diversity determined within populations was comparable or only slightly lower to values obtained by investigating individual plants of outcrossing species such as European globeflower (64%; Despres et al. 2002), white clover (84%; Kölliker et al. 2001) or red clover (80%; Ulloa et al. 2003). In addition, the diversity between bulked samples appeared to reflect the level of human selection that was presumably absent for wild clover, moderate for Mattenklee landraces and most prominent for Mattenklee and field clover cultivars (Table 3.7). A similar effect of human selection has been previously observed for the comparison of landraces and cultivars in red clover (Kölliker et al. 2003) as well as for the comparison of wild and cultivated pepper (Hernandez-Verdugo et al. 2001) and meadow fescue (Kölliker et al. 1998).

Although genetic diversity among Mattenklee landraces was significant, the highest variability among populations was found in Swiss wild clover (Table 3.7). This is in agreement with other comparisons of wild populations and old varieties (Shim and Jorgensen 2000). The high variability among Mattenklee landraces and wild clover populations emphasised their value as genetic resources for red clover breeding as well as for conservation and restoration of biodiversity in pastures and meadows. However, some Mattenklee landraces showed high similarity (Fig. 3.2). Such similar populations might be sufficiently represented by a single landrace. The data presented here facilitate a targeted exploitation of this important gene pool.

A close relationship between Mattenklee landraces and Mattenklee cultivars as predicted by their breeding history (Boller 2000b) was reflected in principal coordinate analysis as well as in cluster analysis (Fig. 3.2 and Fig. 3.3). ‘Leisi’, heretofor identified as a Mattenklee cultivar, is not a true cultivar but rather a landrace commercially distributed for several decades (Boller 2000b). However, PCoA analysis showed that it was correctly allocated to the group of Mattenklee cultivars (Fig 3.2). The field clover ‘Mont Calme’ which grouped close to

Mattenklee landraces (population 2 in Fig. 3.2) was initially bred using Mattenklee germplasm and was included in the group of field clover based on its low persistence and other agronomic traits (Boller 2000b). 'Suez', a field clover cultivar from the Czech Republic, was clearly separated from all other field clover cultivars (population 1 in Fig. 3.2). This separation may be due to an entirely different gene pool or to strong selection with different focus when compared to the other red clover cultivars.

Although only 12% of the total variation was attributed to variation among the 6 red clover groups, variance component and genetic differences between groups were significant for all but 1 pairwise comparison (Tables 3.5 and 3.6). The absence of a significant differentiation between field clover cultivars and Dutch landraces may be explained by the small sample number of the latter or interbreeding with field clover cultivars which can not be completely excluded based on the information available. The large difference in group size of Mattenklee landraces and the remaining 5 groups may have influenced significance of pairwise comparisons. However, calculations using 5 different subsets of 10 randomly selected Mattenklee landraces resulted in similar significance values (data not shown).

The clear separation of Swiss wild clover populations from all other red clover groups (Fig. 3.2, Fig. 3.3, Table 3.6) strongly indicates that introduced red clover cultivars rather than Swiss wild clover populations represent the primary ancestors of Mattenklee. This is in agreement with the history of red clover cultivation which started in Flanders and Brabant and spread across Europe (Taylor and Quesenberry 1996) and the view of Volkart (1929), who proposed that Mattenklee was developed from Brabanterklee. In addition, a clear distinction of wild clover populations and red clover cultivars was also demonstrated using morphological traits (Nüesch 1960) and isozyme markers (Semerikov et al. 2002). Although Mattenklee landraces were generally separated from Swiss wild clover populations, some Mattenklee landraces were grouped close to wild clover populations (Fig. 3.2). This may be an indication for gene flow between these Mattenklee landraces and wild clover which was further supported by redundancy analysis, where the distance to the next Mattenklee landrace collection site had a significant influence on the genetic structure of Swiss wild clover populations (Fig. 3.4). Landraces and wild populations grew next to each other for many decades, allowing for mutual gene flow (Linder et al. 1998). In wild clover, considerable gene flow was observed among neighbouring populations (Mosjidis et al. 2004). In addition, neutral markers have been shown to persist in wild populations (Whitton et al. 1997). Thus, genes introgressed from Mattenklee landraces may have persisted in wild clover populations, although cross pollination has not been possible during the past 30 years.

Genetic diversity among Swiss wild clover populations was also substantially influenced by the longitude of the collection site (Fig. 3.4). Although geographic separation has been shown to support the conservation of genetically distinct populations (Meikle et al. 1999), other ecogeographic factors (Greene et al. 2004) or cultivation practices (Kölliker et al. 1998) may also substantially influence differentiation of populations.

In conclusion, this study provides evidence that Swiss wild red clover and Mattenklee landraces are not closely related and simultaneously indicates that the ancestry of Mattenklee landraces may be found in red clover germplasm introduced from Flanders and Brabant in the 18th century rather than in local wild clover populations. The high diversity observed within and among wild clover populations and Mattenklee landraces emphasises their value as a resource for intraspecific diversity, which may serve as a valuable gene pool for red clover improvement and may contribute to the conservation and restoration of biodiversity in pasture and meadows. The AFLP method applied to DNA templates extracted from bulked leaf samples provided an efficient approach to elucidate diversity, relationship, ancestry, and breeding history of red clover populations. In addition, the method presented may serve as a base for large scale diversity analyses in outcrossing grassland species.

Acknowledgements

The authors would like to thank the companies Fenaco, Otto Hauenstein Samen AG, Eric Schweizer Samen AG and Samen Steffen for providing seed of wild clover accessions; Yvonne Häfele for technical assistance in the laboratory; and Peter Tanner for support in collecting and maintaining the plant material. This study was partially funded by the breeding foundation DSP-BLW and regeneration of Mattenklee landraces was supported by the Swiss Commission for Cultivated Plant Conservation (NAP 13).

4 QTL analysis of seed yield components in red clover (*Trifolium pratense* L.)

D. Herrmann, B. Boller, B. Studer, F. Widmer, R. Kölliker

Theoretical and Applied Genetics (2006) 112:536-545
(with kind permission of Springer Science and Business Media)

Abstract

Cultivars of red clover (*Trifolium pratense* L.), an important forage crop in temperate regions, are often characterised by an unsatisfactory level of seed yield, leading to high production costs. This complex trait is influenced by many components and negatively correlated with other important traits, such as forage yield or persistence. Therefore, seed yield has proven to be difficult to improve. Thus, the objectives of this study were to assess association among seed yield components and to provide the basis for identifying molecular markers linked to QTLs for seed yield components to assist breeding for improved red clover cultivars. A total of 42 SSR and 216 AFLP loci were used to construct a molecular linkage map with a total map length of 444.2 cM and an average distance between loci of 1.7 cM. A total of 38 QTLs were identified for eight seed yield components. The traits of seed number per plant, seed yield per head, seed number per head, head number per plant and percent seed set were highly correlated with seed yield per plant, and QTLs for several of these traits were often detected in the same genome region. Head number per plant may present a particularly useful character for the improvement of seed yield since it can easily be determined before seed maturity. In addition, two genome regions containing four or five QTLs for different seed yield components, respectively, were identified representing candidate regions for further characterisation of QTLs. This study revealed several key components which may facilitate further improvement of seed yield. The QTLs identified represent an important first step towards marker assisted breeding in red clover.

4.1 Introduction

Temperate grasslands play an important role in agriculture, as they cover approximately 8 % of the global land area (White et al. 2000). Red clover (*Trifolium pratense* L.), an outcrossing and diploid (2n = 2x = 14) species with a gametophytic self incompatibility system, is an important component of permanent pastures and meadows as well as of grass-clover leys in temperate regions. Red clover is adapted to a wide range of environmental conditions, has a high nutritive value and, due to its ability to fix atmospheric nitrogen, red clover is of a high value to the environment (Taylor and Quesenberry 1996).

In the last decades, targeted selection has produced red clover cultivars that are considerably improved for traits such as forage yield and quality as well as resistance and persistence (Taylor and Quesenberry 1996). However, these improved cultivars often show an unsatisfactory seed yield that leads to high production costs and limits the success of cultivars in the marketplace (Taylor and Quesenberry 1996). Seed yield may be increased to a certain extent through improved management practices involving irrigation (Oliva et al. 1994) or selection of soil type

(Belzile 1991). However, path coefficient analyses have revealed a causal relationship among several components associated with seed yield, highlighting the complexity of this trait (Montardo et al. 2003; Oliva et al. 1994). Although both studies consistently reported a significant effect of the number of heads per plant on seed yield, they were limited to the phenotypic observation of only a few seed yield components.

The negative correlation of seed yield with other important agronomic traits, such as forage yield (Steiner et al. 1997) represents a further impediment for improving this trait. This negative correlation seems to be particularly pronounced for persistence, i.e. the ability to produce constantly high forage yield across several growing periods. For example, continuous selection from locally adapted Swiss ecotypes led to the development of cultivars with considerably improved persistence, which are commonly referred to as Mattenklee cultivars (Herrmann et al. 2005b). These cultivars show significantly increased forage yield across three or four growing periods when compared to other red clover cultivars (Lehmann and Briner 1998). However, the seed yield of Mattenklee cultivars is considerably lower when compared to less persistent cultivars (Deneufbourg 2004).

Molecular dissection of complex traits and the development of molecular markers linked to genes and quantitative trait loci (QTL) controlling such traits may provide new tools for breeding, which can complement traditional breeding approaches (Newbury 2003). Identification and integration of QTLs in genetic linkage maps is a promising step towards the development of molecular markers for marker assisted breeding. Several examples have been reported for forage crops. In white clover (*Trifolium repens* L.) QTLs for seed yield and other seed yield components were recently identified (Abberton et al. 2005; Barrett et al. 2005). In perennial ryegrass (*Lolium perenne* L.) a genome region associated with four herbage quality traits was located in the vicinity of genes involved in lignin biosynthesis and is therefore a candidate region for more detailed characterisation of QTLs controlling herbage quality (Cogan et al. 2005). In major crops such as soybean (*Glycine max* L.) QTLs have been identified for not only a large number of traits including seed yield (Mansur et al. 1996), and in addition, molecular markers have been developed for application in breeding, for example to select for a soybean cyst nematode resistance allele (Cregan et al. 1999b). However, forage crops in general and red clover in particular have lagged behind in this rapid development. Currently, two linkage maps, a basic prerequisite for identification of markers linked to important traits, have been reported for red clover; one consisting of 157 RFLP markers (Isobe et al. 2003) and the other consisting of 1357 SSR and 148 RFLP markers (Sato et al., submitted). However, so far there is no information available on QTLs controlling important traits, such as seed yield for red clover.

Moreover, there is only limited information available for association among seed yield components, restricted to phenotypic observations of only few seed yield components. The analysis of additional traits such as seed yield per head or time of flowering as well as the molecular dissection of these traits may help to elucidate the association of seed yield components and to develop new strategies for seed yield improvement.

The objectives of the study presented here were: (1) to characterise the association among eight seed yield components (2) to identify components which are easy to score and thus allow for improved selection and (3) to identify genome regions containing QTLs for seed yield components for the future development of molecular markers for marker assisted improvement of seed yield in red clover. For this purpose we established a red clover F_1 population segregating for seed yield components, investigated eight seed yield components in a field experiment, constructed a genetic linkage map using AFLP and SSR markers and performed QTL analysis.

4.2 Materials and Methods

Plant material

A two-way pseudo-testcross population was created performing manual reciprocal pollinations and using two red clover genotypes with contrasting levels of seed yield. One genotype was selected from the cultivar Violetta (pV), a Belgian cultivar characterised by a high seed yield but displaying low persistence (CLO-DvP, Gent, Belgium). The other genotype was selected from the Swiss Mattenklee cultivar Corvus (pC), which is characterised by excellent persistence but rather shows low seed yield (Agroscope FAL Reckenholz, Zurich, Switzerland).

Seeds were harvested from each maternal plant separately and a total of 400 seeds (200 per maternal plant) were germinated on wet filter paper (Schleicher and Schuell, Dassel, Germany) and transferred to soil-filled pots. Plants were grown for 3 months under long-day conditions [16 h light ($\geq$ 100 $\mu Em^{-2}s^{-1}$)] in the greenhouse. To obtain plants with as many shoots as possible, they were then cultivated for 5 months under short-day conditions [9 h light (275 $\mu Em^{-2}s^{-1}$)] in the growth chamber. Clonal replicates were produced by cutting the plant in at least five parts, each containing equally sized longitudinal sections of the taproot. Roots were dipped in synthetic auxin (3-Indol Butyric Acid, Pokon and Chrysal International, Naarden, Netherlands) and re-grown for 4 months in the greenhouse under long-day conditions as described previously.

Experimental conditions and phenotypic evaluation

In spring 2003, four clonal replicates of 280 genotypes (each parental plant serving as maternal genotype for a subset of 140 progenies) were planted in the field in a 4x4 lattice with the blocks arranged as a Latin square design, i.e. each genotype was represented once in each of the four rows and columns of the lattice, respectively. The field experiment was established at Agroscope FAL Reckenholz in Zurich, where temperature and rainfall average 8.5 °C and 1042 mm $year^{-1}$, respectively. Plants were established in a clay soil together with *Poa pratensis* cv. Compact sown at the time of planting.

Phenotypic evaluation of the seed yield components was performed after the first cut in summer 2004. Eight traits were investigated: Seed yield per plant (g; SYP), seed number per plant (SNP), seed yield per head (g; SYH), seed number per head (SNH), head number per plant (HNP), thousand-seed weight (g; TSW), percent seed set (number of seeds per 100 florets; PSS) and time of flowering (days after first cut; TOF). TOF was defined as the day when three heads of a plant were flowering. On maturity, one head per plant was used to count florets and seeds in order to calculate PSS. Seed yield and number of seeds of 20 randomly picked heads, as well as seed yield of the remaining heads were determined to calculate SYP, SNP, SYH, SNH, HNP and TSW.

Analysis of variance was performed using the general linear model (GLM) of the STATISTICA software (version 6.1, StatSoft, Tulsa, OK, USA). Least square means were used for all further calculations. Heritability was calculated according to the formula $h^2 = \sigma^2_{g(mp)}/(\sigma^2_{g(mp)} + \sigma^2_e/r)$, where $\sigma^2_{g(mp)}$ represented the variance component of the genotype nested within the maternal plant, σ^2_e represented the variance component of the error term and r represented the number of replicates (Wricke and Weber 1986).

Genotyping

DNA of 254 genotypes was extracted from fresh or frozen leaf tissue using the DNeasy 96 plant kit (Qiagen, Hilden, Germany). AFLP analysis was performed as described by Herrmann et al. (2005b) using 21 primer combinations. The primer combinations were named according to the standard primer combination code (Keygene, Wageningen, Netherlands; see Fig. 4.1). One hundred and seven SSR primer pairs were screened for polymorphism in the mapping population. SSR assays of primer pairs selected from TPSSR01 to TPSSR57 (Kölliker et al. 2005b) were performed using the protocol of Kölliker et al. (2005b) and 5' pigtail addition to the reverse primer to promote non-templated adenylation of amplicons (Brownstein et al. 1996). Forward primers of SSR loci (RCS004-RCS233) reported by Sato et al. (submitted) were modified by 5’ concatenation of the M13-18-mer 5’-TGTAAAACGACGGCCAGT-3’, which

permitted concurrent fluorescence labelling of PCR products by a third primer (M13) with an incorporated fluorophore (Boutin-Ganache et al. 2001) and reverse primers carried the 5' pigtail. PCR analyses were conducted in a total volume of 20 µl containing 15 ng DNA, 1x PCR buffer, 0.1 µM of forward primer, 0.4 µM of M13 and reverse primer, 2.5 mM $MgCl_2$, 0.25 mM of each dNTP and 0.5 U *Taq* DNA Polymerase (Invitrogen, Carlsbad, CA, USA). PCR conditions consisted of 4 min at 95 °C, 30 cycles of 30 sec 95 °C, 30 sec at 50 °C or 55 °C depending on the primer pair, 30 sec 72 °C, 10 cycles of 30 sec 95 °C, 30 sec 53 °C, 30 sec 72 °C followed by a final extension of 10 min at 72 °C.

AFLP and SSR amplification products were analysed on an ABI Prism 3100 genetic analyser using POP-4 polymer and 36 cm capillaries and scored using the Genotyper 3.6 software (Applied Biosystems, Foster City, CA, USA).

Linkage mapping and QTL analysis

A genetic linkage map was established using JoinMap (version 3.0; Van Ooijen and Voorrips 2001) and the CP (cross pollination) population type. For grouping, a LOD threshold of four or lower was obtained. The order of loci was determined at LOD larger than 1.0, REC smaller than 0.4 and a jump threshold of five using Kosambi's mapping function.

QTL analysis was performed based on least square means of the genotypes using MapQTL (version 5.0, Van Ooijen 2004). In order to reduce computing time necessary for calculations with the CP population type and the MQM (multiple QTL model) algorithm, the number of loci was optimised for each trait as follows:

In a first step two maps were calculated for each parental plant, i.e. one map was calculated with loci heterozygous in pV, the other map contained loci heterozygous in pC. QTL analysis was performed on these maps for all traits using composite interval mapping (CIM) of PlabQTL with the genotypes coded as a doubled haploid (DH) population type (Utz and Melchinger 2003). In a second step, loci near putative QTLs identified in step one on both maps, as well as most bi-parental SSR loci were included to construct one combined reduced map with an average locus distance of approximately 10 cM. QTL analysis was then performed on this reduced map for all eight traits using backward cofactor selection and MQM mapping. In a last step, separate maps were constructed for each trait with increased locus density in regions where putative QTLs were identified in step two. Final QTL mapping was performed with backward cofactor selection and MQM mapping using these maps optimised for each trait. QTLs were taken into account when the LOD score was higher than the LOD threshold derived from the respective map using 10000 permutations.

4.3 Results

Establishment of the mapping population and phenotypic evaluation

Of the 200 seeds harvested from each parental plant, 176 and 172 produced viable genotypes from pV and pC, respectively. Of these, 82 % (140 from pV and 145 pC) yielded at least four equivalent clonal replicates. To include the same number of genotypes from each maternal plant, the final mapping population consisted of 280 F_1 genotypes.

Analyses of variance of the phenotypic data revealed highly significant variation among the 280 genotypes for all eight seed yield components studied (Table 4.1). No significant effect of the maternal plants was observed, i.e. there was no difference between genotypes where seed was harvested from pV when compared to genotypes where seed was harvested from pC. The proportion of variance explained was highest (0.71) for TSW followed by SNH (0.65) and TOF (0.64; Table 4.1). Heritability ranged from 0.51 for SYP to 0.85 for TSW (Table 4.2), whereas coefficients of variation ranged from 0.06 for TOF to 0.33 for SNP.

Table 4.1 *F* values, level of significance and proportion of variance explained (R^2) of analysis of variance for eight seed yield components of a red clover population consisting of 280 F_1 genotypes assessed in a field experiment with four clonal replicates. Abbreviations of seed yield components: *SYP* seed yield per plant, *SNP* seed number per plant, *SYH* seed yield per head, *SNH* seed number per head, *HNP* head number per plant, *TSW* thousand-seed weight, *PSS* percent seed set, *TOF* time of flowering

	df	SYP	SNP	SYH	SNH	HNP	TSW	PSS	TOF
Maternal plant[a]	1	0.7^{ns}	1.9^{ns}	1.0^{ns}	2.2^{ns}	0.7^{ns}	1.6^{ns}	1.0^{ns}	0.4^{ns}
Genotype (maternal plant)	278	1.9^{***}	2.3^{***}	3.1^{***}	4.4^{***}	2.2^{***}	5.9^{***}	3.1^{***}	4.8^{***}
Column	3	18.5^{***}	18.0^{***}	19.1^{***}	23.0^{***}	10.7^{***}	4.8^{**}	11.1^{***}	5.5^{***}
Row	3	13.0^{***}	9.7^{***}	26.7^{***}	10.8^{***}	9.3^{***}	43.6^{***}	3.2^{*}	1.0^{ns}
Error	834								
R^2		0.47	0.50	0.58	0.65	0.48	0.71	0.56	0.64

$^{*}P \leq 0.05$, $^{**}P \leq 0.01$, $^{***}P \leq 0.001$, ^{ns} not significant

[a] The population was based on reciprocal crosses where the parent from the cultivar Violetta (pV) served as maternal plant for one half of the genotypes and pC for the other half

Table 4.2 Key characteristics of eight seed yield components of a red clover population consisting of 280 F_1 genotypes assessed in a field experiment based on least square means of four replicates per genotype

	Mean	Minimum	Maximum	SD[a]	$h^{2\,b}$
SYP[c]	10.0	0.9	19.3	3.1	0.51
SNP	5722	713	11309	1870	0.58
SYH	0.14	0.07	0.21	0.02	0.70
SNH	79	48	110	11	0.79
HNP	72	12	140	21	0.57
TSW	1.75	1.27	2.21	0.15	0.85
PSS	0.76	0.51	0.93	0.08	0.70
TOF	49	38	62	3	0.80

[a] Standard deviation

[b] Heritability

[c] For description of seed yield components, see Table 4.1

Pairwise comparison of SYP with the other seven seed yield components using product moment correlation revealed most of them to be highly significantly ($P < 0.01$) correlated with the exception of TOF which showed an only moderately significant ($P < 0.05$) correlation and TSW which was not significantly correlated to SYP (Table 4.3). Thereby, the correlation coefficients were highest for comparisons of SYP, SNP and HNP. Pairwise comparisons among the other seven seed yield components revealed two thirds of them to be significantly ($P < 0.05$) correlated. For example, TSW and PSS as well as HNP and TOF showed a significantly negative correlation coefficient of -0.28 and -0.26, respectively (Table 4.3).

Table 4.3 Product moment correlation coefficients for pairwise comparisons of eight seed yield components of a red clover population consisting of 280 F_1 genotypes

	SYP[a]	SNP	SYH	SNH	HNP	TSW	PSS
SNP	0.95^{***}						
SYH	0.42^{***}	0.32^{***}					
SNH	0.42^{***}	0.47^{***}	0.79^{***}				
HNP	0.87^{***}	0.89^{***}	0.02^{ns}	0.10^{ns}			
TSW	0.08^{ns}	-0.19^{**}	0.39^{***}	-0.20^{**}	-0.09^{ns}		
PSS	0.22^{***}	0.28^{***}	0.31^{***}	0.47^{***}	0.09^{ns}	-0.28^{***}	
TOF	-0.16^{*}	-0.23^{***}	0.11^{ns}	0.03^{ns}	-0.26^{***}	0.15^{*}	0.11^{ns}

$^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, ^{ns} not significant

[a] For description of seed yield components, see Table 4.1

Linkage mapping

Forty-two (39 %) of the 107 SSR loci analysed yielded polymorphisms among the 254 F_1 genotypes used for mapping (Table 4.4). AFLP analysis yielded a total of 216 polymorphic loci. The nine AFLP primer combinations based on the restriction enzymes *Eco*RI/*Mse*I extended by three selective nucleotides (*Eco*+3/*Mse*+3) yielded 5 - 19 polymorphic loci each, with a total of 117. The 12 *Pst*+3/*Mse*+2 primer combinations yielded 4 - 13 with a total of 99 loci (Table 4.4).

Table 4.4 Number of AFLP and SSR loci of a red clover linkage map based on 254 F_1 genotypes. The percentage of distorted loci is indicated in parentheses

	Bi-parental loci[a]			Mono-parental loci[b]		Total
	abxcd[c]	efxeg	hkxhk	lmxll (pC)	nnxnp (pV)	
SSR	13 (23 %)	10 (30 %)	-	12 (0 %)	7 (14 %)	42 (17 %)
AFLP *Eco*/*Mse*[d]	-	-	19 (5 %)	52 (15 %)	46 (17 %)	117 (15 %)
AFLP *Pst*/*Mse*[d]	-	-	16 (13 %)	47 (6 %)	36 (8 %)	99 (8 %)
Total	13 (23 %)	10 (30 %)	35 (9 %)	111 (10 %)	89 (13 %)	258 (12 %)

[a] Both parents were heterozygous for these loci

[b] Heterozygous in one parent (from the cultivar Corvus (pC) or Violetta (pV)) and homozygous in the other parent

[c] Segregation types according to JoinMap (Van Ooijen and Voorrips 2001)

[d] AFLP primer combinations based on the restriction enzymes *Eco*RI and *Mse*I or *Pst*I and *Mse*I, respectively

All 258 loci were mapped onto seven linkage groups (LG; Fig. 4.1). The length of the LGs ranged from 54 cM (LG 7) to 78 cM (LG 2) with an average of 63.5 cM, resulting in a total map length of 444.2 cM (Fig. 4.1). Average distance between loci varied from 1.5 (LG 2) to 2.2 (LG 6) with an average of 1.7 cM across the entire map.

Of the 258 loci 58 (22.5 %) were bi-parental, i.e. both parents were heterozygous for that locus, whereas the remaining 200 (77.5 %) loci were mono-parental, i.e. heterozygous in one parent and homozygous in the other parent (Table 4.4). No gap larger than 10 cM between two loci across the linkage map and no clustering of loci depending on the marker system or the restriction enzymes used was observed (Fig. 4.1).

The percentage of distorted segregation ($P \leq 0.05$) ranged from 5 % for the 19 mono-parental SSR loci to 26 % for the 23 bi-parental SSR loci resulting in an average of 12 % for the total of 258 loci (Table 4.4). Highly distorted loci ($P \leq 0.001$) were mainly observed on LG 6 where 48 % of the loci were distorted. On LG 7, 47 % of the loci were distorted but distortion was less severe ($P > 0.001$; Fig. 4.1).

QTL analysis

A total of 38 significant QTLs, with a LOD score higher than the LOD threshold calculated from the respective trait and the corresponding map, were detected across all eight seed yield components (Table 4.5, Fig. 4.1). Three to eight QTLs were found per trait explaining together 33.8 - 69.1 % of the total variance with an average of 48 % across all traits. Individual QTLs explained 2.1 - 33.7 % of the variance with an average of 10 %. Two QTLs explained more than 30 % (TSW and PSS), two explained 20 - 30 % (SNH and TOF), whereas eight QTLs explained 10 - 20 % of the total variance (Table 4.5). Seven additional QTLs were identified with LOD scores only slightly lower than the LOD threshold, i.e. two QTLs for SYP (LG 5, 51 cM; LG 1, 44 cM), two for SNP (LG 5, 37 cM; LG 1, 45 cM), one for TSW (LG 6, 48 cM) and two for PSS (LG 2, 78 cM; LG 1, 0 cM; data not shown).

On each LG four to nine significant QTLs were detected and up to five QTLs for different traits were located within 10 cM (LG 6, 44 - 54 cM). QTLs of highly correlated traits (Table 4.3) were often detected within 1 - 10 cM (Fig. 4.1). For example, all QTLs for SYP and for SNP, traits which showed a correlation coefficient of 0.95, were identified within 1 cM (LGs 3 and 6) and 6 cM (LG 4), respectively. In addition, of the seven QTLs identified for SYP and HNP, traits which showed a correlation coefficient of 0.87, two sets of two QTLs were located within 6 cM (LGs 3 and 4). On the other hand, QTLs of insignificantly correlated traits (Table 4.3) were often located on separate LGs or at least were not identified in the same region (Fig. 4.1). For example six QTLs of SYP and TSW, traits which showed an insignificant correlation coefficient of only 0.08, were detected on separate LGs, whereas two QTLs were located 37 cM apart on LG 3 (Fig. 4.1).

Fig. 4.1 Genetic linkage map of a red clover population based on 254 F_1 genotypes, 42 SSR and 216 AFLP loci. Locus names consist of a denomination of the origin of the parental alleles (B = bi-parental locus; C and V = mono-parental locus heterozygous in the parent from the cultivar Corvus and Violetta, respectively), followed by the locus name (standard primer combination code (Keygene, Wageningen, Netherlands) followed by the allele size in relative migration units for AFLP loci or the prefix TPSSR (Kölliker et al., 2005) and RCS (Sato et al., submitted) followed by an identification number for SSR loci, respectively). Significantly distorted loci are indicated by *asterisks* (* $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, **** $P \leq 0.0001$). Positions of QTLs for eight seed yield components were calculated using MQM mapping, the optimised map for the respective trait and least square means of four replicates per genotype (*SYP* seed yield per plant, *SNP* seed number per plant, *SYH* seed yield per head, *SNH* seed number per head, *HNP* head number per plant, *TSW* thousand seed weight, *PSS* percent seed set, *TOF* time of flowering). The maximum LOD score position of each QTL is indicated with an *arrow* and a *bar* represents the interval between two positions obtained at LOD scores two units lower than the maximal score

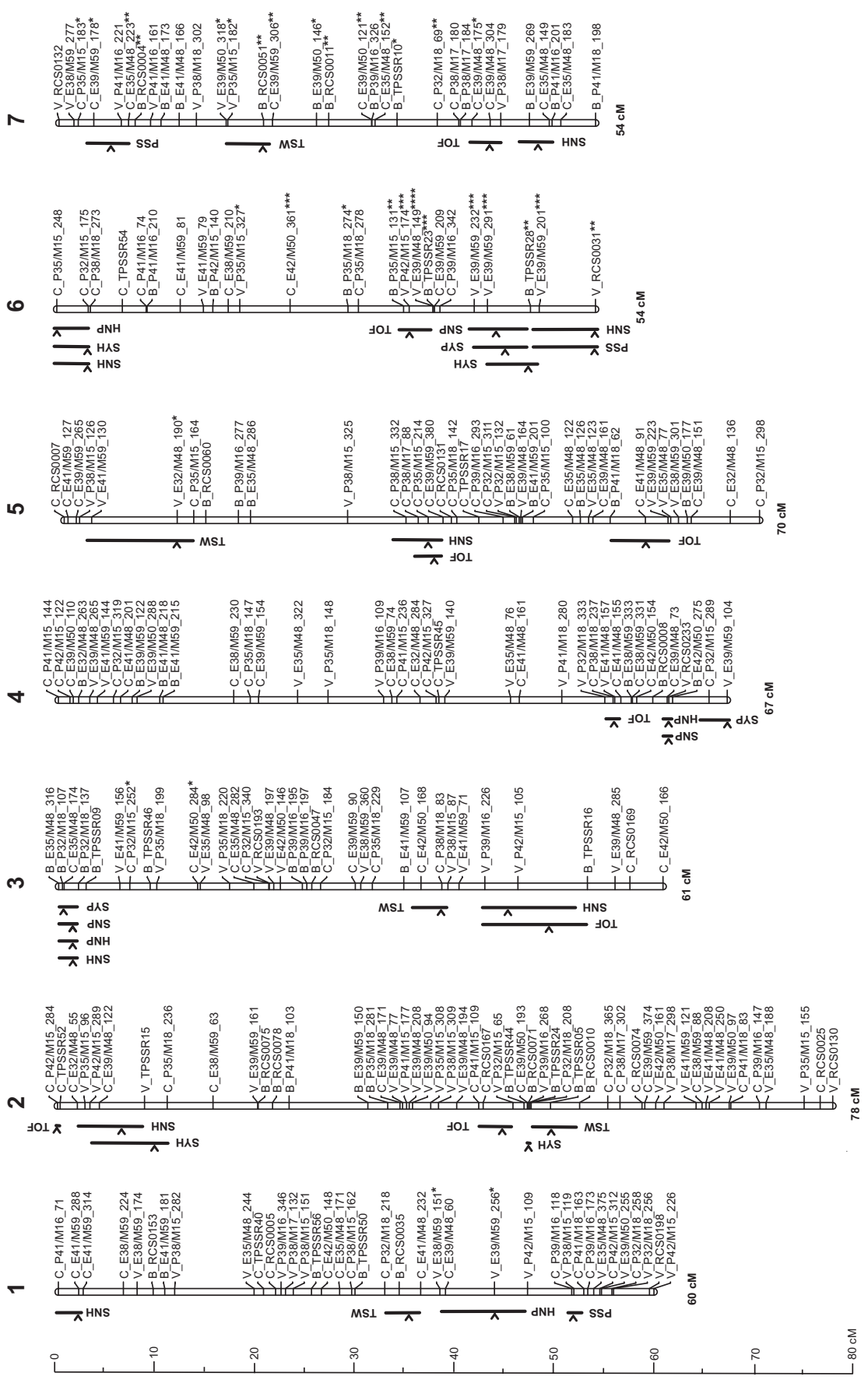


Table 4.5 Position and description of QTLs identified using MQM mapping, the optimised map for the respective trait and least square means of eight seed yield components of a red clover population consisting of 254 F_1 genotypes

Trait	Linkage group	Position (cM)	Closest neighbouring locus	Maximum LOD score[a]	% variance explained
SYP[b]	3	0.6	C_E35/M48_174[c]	10.37	15.3
	4	67.3	V_E39/M59_104	4.90	7.0
	6	45.3	V_E39/M59_291	7.60	11.5
				Total	**33.8**
SNP	3	1.6	B_P32/M18_137	9.89	14.0
	4	61.4	V_RCS0233	6.02	8.1
	6	44.3	V_E39/M59_291	9.65	14.4
				Total	**36.5**
SYH	2	9.8	V_TPSSR15	6.56	8.1
	2	47.6	B_TPSSR24	7.60	6.4
	6	3.2	C_P32/M15_175	3.88	7.7
	6	47.3	B_TPSSR28	11.05	17.5
				Total	**39.7**
SNH	1	2.1	C_E41/M59_288	8.86	8.2
	2	6.2	C_E39/M48_122	4.79	2.6
	3	1.6	B_P32/M18_137	3.94	3.4
	3	44.9	V_P42/M15_105	5.49	5.4
	5	36.6	C_P35/M15_214	5.04	2.1
	6	3.2	C_P32/M15_175	4.81	7.0
	6	54.2	V_RCS0031	21.74	25.7
	7	48.3	C_E35/M48_149	11.25	8.5
				Total	**62.9**
HNP	1	43.9	V_E39/M59_256	5.35	6.7
	3	1.6	B_P32/M18_137	10.55	13.2
	4	61.4	V_RCS0233	7.92	9.9
	6	0.0	C_P35/M15_248	4.34	7.7
				Total	**37.5**
TSW	1	35.3	B_RCS0035	5.77	5.5
	2	49.6	B_TPSSR05	28.23	32.0
	3	38.5	C_P38/M18_83	4.35	4.0
	5	11.4	V_E32/M48_190	12.50	12.3
	7	20.6	B_RCS0051	5.17	4.7
				Total	**58.5**
PSS	1	51.9	V_P38/M15_119	4.63	5.1
	6	54.2	V_RCS0031	18.34	33.7
	7	5.5	V_P41/M16_221	3.93	4.7
				Total	**43.5**
TOF	2	0.0	C_P42/M15_284	9.56	8.4
	2	44.8	V_P32/M15_65	26.26	23.8
	3	49.2	V_P42/M15_105	7.50	6.4
	4	56.0	V_E41/M48_157	7.16	5.2
	5	37.6	C_E39/M59_380	5.17	4.5
	5	58.5	C_E41/M48_91	8.46	7.1
	6	35.5	V_P42/M15_174	12.52	10.5
	7	43.4	C_E39/M48_304	4.47	3.2
				Total	**69.1**

[a]Significant LOD threshold was 3.6 except for SNH where it was 3.7

[b]For description of seed yield components, see Table 4.1

[c]For description of loci, see Fig. 4.1

4.4 Discussion

The explanatory power of trait dissection and QTL analysis largely depends on a reliable assessment of the phenotypic variation for the traits under study. In outcrossing species, where homozygous lines are difficult to obtain, replicated field experiments often rely on clonal replicates of individual genotypes. However, for red clover such replicates are difficult to obtain (Cumming and Steppler 1961), because the tap root system is less suitable for cloning than the fibrous root system of grasses or the stoloniferous growth of white clover. Establishing plants under short day conditions to enhance shoot formation and dipping cutlets in synthetic auxin to enhance root formation was highly successful with 82 % of the genotypes producing four or more clonal replicates. This method provides a valuable prerequisite for further QTL analysis in red clover.

Besides reliable phenotypic data, a robust linkage map with evenly distributed markers is needed. The linkage map developed here consisted of seven linkage groups and was based on 258 SSR and AFLP loci (Fig. 4.1). As the map in this study and the map reported by Sato et al. (submitted) included common SSR loci, and the map reported by Isobe et al. (2003) in turn included RFLP loci present on the map of Sato et al. (submitted), a congruent numbering of LGs was used for all the three maps. The total map with a length of 444.2 cM was only slightly shorter than the 535.7 cM of the red clover map based on 157 RFLP loci reported by Isobe et al. (2003), but exhibited only half the length of a very recently developed map based on 1505 SSR and RFLP loci (850.4 cM; Sato et al., submitted). The large difference in locus number may be one reason for the different map lengths, as was also reported for other species, such as perennial ryegrass (Armstead et al. 2002; Bert et al. 1999; Jones et al. 2002b). Other factors, such as the heterogeneity of the parents, i.e. the particular parental genotype, may also play an important role. Since longer maps only rarely covered additional genome regions, when compared to shorter maps of the same species (Cregan et al. 1999a; Freyre et al. 1998), the 258 loci used in this study are likely to cover the majority of the red clover genome and to be sufficient for QTL analysis. However, complete genome coverage of the map presented here can not be assumed.

For enhanced accuracy and power of QTL detection, increased number of genotypes rather than the number of loci used for analysis are crucial. A population size of at least 200 genotypes is needed to detect a QTL with an explained variance of 5 % (Van Ooijen 1992). On the other hand, the power of detecting a QTL remains virtually the same no matter whether a map with an average locus distance of 10 cM or with an infinite number of loci is used (Darvasi et al. 1993). The average locus distance of 1.7 cM obtained in the present map therefore provided a good basis for QTL analysis.

In addition, the two marker systems used in this study complemented one another and allowed the construction of a meaningful map. On the one hand, the integrated SSR loci can be used to link the map with the other published red clover maps (Isobe et al. 2003; Sato et al., submitted) providing a valuable basis for further investigation of genome regions of interest. On the other hand, the AFLP loci offer a powerful tool to quickly fill gaps between SSR loci on linkage maps. One drawback of the AFLP marker system might be the clustering of loci in centromeric regions observed using the restriction enzyme *EcoR*I (Bert et al. 1999; Vuylsteke et al. 1999). Thus, to avoid clustering of AFLP loci but still ensuring coverage of centromeric regions, both rare cutting restriction enzymes, *EcoR*I and *Pst*I, were used. Although, the average distance between two loci of the map was comparable to maps in which clustering of *EcoR*I loci was observed (Bert et al. 1999), we observed no such clustering (Fig. 4.1).

Due to the manifestation of deleterious recessives, inbreeding depression may lead to substantial segregation distortion and may negatively influence the stability of genetic maps as well as the accuracy of QTL analyses (Brummer et al. 1993; Echt et al. 1994). The use of non-inbred F_1 populations may be an effective way to avoid distortion in self-incompatible species such as red clover (Tavoletti et al. 1996). Indeed, the proportion of significantly ($P \leq 0.05$) distorted loci observed (Table 4.4) was only 12 %, which is considerably lower than the 37 % observed in a back cross population of red clover (Isobe et al. 2003). Self incompatibility may be another reason for segregation distortion and may result in a clustering of distorted loci around a self incompatibility locus (Bert et al. 1999). Thus, the region on LG 6, where highly distorted loci were observed, might correspond to the single self incompatibility locus reported in red clover (Lawrence 1996). Segregation distortion based on self incompatibility can only occur if the parents share a common self incompatibility allele. If this is the case, not all of the three resulting genotypes would have the same probability to be successfully pollinated. If pollen is scarce, this may lead to incomplete pollination of florets resulting, first of all, in a negative effect onto PSS. As a consequence, ghost QTLs for seed yield components might be observed near the self incompatibility locus. However, a lack of pollen is very unlikely for the present study since three quarters of the 1120 individuals flowered within 10 days (data not shown) and PSS was comparable to other studies (Oliva et al. 1994). Moreover, the theoretical proportion of compatible pollen in an isolated progeny of a cross between two parents sharing one self incompatibility allele varies only between 31.25 % and 37.5 % for the handicapped and favoured genotypes, respectively.

The overall aim of this study was not only to identify QTLs for the development of molecular markers linked to seed yield, but also to elucidate the association among seed yield components. According to the correlation coefficients, SNP and HNP showed the largest effect on SYP

(Table 4.3). This result is congruent with path coefficient analyses for seed yield in red clover, for which the number of heads was identified as the primary component affecting seed yield (Montardo et al. 2003; Oliva et al. 1994). Some of the seed yield components were not determined independently and showed some mathematical causality, partially explaining their correlation. However, detailed QTL analyses confirmed these associations since QTLs for SNP, HNP, SNH, SYH and PSS were detected in the same regions (≤ 10 cM) as the three QTLs for SYP (Fig 4.1, Table 4.5). Although all five factors substantially influence SYP, only HNP offers an advantage for phenotypic selection compared to SYP. HNP is comparatively easy to determine and can be assessed in the field earlier before seed maturity. Therefore, the selection for increased HNP may present a valuable strategy to improve seed yield in red clover.

On the other hand, TSW showed insignificant correlation with HNP and SYP (Table 4.3), which was in congruence with path coefficient analysis in red clover where the influence of thousand-seed weight was minor (Montardo et al. 2003; Oliva et al. 1994) or with white clover where thousand-seed weight was not correlated with seed yield and inflorescence density (Barrett et al. 2005). Since only two of a total of nine QTLs for HNP and TSW were detected in the same region, successful selection for TSW may be possible independently of the proposed selection for HNP. The QTL located on LG 2 is of particular interest for further investigations to improve TSW, as it explains 32 % of the variation.

We were able to identify two regions covering less than 10 cM, where five (LG 6) and four (LG 3) QTLs of different seed yield components were clustered, respectively (Fig. 4.1). All but one QTL in these two regions explained more than 10 % of the total variation (Table 4.5). However, QTL analysis based on segregating populations derived from parents with contrasting phenotypes has several limitations. The precision and accuracy of QTL detection depends on a large population size. Small populations lead to an underestimation of the number of QTLs and an overestimation of the explained variance (Schön et al. 2004) as well as to a limited precision regarding the QTL position (Visscher and Goddard 2004). This is particularly true when assessing traits influenced by a high number of QTLs with small effects (Schön et al. 2004). Moreover, in contrast to more recent approaches like association mapping, only two alleles at a given locus can be studied simultaneously. However, association mapping relies on information about the nature of linkage disequilibrium within the genome of the respective plant species (Flint-Garcia et al. 2003), which is currently not available. Taken these limitations into account, the presented results may still serve as a valuable base for further molecular dissection of seed yield in red clover.

For future studies, the following approaches may be considered. Fine mapping of the detected QTL, using the information of the recently published high-density SSR map (Sato et al.,

submitted), offers a promising possibility to identify closely linked markers for marker assisted breeding or even the identification of genes involved in the control of seed yield by map based cloning. The recently published EST resources of the model legume *Medicago truncatula* (Cannon et al. 2005) offer an additional possibility to further explore genetic control of seed yield in red clover. A prerequisite to apply such information is a certain degree of synteny between the target species (i.e. red clover) and the species for which genetic information is available. The existence of syntenic relationships between a number of legume species including *Medicago truncatula,* alfalfa (*Medicago sativa*), soybean, pea and *Lotus japonicus* has been demonstrated (Choi et al. 2004). Thus, comparative techniques such as comparative anchor marker tag sequences (CATS; Schauser et al. 2005) or single nucleotide polymorphisms (SNP; Andersen and Lübberstedt 2003) may be alternative approaches for future investigations. Based on genes and QTLs associated with seed yield in other species, such as *Medicago truncatula* (Cannon et al. 2005) or white clover (Barrett et al. 2005), these approaches may help to elucidate genetic control of seed yield in red clover.

In conclusion, with the stable linkage map obtained using SSR and AFLP loci and 254 genotypes of a F_1 population as well as with the field analysis based on four clonal replicates, a solid basis for QTL analysis for seed yield components was provided. A total of 38 QTLs were detected for the eight seed yield components. The associations among seed yield components allowed the identification of head number per plant as an easy to determine, indirect character to select for seed yield. Furthermore, two genome regions rich in QTLs for seed yield components were identified with great potential for future characterisation and the development of markers closely linked to seed yield components. To the best of our knowledge, this is the first report on QTL analysis in red clover, which presents an important first step towards marker assisted selection and may help to implement new breeding strategies to complement breeding for complex traits such as seed yield.

Acknowledgements

The authors would like to thank Yvonne Häfele for technical assistance in the laboratory, Simone Günter, Philipp Streckeisen and Peter Tanner for support in the field, Eva Bauer of the State Plant Breeding Institute at the University of Hohenheim, Germany, for assistance with linkage mapping and QTL analysis and Sachiko Isobe of the National Agricultural Research Centre for Hokkaido Region, Japan for information on SSR primer sequences and map location. This study was funded by the breeding foundation DSP-BLW.

5 Stem length – an indicator for persistence and seed yield of red clover (*Trifolium pratense* L.) identified by QTL analysis

D. Herrmann, B. Boller, B. Studer, F. Widmer, R. Kölliker

Plant Breeding (in preparation)

Abstract

Cultivars of red clover (*Trifolium pratense* L.) often show an insufficient persistence leading to a lack of temporally stable proportions of red clover in forage stands. Persistence is a complex trait, which was reported to be negatively correlated with other important traits such as seed yield. Therefore, it has proven to be difficult to improve and molecular markers linked to persistence would represent a helpful tool to complement traditional breeding procedures. Thus, the objectives of this study were to identify QTLs for persistence and additional morphological traits and to investigate the association of these traits with seed yield. A total of 19 QTLs were identified for persistence, number of stems, length of stem, width of medial leaflet and length of medial leaflet. While there was no negative correlation between persistence and seed yield, persistence was highly correlated with length of stem and thousand-seed weight. Furthermore, length of stem was also highly correlated with seed yield and QTLs of persistence, length of stem and seed yield were detected in the same genome region. This genome region represents a candidate region for further characterisation and development of markers closely linked to persistence and seed yield. Our results indicate that improvement of persistence should be possible without adverse effects on seed yield. Moreover, length of stem was identified as an easy to determine indirect character to select for improved persistence and seed yield in red clover. This study forms an important step towards marker assisted breeding in red clover.

5.1 Introduction

Persistent red clover cultivars are an important prerequisite to maintain an adequate legume component in a forage stand. Such a component allows to enhance the quality of forage as well as to furnish fixed nitrogen to the other plant species. This is particularly important in organic farming systems, where nitrogen supply is often limiting (Taylor and Quesenberry 1996). As the relevance of organic farming or sustainable agriculture in general increased in the last decades, selection for persistence of red clover became more important. However, many red clover cultivars show an insufficient persistence, i.e. a still satisfying forage yield in the third or fourth growing season. Persistence is the result of adaptation of the crop to its stress load, i.e. physiogenic or pathogenic factors, that contribute to less than luxuriant growth (Taylor and Quesenberry 1996). Therefore, persistence is a very complex trait, which is influenced by many factors. Furthermore, negative correlations between persistence and other important agronomic traits were reported. For example, Mattenklee cultivars, red clover cultivars developed from locally adapted Swiss landraces, often show significantly increased persistence compared to other red clover cultivars (Lehmann and Briner 1998). On the other hand, seed yield of these

Mattenklee cultivars is considerably lower when compared to less persistent cultivars (Deneufbourg 2004), suggesting a negative correlation between persistence and seed yield. Thus, persistence is an important trait but difficult to improve in breeding programs.

Development of molecular markers linked to genes and QTLs controlling complex traits may provide new tools for breeding, which can complement traditional breeding approaches (Newbury 2003). Identification and integration of QTLs in genetic linkage maps is a promising step towards the development of molecular markers for marker assisted breeding. Several examples for forage crops were reported. In alfalfa (*Medicago sativa*) and ryegrass (*Lolium perenne*) QTL analyses for winter survival, a trait influencing persistence, were performed (Brouwer et al. 2000; Yamada et al. 2004). In red clover, two linkage maps have been reported (Isobe et al. 2003; Sato et al., submitted), but with the exception of the QTL analyses performed in this thesis (Herrmann et al. 2006), there is no information available on molecular dissection of agronomic traits in red clover.

For the successful development and application of markers linked to QTLs, it is essential not only to perform QTL analysis of each trait separately, but also to elucidate associations among the targeted traits. This allows for the identification of genome regions, where QTLs of two traits that have adverse effects on each other can be detected independently, enabling an independent improvement of the two traits. Similarly, when two traits are positively correlated, regions can be identified including QTLs of both traits, allowing for simultaneous improvement. Furthermore, positive association between a complex target trait and a trait, which is easy to score, may be detected, enabling an indirect selection for this complex target trait.

Therefore, the objective of this study was to investigate the association of persistence and seed yield as well as associations of these two traits to additional morphological traits. Molecular dissection of these associations was performed as a first step towards marker assisted improvement of persistence and seed yield in red clover. For this purpose we used the F_1 population, linkage map and the QTLs of seed yield per plant and thousand-seed weight reported by Herrmann et al. (2006) and performed QTL analysis for the five additional traits persistence, number of stems, length of stem, width of medial leaflet and length of medial leaflet.

5.2 Materials and Methods

Plant material, experimental conditions and linkage mapping

A segregating F_1 population derived from diploid parents with contrasting phenotypes for persistence and seed yield was established. In spring 2003, four clonal replicates of 280

genotypes were planted to the field. DNA of 254 genotypes was extracted and a genetic linkage map based on AFLP and SSR markers was constructed (Herrmann et al. 2006).

Phenotypic evaluation and QTL analysis

Five traits were investigated: Persistence (PT), number of stems (NS), length of stem (cm, LS), width of medial leaflet (mm, WL) and length of medial leaflet (mm, LL). Vigour was scored visually with a scale from 1 (dead) to 9 (very vigorous) and persistence was calculated as a weighted sum of 12 scores of vigour (four scores in each growing season of 2003, 2004 and 2005). Scores were weighted according to the numbers of days between the first score on 11th July 2003 (weight = 1) and the last score on 12th August 2005 (weight = 4). Only days from 29th March to 26th November were taken into account. Phenotypic evaluations of the other four traits were performed according to the guidelines of the International Union for the Protection of New Varieties of Plants (UPOV 2001) during re-growth after the first cut in summer 2004. NS was determined 40 days after cutting and LS, WL and LL were measured at full flowering (14 days after three heads of a plant had begun to flower). The longest stem of each plant was selected for the measurement of LS and the average of three normally developed top leaves of three randomly selected stems was used to determine WL and LL.

STATISTICA software (version 6.1, StatSoft, Tulsa, OK) was used to perform statistical analyses of phenotypic data. QTL analysis was performed using MapQTL (version 5.0, Van Ooijen 2004) based on least square means of the genotypes and a map with an optimised number of loci for each trait (Herrmann et al. 2006).

For pairwise comparison of traits using product moment correlation and for comparisons of QTL locations, the two seed yield components seed yield per plant (SYP) and thousand-seed weight (TSW) were used as determined by Herrmann et al. (2006).

5.3 Results

Phenotypic evaluation

Analyses of variance revealed highly significant variation among the 280 F_1 genotypes for the five traits PT, NS, LS, WL and LL (Table 5.1). No significant effect of the maternal plant was observed, i.e. there was no difference whether seed was harvested from the one or the other parental plant. The proportion of variance explained was higher for LS, WL and LL than for PT and NS ranging from 0.39 to 0.57 (Table 5.1). Heritability was highest for LL (0.69) followed

by LS (0.68) and WL (0.67), whereas heritability for NS and PT was lower with 0.44 and 0.41, respectively (Table 5.2).

Table 5.1 F-value, level of significance and proportion of variance explained (R^2) of analysis of variance for five traits of a red clover population consisting of 280 F_1 genotypes assessed in a field experiment with four clonal replicates. PT = persistence, NS = number of stems, LS = length of stem, WL = width of medial leaflet, LL = length of medial leaflet

	df[a]	PT	NS	LS	WL	LL
Maternal plant[b]	1	0.8^{ns}	1.6^{ns}	3.0^{ns}	0.2^{ns}	1.6^{ns}
Genotype (maternal plant)	278	1.7^{***}	1.8^{***}	3.0^{***}	2.9^{***}	3.1^{***}
Column[c]	3	17.9^{***}	19.4^{***}	23.2^{***}	18.9^{***}	21.5^{***}
Row[c]	3	2.5^{ns}	6.6^{***}	36.6^{***}	7.7^{***}	13.0^{***}
Error	834					
R^2		0.39	0.43	0.57	0.54	0.56

$^{***}P \leq 0.001$, ^{ns} not significant

[a]Degrees of freedom

[b]The population was based on reciprocal crosses where the parent from the cultivar Violetta served as maternal plant for one half of the genotypes and the parent from the cultivar Corvus for the other half

[c]Each genotype was represented once in each of the four rows and columns of the field design, respectively

Table 5.2 Key characteristics for five traits of a red clover population consisting of 280 F_1 genotypes assessed in a field experiment based on least square means of four replicates per genotype

	Mean	Minimum	Maximum	SD[a]	$h^{2\,b}$
PT[c]	129.3	64.7	217.4	28.1	0.41
NS	37.4	9.4	69.1	7.9	0.44
LS	84.3	56.3	112.8	8.6	0.68
WL	26.4	19.6	31.8	2.0	0.67
LL	45.3	34.6	55.4	3.7	0.69

[a] Standard deviation

[b] Heritability

[c] For description of the traits, see Table 5.1

For pairwise comparison of traits using product moment correlation, the five traits PT, NS, LS, WL and LL and the two seed yield components SYP and TSW (Herrmann et al. 2006) were used (Table 5.3). PT was highly correlated to the other six traits with correlation coefficients ranging from 0.21 to 0.37 with the exception of SYP, which showed a low but still significant

correlation coefficient of 0.13 to PT. For SYP highest correlation coefficients were detected to LS (0.48) and NS (0.40), although WL and LL were also significantly ($P \leq 0.05$) correlated showing correlations coefficient of 0.21 and 0.27 to SYP, respectively. Pairwise comparisons among the other five traits revealed two thirds of them to be significantly correlated. The correlation coefficient was highest (0.74) between the two leaf size characters, WL and LL (Table 5.3).

Table 5.3 Product moment correlation coefficients for five traits of a red clover population consisting of 280 F_1 genotypes. Calculations are based on least square means obtained from four replicates per genotype

	PT	NS	LS	WL	LL	TSW
NS[a]	0.21 ***					
LS	0.32 ***	0.14 *				
WL	0.32 ***	0.01 ns	0.36 ***			
LL	0.28 ***	0.01 ns	0.51 ***	0.74 ***		
TSW	0.37 ***	-0.10 ns	0.20 **	0.35 ***	0.42 ***	
SYP	0.13 *	0.40 ***	0.48 ***	0.21 ***	0.27 ***	0.08 ns

*P ≤ 0.05, **P ≤ 0.01, ***P ≤ 0.001, ns not significant

[a] For description of the traits, see Table 5.1; SYP = seed yield per plant, TSW = thousand-seed weight

QTL analysis

For the five traits PT, NS, LS, WL and LL a total of 19 QTLs with a LOD score higher than the LOD threshold were detected (Table 5.4; Fig. 5.1). One to seven QTLs were found per trait explaining 7.5 % to 58.6 % of the total variation with an average of 33 % across all traits. Individual QTLs explained 4.2 to 20.7 % of the total variation with an average of 8.7 %. One QTL explained more than 20 % (LS), five QTLs explained 10 to 20 %, whereas the other 13 QTLs explained less than 10 % (Table 5.4). Three additional QTLs were identified with LOD scores only slightly lower than the significant LOD threshold, i.e. one QTL for NS (LG 3; 9 cM), one for LS (LG 7; 43 cM) and one for LL (LG 7; 47 cM).

In addition to the QTLs of the five traits PT, NS, LS, WL and LL, QTLs of the two seed yield components SYP and TSW (Herrmann et al. 2006) were included for comparisons (Fig. 5.1). Of the total of 27 QTLs identified for these seven traits, two to seven significant QTLs were detected on each linkage group (LG). Five QTLs were located within 15 cM on LG 2, whereas three QTLs were identified within 11 cM on LG 3 and 4 (Fig. 5.1). QTLs of highly correlated traits were often detected within 1 to 15 cM (Fig. 5.1). For example, of the twelve QTLs

identified for WL and LL, traits which showed a correlation coefficient of 0.74 (Table 5.3), three pairs of QTLs were located within no more than 12 cM on LGs 2, 4 and 5 (Fig. 5.1).

Table 5.4 Position and description of QTLs identified using MQM mapping, the optimised map for the respective trait and least square means for five traits of a red clover population consisting of 254 genotypes

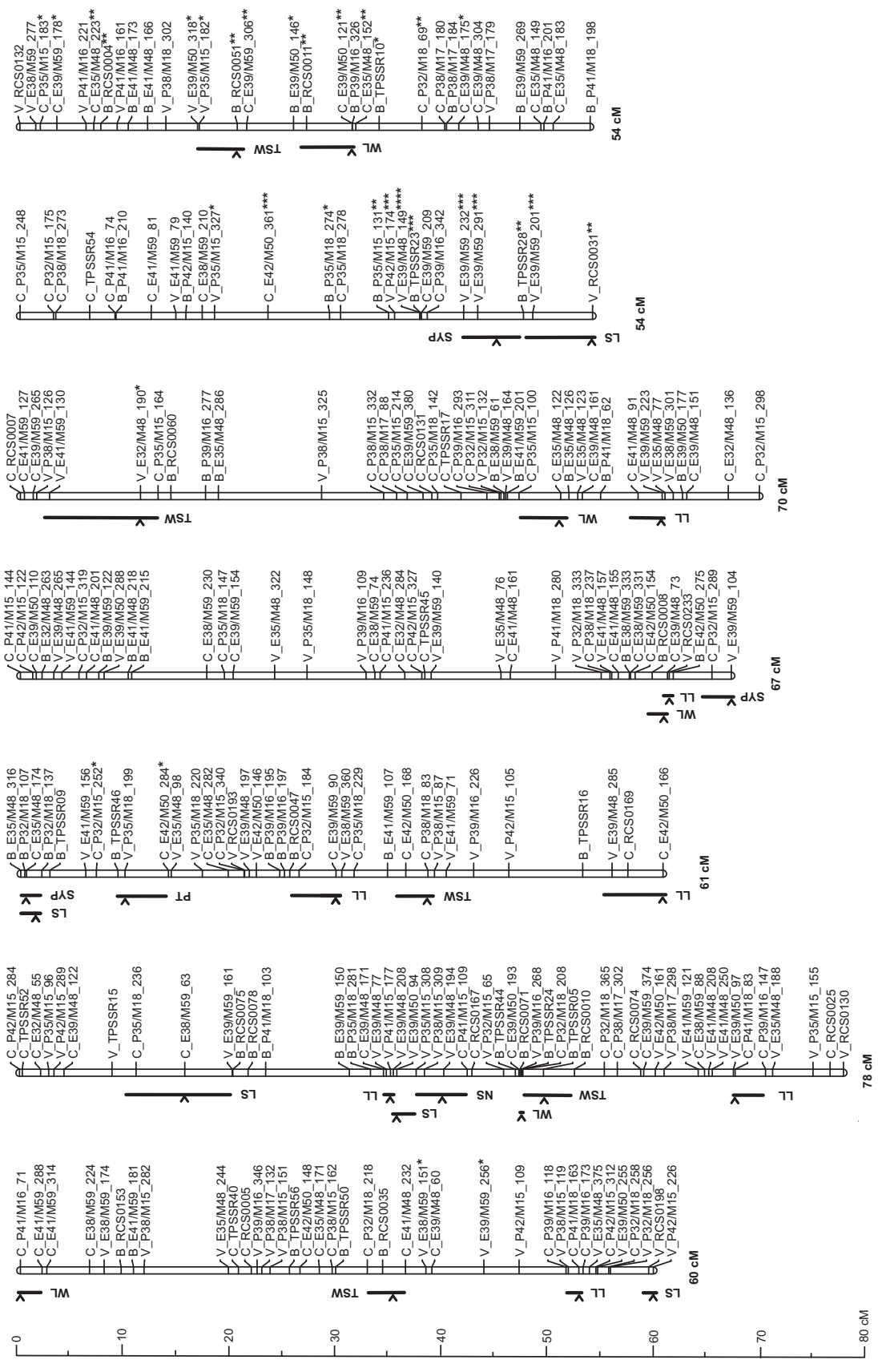
Trait	Linkage group	Position (cM)	Closest neighbouring locus	Maximum LOD score[a]	% variance explained
PT[b]	3	10.9	V_P35/M18_199[c]	5.90	11.0
				Total	**11.0**
NS	2	40.2	V_E39/M48_194	3.66	7.5
				Total	**7.5**
LS	1	59.9	V_P42/M15_226	3.99	4.5
	2	15.6	C_E38/M59_63	3.60	4.4
	2	35.8	V_E39/M50_94	7.43	8.0
	3	1.6	B_P32/M18_137	16.57	20.7
	6	54.2	V_RCS0031	6.20	8.7
				Total	**46.3**
WL	1	0.0	C_P41/M16_71	4.87	5.4
	2	47.6	B_TPSSR24	11.26	11.8
	4	60.8	C_E39/M48_73	10.32	11.4
	5	51.2	C_E35/M48_122	7.27	7.8
	7	31.5	C_E39/M50_121	5.29	5.4
				Total	**41.8**
LL	1	52.9	C_P41/M18_163	5.52	5.8
	2	35.2	V_P41/M15_177	13.47	14.4
	2	67.8	C_P41/M18_83	9.10	7.8
	3	28.4	C_E39/M59_90	8.88	10.9
	3	60.8	C_E42/M50_166	6.39	8.3
	4	61.4	RCS0233	7.61	7.2
	5	60.8	V_E39/M59_223	3.73	4.2
				Total	**58.6**

[a]Significant LOD threshold was 3.6

[b]For description of the traits, see Table 5.1

[c]For description of loci, see Fig. 5.1

Fig. 5.1 Genetic linkage map of a red clover population based on 254 F_1 genotypes, 42 SSR and 216 AFLP loci. Locus names consist of a denomination of the origin of the parental alleles (B = bi-parental locus; C and V = mono-parental locus heterozygous in the parent from the cultivar Corvus and Violetta, respectively), followed by the locus name (standard primer combination code (Keygene, Wageningen, Netherlands) followed by the allele size in relative migration units for AFLP loci or the prefix TPSSR (Kölliker et al., 2005) and RCS (Sato et al., submitted) followed by an identification number for SSR loci, respectively). Significantly distorted loci are indicated by asterisks (* $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, **** $P \leq 0.0001$). Positions of QTLs for seven traits were calculated using MQM mapping, the optimised map for the respective trait and least square means of four replicates per genotype (PT = persistence, NS = number of stems, LS = length of stem, WL = width of medial leaflet, LL = length of medial leaflet, SYP = seed yield per plant, TSW = thousand-seed weight). The maximum LOD score position of each QTL is indicated with an arrow and a bar represents the interval between two positions obtained at LOD scores two units lower than the maximal score


1
C_P41/M16_71
C_E41/M59_288
C_E41/M59_314
C_E38/M59_224
V_E38/M59_174
B_RCS0153
B_E41/M59_181
V_P38/M15_282
V_E35/M48_244
C_TPSSR40
C_RCS0005
V_P39/M16_346
V_P38/M17_132
V_P38/M15_151
B_TPSSR56
C_E42/M50_148
C_E35/M48_171
C_P38/M15_162
B_TPSSR50
C_P32/M18_218
B_RCS0035
C_E41/M48_232
V_E38/M59_151*
C_E39/M48_60
V_E39/M59_256*
V_P42/M15_109
C_P39/M16_118
V_P38/M15_119
C_P41/M18_163
C_P39/M16_173
V_E35/M48_375
C_P42/M15_312
V_E39/M50_255
C_P32/M18_258
V_P32/M18_256
V_RCS0198
V_P42/M15_226
60 cM
WL
TSW
LL
LS
2
C_P42/M15_284
C_TPSSR52
C_E32/M48_55
V_P35/M15_96
V_P42/M15_289
C_E39/M48_122
V_TPSSR15
C_P35/M18_236
C_E38/M59_63
V_E39/M59_161
B_RCS0075
B_RCS0078
B_P41/M18_103
B_E39/M59_150
B_P35/M18_281
C_E39/M48_171
V_E39/M48_77
V_P41/M15_177
V_E39/M48_208
V_E39/M50_94
V_P35/M15_308
V_P38/M15_309
V_E39/M48_194
C_P41/M15_109
C_RCS0167
V_P32/M15_65
B_TPSSR44
C_E39/M50_193
B_RCS0071
V_P39/M16_268
B_TPSSR24
C_P32/M18_208
B_TPSSR05
B_RCS0010
C_P32/M18_365
C_P38/M17_302
C_RCS0074
C_E39/M59_374
V_E42/M50_161
V_P38/M17_298
V_E41/M59_121
C_E38/M59_88
V_E41/M48_208
V_E41/M48_250
V_E39/M50_97
C_P41/M18_83
C_P39/M16_147
V_E35/M48_188
V_P35/M15_155
C_RCS0025
V_RCS0130
78 cM
LS
LL
NS
TSW
WL
3
B_E35/M48_316
B_P32/M18_107
C_E35/M48_174
B_P32/M18_137
B_TPSSR09
V_E41/M59_156
C_P32/M15_252*
B_TPSSR46
V_P35/M18_199
C_E42/M50_284*
V_E35/M48_98
V_P35/M18_220
C_E35/M48_282
C_P32/M15_340
V_RCS0193
V_E39/M48_197
V_E42/M50_146
B_P39/M16_195
B_P39/M16_197
B_RCS0047
C_P32/M15_184
C_E39/M59_90
V_E38/M59_360
C_P35/M18_229
B_E41/M59_107
C_E42/M50_168
C_P38/M18_83
V_P38/M15_87
V_E41/M59_71
V_P39/M16_226
V_P42/M15_105
B_TPSSR16
V_E39/M48_285
C_RCS0169
C_E42/M50_166
61 cM
LS
SYP
PT
LL
TSW
4
C_P41/M15_144
C_P42/M15_122
C_E39/M50_110
B_E32/M48_263
V_E39/M48_265
V_E41/M59_144
C_P32/M15_319
C_E41/M48_201
B_E39/M59_122
V_E39/M50_288
B_E41/M48_218
B_E41/M59_215
C_E38/M59_230
C_P35/M18_147
C_E39/M59_154
V_E35/M48_322
V_P35/M18_148
V_P39/M16_109
C_E38/M59_74
C_P41/M15_236
C_E32/M48_284
C_P42/M15_327
C_TPSSR45
V_E39/M59_140
V_E35/M48_76
C_E41/M48_161
V_P41/M18_280
V_P32/M18_333
C_P38/M18_237
V_E41/M48_157
C_E41/M48_155
B_E38/M59_333
C_E38/M59_331
C_E42/M50_154
B_RCS0008
C_E39/M48_73
V_RCS0233
B_E42/M50_275
C_P32/M15_289
V_E39/M59_104
67 cM
WL
LL
SYP
5
C_RCS0007
C_E41/M59_127
C_E39/M59_265
V_P38/M15_126
V_E41/M59_130
V_E32/M48_190*
C_P35/M15_164
B_RCS0060
B_P39/M16_277
B_E35/M48_286
V_P38/M15_325
C_P38/M15_332
C_P38/M17_88
C_P35/M15_214
C_E39/M59_380
C_RCS0131
C_P35/M18_142
C_TPSSR17
C_P39/M16_293
C_P32/M15_311
V_P32/M15_132
B_E38/M59_61
V_E39/M48_164
B_E41/M59_201
C_P35/M15_100
C_E35/M48_122
B_E35/M48_126
V_E35/M48_123
C_E39/M48_161
B_P41/M18_62
C_E41/M48_91
V_E39/M59_223
V_E35/M48_77
V_E38/M59_301
B_E39/M50_177
C_E39/M48_151
C_E32/M48_136
C_P32/M15_298
70 cM
TSW
WL
LL
6
C_P35/M15_248
C_P32/M15_175
C_P38/M18_273
C_TPSSR54
C_P41/M16_74
B_P41/M16_210
C_E41/M59_81
V_E41/M59_79
B_P42/M15_140
C_E38/M59_210
V_P35/M15_327*
C_E42/M50_361***
B_P35/M18_274*
C_P35/M18_278
B_P35/M15_131**
V_P42/M15_174***
V_E39/M48_149****
B_TPSSR23***
C_E39/M59_209
C_P39/M16_342
V_E39/M59_232***
V_E39/M59_291***
B_TPSSR28**
V_E39/M59_201***
V_RCS0031**
54 cM
SYP
LS
7
V_RCS0132
V_E38/M59_277
C_P35/M15_183*
C_E39/M59_178*
V_P41/M16_221
C_E35/M48_223**
B_RCS0004**
V_P41/M16_161
B_E41/M48_173
B_E41/M48_166
V_P38/M18_302
V_E39/M50_318*
V_P35/M15_182*
B_RCS0051**
C_E39/M59_306**
B_E39/M50_146*
B_RCS0011**
C_E39/M50_121**
B_P39/M16_326
C_E35/M48_152**
B_TPSSR10*
C_P32/M18_69**
C_P38/M17_180
B_P38/M17_184
C_E39/M48_175*
C_E39/M48_304
V_P38/M17_179
B_E39/M59_269
C_E35/M48_149
B_P41/M16_201
C_E35/M48_183
B_P41/M18_198
54 cM
TSW
WL
0
10
20
30
40
50
60
70
80 cM

5.4 Discussion

Phenotypic traits which are easier to score than complex target traits but are closely correlated to the latter may help to greatly facilitate and enhance selection. Therefore, one aim of this study was to identify traits closely correlated to persistence, a trait which is laborious to determine and difficult to improve. The traits NS, LS, LW, and LL, which are known to influence plant vigour, showed a close relationship to persistence (PT; Table 5.3). This is in congruence with phenotypic observations in red clover, where vigour and leafiness in the second year were highly correlated with persistence, i.e. the percentage of surviving plants in the third year (Taylor et al. 1962). A plant, which showed poor vigour already in the second year, may be limited in its ability to mobilise energy for seed growth resulting in a poor TSW. This may explain the close relationship of PT and TSW. On the other hand, SYP showed low, but still significant positive association to PT (Table 5.3). This is in contrast to the hypothesis of a negative correlation between persistence and seed yield, based on the fact that the highly persistent Mattenklee cultivars often show a decreased seed yield (Deneufbourg 2004; Lehmann and Briner 1998). Therefore, selection for increased seed yield in Mattenklee and other red clover cultivars may be possible without decreasing persistence. However, investigation of seed yield and persistence was performed in the same field trial, which may result in an interference of persistence by seed yield. The investigation of seed yield changed frequency and timing of cuts, which was reported to have an influence on persistence (Wiersma et al. 1998). In addition, seed development in general represents a stress for a plant. Thus, to further elucidate association between PT and SYP independent field trials are necessary.

Persistence is the result of physiogenic and pathogenic stress load, i.e. influenced by numerous factors (Taylor and Quesenberry 1996) and showed a low heritability (Table 5.2). Therefore, it was expected that QTLs will only explain a low percentage of total variation (Table 5.4). Such conditions render detection of QTLs difficult, which has been demonstrated in a QTL study in perennial ryegrass, where no QTL for winter survival, a trait influencing persistence, could be detected (Yamada et al. 2004). Nevertheless, we were able to identify one QTL for persistence allowing to partly confirm the close relationships of PT to other traits such as SYP or LS. The QTL identified for PT was detected in the same region as a QTL for LS (LG 3; Fig. 5.1), which confirmed the high correlation between these two traits (Table 5.3). Thus, LS may present a time-saving indirect character to select for PT.

There was a third QTL for SYP located in this region (LG 3; Fig. 5.1), which confirmed the high correlation between LS and SYP (Table 5.3). These positive associations between LS and SYP as well as PT, made a negative association between the two QTLs of PT and SYP on LG 3 very unlikely. Moreover, these three QTLs explained between 11 and 21 % of the total variation

of the respective traits (Table 5.4). Therefore, further investigations of this region on LG 3 may provide a promising opportunity to increase SYP and PT simultaneously.

The close relationship of LS to SYP was further confirmed by a second genome region on LG 6, where a QTL of LS and SYP were detected, respectively (Fig. 5.1). Indeed, Mattenklee cultivars, which showed a low seed yield (Deneufbourg 2004), showed also shorter stems compared to other red clover cultivars (Bundessortenamt 2003). An association between the location of QTLs and the correlation of the two traits length of stem and seed yield has also been reported for soybean (*Glycine max*; Mansur et al. 1996). Therefore, LS may also present an easy and time-saving trait to be included in a strategy for improvement of seed yield in red clover. However, length of stem is a trait which is correlated with lodging (Inoue et al. 2004) and therefore longer stems are only desirable to a certain extent. Thus, the QTL of SYP detected on LG 4 (Fig. 5.1) may be interesting for further characterisation although it explains less than the other two QTLs (Table 5.4). Moreover, in this region on LG 4 QTLs for WL and LL were identified, traits which may have positive effect on forage yield (Nüesch 1960).

In conclusion, QTL analysis of persistence, number of stems, length of stem, width of medial leaflet and length of medial leaflet allowed for the identification of 19 QTLs. The associations among these five traits and two seed yield components (seed yield per plant and thousand-seed weight) indicated that improvement of seed yield should be possible without decreasing persistence. It allowed for the identification of length of stem as an easy to determine indirect character to improve seed yield and persistence in red clover. Furthermore, several genome regions were identified including promising QTLs for further characterisation and development of markers.

Acknowledgements

The authors would like to thank Simone Günter, Philipp Streckeisen and Peter Tanner for support in the field. This study was funded by the breeding foundation DSP-BLW.

6 General Discussion

6.1 Analysis of genetic diversity – methods and applications

Detailed knowledge of genetic diversity is an important prerequisite for the efficient protection and conservation of genetic resources and for their targeted exploitation in plant breeding or can contribute to ecosystem stability (Becker 1993; Jarvis et al. 2000). Molecular markers in general and AFLP markers in particular represent a valuable method for a rapid and reliable characterisation of genetic diversity (Beebe et al. 2001; Kölliker et al. 2001). Diversity studies conducted in the framework of this thesis showed that Mattenklee landraces and Swiss wild clover populations are distinct and diverse genetic resources for red clover improvement and may contribute to the conservation and restoration of biodiversity in pastures and meadows (chapter 2 and 3). The clear separation of Swiss wild red clover populations from the other red clover groups emphasises that the ancestry of Mattenklee is primarily found in introduced rather than in natural wild clover populations (chapter 3). In addition to the investigation of the over all value of Mattenklee landraces as a genetic resource, the determination of distinctness and diversity of specific populations may play an important role in selecting populations for the introgression into breeding material. Furthermore, a comparison of different methods to investigate genetic diversity, i.e. investigation of individual plants (chapter 2) or bulked samples (chapter 3), may allow to facilitate and optimise further investigation of diversity in outcrossing species.

6.1.1 Methodological considerations

A single population of an outcrossing species consists of many different genotypes and therefore a large number of individuals have to be examined to represent such a population (Crossa 1989). An alternative to analyses of individual plants of a population can be the combination of plant material from individual plants into one or several bulked samples (Michelmore et al. 1991). If individual plants are analysed rare alleles can be also detected, which makes this method particularly valuable to investigate diversity within a population, but also diversity among populations and groups can be reliably estimated. However, this method often produces large numbers of samples to be investigated. As an alternative, bulked samples reduce the number of samples to be analysed and population specific markers may be accumulated (Michelmore et al. 1991), which makes this method particularly useful to determine diversity among populations and groups in large scale studies. If two or several bulked samples are analysed, even an estimation of diversity within populations becomes possible. However, rare alleles may not be detectable with this approach (Gilbert et al. 1999) and if only one bulked sample per population is investigated, diversity within populations can not be determined. Therefore, investigation of individual plants as well as of bulked samples are

valuable and comparable methods and whether one or the other method is applied mainly depends on the specific research question.

Comparison of results of the diversity studies based on 24 individual plants per population (chapter 2) or based on two bulked leaf samples of 20 plants each (chapter 3) per population, confirmed the comparability of these two methods. Both studies revealed high genetic diversity within Mattenklee landraces, which was clearly higher than diversity within Mattenklee or field clover cultivars (Table 2.5; Table 3.7). In addition, both studies indicated that Mattenklee landraces represent a distinct group, clearly separated from Mattenklee cultivars and field clover cultivars (Fig. 2.2; Table 3.6; Fig. 3.2). Therefore, the value of Mattenklee landraces as a genetic resource was demonstrated independently of whether individual plants or bulked samples were used.

However, there was also some discrepancy between the two studies. In the study with individual plants, Mattenklee landraces were grouped closer to field clover cultivars when compared to Mattenklee cultivars (Fig. 2.4), whereas in the study with bulked samples, Mattenklee landraces and cultivars were grouped together when compared to field clover cultivars (Fig. 3.2; Fig. 3.3). One reason for this observation may be the fact that a different number of populations was analysed in both studies. In the study with bulked samples, 89 Mattenklee landraces and six field clover cultivars were included, while in the study with individual plants only eight Mattenklee landraces and three field clover cultivars were used. The larger the number of plants investigated of a population, the better the population is represented (Crossa 1989). This may analogously be valid for the number of populations representing a group of a species. A number too small, e.g. three field clover cultivars, may therefore not represent the group investigated, but rather the individual population.

Another reason for the discrepancy between the two studies may be the fact that only few populations of Mattenklee cultivars were included in both studies. Two cultivars (Pica and Pavo), exclusively analysed in the study with individual plants, were developed using not only Mattenklee germplasm but also wild clover ecotypes collected in Croatia (Boller 2000a; Boller et al. 2004). This may have resulted in a more distant relationship of Mattenklee landraces and Mattenklee cultivars. On the other hand, the inclusion of the two cultivars Changins and Leisi, exclusively investigated in the study with bulked samples, may have resulted in a closer relationship of Mattenklee landraces and Mattenklee cultivars. Changins is an old cultivar (listed in 1970) no longer on the market and Leisi was regarded as a Mattenklee cultivar although it is rather a Mattenklee landrace commercially distributed for several decades (Boller 2000b).

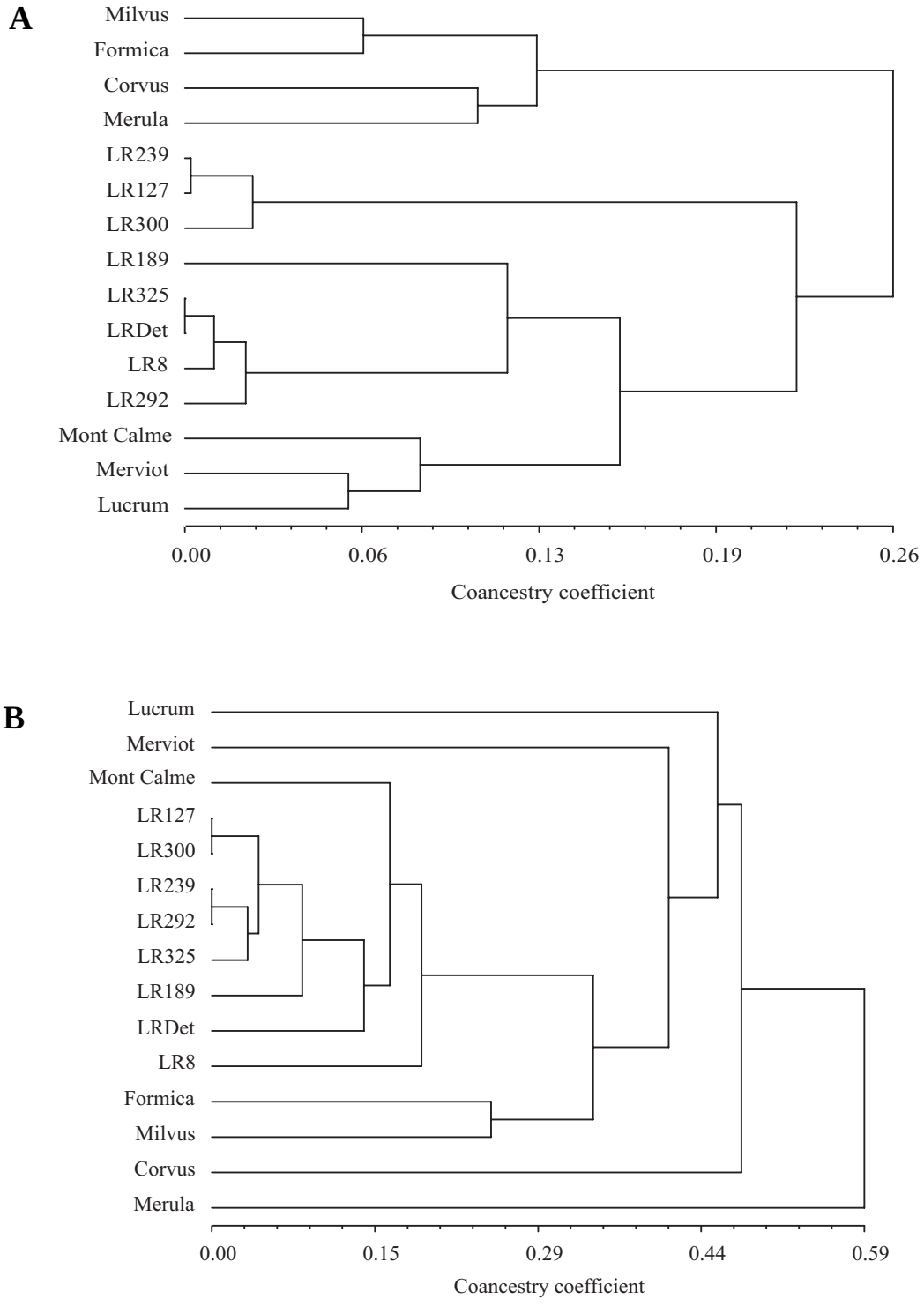


Fig. 6.1 UPGMA clustering of eight Mattenklee landraces, four Mattenklee cultivars and three field clover cultivars. Analysis was based on on coancestry coefficients derived from 24 individual plants per population and 126 AFLP markers (**A**; chapter 2) or coancestry coefficients derived from two bulked samples per population and 410 AFLP markers (**B**; chapter 3).

Therefore, disagreement between the two studies is probably rather due to different number and nature of populations rather than due to the different methods applied. Comparison of topologies of two dendrograms (Fig. 6.1) partly confirmed this hypothesis. The two dendrograms were based on coancestry coefficients (chapters 2 and 3) using the unweighted pair-group method with an arithmetic mean (UPGMA) and included the common populations investigated in the study with individual plants (Fig. 6.1A) as well as with bulked samples (Fig. 6.1B). Both dendrograms clearly separated the four Mattenklee cultivars Corvus, Merula, Formica and Milvus from the eight Mattenklee landraces. Furthermore, the field clover cultivar Mont Calme grouped in both studies within the cluster comprising the eight Mattenklee landraces. However, in the dendrogram based on individual plants the two other field clover cultivars Merviot and Lucrum were also grouped within the cluster comprising the Mattenklee landraces (Fig. 6.1A), whereas in the dendrogram based on bulked samples these two field clover cultivars were clearly separated from the Mattenklee landraces (Fig. 6.2B).

Over all, the investigation of individual plants and the investigation of bulked samples are two comparable methods to determine distinctness among groups as well as diversity within populations of a group. The advantage of bulked samples thereby is the reduction of the sample number. It therefore provides an efficient method for large scale studies not only for red clover, but also for diversity studies in other outcrossing grassland species. However, if determining of diversity within individual populations plays a decisive role in the objectives of a study, individual plants of a population have to be investigated.

6.1.2 The value of diversity studies for the improvement of red clover

The knowledge of diversity and relationship among Mattenklee landraces and Swiss wild red clover populations as well as of their relationship to other red clover groups may be applied to a targeted improvement of red clover cultivars (chapter 2 and 3). On the one hand, such information may be applied for an over all enhancement of diversity in cultivars or breeding material. Such an enhancement may have a positive effect on forage yield and quality (Kölliker et al. 2005a). On the other hand, the high diversity of landraces and wild populations makes it likely that a trait of interest, for example resistance to biotic or abiotic stress, is present. Introgression of such a population into breeding material may result in an improvement of the trait of interest in a cultivar (Rao et al. 2003).

Swiss wild clover populations formed a clearly separated distinct group with a different genetic background compared to the other red clover groups. In addition, this group showed a very high diversity among populations. Therefore, each of the thirteen investigated Swiss wild clover populations form a valuable and different genetic resource potentially useful in a breeding

program for the over all enhancement of genetic diversity. Furthermore, there is high potential, that a trait of interest is particularly well expressed in some of these populations (Boller 2000a; Boller et al. 2004). However, compared to cultivars, wild clover populations have several disadvantages, which limit their direct use in breeding programs. The trait of interest is often linked to undesired traits such as sensitivity to frequent cutting or low forage yield (Nüesch 1960).

In contrast to wild clover populations, Mattenklee landraces may have the advantage that certain qualities for cultivation are already present, since they have been cultivated and propagated over many decades. For example, seeds of these landraces were traditionally harvested in the third year, indicating that a certain selection for persistence was already performed (Boller 2000b). Phenotypic investigation of a subset of 33 Mattenklee landraces further confirmed such qualities. Although average agronomic performance of the landraces was poorer when compared to the Mattenklee cultivar Milvus, there was, for most of the traits, at least one Mattenklee landrace, which showed better performance (Herrmann et al. 2005a). Therefore, the large pool of 89 Mattenklee landraces harbours a high potential to improve red clover cultivars.

In order to increase over all diversity in the breeding material, landraces, which differentiated from other landraces as well as from Mattenklee cultivars, are of particular interest. Coancestry coefficients and principal coordinate analysis are valuable criteria to select Mattenklee landraces and to estimate diversity among populations (Reynolds et al. 1983). In this respect, LR329 is most interesting as it has an average coancestry coefficient of 0.59 to all other landraces, whereas the average coancestry coefficient was 0.22 for all landraces. In addition, the average coancestry coefficient of LR329 to Mattenklee cultivars was 0.78 compared to the average coefficient of Mattenklee landraces to Mattenklee cultivars of 0.39 (chapter 3).

Mattenklee landraces grouping close to Swiss wild clover populations such as LR62 (Fig. 3.2) may represent another group of landraces particularly interesting for red clover breeding. Gene flow between Mattenklee landraces and Swiss wild clover populations was postulated in this thesis (Fig. 3.4). Genes from introgressions have been shown to persist for at least five generations, if the genes were either neutral or advantageous for the plant (Linder et al. 1998). Valuable characteristics such as resistance to powdery mildew (*Erysiphe polygoni*; Boller 2000a; Boller et al. 2004) of wild populations have a good chance to persist in landraces. Therefore, such landraces may possess valuable characteristics of wild populations, but do not have the disadvantages of wild populations, for example low forage yield (Nüesch 1960). However, further and more detailed investigations of landraces and wild populations have to be performed to confirm this hypothesis.

Coancestry coefficients allow also to identify landraces of less interest such as LR222, which was grouped close to cultivars (Fig. 3.2). For such landraces, it can not be excluded that seed of cultivars have been used to complement Mattenklee landraces in times of seed scarcity and they have therefore a high potential to represent a similar genetic background as the cultivars (Boller 2000b). In addition, up to nine landraces were identified, which were not distinguishable from each other based on coancestry coefficients (data not shown). At least when more than one landrace is planned to be integrated in a breeding program, no closely related landraces should be selected. Moreover, identification of closely related landraces or in general redundancy in gene banks could help to reduce the number of samples, the cost of maintenance and, therefore, may help to more efficiently manage germplasm collections (Becker 1993).
In summary, the diversity studies performed in this thesis (chapter 2 and 3) provide a tool for selection of specific Mattenklee landraces for the introgression into the breeding material in order to enhance the diversity in cultivars. Thus, the diversity studies yielded directly applicable results to improve red clover cultivars.

6.2 Improvement of seed yield and persistence in red clover using molecular markers – prospects and next steps

Seed yield and persistence of red clover are two important traits, which are difficult to improve, because of their complexity. Identification of molecular markers closely linked to these two traits for application in marker assisted selection (MAS) would therefore be a very helpful tool to improve red clover cultivars (Newbury 2003). QTL analysis of seed yield and persistence allowed for the identification of head number per plant and length of stem as two easily determinable, indirect characters to select for seed yield (chapter 4 and 5). The results further suggested that increasing seed yield by selection should be possible without decreasing persistence. In addition, length of stem may also represent an indirect character to select for persistence (chapter 5). However, further detailed analyses are needed before MAS for seed yield and persistence is applicable. The detected QTLs have to be stable and therefore the identified QTLs have to be confirmed in additional populations and environments. Furthermore, reliable markers closely linked to these QTLs have to be developed and confirmed (Young 1999a).

6.2.1 From QTL analysis to marker assisted selection – a long way to go

A total of 57 QTLs for persistence, eight seed yield components as well as for four additional morphological traits were detected in 2004 in one environment (chapter 4 and 5). However, an

important step towards MAS is the confirmation of the QTLs over several years, in several environments and in other populations (Young 1999a), which is beyond the scope of a PhD thesis. In addition to the resources and the time necessary for confirmation, one major problem of red clover is that the required clonal replicates are difficult to obtain and maintenance of red clover in the greenhouse is limited (Cumming and Steppler 1961). Only few clonal replicates can be obtained by cloning individual plants and therefore for each additional field trial cloning has to be performed separately. However, before a clonal replicate can again be cloned at least four months re-growths under longday conditions and four months under shortday conditions to obtain as many shoots as possible are necessary (chapter 4). Thus, confirmation of QTLs over different years and environments is very difficult to perform. However, major QTLs appear to be stable across different environments and years as was found for example in pea (*Pisum sativum* L.), where four of nine QTLs for thousand seed weight and three of nine for seed number were detected in three different years (Timmermann-Vaughan et al. 2005). Confirmation of the detected QTLs over different populations is another important point towards MAS. The map described in chapter 4 was linked to other red clover maps (Isobe et al. 2003; Sato et al., submitted), where QTL analyses are currently performed (S. Isobe, personal communication). Therefore, some confirmation of location, number and percentage of explained variance of the identified QTLs should be possible in the near future.

The greatest prospect of success for the confirmation of QTLs and therefore for the identification of stable QTLs is obtained, when a trait is controlled by only few QTLs and has a high heritability, i.e. a high proportion of the total observed phenotypic variance can be explained by the genotypic variance (Beavis 1998). For example, in pea the decrease of confirmation of QTLs across different years and environments corresponded to the decrease of heritability of the traits (Timmermann-Vaughan et al. 2005). Therefore, for the improvement of red clover cultivars, i.e. for a high response to selection using MAS, high heritability is highly useful. However, traits with high heritability show also highest response to traditional phenotypic selection (Edwards and Page 1994). As markers for MAS have to be developed before the application in breeding is possible, response to selection using markers should be better than that for phenotypic selection alone.

Although high response to selection is the main objective in breeding, factors such as time and cost to obtain this response have also to be taken into account. For example, phenotypic evaluation of seed yield in red clover is only possible in the second year and is very laborious, whereas MAS could be performed a few weeks after germination with significantly lower effort. Moreover, although MAS is quite expensive, phenotypic analysis appeared to be more costly for quality traits such as kernel sucrose concentration (Yousef and Juvik 2001). Therefore, even for

simple traits with a high heritability, which are controlled by only few QTLs, MAS may be superior to phenotypic selection alone. For example, control of resistance against the soybean cyst nematode (*Heterodera glycines*) is controlled by only a few loci and it was demonstrated that one locus (*rhg1*) explains more than 50 % of the variation for resistance (Cregan et al. 1999b). The soybean cyst nematode is one of the most destructive pests of soybean (*Glycine max*), for which the use of resistant cultivars is very effective to reduce crop losses. However, resistance against cyst nematode is difficult and time consuming to score with conventional techniques. Therefore, MAS with SSR markers closely linked to the *rhg1* resistance locus form the basis of many public and commercial breeding efforts (Young 1999a).

However, greatest potential for successful and efficient application of MAS can be expected for those quantitative traits that exhibit low heritability, where phenotypic selection is limited (Beavis 1998; Edwards and Page 1994). MAS will offer a primary advantage over phenotypic selection for the first two to three cycles of selection (Edwards and Page 1994; Yousef and Juvik 2001). Seed yield and persistence are such traits, which are difficult and laborious to improve by phenotypic selection. Both traits exhibited low heritability when compared to the other traits investigated (chapter 4 and 5), which is in congruence with the findings of other studies investigating for example seed yield in pea (*Pisum sativum*; Timmermann-Vaughan et al. 2005) or winter survival, a trait influencing persistence, in perennial ryegrass (*Lolium perenne*; Yamada et al. 2004). Moreover, seed yield and persistence are associated with many other factors (chapter 4 and 5; Oliva et al. 1994; Taylor and Quesenberry 1996) and negative correlation to other important traits such as forage yield (Steiner et al. 1997) have been reported. These factors represent further impediments for improving these two traits with traditional phenotypic selection.

However, for the identification of stable QTLs for complex traits with a low heritability and negative correlation to other traits a large number of individuals have to be evaluated. Particularly for such traits, the number of detected QTLs is underestimated and the variance explained of a QTL is often overestimated, when only a small number of individuals are investigated (Beavis 1998). Nevertheless, significant response to selection based on MAS has been achieved. For example, for sweet corn (*Zea mays* L.) it has been reported that out of 52 pairwise comparisons of MAS and phenotypic selection, MAS resulted in a significantly higher response to selection than phenotypic selection in 38 % of the comparisons, while phenotypic selection was only superior to MAS in 4 % of the comparisons (Yousef and Juvik 2001). Therefore, taken these limitations into account, molecular dissection of seed yield and persistence still harbours a high potential for improvement of red clover cultivars, although there is a long way to go until MAS for these characters can be applied in breeding programs.

6.2.2 Next steps towards markers linked to seed yield and persistence

In a next step, the identified interesting genome regions (chapter 4 and 5) need to be characterised in more detail. For seed yield, the focus may lay on the region identified on LG 3. This region not only included four QTLs within 1 cM mostly explaining more than 10 % of variation of seed yield components, it also contained a QTL for length of stem explaining more than 20 % of the variation. Furthermore, the only identified QTL for persistence, which was positively correlated with QTLs for seed yield components and length of stem, was detected within 11 cM (Fig. 4.1; Fig. 5.1).

Fine mapping of QTLs offers a promising possibility to more precisely characterise this region with regard to identification of reliable markers for MAS. Use of the information of the recently published high-density SSR map (Sato et al., submitted) may on the one hand allow to identify markers more closely linked to the QTLs. On the other hand, these SSR markers have the advantage over AFLP markers that they are user-friendly, i.e. for each SSR marker a specific primer pair is available, which allows for an efficient screening of populations. An AFLP marker closely linked to a QTL has first to be converted into a STS marker to allow for efficient screening.

Identification of closely linked markers on either side of the QTL of a target trait is the first step towards map based cloning (Kumar 1999). Map based cloning allows not only to verify the existence of genes of interest in a specific plant, but to identify the exact location and to isolate the gene of interest. Cloning of a gene includes several steps, i.e. to develop a physical map including the two flanking markers and the gene of interest, to construct a gene library and to perform chromosome walking starting from the two flanking markers using this library. This results in cloning the entire DNA sequence between the closest markers, including the gene of interest. The last step will be to identify the target gene from a number of candidate genes derived from the cloned region (Kumar 1999). Map based cloning has been used in various crop species to isolate for example the *Vf*- resistance gene against scab (*Venturia inaequalis*) in apple (*Malus xdomestica*; Barbieri et al. 2003) or the *Lr10* resistance gene against leaf rust (*Puccinia triticina*) in wheat *(Triticum aestivum;* Feuillet et al. 2003). However, the major limitation of map based cloning lies in the laborious and time consuming steps involved in chromosome walking. Comparative mapping, i.e. the alignment of chromosomes of related species based on common molecular markers, provides a possibility to assist in the determination and mapping of candidate genes and therefore to clone the gene of interest. Comparative mapping is well established in several plant families such as *Poaceae*, where initial studies predicted that comparative mapping would greatly facilitate gene discovery among related species (Devos and Gale 2000). Extensively investigated model plants such as *Medicago truncatula*, where an

international consortium is sequencing the genome (Cannon et al. 2005), have a great potential as sources of candidate genes for legumes. However, up to now no genes for seed yield or persistence have been isolated in legumes. In contrast, such genes have been identified in *Arabidopsis* (Jofuku et al. 2005). The gene *APETALA2*, which has long been known for its role in flower organ development, has been reported to also play an important role in determining seed size, seed weight, and the accumulation of seed oil and protein (Jofuku et al. 2005). However, several studies have suggested that the practical utility of comparative mapping in plants may be limited to within-family comparisons, which seems also to be true for legumes (Choi et al. 2004).

Although the application of comparative mapping will be limited in the near future with regard to genes directly associated with seed yield and persistence, comparative mapping will be a very helpful tool to identify conserved QTL locations between red clover and other legumes. QTL analyses for seed yield were also conducted in white clover (Barrett et al. 2004), soybean (Mansur et al. 1996) or pea (Timmermann-Vaughan et al. 2005). An alignment of the maps of these species could therefore help to identify common QTL regions, which may allow for the confirmation, more detailed description and the identification of additional locations of QTLs in red clover. Several important crop legumes such as alfalfa or soybean have either a very large genome or are polyploid in nature, which limits their utility as experimental systems. In contrast, red clover is comparable to model legumes with favourable genetic attributes such as a relative small genome size, only seven chromosomes and a diploid nature. Moreover, as the model plant *Medicago truncatula* and red clover belong to the same tribe, have a similar genome size and only one chromosome difference, a high level of synteny may be expected (Choi et al. 2004). Thus, comparative mapping probably forms the most promising possibility for further investigations towards markers closely linked to seed yield and persistence in red clover.

6.3 Conclusions

This thesis clearly shows that Mattenklee landraces and Swiss wild clover populations form distinct and very diverse genetic resources of red clover. Moreover, it demonstrates that the ancestry of Swiss Mattenklee is primarily found in introduced germplasm rather than in Swiss wild clover populations. These findings will not only assist an efficient and targeted conservation and maintenance of Mattenklee landraces and wild clover populations, it will also help to broadening the diversity of red clover breeding material. Specific Mattenklee landraces for the introgression into the breeding material were identified, which allows for the direct application of the results to improve red clover cultivars. The optimised bulking strategy developed in this thesis allows for an efficient assessment of genetic diversity in outcrossing species. In addition, the comparison of investigations based on individual plants and bulked samples helps to optimise analysis methods depending on the specific research question. The established linkage map together with the detected QTL provide a valuable basis for the further development of molecular markers linked to seed yield and persistence of red clover and present an important step towards marker assisted selection in this species. In addition, the thorough phenotypic and genetic analysis allowed to identify several traits such as head number per plant or length of stem which are highly correlated to seed yield and persistence but are far easier to score than the target traits. These traits may directly be applied in breeding programs for the further improvement of red clover cultivars. In summary, this thesis provides the basis for a systematic exploration of the red clover genome, helps to develop marker assisted breeding strategies and contributes to the efficient maintenance of red clover gene pools.

7 References

Abberton MT, Cogan NOI, Smith KF, Kearney G, Marshall AH, Williams A, Michaelson-Yeates TPT, Bowen C, Jones ES, Veccies AC, Forster JW (2005) Quantitative trait locus analysis of morphogenetic and developmental traits in an SSR- and AFLP-based genetic map of white clover (*Trifolium repens* L.). In: Humphreys MO (ed) Proceedings of the 4th international symposium on the molecular breeding of forage crops and turf, a satellite workshop of the XXth International Grassland Congress, July 2005, Aberystwyth, Wales. Academic, Wageningen, pp 150

Andersen IP, Lübberstedt T (2003) Functional markers in plants. Trends Plant Sci 8:554-560

Annicchiarico P, Piano E (1997) Effect of selection under cultivation on morphological traits and yield of ladino white clover landraces. Genet Res Crop Evol 44:405-410

Armstead IP, Turner LB, King IP, Cairns AJ, Humphreys MO (2002) Comparison and integration of genetic maps generated from F2 and BC1-type mapping populations in perennial ryegrass. Plant Breed 121:501-507

Badoux S, Guyer H, Lehmann J (1967) Red clover trials 1964-66. Mitt Schweiz Landw 15:140-146

Baltisberger M (1995) Introduction in the systematic of plants. VDF Hochschulverlag AG ETH, Zurich

Barbieri M, Belfanti E, Tartarini S, Vinatzer BA, Sansavini S, Silfverberg-Dilworth E, Gianfranceschi L, Hermann D, Patocchi A, Gessler C (2003) Progress of map-based cloning of the *Vf*-resistance gene and functional verification: Preliminary results from expression studies in transformed apple. Hortscience 38:329-331

Barrett B, Griffiths A, Schreiber M, Ellison N, Mercer C, Bouton J, Ong B, Forster J, Sawbridge T, Spangenberg G, Bryan G, Woodfield D (2004) A microsatellite map of white clover. Theor Appl Genet 109:596-608

Barrett BA, Baird IJ, Woodfield D. R. (2005) A QTL analysis of white clover seed production. Crop Sci 45:1844-1850

Beavis WD (1998) QTL analyses: power, precision, and accuracy. In: Paterson AH (ed) Molecular dissection of complex traits. CRC Press LLC, Boca Raton, pp 145-162

Becker H (1993) Plant Breeding. UTB für Wissenschaft, Stuttgart

Beebe S, Rengifo J, Gaitan E, Duque MC, Tohme J (2001) Diversity and origin of Andean landraces of common bean. Crop Sci 41:854-862

Belzile L (1991) The effect of soil type on red clover seed yield. Can J Plant Sci 71:1039-1046

Bert PF, Charmet G, Sourdille P, Hayward MD, Balfourier F (1999) A high-density molecular map for ryegrass (*Lolium perenne*) using AFLP markers. Theor Appl Genet 99:445-452

Bingefors S, Akerberg E (1961) Swedish land-races of red clover. Euphytica 10:147-151

Boller B (2000a) Corvus and Pica: two new red clover cultivars of the persistent Mattenklee type. Agrarforschung 7:484-489

Boller B (2000b) History and development of the Swiss Mattenklee, a persistent form of cultivated red clover. Vierteljahresschr Nat Forsch Ges Zürich 145:143-151

Boller B, Tanner P, Schubiger F (2004) Merula and Pavo: new, persistent red clover cultivars of the Mattenklee type. Agrarforschung 11:162-167

Boller BC, Nösberger J (1987) Symbiotically fixed nitrogen from field-grown white and red-clover mixed with ryegrasses at low levels of N-15-fertilization. Plant Soil 104:219-226

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314-331

Boutin-Ganache I, Raposo M, Raymond M, Deschepper CF (2001) M13-tailed primers improve the readability and usability of microsatellite analyses performed with two different allele-sizing methods. BioTechniques 31:24-28

Brouwer DJ, Duke SH, Osborn TC (2000) Mapping genetic factors associated with winter hardiness, fall growth, and freezing injury in autotetraploid alfalfa. Crop Sci 40:1387-1396

Brouwer DJ, Osborn TC (1999) A molecular marker linkage map of tetraploid alfalfa (*Medicago sativa* L.). Theor Appl Genet 99:1194-1200

Brown SM, Szewc-McFadden AK, Kresovich S (1996) Development and application of simple sequence repeat (SSR) loci for plant genome analysis. In: Jauhar PP (ed) Methods of genome analysis in plants. CRC Press Inc., Boca Raton, pp 147-159

Brownstein MJ, Carpten JD, Smith JR (1996) Modulation of non-templated nucleotide addition by *Taq* DNA polymerase: Primer modifications that facilitate genotyping. BioTechniques 20:1004-1010

Brummer EC, Bouton JH, Kochert G (1993) Development of an RFLP map in diploid alfalfa. Theor Appl Genet 86:329-332

Bundesamt für Statistik (2004) Statistical Yearbook of Switzerland 2004. Neue Zürcher Zeitung, Zurich

Bundessortenamt (2003) Descriptive list of varieties. Deutscher Landwirschaftsverlag GmbH, Hannover

Campos-De-Quiroz H, Ortega-Klose F (2001) Genetic variability among elite red clover (*Trifolium pratense* L.) parents used in Chile as revealed by RAPD markers. Euphytica 122:61-67

Cannon SB, Crow JA, Heuer ML, Wang XH, Cannon EKS, Dwan C, Lamblin AF, Vasdewani J, Mudge J, Cook A, Gish J, Cheung F et al. (2005) Databases and information integration for the *Medicago truncatula* genome and transcriptome. Plant Physiol 138:38-46

Castiglioni P, Pozzi C, Heun M, Terzi V, Muller KJ, Rohde W, Salamini F (1998) An AFLP-based procedure for the efficient mapping of mutations and DNA probes in barley. Genetics 149:2039-2056

Choi HK, Mun JH, Kim DJ, Zhu HY, Baek JM, Mudge J, Roe B, Ellis N, Doyle J, Kiss GB, Young ND,

Cook DR (2004) Estimating genome conservation between crop and model legume species. Proc Natl Acad Sci USA 101:15289-15294

Choo TM (1984) Association between growth habit and persistence in red clover. Euphytica 33:177-185

Christie BR, Martin RA (1999) Selection for persistence in red clover. Can J Plant Sci 79: 357-359

Cogan N, Smith K, Yamada T, Francki M, Vecchies A, Jones E, Spangenberg G, Forster J (2005) QTL analysis and comparative genomics of herbage quality traits in perennial ryegrass (*Lolium perenne* L.). Theor Appl Genet 110:364-380

Cregan PB, Jarvik T, Bush AL, Shoemaker RC, Lark KG, Kahler AL, Kaya N, Vantoai TT, Lohnes DG, Chung L, Specht JE (1999a) An integrated genetic linkage map of the soybean genome. Crop Sci 39:1464-1490

Cregan PB, Mudge J, Fickus EW, Danesh D, Denny R, Young ND (1999b) Two simple sequence repeat markers to select for soybean cyst nematode resistance conditioned by the *rhg1* locus. Theor Appl Genet 99:811-818

Cresswell A, Hamilton NRS, Roy AK, Viegas BMF (2001) Use of amplified fragment length polymorphism markers to assess genetic diversity of *Lolium* species from Portugal. Mol Ecol 10:229-241

Crossa J (1989) Methodologies for estimating the sample size required for genetic conservation of outbreeding crops. Theor Appl Genet 77:153-161

Crouch HK, Crouch JH, Madsen S, Vuylsteke DR, Ortiz R (2000) Comparative analysis of phenotypic and genotypic diversity among plantain landraces (*Musa* spp., AAB group). Theor Appl Genet 101:1056-1065

Cumming BG, Steppler HA (1961) The control of growth and development in red clover (*Trifolium pratense* L.) IV: Vegetative propagation and the use of growth regulators. Can J Plant Sci 836-848

Darvasi A, Weinreb A, Minke V, Weller JI, Soller M (1993) Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. Genetics 134:943-951

Deneufbourg F (2004) Seed yield of forage crop cultivars. Bull Semen 175:20-21

Despres L, Loriot S, Gaudeul M (2002) Geographic pattern of genetic variation in the European globeflower *Trollius europaeus* L. (Ranunculaceae) inferred from amplified fragment length polymorphism markers. Mol Ecol 11:2337-2347

Devos KM, Gale MD (2000) Genome relationships: The grass model in current research. Plant Cell 12:637-646

Echt CS, Kidwell KK, Knapp SJ, Osborn TC, Mccoy TJ (1994) Linkage mapping in diploid alfalfa (*Medicago sativa*). Genome 37:61-71

Edwards MD, Page NJ (1994) Evaluation of marker-assisted selection through computer-simulation.

Theor Appl Genet 88:376-382

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes - application to human mitochondrial-DNA restriction data. Genetics 131:479-491

Fahima T, Roder MS, Wendehake K, Kirzhner VM, Nevo E (2002) Microsatellite polymorphism in natural populations of wild emmer wheat, *Triticum dicoccoides*, in Israel. Theor Appl Genet 104:17-29

Fahima T, Sun GL, Beharav A, Krugman T, Beiles A, Nevo E (1999) RAPD polymorphism of wild emmer wheat populations, *Triticum dicoccoides*, in Israel. Theor Appl Genet 98:434-447

Felsenstein J (2002) PHYLIP (Phylogeny Inference Package) Version 3.6a3. Distributed by the author, Department of Genome Sciences, University of Washington, Seattle

Feltus FA, Wan J, Schulze SR, Estill JC, Jiang N, Paterson AH (2004) An SNP resource for rice genetics and breeding based on subspecies indica and japonica genome alignments. Genome Res 14:1812-1819

Feuillet C, Travella S, Stein N, Albar L, Nublat A, Keller B (2003) Map-based isolation of the leaf rust disease resistance gene *Lr10* from the hexaploid wheat (*Triticum aestivum* L.) genome. Proc Natl Acad Sci USA 100:15253-15258

Fischer M, Matthies D (1998) RAPD variation in relation to population size and plant fitness in the rare *Gentianella germanica* (Gentianaceae). Am J Bot 85:811-819

Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. Annu Rev Plant Biol 54:357-374

Forster BP, Ellis RP, Thomas WTB, Newton AC, Tuberosa R, This D, El-Enein RA, Bahri MH, Ben Salem M (2000) The development and application of molecular markers for abiotic stress tolerance in barley. J Exp Bot 51:19-27

Freyre R, Skroch PW, Geffroy V, Adam-Blondon AF, Shirmohamadali A, Johnson WC, Llaca V, Nodari RO, Pereira PA, Tsai SM, Tohme J, Dron M et al. (1998) Towards an integrated linkage map of common bean. 4. Development of a core linkage map and alignment of RFLP maps. Theor Appl Genet 97:847-856

Gaston KJ (1996) Biodiversity: A biology of numbers and difference. Blackwell Science Ltd, Oxford

Gaudeul M, Taberlet P, Till-Bottraud I (2000) Genetic diversity in an endangered alpine plant, *Eryngium alpinum* L. (Apiaceae), inferred from amplified fragment length polymorphism markers. Mol Ecol 9:1625-1637

Gilbert JE, Lewis RV, Wilkinson MJ, Caligari PDS (1999) Developing an appropriate strategy to assess genetic variability in plant germplasm collections. Theor Appl Genet 98:1125-1131

Greene SL, Gritsenko M, Vandemark G (2004) Relating morphologic and RAPD marker variation to

collection site environment in wild populations of red clover (*Trifolium pratense* L.). Genet Res Crop Evol 51:643-653

Grenier C, Deu M, Kresovich S, Bramel-Cox PJ, Hamon P (2000) Assessment of genetic diversity in three subsets constituted from the ICRISAT sorghum collection using random vs non-random sampling procedures. B. Using molecular markers. Theor Appl Genet 101:197-202

Gustafson DJ, Gibson DJ, Nickrent DL (1999) Random amplified polymorphic DNA variation among remnant big bluestem (*Andropogon gerardii* Vitman) populations from arkansas' grand prairie. Mol Ecol 8:1693-1701

Gustine DL, Voigt PW, Brummer EC, Papadopoulos YA (2002) Genetic variation of RAPD markers for North American white clover collections and cultivars. Crop Sci 42:343-347

Guthridge KM, Dupal MP, Kolliker R, Jones ES, Smith KF, Forster JW (2001) AFLP analysis of genetic diversity within and between populations of perennial ryegrass (*Lolium perenne* L.). Euphytica 122:191-201

Hamrick JL, Linhart YB, Mitton JB (1979) Relationships between life history characteristics and electrophoretically detectable genetic variation in plants. Annu Rev Ecol Syst 10:173-200

Hernandez-Verdugo S, Luna-Reyes R, Oyama K (2001) Genetic structure and differentiation of wild and domesticated populations of *Capsicum annuum* (*Solanaceae*) from Mexico. Plant Syst Evol 226:129-142

Herrmann D, Boller B, Widmer F, Kölliker R (2005a) Genetic and phenotypic diversity of Swiss red clover landraces. In: Humphreys MO (ed) Proceedings of the 4th international symposium on the molecular breeding of forage crops and turf, a satellite workshop of the XXth International Grassland Congress, July 2005, Aberystwyth, Wales. Academic, Wageningen, pp 261

Herrmann D, Boller B, Widmer F, Kölliker R (2005b) Optimization of bulked AFLP analysis and its application for exploring diversity of natural and cultivated populations of red clover. Genome 48:474-486

Herrmann D, Boller B, Studer B, Widmer F, Kölliker R (2006) QTL analysis of seed yield components in red clover (*Trifolium pratense* L.). Theor Appl Genet 112:536-545

Hess HE, Landolt E, Hirzel R (1970) Flora of Switzerland. Band 2: Nymphaceae to Primulaceae. Birkhäuser, Basel

Hill RR, Shenk JS, Barnes RF (1988) Breeding for yield and quality. In: Hanson AA, Barnes DK, Hill RR (eds) Alfalfa and alfalfa improvement. ASA, CSSA and SSSA, Madison, pp 809-825

Hollingsworth PM, Ennos RA (2004) Neighbour joining trees, dominant markers and population genetic structure. Heredity 92:490-498

Huang N, Angeles ER, Domingo J, Magpantay G, Singh S, Zhang G, Kumaravadivel N, Bennett J, Khush GS (1997a) Pyramiding of bacterial blight resistance genes in rice: marker-assisted selection using

RFLP and PCR. Theor Appl Genet 95:313-320

Huang XQ, Hsam SLK, Zeller FJ (1997b) Chromosomal location of genes for resistance to powdery mildew in common wheat (*Triticum aestivum* L. Em. Thell.). 4. Gene *Pm24* in Chinese landrace Chiyacao. Theor Appl Genet 95:950-953

Huff DR, Peakall R, Smouse PE (1993) RAPD variation within and among natural populations of outcrossing buffalograss [*Buchloe dactyloides* (Nutt.) Engelm.]. Theor Appl Genet 86:927-934

Inoue M, Gao ZS, Cai HW (2004) QTL analysis of lodging resistance and related traits in Italian ryegrass (*Lolium multiflorum* Lam.). Theor Appl Genet 109:1576-1585

Isobe S, Klimenko I, Ivashuta S, Gau M, Kozlov NN (2003) First RFLP linkage map of red clover (*Trifolium pratense* L.) based on cDNA probes and its transferability to other red clover germplasm. Theor Appl Genet 108:105-112

Jain SM, Brar DS, Ahloowalia BS (2002) Molecular techniques in crop improvement. Kluwer, Dordrecht

Jansen RC, Stam P (1994) High resolution of quantitative traits into mutiple loci via interval mapping. Genetics 136:1447-1455

Jarvis D, Sthapit B, Sears L (2000) Conserving agricultural biodiversity in situ: A scientific basis for sustainable agriculture. International Plant Genetic Resources Institute, Rome

Jofuku KD, Omidyar PK, Gee Z, Okamuro JK (2005) Control of seed mass and seed yield by the floral homeotic gene APETALA2. Proc Natl Acad Sci USA 102:3117-3122

Jones CJ, Edwards KJ, Castaglione S, Winfield MO, Sala F, Vandewiel C, Bredemeijer G, Vosman B, Matthes M, Daly A, Brettschneider R, Bettini P et al. (1997a) Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. Mol Breed 3:381-390

Jones ES, Dupal MP, Dumsday JL, Hughes LJ, Forster JW (2002a) An SSR-based genetic linkage map for perennial ryegras (*Lolium perenne* L.). Theor Appl Genet 105:577-584

Jones ES, Hughes LJ, Drayton MC, Abberton MT, Michaelson-Yeates TPT, Bowen C, Forster JW (2003) An SSR and AFLP Molecular marker-based genetic map of white clover (*Trifolium repens* L.). Plant Science 165:531-539

Jones ES, Mahoney NL, Hayward MD, Armstead IP, Jones JG, Humphreys MO, King IP, Kishida T, Yamada T, Balfourier F, Charmet G, Forster JW (2002b) An enhanced molecular marker based genetic map of perennial ryegrass (*Lolium perenne*) reveals comparative relationships with other Poaceae genomes. Genome 45:282-295

Jones N, Ougham H, Thomas H (1997b) Markers and mapping - we are all geneticists now. New Phytol 137:165-177

Julier B (1996) Traditional seed maintenance and origins of the French lucerne landraces. Euphytica 92:353-357

Keim P, Schupp JM, Travis SE, Clayton K, Zhu T, Shi LA, Ferreira A, Webb DM (1997) A high-density soybean genetic map based on AFLP markers. Crop Sci 37:537-543

Keller M, Keller B, Schachermayr G, Winzeler M, Schmid JE, Stamp P, Messmer MM (1999) Quantitative trait loci for resistance against powdery mildew in a segregating wheat x spelt population. Theor Appl Genet 98:903-912

Koblet R, Nüesch B (1960) Physiological and genetical studies on Swiss Mattenklee, a persistent type of early-flowering red clover. Proceedings of the 8th International Grassland Congress in England, Great Britain, pp 422-425

Kölliker R, Boller B, Widmer F (2005a) Marker assisted polycross breeding to increase diversity and yield in perennial ryegrass (*Lolium perenne* L.). Euphytica (in press)

Kölliker R, Enkerli J, Widmer F (2005b) Characterisation of novel microsatellite loci for red clover (*Trifolium pratense* L.) from enriched genomic libraries. Mol Ecol Notes DOI 10.1111/j.1471-8286.2005.01133.x

Kölliker R, Herrmann D, Boller B, Widmer F (2003) Swiss Mattenklee landraces, a distinct and diverse genetic resource of red clover (*Trifolium pratense* L.). Theor Appl Genet 107:306-315

Kölliker R, Jones ES, Jahufer MZZ, Forster JW (2001) Bulked AFLP analysis for the assessment of genetic diversity in white clover (*Trifolium repens* L.). Euphytica 121:305-315

Kölliker R, Stadelmann FJ, Reidy B, Nösberger J (1998) Fertilization and defoliation frequency affect genetic diversity of *Festuca pratensis* Huds. in permanent grasslands. Mol Ecol 7:1557-1567

Kongkiatngam P, Waterway MJ, Coulman BE, Fortin MG (1996) Genetic variation among cultivars of red clover (*Trifolium pratense* L.) detected by RAPD markers amplified from bulk genomic DNA. Euphytica 89:355-361

Kongkiatngam P, Waterway MJ, Fortin MG, Coulman BE (1995) Genetic variation within and between two cultivars of red clover (*Trifolium pratense* L.) - comparisons of morphological, isozyme, and RAPD markers. Euphytica 84:237-246

Kouame CN, Quesenberry KH (1993) Cluster analysis of a world collection of red clover germplasm. Genet Res Crop Evol 40:39-47

Kraft T, Sall T (1999) An evaluation of the use of pooled samples in studies of genetic variation. Heredity 82:488-494

Kubik C, Meyer WA, Gaut BS (1999) Assessing the abundance and polymorphism of simple sequence repeats in perennial ryegrass. Crop Sci 39:1136-1141

Kumar LS (1999) DNA markers in plant improvement: An overview. Biotechnol Adv 17:143-182

Lakew B, Semeane Y, Alemayehu F, Gebre H, Grando S, Vanleur JAG, Ceccarelli S (1997) Exploiting the diversity of barley landraces in Ethiopia. Genet Res Crop Evol 44:109-116

Lander ES, Green P, Abrahamson J, Barlow A, Daly ML, Lincoln SE, Newburg L (1987) MAPMAKER: an interactive computer package for constructing primary linkage maps of experimental and natural populations. Genomics 1:174-181

Lawrence MJ (1996) Number of incompatibility alleles in clover and other species. Heredity 76:610-615

Lehmann J, Briner HU (1998) Varieties of red clover and meadow fescue in tests. Agrarforschung 5:177-180

Linder CR, Taha I, Seiler GJ, Snow AA, Rieseberg LH (1998) Long-term introgression of crop genes into wild sunflower populations. Theor Appl Genet 96:339-347

Liu ZH, Anderson JA, Hu J, Friesen TL, Rasmussen JB, Faris JD (2005) A wheat intervarietal genetic linkage map based on microsatellite and target region amplified polymorphism markers and its utility for detecting quantitative trait loci. Theor Appl Genet 111:782-794

Maliepaard C, Jansen J, Van Ooijen JW (1997) Linkage analysis in a full-sib family of an outbreeding plant species: Overview and consequences for applications. Genet Res 70:237-250

Mansur LM, Orf JH, Chase K, Jarvik T, Cregan PB, Lark KG (1996) Genetic mapping of agronomic traits using recombinant inbred lines of soybean. Crop Sci 36:1327-1336

Meikle A, Paterson S, Finch RP, Marshall G, Waterhouse A (1999) Genetic characterization of heather [*Calluna vulgaris* (L.) Hull] subject to different management regimes across Great Britain. Mol Ecol 8:2037-2047

Menges ES (1991) Seed-germination percentage increases with population-size in a fragmented prairie species. Conserv Biol 5:158-164

Mengoni A, Gori A, Bazzicalupo M (2000) Use of RAPD and microsatellite (SSR) variation to assess genetic relationships among populations of tetraploid alfalfa, *Medicago sativa*. Plant Breed 119:311-317

Merkenschlager F (1934) Migration and distribution of red clover in Europe. Herb Rev 2:88-92

Mian MAR, Saha MC, Hopkins AA, Wang ZY (2005) Use of tall fescue EST-SSR markers in phylogenetic analysis of cool-season forage grasses. Genome 48:637-647

Michelmore RW, Paran I, Kesseli RV (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis - a rapid method to detect markers in specific genomic regions by using segregating populations. Proc Natl Acad Sci USA 88:9828-9832

Montardo DP, Dall'agnol M, Crusius AF, Paim NR (2003) Path analysis for seed production in red clover (*Trifolium pratense* L.). Braz J Anim Sci 32:1076-1082

Morgan TH (1911) Random segregation versus coupling in Mendelian inheritance. Science 34:384

Morris JB, Greene SL (2001) Defining a multiple-use germplasm collection for the genus *Trifolium*. Crop Sci 41:893-901

Mosjidis JA, Greene SL, Klingler KA, Afonin A (2004) Isozyme diversity in wild red clover populations from the Caucasus. Crop Sci 44:665-670

Mullis KB, Faloona F (1987) Specific synthesis of DNA in vitro via a polymerase catalyzed chain reaction. Methods Enzymol 155:335-350

Muluvi GM, Sprent JI, Soranzo N, Provan J, Odee D, Folkard G, Mcnicol JW, Powell W (1999) Amplified fragment length polymorphism (AFLP) analysis of genetic variation in *Moringa oleifera* Lam. Mol Ecol 8:463-470

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

Newbury HJ (2003) Plant Molecular Breeding. Blackwell, Oxford

Nilsson-Ehle H (1909) Investigations of crossings of oat and wheat. Lund Univ Arsskrift, ns, series 2 5:1-122

Nüesch B (1976) Investigations on local varieties of Swiss cultivated red clover (*Trifolium pratense* L.). Schweiz Landw Forsch 15:401-410

Nüesch BE (1960) Investigations of red clover populations in order to improve breeding of Mattenklee cultivars. PhD thesis 2992, Swiss Federal Institute of Technology, Zurich

Oliva RN, Steiner JJ, Young WC (1994) Red clover seed production: 2. Plant water status on yield and yield components. Crop Sci 34:184-192

Ouendeba B, Ejeta G, Hanna WW, Kumar AK (1995) Diversity among African pearl-millet landrace populations. Crop Sci 35:919-924

Paterson AH (1998) Molecular dissection of complex traits. CRC Press LLC, Boca Raton

Paterson AH, Lander ES, Hewitt JD, Peterson S, Lincoln SE, Tanksley SD (1988) Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 335:721-726

Pejic I, Ajmone-Marsan P, Morgante M, Kozumplick V, Castiglioni P, Taramino G, Motto M (1998) Comparative analysis of genetic similarity among maize inbred lines detected by RFLPs, RAPDs, SSRs, and AFLPs. Theor Appl Genet 97:1248-1255

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol Breed 2:225-238

Powell W, Thomas WTB, Baird E, Lawrence P, Booth A , Harrower B, McNicol JW, Waugh R (1997) Analysis of quantitative traits in barley by the use of amplified fragment length polymorphisms. Heredity 79:48-59

Prashanth SR, Parani M, Mohanty BP, Talame V, Tuberosa R, Parida A (2002) Genetic diversity in cultivars and landraces of *Oryza sativa* subsp. *indica* as revealed by AFLP markers. Genome

45:451-459

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945-959

Pritchard JK, Wen W (2002) Documentation for structure software: Version 2. http://pritch.bsd.uchicago.edu

Puecher DI, Robredo CG, Rios RD, Rimieri P (2001) Genetic variability measures among *Bromus catharticus* Vahl. populations and cultivars with RAPD and AFLP markers. Euphytica 121:229-236

Pujar S, Tamhankar SA, Rao VS, Gupta VS, Naik S, Ranjekar PK (1999) Arbitrarily primed PCR based diversity assessment reflects hierarchical groupings of Indian tetraploid wheat genotypes. Theor Appl Genet 99:868-876

Rao NK, Reddy LJ, Bramel PJ (2003) Potential of wild species for genetic enhancement of some semi-arid food crops. Genet Res Crop Evol 50:707-721

Reed DH, Frankham R (2003) Correlation between fitness and genetic diversity. Conserv Biol 17:230-237

Reynolds J, Weir BS, Cockerham CC (1983) Estimation of the coancestry coefficient: basis for a short-term genetic distance. Genetics 105:767-779

Rohlf FJ (2000) NTSYS-pc: Numerical Taxonomy and Multivariate Analysis System, Version 2.1. Exter, Setauket

Sato S, Isobe S, Asamizu E, Ohmido N, Nakamura Y, Sakurai N, Okumura K, Klimenko I, Sasamoto S, Wada T, Watanabe A, Fujishiro T et al. Comprehensive structural analysis of the genome of red clover (*Trifolium pratense*). (submitted)

Sawkins MC, Maass BL, Pengelly BC, Newbury HJ, Ford-Lloyd BV, Maxted N, Smith R (2001) Geographical patterns of genetic variation in two species of *Stylosanthes* Sw. using amplified fragment length polymorphism. Mol Ecol 10:1947-1958

Sax K (1923) The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. Genetics 8:552-560

Schauser L, Fredslund J, Heegard Madsen L, Sandal N, Stougaard J (2005) A computational pipeline for the development of comparative anchor tagged sequence (CATS) markers. In: Humphreys MO (ed) Proceedings of the 4th international symposium on the molecular breeding of forage crops and turf, a satellite workshop of the XXth International Grassland Congress, July 2005, Aberystwyth, Wales. Academic, Wageningen, pp 73-81

Schneider S, Roessli D, Excoffier L (2000) Arlequin: a Software for Population Genetics Data Analysis. Genetics and Biometry Laboratory, Department of Anthropology and Ecology, University of Geneva, Geneva

Schön CC, Utz HF, Groh S, Truberg B, Openshaw S, Melchinger AE (2004) Quantitative trait locus

mapping based on resampling in a vast maize testcross experiment and its relevance to quantitative genetics for complex traits. Genetics 167:485-498

Segovia-Lerma A, Cantrell RG, Conway JM, Ray IM (2003) AFLP-based assessment of genetic diversity among nine alfalfa germplasms using bulk DNA templates. Genome 46:51-58

Semagn K, Bjornstad A, Stedje B, Bekele E (2000) Comparison of multivariate methods for the analysis of genetic resources and adaptation in *Phytolacca dodecandra* using RAPD. Theor Appl Genet 101:1145-1154

Semerikov VL, Belyaev AY, Lascoux M (2002) The origin of Russian cultivars of red clover (*Trifolium pratense* L.) and their genetic relationships to wild populations in the Urals. Theor Appl Genet 106:127-132

Shim SI, Jorgensen RB (2000) Genetic structure in cultivated and wild carrots (*Daucus carota* L.) revealed by AFLP analysis. Theor Appl Genet 101:227-233

Skovmand B, Reynolds MP, Delacy IH (2001) Mining wheat germplasm collections for yield enhancing traits. Euphytica 119:25-32

Smith RR, Taylor NL, Bowley SR (1985) Red clover. In: Taylor NL (ed) Clover Science and Technology. The American Society of Agronomy, Madison, pp 457-470

Staub JE, Serquen FC, Gupta M (1996) Genetic markers, map construction, and their application in plant breeding. Hortscience 31:729-741

Steiner JJ, Smith RR, Alderman SC (1997) Red clover seed production: 4. Root rot resistance under forage and seed production systems. Crop Sci 37:1278-1282

Tavoletti S, Veronesi F, Osborn TC (1996) RFLP linkage map of an alfalfa meiotic mutant based on an F1 population. J Hered 87:167-171

Taylor NL, Quesenberry KH (1996) Red Clover Science. Kluwer, Dordrecht

Taylor NL, Stroube WH, Kendall WA, Fergus EN (1962) Variation and relation of clonal persistence and seed production in red clover. Crop Sci 2:303-305

ter Braak CJF, Smilauer P (2002) CANOCO Reference Manual and CanoDraw for Windows User's Guide: Software for Canonical Community Ordination (Version 4.5). Microcomputer Power, Ithaca

Thoday JM (1961) Location of polygenes. Nature 191:368-370

Tilman D, Downing JA (1994) Biodiversity and stability in grasslands. Nature, 367:363-365

Timmermann-Vaughan GM, Mills A, Whitfield C, Frew T, Butler R, Murray S, Lakeman M, McCallum J, Russell A, Wilson D (2005) Linkage mapping of QTL for seed yield, yield components, and developmental traits in pea. Crop Sci 45:1336-1344

Tosti N, Negri V (2002) Efficiency of three PCR-based markers in assessing genetic variation among cowpea (*Vigna unguiculata* subsp. *unguiculata*) landraces. Genome 45:268-275

Tyrka M, Blaszczyk L, Chelkowski J, Lind V, Kramer I, Weilepp M, Wisniewska H, Ordon F (2004) Development of the single nucleotide polymorphism marker of the wheat *Lr1* leaf rust resistance gene. Cell Mol Biol Lett 9:879-889

Ulloa O, Ortega F, Campos H (2003) Analysis of genetic diversity in red clover (*Trifolium pratense* L.) breeding populations as revealed by RAPD genetic markers. Genome 46:529-535

UPOV (2001) Guidelines for the conduct of tests for distinctness, uniformity and stability: Red clover (*Trifoilum pratense* L.). International Union for the Protection of New Varieties of Plants, Geneva

Utz HF, Melchinger AE (2003) PLABQTL, version 1.2, a computer program to map QTL. Institute of Plant Breeding, Seed Science and Population Genetics, University of Hohenheim, Stuttgart

Van Ooijen JW (1992) Accuracy of mapping quantitative trait loci in autogamous species. Theor Appl Genet 84:803-811

Van Ooijen JW (2004) MapQTL 5, software for the mapping of quantitative trait loci in experimental populations. Kyazma B.V., Wageningen

Van Ooijen JW, Voorrips RE (2001) JoinMap 3.0, software for the calculation of genetic linkage maps. Kyazma B.V, Wageningen

Visscher PM, Goddard ME (2004) Prediction of the confidence interval of quantitative trait loci location. Behav Genet 34:477-482

Volkart A (1929) The Swiss forage production. 4. Fortbildungslehrgang des Deutschen Grünlandbundes e.V. in Sachsen, Germany, pp 1-20

Vos P, Hogers R, Bleeker M, Reijans M, Vandelee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP- a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407-4414

Vuylsteke M, Mank R, Antonise R, Bastiaans E, Senior ML, Stuber CW, Melchinger AE, Lübberstedt T, Xia XC, Stam P, Zabeau M, Kuiper M (1999) Two high-density AFLP linkage maps of *Zea mays* L.: Analysis of distribution of AFLP markers. Theor Appl Genet 99:921-935

Warnke SE, Barker RE, Jung G, Sim SC, Mian MAR, Saha MC, Brilman LA, Dupal MP, Forster JW (2004) Genetic linkage mapping of an annual x perennial ryegrass population. Theor Appl Genet 109:294-304

White RP, Murray S, Rohweider M (2000) Pilot analysis of global ecosystems: Grassland Ecosystems. World Resources Institute, Washington DC

Whitton J, Wolf DE, Arias DM, Snow AA, Rieseberg LH (1997) The persistence of cultivar alleles in wild populations of sunflowers five generations after hybridization. Theor Appl Genet 95:33-40

Wiersma DW, Smith RR, Mlynarek MJ, Rand RE, Sharpee DK, Undersander DJ (1998) Harvest management effects on red clover forage yield, quality, and persistence. J Prod Agric 11:309-313

Williams JGK, Reiter RS, Young RM, Scolnik PA (1993) Genetic mapping of mutations using

phenotypic pools and mapped RAPD markers. Nucleic Acids Res 21:2697-2702

Woodfield DR, Caradus JR (1994) Genetic improvement in white clover representing six decades of plant breeding. Crop Sci 34:1205-1213

Wricke G, Weber WE (1986) Quantitative genetics and selection in plant breeding. Walter de Gruyter, Berlin

Xu S (1995) A comment on the simple regression method for interval mapping. Genetics 141:1657-1659

Yamada T, Jones ES, Cogan NOI, Vecchies AC, Nomura T, Hisano H, Shimamoto Y, Smith KF, Hayward MD, Forster JW (2004) QTL analysis of morphological, developmental, and winter hardiness-associated traits in perennial ryegrass. Crop Sci 44:925-935

Yang GP, Maroof MAS, Xu CG, Zhang QF, Biyashev RM (1994) Comparative analysis of microsatellite DNA polymorphism in landraces and cultivars of rice. Mol Gen Genet 245:187-194

Young ND (1996) QTL mapping and quantitative disease resistance in plants. Annu Rev Phytopathol 34:479-501

Young ND (1999a) A cautiously optimistic vision for marker-assisted breeding. Mol Breed 5:505-510

Young WP, Schupp JM, Keim P (1999b) DNA methylation and AFLP marker distribution in the soybean genome. Theor Appl Genet 99:785-792

Yousef GG, Juvik JA (2001) Comparison of phenotypic and marker-assisted selection for quantitative traits in sweet corn. Crop Sci 41:645-655

Yu J, Mosjidis JA, Klingler KA, Woods FM (2001) Isozyme diversity in North American cultivated red clover. Crop Sci 41:1625-1628

Yu KF, Pauls KP (1993) Rapid estimation of genetic relatedness among heterogeneous populations of alfalfa by random amplification of bulked genomic DNA samples. Theor Appl Genet 86:788-794

Zeng ZB (1994) Precision mapping of quantitative trait loci. Genetics 136:1457-1468

Zeven AC (1998) Landraces: A review of definitions and classifications. Euphytica 104:127-139

Zeven AC (2000) Traditional maintenance breeding of landraces: 1. Data by crop. Euphytica 116:65-85

Zhu J, Gale MD, Quarrie S, Jackson MT, Bryan GJ (1998) AFLP markers for the study of rice biodiversity. Theor Appl Genet 96:602-611

Zhu YL, Song QJ, Hyten DL, Van Tassell CP, Matukumalli LK, Grimm DR, Hyatt SM, Fickus EW, Young ND, Cregan PB (2003) Single-nucleotide polymorphisms in soybean. Genetics 163:1123-1134

Zohary M, Heller D (1984) The Genus Trifolium. The Israel Academy of Sciences and Humanities, Jerusalem

8 Acknowledgements

I would like to thank…

Dr. Roland Kölliker for his excellent supervision, profound expertise, endless help and patience.

Prof. Dr. Beat Keller for taking the responsibility as the supervisor of the University of Zurich and for the critical review of my thesis.

Dr. Beat Boller for assistance in field trials, statistical questions and revisions of manuscripts.

Dr. Franco Widmer for giving me the opportunity to carry out this thesis in his research group and for suggestions to improve my talks and manuscripts.

Prof. Dr. Ueli Grossniklaus and Dr. Michael Abberton, IGER, Aberystwyth, UK, for their expertise and critical comments on my thesis.

The whole Molecular Ecology group for scientific and non-scientific support, particularly Yvonne Häfele and Jaqueline Bernet for technical assistance in the laboratory and Bruno Studer for extensive 'QTL analysis' discussions.

The group of fodder crop breeding for the many hours spent in the field and threshing red clover seeds, especially Simone Günter, Philipp Streckeisen and Peter Tanner.

My parents and my friends for supporting me during my whole thesis and encouraging me not to give up and to stay the course.

The Breeding Foundation DSP-BLW for founding this thesis and the Swiss Federal Office for Agriculture (NCP 13) for support of the regeneration of Mattenklee landraces.

9 Curriculum vitae

Name:	HERRMANN Doris
Date of birth:	11th March 1975
Place of birth:	Langnau BE
Marital status:	unmarried

EDUCATION:

2002 – 2005	PhD studies at Agroscope FAL Reckenholz, Molecular Ecology, Zurich in cooperation with the University of Zurich, Institute of Plant Biology, Department of Plant Molecular Biology
2001	Trainee at Agroscope FAL Reckenholz, Seed Ecology, Zurich Teaching certificate in agricultural sciences at the Swiss Federal Institute of Technology (ETH) Zurich
2000	Diploma in agriculture (Major: Biotechnology of agriculture) at the ETH Zurich: Dipl. Ing. Agr. ETH
1995 - 2000	Studies at the Department of Agricultural Sciences, ETH Zurich
1991 – 1995	High school in Burgdorf BE: Matura (Type C)
1982 – 1991	Primary school in Buttenried BE and Koppigen BE

PUBLICATIONS:

D. Herrmann, B. Boller, B. Studer, F. Widmer, R. Kölliker (2006) QTL analysis of seed yield components in red clover (*Trifolium pratense* L.). *Theoretical and Applied Genetics* 112:536-545

D. Herrmann, B. Boller, F. Widmer, R. Kölliker (2005) Optimization of bulked AFLP analysis and its application for exploring diversity of natural and cultivated populations of red clover. *Genome* 48: 474-486

B. Boller, F.X. Schubiger, P. Tanner, P. Streckeisen, **D. Herrmann**, R. Kölliker (2005) La diversité génétique dans les prairies naturelles suisses et son utilisation en sélection. *Fourrages* 182: 263-274

R. Kölliker, **D. Herrmann**, B. Boller, F. Widmer (2003) Swiss Mattenklee landraces, a distinct and diverse genetic resource of red clover (*Trifolium pratense* L.). *Theoretical and Applied Genetics* 107: 306-315

D. Herrmann, B. Boller, F. Widmer, R. Kölliker (2003) Genetic characterisation and potential origin of Swiss red clover landraces (*Trifolium pratense* L.). Proceedings of 25th EUCARPIA Fodder Crops and Amenity Grasses Section Meeting. *Special Issue: Czech Journal of Genetics and Plant Breeding* 39: 120-124

R. Kölliker, B. Boller, Y. Häfele, S. Günter, **D. Herrmann**, F. Widmer (2003) Genetische Diversität im Schweizer Mattenklee. *Agrarforschung* 10: 361-368

ORAL PRESENTATIONS:

D. Herrmann, B. Boller, F. Widmer, R. Kölliker (2004) Diversity and potential ancestry of red clover populations. Annual Symposium of the PhD Program in Agroecology (ASPPA), Zurich

D. Herrmann, B. Boller, F. Widmer, R. Kölliker (2003) Genetic characterisation and potential origin of Swiss red clover landraces (*Trifolium pratense* L.). 25th EUCARPIA Fodder Crops and Amenity Grasses Section Meeting, Brno, CZ

POSTERS:

D. Herrmann, B. Boller, B. Studer, F. Widmer, R. Kölliker (2006) Identification of QTLs for seed yield and its components in red clover (*Trifolium pratense* L.). Plant and Animal Genome XIV Conference, San Diego, USA

D. Herrmann, B. Boller, F. Widmer, R. Kölliker (2005) Genetic and phenotypic diversity of Swiss red clover landraces. Satellite workshop of the XX International Grassland Congress, Aberystwyth, UK

D. Herrmann, B. Boller, B. Studer, F. Widmer, B. Keller, R. Kölliker (2005) Molecular dissection of seed yield and persistence in red clover. Annual Symposium of the PhD Program in Sustainable Agriculture (ASPSA), Zurich

D. Herrmann, B. Boller, F. Widmer, R. Kölliker (2004) Mattenklee - Swiss red clover of foreign origin? Symposium Welternährungstag 2004, Biodiversität im Dienste der Ernährungssicherheit, Zurich

D. Herrmann, B. Boller, F. Widmer, R. Kölliker (2004) Mattenklee - Swiss red clover of foreign origin? 12. Jahrestagung der Schweizerischen Gesellschaft für Pflanzenbauwissenschaften (SGPW), Neuenburg