Genetic analysis of the arable weeds *Arabidopsis thaliana* and *Viola arvensis* from agricultural fields with different use-histories

Robert O'Neill

Contents

List of Figures

List of Tables

List of Terms and Abbreviations

IBD – Identical-By-Descent

IIS – Identical-In-State

Homoplasy - alleles that are identical-in-state, not identical-by-descent

SMM – Stepwise Mutation Model

IAM – Infinite Allele Model

$F_{ST} = H_t - H_s / H_t$
where H_s represents the expected level of heterozygosity in a subpopulation, and H_t the expected level of heterozygosity if all subpopulations were pooled together (i.e. in the total population).

d-F_{ST} - The average number of different alleles between individuals within a population

$R_{ST} = S - S_w / S$
where S_w and S are the average sum of squares of the difference in allele size within a subpopulation and for the entire population, respectively

d-R_{ST} - the average squared difference in allele size (measured as microsatellite product length) between individuals within a population

Singletons – a unique AFLP band (occurring only once throughout all samples)

M - the ratio of number of alleles (*k*) to range in allele size (*r*) for any given population

m – absolute number of migrants

θ - equal to $4N\mu$, where N is the population size, and μ is the mutation rate

θ_{Hom} – an estimation of θ based on expected homozygosity values

1 Introduction

During the last half century there has been a marked intensification in agriculture across much of Europe (Stoate et al. 2001). Simultaneously, a decline in flora and fauna associated with the farmland habitat has been recorded (Donald et al. 2001) to the extent that some arable weed populations have markedly declined or even become extinct (Marshall et al. 2003). Evidence points to a link between arable weed depletion on agricultural land and the decline in abundance of life forms associated with them, such as farmland birds (Chamberlain et al. 2000).

Intensification has resulted in the fragmentation of many populations, creating barriers to gene flow and splitting large populations into many smaller ones. Smaller, isolated populations are particularly susceptible to genetic drift, a process which, through either the fixation or loss of alleles at any given locus, reduces their within-population genetic diversity. The chance of inbreeding also increases as population size decreases, resulting in a decrease of heterozygosity. As homozygotes are considered in general to show a reduced fitness in comparison to heterozygotes (Parsons 1996), such intensification could possibly reduce the evolutionary potential of arable weed populations.

The reduction in abundance of primary producers in the food web can have only a negative impact on the populations of associated plants and animals. In order to prevent this loss of biodiversity it is necessary to understand the population genetic structure of arable weeds, so that measures can be undertaken to ensure that populations maintain a genetic diversity which will allow them to survive on intensively managed land.

The objective of this study is to address this issue. Through genetic analysis of arable weed populations from intensive and extensive agricultural fields, it is hoped that the underlying trends affecting the genetic structure of these populations can be elucidated. Only through understanding these processes can mitigating measures be undertaken to protect the biodiversity surrounding Europe’s farmlands.

2 Literature Review

2.1 Agricultural Intensification

During the last half of the 20th Century, the European agricultural landscape has undergone a period of agricultural intensification (Stoate et al. 2001; Environmental Signals 2002). Many factors have contributed to this intensification. Not only have previously unused wetlands, grasslands and forested areas been converted into arable pastures (Stra-Co 2002), but the advances in technology have led to the widespread application of herbicides, pesticides and inorganic fertiliser. There has also been a move towards specialisation, with either purely crop or livestock farms replacing mixed farms, a strategy which allows for higher intensity of land use (Signals 2002).

Such intensive agricultural practices have resulted in large scale changes to the environment, including a notable loss of non-crop habitats (Stoate et al. 2001). Natural habitats have been either reduced in size or fragmented, water bodies have become polluted, and the biodiversity of Europe's farmlands has declined sharply (Donald et al. 2001).

Europe's Common Agricultural Policy (CAP) has been criticised by the World Wildlife Fund (WWF) for focussing too strongly on productivity at a cost to biodiversity. However, due to growing environmental concerns approximately 10-15% of the European Union (EU) budget is now being made available for rural development programmes (Stra-Co 2002). The objective of these programmes is to conserve, and reverse the decline in, biodiversity by encouraging extensive and organic agricultural practices. Indeed, during the past five to ten years there has been a five- to ten-fold increase in organic agriculture in most western European countries, and it has been estimated that organic farming will account for 5-10% of total farming in the EU by 2005 (Signals 2002). Such developments should help preserve current biodiversity and allow the opportunity for an increase in biodiversity. The major aim of conservation biology is the preservation of biological resources at all orders of complexity, i.e. ranging from genes to ecosystems, in order to preserve or restore the evolutionary potential of

species (Parsons 1996). Studies by Wilson et al. (1999) and Hyvönen et al. (2003) have found evidence that a reversal of intensification and organic cropping systems can promote biodiversity recovery.

2.2 Biodiversity

The term 'biodiversity' can convey a variety of meanings depending on the focus. The most common description is simply the "variety of life" (Huston 1994; Gaston 1996; Waldman and Shevah 2000). As biological diversity can be assessed on a number of levels, an objective measurement for biodiversity is difficult to calculate. Often biodiversity is measured as the number of species present in any given environment. Despite the usefulness of this definition in many situations, it unfortunately neglects the extent of the genetic diversity present in the populations of the entities in question. The incorporation of genetic diversity into the meaning of biodiversity is, of course, only possible when the detection of any underlying genetic diversity is possible.

Temporal and spatial factors also play a part in determining biodiversity. The vast range of literature devoted to biodiversity reveals a division of opinion as to whether the concept represents simply a 'snapshot' of the variety of biological entities at any one time, or if it also encompasses the processes and complexes formed as a result of the interaction between landscapes and living organisms, and among the living organisms themselves (Gaston 1996). Indeed, it could be viewed that the number of species in an environment, the genetic structure of these species, and the existing temporal trends that effect these species in the future, are a direct reflection of the landscape and climate of that environment, thus forming an inseparable relationship.

2.3 Arable Weeds and Their Role in Biodiversity

Since the advent of agriculture more than 10 000 years ago (Mannion 1997; Mannion 1999), arable weeds have primarily been viewed as a major source of yield loss in crops (Kudsk and Streibig 2003; Marshall et al. 2003).

Given the importance of crop yield to civilisations, considerable effort has understandably been invested into weed control. Until the mid-19th century, weeds were controlled using a combination of manual weeding, crop rotation, appropriate tillage and fallow systems (Kudsk and Streibig 2003). However, the technological advances made during the Second World War gave rise to the widespread introduction of herbicides throughout Europe[1] in the 1940's (Kudsk and Streibig 2003). The increase in herbicide use during the following decades led to a marked decline in the abundance of most arable weed species (Erviö and Salonen 1987; Albrecht 1995; Andreasen et al. 1996; Lutman et al. 2002), and even led to the extinction of some (Marshall et al. 2003). However, this decline has also been attributed to the increasing application of nitrogen fertilisers (combined with improved fertilising methods), and the increased use of crop monocultures (Hyvönen 2003).

Such a loss of arable weeds from crop fields initially appears to be agriculturally beneficial, however it was only later that the implications of a reduction in weed abundance became apparent. Arable weeds, like all plants, are essential to terrestrial ecosystems, being the primary producers upon which food chains are built (Marshall et al. 2003). Not only do they provide food sources for many species (e.g. the browsing of leaves and stems, use of pollen and nectar), but they also provide habitats for reproduction and protection. A study of the Phytophagous Insect Data Base (PIDB) by Marshall et al. (2003) revealed the importance of many weed species to phytophagous insects.

Therefore, reducing the abundance of arable weeds in crop fields can lead to a subsequent reduction in the available habitat for many species further

[1] http://www.ecochem.com/t_organic_fert.html

along the food chain, having negative implications for biodiversity in these regions. This has been seen in the marked decline in the populations of farmland birds in Europe over the last quarter of a century, a decline which has been more significant in countries employing a higher proportion of intensive agricultural practices (Chamberlain et al. 2000; Donald et al. 2001; Stoate et al. 2001). A variety of mechanisms have been proposed to have contributed to the decline in farmland bird populations. The main suggested causes are bird mortality due to farming operations (Green 1995), a reduction in suitable nesting habitat (Chamberlain and Crick 1999), and diminishing food supplies (Evans et al. 1997; Brickle et al. 2000).

The loss of hedgerows and uncultivated field margins not only reduces potential breeding grounds for birds, but simultaneously removes feeding habitats (Wilson et al. 1999). The main sources of nourishment for farmland birds are invertebrates and plant seeds. Changes in the abundance of arable weeds therefore have a potentially large effect on the whole food web of the farmland community.

2.4 Genetic Diversity and Biodiversity

Biodiversity and genetic diversity share a close, albeit complex, relationship. Consider the following two scenarios: a) an environment with a large number of species, the populations of which all show low genetic diversity, and b) an environment containing relatively few species, but whose populations exhibit a large genetic diversity. It is difficult to say which scenario has a greater overall biodiversity. In this situation the temporal aspect of biodiversity becomes important. Despite the abundance of species in the first scenario, the lack of genetic diversity within species' populations may render them particularly susceptible to environmental changes, which, in the future, could lead to population reduction or even extinction. It is generally accepted that populations containing a relatively high genetic diversity have a higher vitality and the ability to more successfully re-

spond to environmental change (Mackill et al. 1999; Cruzan 2001). Therefore, the higher genetic diversity present in populations in the second scenario could ensure the survival of these species despite environmental stresses that may lead to the extinction of species in the first scenario. Therefore in assessing biodiversity, such as in the flora and fauna connected with Europe's farmland, genetic diversity should be taken into account if possible.

2.5 Population Genetics

Population genetics is the study of how genetic principles (such as Mendel's laws) apply to entire populations (Hartl and Clark 1997). The populations of most, if not all, species exhibit some form of genetic structuring (Balloux and Lugon-Moulin 2002a). A population showing no genetic structure would imply that there are no barriers to gene flow between its individuals. However, such an environment is highly unlikely to exist naturally due to the effect of environmental barriers, historical processes and life histories (e.g. mating systems), which may all contribute to shaping the genetic structure of a population through imposing barriers to gene flow. Even if such barriers do not exist, a species' geographical distribution is usually greater than an individual's dispersal capacity, leading to genetic isolation by distance (Balloux and Lugon-Moulin 2002a). According to the Hardy-Weinberg principle, a high gene flow between populations will homogenise allele frequencies between them, preventing local adaptation and, ultimately, speciation (Barton and Hewitt 1985). A population receiving no migrants (i.e. no gene flow) must rely on mutations to introduce novel alleles into the population. These factors oppose the effects of genetic drift – a process which leads to the fixation or loss of alleles in local populations. Small populations are particularly subject to the effects of genetic drift, and their evolutionary potential can be affected by the fixation of deleterious mutations (Wright 1977).

By studying the degree of genetic structuring in populations, inferences can be made about the current state of populations and of the evolution-

ary forces which have shaped them, allowing estimates of the evolutionary potential of populations. This is of particular interest for the conservation of biodiversity. The abiotic factors which have been forced upon many European farmland fauna and flora over the last half century due to agricultural intensification have resulted in habitat fragmentation and more extreme environments. Under such conditions gene flow between local populations may be reduced or even wholly prevented, giving rise to a number of smaller populations instead of one metapopulation. Each of the smaller populations is likely to posses less genetic diversity than that of the metapopulation, making them more susceptible to genetic drift.

Populations of smaller size are also more prone to the effects of inbreeding, the mating of related individuals. The primary effect of inbreeding is a departure from Hardy-Weinberg genotype frequencies, resulting in an excess of homozygotes compared to what would be expected under panmictic conditions. Inbreeding alone does not alter the allele frequencies within a population, but results in changes to the genotype frequencies. It has been well documented that fitness is reduced when inbreeding increases (Parsons 1996); this phenomenon is known as *inbreeding depression*.

Two main hypotheses have been put forward to explain this effect. The *partial dominance theory* explains this reduction in fitness as the increase in frequency of individuals homozygous for deleterious alleles (known as the genetic load). Populations can, over time, purge their genetic load through a combination of successive inbreeding together with selection against those with deleterious alleles (Oostermeijer 1996). In this situation genetic diversity is continually lost but with no reduction in fitness. In the *overdominance hypothesis*, it is considered that a heterozygote exhibits higher fitness than both of the homozygotes, and thus if the frequency of homozygotes increases, a reduction in fitness occurs.

Inbreeding depression is closely related to the opposite phenomenon of *heterosis*, the increase in fitness of vigour resulting from outcrossing among genetically different individuals.

2.6 Detecting Genetic Diversity

Before the advent of molecular genetic techniques, the only method by which genetic information about individuals (and therefore populations) could be elucidated was through the observation of phenotypic data (morphological markers). Differences in morphological characteristics between individuals are a visible manifestation of the differences in their genome. However, when one compares the possible number of phenotypes which can be effectively recorded to the estimated 25 000 genes in *Arabidopsis thaliana* (Alonso-Blanco and Koornneef 2000), it becomes apparent that the amount of information that can be gleaned from morphological marker analysis is limited. Added to this is the disadvantage that, when using morphological markers, the non-coding sequences of the genome (which can often account for more than 95% of the genome of higher plants) are immediately excluded from analysis. However, mutations in the non-coding sequences of the genome are assumed to confer little, or no, selective advantage or disadvantage, thus their inheritance can generally be assumed to conform to Mendelian laws.

The introduction of allozyme electrophoresis by Lewontin and Hubby (1996) and the development of the Polymerase Chain Reaction (PCR) in the mid-1980's (Saiki et al. 1985) paved the way for many advances in molecular technology over the past decade (Holsinger et al. 2002) which have enabled access to a vast array of molecular markers.

Molecular markers reveal mostly neutral sites of variation at the DNA sequence level (Jones et al. 1997b). The term 'neutral' stems from the fact that, unlike morphological markers, these variations are not revealed in the phenotype.

2.6.1 Measures of Genetic Diversity

There are a number of ways through which genetic diversity can be expressed statistically. The possibilities are, however, limited by the marker system employed. Co-dominant marker systems (e.g. microsatellites, isozymes, RFLPs) allow the detection of individual alleles at each locus,

thus heterozygotes and homozygotes can be distinguished. From this information the number of loci that are polymorphic (*P*) and the number of alleles at each locus (*NA*) can be calculated, which themselves are an indication of genetic diversity.

Potentially more informative measurements can be derived from the observed allele frequencies, which enable comparisons between observed and expected heterozygosity. Observed heterozygosity is simply the proportion of individual loci that exhibit two different alleles (instead of the two identical alleles seen in homozygotes). Expected heterozygosity, also referred to as 'gene diversity' (Nei 1987), is defined as the amount of heterozygosity that would be present in a population, given the observed allele frequencies, under Hardy-Weinberg Equilibrium (HWE) (i.e. a large, randomly mating population with negligible selection, mutation and migration). Significant deviation from HWE suggests that there are barriers to gene flow (e.g. geographical barriers, non-random mating) operating within the population.

Dominant marker systems (e.g. AFLPs, RAPDs) cannot distinguish between homozygotes and heterozygotes and therefore cannot be used to calculate allele frequencies. Instead, information from dominant loci are recorded as presence or absence of a band, and a pairwise comparison for all bands across all individuals can be performed to produce an average Eucledian distance between samples within and between populations.

While a higher polymorphism and heterozygosity are considered a reflection of the genetic capacity to resist or adapt to change (either for individuals or populations), another measurement which can also be used as an indicator for individuals or populations to resist environmental change is the effective population size (N_e). Initially introduced by Wright (1931) to link real populations to the theory developed for ideal populations, N_e is defined as "the size of an ideal population whose genetic composition is influenced by random processes in the same way as a real population of size *N*" (Nunney 2002). An increase in effective population size represents the ability of the population to retain neutral and nearly

neutral genetic variation. This is important for evolutionary potential, as much neutral or nearly neutral variation may become important in response to environmental change (Lynch 1996).

A similar concept is also derived from allele frequencies. The effective number of alleles in a population (A_e) can be considered as "the number of equally frequent alleles in an ideal population that would be required to produce the same homozygosity as in an actual population" (Hartl and Clark 1997). The interpretation of A_e values is much the same as for N_e: The more alleles present in a population, the greater the adaptive ability of the population to environmental changes.

In general, it can be assumed that reductions in P and A_e are a result of genetic drift, while a reduction in heterozygosity is generally caused by inbreeding.

2.7 Microsatellites (Simple Sequence Repeats – SSRs)

Microsatellites, also known as simple sequence repeats (SSR), are a class of variable number tandem repeats (VNTRs) comprising of a short (1-6 bp) nucleotide motif repeated up to 100 times or more. They are found in eukaryotic nuclear genomes and in the chloroplast genome of some plants (Powell et al. 1995). Due to their high mutation rate, which has been estimated at 10^{-2} (Dallas 1992), 10^{-4} (Henderson and Petes 1992) or 10^{-6} (Udupa and Baum 2001), they are of great interest to population geneticists. Such a high rate of mutation results in highly polymorphic markers, allowing genetic discrimination of very closely related individuals with only a relatively low number of markers. Because of the ultimate restriction in the length of microsatellites (and therefore the number of possible alleles) they are not so informative for distantly related taxa (Goldstein et al. 1995), but for studies of genetic variation within and among populations of the same species microsatellite markers are an extremely valuable tool. Even highly inbred plant species are often found to obtain sufficient allelic variation at microsatellite loci to enable the elucidation of phylogenetic relationships.

The co-dominant nature of microsatellite markers allows the distinction between heterozygotes from homozygotes (homozygotes display a single band, heterozygotes two bands in diploid organisms), and their dimeric phenotype makes the scoring of genotypes and calculation of allele frequencies extremely simple. Moreover, the high mutation rate at microsatellite loci gives a high number of alleles per locus. For genetic diversity assessment at the subpopulation level this is a significant advantage, particularly in species with a high degree of inbreeding and in small or fractured populations.

A large number of studies have demonstrated the utility of microsatellite variation for investigation and analysis of population structure and genetic diversity in various plant species. Diverse selected examples from the recent literature include evaluations of diversity in endemic and introduced soybean (Brown-Guedira et al. 2000; Narvel et al. 2000), variation in apple hybrids (Hokanson 2001), a study of clonal diversity and population substructure in the marine flowering plant *Zostera marina* (Reusch et al. 2000), and studies of natural variation in rice (Akagi et al. 1996; Panaud et al. 1996), barley (Saghai Maroof et al. 1994), and many on the model plant *A. thaliana* (Todokoro et al. 1995; Innan et al. 1997; Virk et al. 1999; Loridon et al. 1998).

2.7.1 Microsatellite Models of Evolution

The molecular mechanisms through which new alleles are produced in microsatellites are not fully understood (Innan et al. 1997). Several mechanisms have been proposed: replication slippage, sister chromatid exchange, unequal crossing over, and gene conversion (Drake et al. 1983). From these, it appears that replication slippage plays a major role in the formation of new alleles (Wolff et al. 1991). A variety of models have been proposed in an attempt to explain the mutation behaviour of microsatellites. Understanding these models is important for the development of statistical methods which accurately reflect the underlying genetic structure of populations. Two mutation models, initially designed to explain allozyme polymorphism, have been applied to microsatellites: the infinite alleles model (IAM; Kimura and Crow 1964), and the Stepwise Mutation Model (SMM; Kimura and Otha 1978). Under the IAM, each mutation event gives rise to a novel allele. This model does not allow for homoplasy (alleles that are identical-in-state; IIS, as opposed to identical-by-descent; IBD), and therefore any two identical alleles in the population are assumed to be IBD. A variation of the IAM is the *K*-allele model (KAM; Crow and Kimura 1970). It is in principle the same as the IAM except that the total number of possible allelic states is limited to *K*, and thus allows for homoplasy. The IAM is a special case of the KAM with $K = \infty$ (i.e. no homoplasy). In both these models, the coalescence time for two different alleles is considered equal irrespective of the difference in length of the microsatellite repeat number.

However, with microsatellite loci there is evidence that the size of a new mutant allele is closely dependant on the size of the allele that mutated (Slatkin 1995). Weber and Wong (1993) showed that almost all microsatellite mutations in human families differ from their ancestor by one or two repeat units. The SMM simulates an extreme (if the most common) case of microsatellite mutation where each new mutation creates a novel allele by addition or deletion of one repeat unit with equal probability. The SMM therefore incorporates a memory of allele size as alleles of similar

size will be considered closer from an evolutionary point of view than two alleles of vastly different size.

A hybrid model incorporating aspects of the models above has been developed by Di Rienzo (Di Rienzo et al. 1994), the Two-Phase Model (TPM). This is in essence the same as the SMM, but also takes into account the addition or deletion of several units. The probability of a mutation which increases or decreases the allele size by one repeat occurs with probability p, while an increase or decrease in allele size of k repeats has the probability (1-p), with k following a specified probability distribution.

None of the above models, however, appear to perfectly explain the mutation behaviour of all microsatellite loci (Balloux and Lugon-Moulin 2002a). This has resulted in a number of statistical methods being developed to interpret allele frequency results from studies using microsatellites.

2.7.2 Population Genetics Statistics for Microsatellites

In 1921 Sewell Wright developed a fixation index to account for the effect of inbreeding within a population, defined as:

$$F = 1 - H_{obs} / H_{exp}$$

Where H_{obs} is the observed heterozygosity within a population and H_{exp} is the heterozygosity that would be expected under HWE (Wright 1921). The inbreeding coefficient F is the probability that the two alleles of a gene in an individual are identical by descent. As a population moves away from panmixia, the possibility of non-random mating increases, and the difference between H_{obs} and H_{exp} will become greater. When H_{obs} becomes zero, i.e. complete homozygosity, F is equal to one. Conversely, when H_{obs} equals that what is expected (such as for a population in HWE), F becomes zero. Therefore, F is a measure of the non-random association of alleles in an individual.

This concept was further expanded (Wright 1951) so that it could be applied to a population subdivided into a set of subpopulations. This led to the development of the commonly-used traditional hierarchical F-statistics, F_{IS}, F_{ST} and F_{IT} (where I stands for individuals, S for subpopula-

tions and *T* for the total population). The most commonly applied of these statistics is F_{ST} (comparing subpopulations to the total population), which can be defined as the correlation between two alleles chosen at random within subpopulations relative to alleles sampled at random from the total population (Wright 1965). Wright developed his *F*-statistics for a scenario of one locus with two alleles. A formulation of his work for multiallelic data was presented by Nei (1977) and is calculated as:

$$F_{ST} = H_t - H_s / H_t$$

where H_s represents the expected level of heterozygosity in a subpopulation, and H_t the expected level of heterozygosity if all subpopulations were pooled together (i.e. in the total population). Given two subpopulations and a two-allele locus, this value reaches zero when the allele frequencies in the subpopulations are identical. However, as the subpopulations diverge this value moves towards one, which occurs when both are completely homozygous and fixed for the alternative allele. Therefore, F_{ST} measures the heterozygote deficiency (or the inbreeding effects) due to population subdivision relative to its expectation under Hardy-Weinberg equilibrium (Hartl and Clark 1997), and thus can be used as a guideline as to how genetically distant two populations are from each other.

In an ideal population (island model of migration; Wright 1931), and assuming mutation follows the IAM, F_{ST} can also be described as a function of $N(m + \mu)$, the product of local population size (N) and the sum of migration (m) and mutation (μ) (Hartl and Clark 1997). If mutation is negligible, F_{ST} becomes simply a measurement of migration. However, given the rate of microsatellite mutation and the fact that the majority of mutations do not follow the IAM but the SMM (i.e. can generate homoplasy), the relationship between F_{ST} and the number of migrants plus mutations does not always hold.

This led Slatkin (Slatkin 1995) to develop *R*-statistics, analogues of Wrights *F*-statistics developed specifically for microsatellite data through measuring the average sum of squares of the difference in allele size.

From coalescent theory it can be shown that a relationship exists between F_{ST} and average coalescence times:

$$F_{ST} = t - t_o / t$$

where t is the average time to coalescence of any two randomly chosen alleles from the entire population, and t_o is the average time to coalescence of any two randomly chosen alleles drawn from the same subpopulation (Slatkin 1991). Thus the average coalescence time for two alleles drawn randomly from a subpopulation will be less than the average coalescence time for any two alleles drawn from the entire population. Under the SMM with constant mutation rate, the difference in size between any two microsatellite alleles is therefore proportional to the time, in generations, since they shared a common ancestor. Based on this theory, Slatkin (Slatkin 1995) developed an estimation of F_{ST} which incorporates the mutational history contained within the alleles under the SMM, denoted R_{ST}:

$$R_{ST} = S - S_w / S$$

where S_w and S are the average sum of squares of the difference in allele size within a subpopulation and for the entire population, respectively.

Under a strict SMM, R_{ST} is independent of the mutation rate, and estimates are, on average, expected to reveal population differentiation more accurately than *F*-statistics.

For microsatellites (as opposed to allozyme data) the main drawback when using *F*-statistics is their sensitivity to the mutation rate when migration is low (Balloux and Lugon-Moulin 2002a).

While the range of variation in microsatellites alleles is large, it is ultimately restricted (Bowcock et al. 1994). Therefore, after a sufficient time period has elapsed, any distance measure applied to these loci will reach a maximal value (Goldstein et al. 1995).

2.8 Amplified Fragment Length Polymorphism

First described by Zabeau and Vos in 1993, Amplified fragment length polymorphisms (AFLP) is a method based on selective PCR amplification of

restriction fragments. This is achieved by annealing known adapter sequences to the fragment ends, which then act as highly-specific primer binding sites for subsequent PCR amplification (see Vos et al. 1995). Selectivity is achieved by use of primers which are extended by extra nucleotides (usually between one and four) in addition to the adapter sequence, so that only the subset of restriction fragments containing these bases directly adjacent to the restriction site are amplified. For most plant genomes the genomic DNA is double-digested using both a frequently cutting restriction enzyme (with a 4 bp restriction site) and also a rare cutter (with a 6 bp restriction site). A restriction site for a 4-cutter enzyme occurs on average every 256 bp in the genome, hence the resulting restriction fragments have a length suitable for PCR amplification (< 1000 bp). Furthermore, on average $1/16^{th}$ of the products can be expected to have one end digested with the 6-cutting enzyme. For plant genomes with a DNA content between about 800 Mbp (e.g. diploid *Brassica* species) and 5000 Mbp (e.g. barley), this results in a total of between 3.1 million and 20 million fragments, of which around 195000 to 1.2 million have one end cut with the rare-cutting enzyme. Selective PCR with primers containing three extra, randomly selected nucleotides ("+3-primer") will amplify on average $1/4096^{th}$ of these. In order to visualise only those amplification products from restriction fragments cut with the rare-cutting enzyme, and not the considerably more numerous fragments cut at both ends with the 4-cutter, the primer specific to the adapter sequence of the rare-cutter restriction site is labelled with either a fluorescent or radioactive marker. For the above examples this can be expected to give an average of between 50 (Brassica) to 300 (barley) labelled amplification products for any given combination of +3-primers (Breyne et al. 1999; Erschadi et al. 2000; Ren and Timko 2001). The size and number of the resulting AFLP products makes them ideally suited to size-fractionation and visualisation as bands by polyacrylamide gel electrophoresis. The choice of restriction enzymes can be important. Breyne et al. (1999) revealed that *Sac*I/*Mse*I digestion gave rise to fewer polymorphic markers in *Arabidopsis thaliana*

ecotypes than *Eco*RI/*Mse*I digestion. This finding reflected the GC content of the genome. However, both digestions produced similar dendrograms when genetic variation was calculated from the data.

Depending on the structure of the species or population under investigation, up to around 80% of all AFLP bands can show length polymorphisms (see Miyashita et al. 1999; Krauss 2000; Sawkins et al. 2001), hence an extremely large number of polymorphic markers can be obtained in a single experiment. Moreover, the use of long (approx. 20 bp), restriction-site specific PCR primers makes AFLP amplification products highly reproducible, both between experiments in the same laboratory and also between different laboratories. A comparative study by Jones et al. (1997a) in eight different European laboratories revealed only one single scoring difference out of a total of 172 markers when all laboratories analysed the same samples with the same primer combinations.

The amount of genomic DNA required for AFLP analysis is very small and, because a single restriction digest can be used for hundreds of amplifications with different primer combinations, the number of markers that can be produced is extremely high. The requirement for DNA restriction and ligation enzymes makes AFLP analyses relatively expensive, particularly for extensive population studies. Also, special equipment or facilities are required for handling and analysis of fluorescent or radioactively labelled amplification products. Despite these drawbacks, however, the quantity and quality of genotype information that can be obtained using this marker system with a relatively low experimental input is extremely high in comparison with other methods. Although AFLP markers are co-dominantly inherited, the sheer volume of reproducible, polymorphic marker data that can be achieved by AFLP analysis makes this technique highly applicable not only for genetic mapping studies, e.g. in oat, (Groh et al. 2001), rice (Chen et al. 2001), wheat (Nachit et al. 2001), maize (Castiglioni et al. 1999), barley (Lahaye et al. 1998), oilseed rape (Snowdon and Friedt 2004), but also for in-depth studies of genetic diversity at the population and sub-population level. In recent years an enormous

wealth of information has been obtained by AFLP analysis to describe genetic diversity in a vast number of plant species ranging from the model plant *Arabidopsis thaliana* (Miyashita et al. 1999; Erschadi et al. 2000; Breyne et al. 1999), to major agricultural plants including rice (Fuentes et al. 1999; Katiyar et al. 2000), wheat (Manifesto et al. 2001), barley (Pakniyat et al. 1997) and rapeseed (Snowdon and Friedt 2004).

3 Materials and Methods

3.1 Sampling

In order to investigate the potential long-term effects of different agricultural regimes on the phenotype and genotype of arable weed populations, it was necessary to utilise the persistent seed bank to allow the comparison of older and contemporary populations. Therefore, species were selected that build persistent seed banks. Based on previous research conducted under the SFB 299 research programme "Land-use Concepts for Peripheral Regions", contemporary and historical agricultural sites within and around the Lahn-Dill-Bergland of central Germany were able to be identified (Fuhr-Boßdorf et al. 1999). From this information the species *Arabidopsis thaliana* and *Viola arvensis* were selected as model species.

3.1.1 Arabidopsis thaliana (L.) HEYNH.

The wild crucifer *Arabidopsis thaliana* (also known as "mouse ear cress") is a predominantly selfing, small, white flowered annual, belonging to the *Brassicaceae* family. It is therefore closely related to other crucifers such as mustard, rapeseed, swedes, turnip, rape, broccoli, and the extremely diverse cabbage family.

Fig. 1: Photo of *Arabidopsis thaliana* (2n = 10).
William S. Justice @ USDA-NRCS PLANTS Database

Since being adopted as a model system in plant genetics in the 1950's (Redei 1992) it has become the reference genome for molecular biological analysis of dicotyledonous plants (Goodman et al. 1995) due to its small genome and small number of chromosomes (2n=10), easy genetic transformation, ability to grow at high density (small growth habitat), and fast life-cycle (Koornneef et al. 1997; Baumann et al. 1998; Pigliucci 1998).

A. thaliana is considered a typical annual weed. It

has a filamentous root system and a basal rosette of leaves from which the main flowering stem develops and grows to about 40 cm in height (usually developing lateral branches). Under favourable conditions, additional basal meristems can develop into inflorescences, giving rise to a large variation in seed output between individuals (Pigliucci 1998).

A. thaliana exhibits a simple life-cycle. Initially the seed germinates releasing the cotyledons which form the first leaves. The basal rosette then grows a number of leaves before the main stem begins to elongate (bolting) and produce flowers. Finally, after between 30 and 50 days (depending on the ecotype), the leaves and fruits begin to senesce and the seeds are released, either to germinate immediately or remain dormant.

According to Napp-Zinn (1985), four distinct morphotypes of *A. thaliana* can be physiologically distinguished:

- A very early flowering type (probably a spring ephemeral) that bolts within 40 days of sowing (nearly all the ecotypes used in molecular genetic research are of such a kind; Pigliucci 1998)
- A type that flowers slightly later than those belonging to the group above and responds slightly to vernalisation (probably an early summer annual)
- A third type that behaves as a late summer annual (bolting is delayed even longer) and has a greater response to vernalisation
- A winter annual type that flowers after four to six months and requires vernalisation.

Despite the fact that it is a poor competitor, preferring to inhabit disturbed habitats in which it can quickly establish and produce seeds before competition arises, *A. thaliana* maintains a remarkable world distribution (Fig. 2).

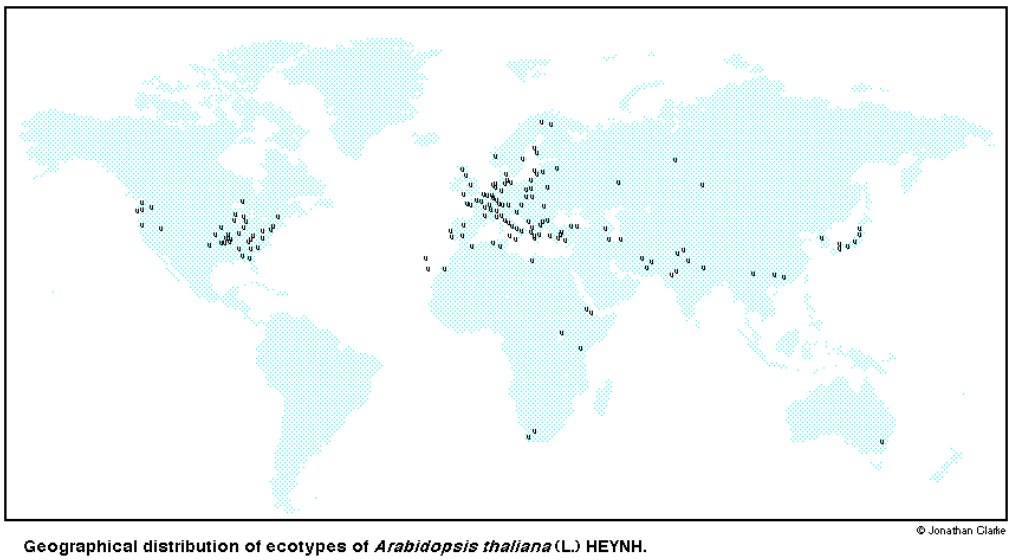


Fig. 2: Geographical distribution of known natural populations of *A. thaliana* (L.) Heynh. throughout the world. This map has been reproduced courtesy of the TAIR Arabidopsis web resource (http://www.arabidopsis.org/info/aboutarabidopsis.jsp).

There appears to be, however, no clear association between the flowering behaviour of *A. thaliana* and its geographical distribution. As Pigliucci (1998) mentions, quite different environments (such as Tennessee compared to Scandinavia) maintain populations exhibiting the same flowering behaviour, whereas similar environments (with respect to temperature and precipitation) such as Cologne and Grenoble are characterised by summer annual and winter annual populations, respectively. It has been shown that flowering times are linked strongly to the dominant FRIGIDA (FRI) gene and that variants exhibiting early flowering times contain loss-of-function FRI alleles (Napp-Zinn 1985; Clarke and Dean 1994; Johanson et al. 2000).

The family *Brassicaceae* has been classified as being "very important" for the diet of farmland birds (Marshall et al. 2003), although no data is available specifically for *A. thaliana*.

3.1.2 *Viola arvensis* (MURR.)

In contrast to *A. thaliana*, *Viola arvensis* has been relatively little studied. It is a member of the *Violaceae* family and is commonly known as the European field pansy. It is a winter annual, growing to a height of 30 cm. Flowering usually occurs between April and October, although it can flower at any given time of year. Unlike many *Viola* species, *V. arvensis* belongs to the *Melanium* group and does not produce cleistogamous flowers, however it can reproduce asexually. It displays a varied dormancy cycle: once the primary dormancy is broken, the seeds cycle between non-dormancy and conditional dormancy, resulting in two types of seeds possessing either an annual conditional dormancy/non-dormancy cycle or a dormancy/non-dormancy cycle. Persistent seed banks of *V. arvensis* therefore contain a mixture of seeds. In Autumn, seeds from both types are non-dormant, germinating over a wide range of temperatures. In Spring, only those seeds possessing a conditional dormancy/non-dormancy cycle can be stimulated to germinate if the conditions are appropriate (Baskin and Baskin 1995).

Fig. 3: Photo of *Viola arvensis* (2n = 34).

William S. Justice @ USDA-NRCS PLANTS Database

Despite also being considered a poor competitor (Holzmann 1987), *V. arvensis* has been somewhat more successful on arable land than *A. thaliana*, having established a near world-wide distribution through survival on cereal crop fields over the last many decades (Eggers 1978; Bachthaler et al. 1986) due to the lack of effective herbicides (Christen et al. 1999). This was supported by findings from a study in Finland (Hyvönen et al. 2003) where a 16% increase in the number of spring cereal fields which were inhabited by *V. arvensis* was found between 1982-84 and 1997-99.

However, recent work in England has shown that in recent years the abundance of *V. arvensis* in arable fields has remained stable[2]. The same research group has also classified the plant as being of "intermediate importance" for the purposes of supporting insect and bird species. This classification was supported by Marshall (2003), who categorised the family *Violaceae* as being "important" for farmland bird diet.

3.1.3 Galeopsis tetrahit

Initially *Galeopsis tetrahit* was selected as the third arable weed species to be used for genetic analysis. As shown in the results (Tab. 14), extremely few *G. tetrahit* plants germinated from the soil samples, and therefore the species was removed from the study.

3.2 Sampling Locations

The selection of locations from which to obtain *A. thaliana* and *V. arvensis* was achieved using information provided by the SFB 299 project "Land-use Concepts for Peripheral Regions" combined with aerial photos taken from 1945-1998 (Fuhr-Boßdorf et al. 1999).

Three locations were selected from which soil samples were taken: Erda and Eibelshausen (both in the Lahn-Dill-Bergland region of Hesse, central Germany), and Amöneburg, which lies approximately 30 km east of the other two locations (average distance in meters between the sampling sites is given in Tab. 1).

Four types of agricultural regime were identified from which soil was to be taken: intensive cropping, extensive cropping, meadow, and fallow.

The intensive cropping locations were all situated at Amöneburg. This site lies 200 to 295 m above sea level and is used predominantly for the production of wheat, barley, rapeseed, and sugar beet. The fields receive herbicide treatments twice per year and are fertilised with between 120

[2] http://www.rothamsted.bbsrc.ac.uk/pie/PeterGrp/bcpc/bcpcviola.htm

and 170 kg N/ha. The average yearly temperature is 7.8 °C with a precipitation of 576 mm per year[3].

The other two locations both contained sampling sites for extensive cropping, meadow and fallow populations. These sites receive 650 to 1100 mm of precipitation per year and lie 285 to 500 m above sea level. The average yearly temperature is 5 to 8 °C[4]. The area is considered to be less viable for agriculture than the Amöneburg region. The term extensive cropping was used in this study to describe fields on which extensive agriculture (without the addition of herbicides or fertiliser) has continually taken place over an extended period of time. The fields termed meadow (70's) were used for extensive agriculture up until the 1970's only. However, since the year 2000 they have been used as grassland, being mown twice a year. Fallow areas (50's) have been left untended since extensive agriculture was halted in these areas during the 1950's.

Tab. 1: Average distance in meters between sampling sites at each location.

	Amöneburg	Erda	Eibelshausen
Amöneburg	-		
Erda	25892	-	
Eibelshausen	37967	20441	-

[3] http://www.uni-giessen.de/tbe/home_rh.htm
[4] http://www.uni-giessen.de/sfb299/uis/U_Region/natur/natrsfb.html

Tab. 2: A list of the sampling sites from which soil samples were taken. "Field number" is the designation of the field from which the samples came. "Short label" is an artificial label given for quicker recognition: first letter = location (A – Amöneburg, R – Erda, I – Eibelshausen), second letter = regime type (I – intensive, E – extensive, M – meadow, F – fallow).

Amöneburg			**Erda**			**Eibelshausen**		
Regime Type	Field Number	Short Label	Regime Type	Field Number	Short Label	Regime Type	Field Number	Short Label
Int	Am1	AIa	Ext	Er124	REa	Ext	Ei7641	IEa
Int	Am2	AIb	Ext	Er1176	REb	Meadow	Ei6044	IMa
Int	Am3	AIc	Ext	Er1747	REc	Meadow	Ei6052	IMb
Int	Am4	AId	Ext	Er1955	REd	Meadow	Ei7625	IMc
Int	Am5	AIe	Ext	Er3101	REe	Meadow	Ei11219	IMd
Int	Am5	AIf	Meadow	Er165	RMa	Fallow	Ei4429	IFa
			Fallow	Er91	RFa			
			Fallow	Er148	RFb			
			Fallow	Er1985	RFc			
			Fallow	Er2021	RFd			

3.2.1 Sampling Technique

Soil samples were taken during September 1999. Five samples were taken at each of the 21 sample sites (Tab. 2). Each sample was a mixture of 25 core samples taken within a $25m^2$ area with a corer measuring 3.5 cm in diameter and taking soil from 5-10 cm under the surface. For the undisturbed habitats (meadow and fallow) this procedure taps into the persistent seed bank. For the agricultural fields, which undergo regular ploughing, this technique samples both the persistent and transient seed banks.

In Spring 2000 and 2001 the samples were placed in a glasshouse to allow germination of viable seeds. The soil samples were stored indoors over winter.

Seeds which germinated were carefully removed and replanted in individual pots. To prevent outcrossing, all plants were encapsulated by Crispack selfing bags. The individuals obtained using this sampling technique were

also used by Schubert (2003), who carried out a study concentrating on the phenotypic characteristics of these populations.

It can be seen in Tab. 2 that the sample area Am5 occurs twice (as AIe and AIf). As will be seen in the results, a total of only 23 individuals was obtained from all of the intensive sample sites. Therefore, in spring 2002, all of the intensive sample sites were searched for living *A. thaliana* specimens. A total of 44 plants were located within an area of approximately 50 m^2 on the sample site Am5.

In addition to this, a variety of *A. thaliana* ecotypes were also included for the cluster analysis and Principal Coordinate Analysis (PCA). Eight ecotypes provided by the Nottingham Arabidopsis Stock Centre (NASC) were analysed: Corscalla (Italy), Köln (Germany), Devon (UK), Keukenhof (Netherlands), Orge (France), Kelsterbach (Germany), Seattle (Washington, USA), and Kazahkstan[5]. Also, five individuals found in close proximity to each other in the Rhön area of central Germany were also included.

3.2.2 Plant Material for Genetic Analysis

The original plants that germinated from the seed bank were not able to be used for the genetic analysis. This was due to the fact that they were oven dried for measurement of dry matter which can cause damage to the genomic DNA. Therefore, the seeds produced by selfing of the original plants were collected, replanted into individual pots, and again placed in a glasshouse during spring and summer to allow germination and growth.

Several seeds from the same parent were planted in each pot. The fresh leaf material from between four and ten seedlings originating from a single parent was taken for DNA extraction. The need to use multiple offspring was due to the fact that a heterozygote parent can produce both homozygous and heterozygous offspring (Loridon et al. 1998). By sampling at least four plants the probability of selecting four similarly homozygous plants is reduced to at least 0.4% compared to 50% if taking only one offspring individual.

[5] NASC On-Line Catalogue: http://nasc.nott.ac.uk/home.html

3.3 Molecular Marker Techniques

3.3.1 DNA Extraction

DNA was extracted according to Doyle and Doyle (1987), using the reagents shown in Tab. 3. Fresh leaf material was crushed to a powder using a pestle and mortar containing liquid nitrogen and a small amount of sand. The resulting material was placed into an Eppendorf tube containing 1 mL of cetyl-trimethylammonium bromide (CTAB) extraction buffer and placed in a 65^{o} C water bath for 30 minutes. For the *V. arvensis* plant material the extraction buffer also included 2% polyvinylpyrollidone (PVP) to reduce viscosity. After incubation, 1 mL from a 24:1 mixture of chloroform and isoamylalcohol (CiA) was added and, after inverting several times, was centrifuged at 12 000 g for 10 minutes.

The supernatant was transferred to a new tube and CiA was added at a 1:1 ratio with the inverting and centrifuging being repeated. The resulting supernatant from this step was again transferred to a new tube to which 67 µL of both sodium acetate (3 M, pH 5.5) and ammonium acetate (10 M) were added. Iso-propylalcohol was then added at a 2:3 ratio of the volume in the tube (usually around 400 µL), mixed by inversion, and left for 15 minutes to allow DNA precipitation. After further centrifuging for 5 minutes at 12 000 g the supernatant was removed and the remaining DNA pellet was washed for at least 10 minutes with 70% ethanol containing 10 mM ammonium acetate. The pellet was dried at room temperature after tipping off the ethanol and was then resuspended in 100 µL TE buffer. To prevent interference from RNA in the subsequent PCR and restriction digestions, 1 µL of RNAse was added to each sample.

Tab. 3: Reagents used for DNA extraction.

Doyle & Doyle extraction buffer		**Alcohol for Washing**	
NaCl	1.4 M	Ethanol	70%
Tris HCl (pH 8.0)	0.1 M	NH_4OAc	10 mM
EDTA (pH 8.0)	20 mM		
CTAB	2%	**TE Buffer**	
$Na_2S_2O_5$	1%	Tris HCl (pH 8.0)	10 mM
Mercaptoethanol (autoclaved)	0.2%	EDTA (pH 8.0)	1 mM

3.3.2 DNA Quantification

Extracted DNA was quantified using a DyNAQuant fluorometer 200 (Hoefer). Sample DNA was added to a solution containing H33258 (Hoechst), which binds to the DNA. H33258 fluoresces when exposed to a wavelength of 365 nm, emitting light at 460 nm which is measured by the fluorometer to give the DNA concentration in comparison to a known standard.

3.3.3 Microsatellite Analysis

Microsatellite analysis was carried out only on *A. thaliana*. This was due to the fact that microsatellite markers have yet to be developed for *V. arvensis*. A total of seven microsatellite markers were used (Tab. 4), based on protocols used by previous *A. thaliana* researchers (Bell and Ecker 1994; Innan et al. 1997; Todokoro et al. 1995; Virk et al. 1999). Primers were synthesised by MWG Biotech, with the reverse primer of each primer pair being labelled IRD-800 to enable fluorescent detection.

Tab. 4: Summary of microsatellite primers used for analysis on *A. thaliana*. All primers had an annealing temperature of 55°C.

Primer	Forward sequence	Reverse sequence	Chromosome
nga59	GCATCTGTGTTCACTCGCC	TTAATACATTAGCCCAGACCCG	1
nga106	GTTATGGAGTTTCTAGGGCAC	TGCCCCATTTTGTTCTTCTC	5
nga111	CTCCAGTTGGAAGCTAAAGGG	TGTTTTTTAGGACAAATGGCG	1
nga128	GGTCTGTTGATGTCGTAAGTCG	ATCTTGAAACCTTTAGGGAGGG	1
nga139	AGAGCTACCAGATCCGATGG	GGTTTCGTTTCACTATCCAGG	5
nga151	GTTTTGGGAAGTTTTGCTGG	CAGTCTAAAAGCGAGAGTATGATG	5
nga168	TCGTCTACTGCACTGCCG	GAGGACATGTATAGGAGCCTCG	2

PCR reactions were carried out using a Primus 96^{Plus} thermocycler (MWG Biotech) in a 20 µL volume containing 10-30 ng of template DNA, 200 µM dNTPs, 5 pmol of each primer, 1x PCR reaction buffer (Qiagen), 0.5 units *Taq*-polymerase (Qiagen), and 2.5 mM MgCl. The thermocycler protocol used is shown in Tab. 5.

Tab. 5: Thermocycler programme for PCR amplification of microsatellites.

Step	Temperature	Length of time	Number of Cycles
Lid Heating	110° C	Until 110° C	
Denaturation	94° C	2 min	1
Denaturation	94° C	15 sec	
Annealing	55° C	15 sec	25
Polymerisation	72° C	30 sec	
Final Polymerisation	72° C	5 min	1

The amplified products were resolved by polyacrylamide gel electrophoresis (PAGE) as described in section 4.2.5.

4.2.4 AFLP Analysis

AFLP analysis was performed essentially as described by Vos et al. (1995). Both *A. thaliana* and *V. arvensis* were analysed using the AFLP technique. To the author's knowledge, no previous AFLP analysis has been carried out on *V. arvensis*. For all samples, AFLP Core Reagent Kits (Gibco Life Technologies, Karlsruhe) were used.

For each sample 60 ng of DNA was digested with the restriction enzymes *Mse*I and *Eco*RI in a volume of 25 µL containing 1x reaction buffer. This was achieved by incubating these samples for two hours at 37° C followed by a final step of 70° C for 15 min. Following this step, 24 µL of double-stranded adapter-ligation solution along with 1 µL of the enzyme T4 DNA ligase were added to each sample and incubated for two hours at 20° C. The samples were stored at –20° C to halt this reaction. With the given samples, it was found that a preamplification step with +0-primers did not noticeably enhance the results and so this step was removed from the standard protocol.

For the +1-amplification, +1-primers (MWG Biotech) were chosen whose nucleotide sequence is complementary to the DNA fragments from the ligation step, which themselves bound to the sticky ends caused by the restriction digestion. Each +1-primer contains a selective base at the 3'-end. A summary of the primers chosen for the +1-amplification is shown in Tab. 6. After the +1-primer solution had been made (Tab. 7) PCR amplification was carried out (Tab. 8).

Tab. 6: Sequences of primers used for +1-amplification of both species. Primers can also be referred to as, for example, M-A (equivalent to *Mse*I-A) or E-C (equivalent to *Eco*RI-C), etc.

Species	Primer	Label	Sequence
A. thaliana	+1(C)-*Eco*RI-Primer	E02	5' – GAC TGC GTA CCA ATT C**C** – 3'
	+1(C)-*Mse*I-Primer	M02	5' – GAT GAG TCC TGA GTA A**C** – 3'
V. arvensis	+1(A)-*Eco*RI-Primer	E01	5' – GAC TGC GTA CCA ATT C**A** – 3'
	+1(A)-*Mse*I-Primer	M01	5' – GAT GAG TCC TGA GTA A**A** – 3'

Tab. 7: Protocol for the set-up of +1-amplification solutions.

Ingredient	Amount
DNA	5 µL (from 1:10 dilution of adapter-ligation product)
dNTPs (10 mM)	1 µL
PCR buffer 10x	5 µL
*Mse*I-primer (100 ng/µL)	0.75 µL
*Eco*RI-primer (100 ng/µL)	0.75 µL
Taq polymerase (5 U/ µL)	0.2 µL
Milli-Q H_2O	Add to 50 µL

Tab. 8: Thermocycler programme for +1-amplification.

Step	**Temperature**	**Length of time**	**Number of Cycles**
Lid Heating	110°C	Until 110°C	
Denaturation	94°C	3 min	1
Denaturation	94°C	30 sec	
Annealing	56°C	1 min	20
Polymerisation	72°C	1 min	
Final Polymerisation	72°C	5 min	1

The products of the +1-amplification were also diluted 1:10 before being used as template DNA for either +3-amplification (*V. arvensis*) or +2-amplification (*A. thaliana*). During primer screening for *V. arvensis* the +2-amplification produced too many bands for effective scoring, hence the use of +3-primers. However, for *A. thaliana* (which has a small genome) the +2-amplification produced a sufficient number of bands that were easy to score, whereas with +3-primers too few bands were amplified for *A. thaliana*. A total of seven +3-primer combinations were used for *V. arvensis* (Tab. 9) and six +2-primer combinations for *A. thaliana* (Tab. 10).

Tab. 9: AFLP analysis: +3-primer combinations used for *V. arvensis*. Shaded cells represent the primer combinations that were used.

*Eco*RI +3-primers	*Mse*I +3-primers				
	M-ACG	M-ACT	M-ATC	M-AGG	M-ATT
E-ACT					
E-ATC					

Tab. 10: AFLP analysis: +2-primer combinations used for *A. thaliana*. Shaded cells represent the primer combinations that were used.

*Eco*RI +3-primers	*Mse*I +3-primers			
	M-CA	M-CC	M-CG	M-CT
E-CA				
E-CC				
E-CG				
E-CT				

The selective amplification (with +2 or +3 primers) was carried out using the same ingredients as for the +1-amplification except that MgCl was added to the mixture. EcoRI selective primers were fluorescently labelled with IRD800. Tables 11 and 12 show the protocol for sample production and the thermocycler programme followed for the selective amplification, respectively.

Tab. 11: Protocol for the set-up of +3-amplification solutions.

Ingredient	Amount
DNA	5 µL (from 1:10 dilution of +1-amplification product)
dNTPs (10 mM)	0.4 µL
PCR buffer 10x	2 µL
MgCl (25 mM)	1 µL
*Mse*I-primer (100 ng/µL)	0.2 µL
*Eco*RI-primer (100 ng/µL)	0.15 µL
Taq polymerase (5 U/ µL)	0.08 µL
Milli-Q H_2O	Add to 20 µL

Tab. 12: Thermocycler programme for +3-amplification.

Step	Temperature	Length of time	Number of Cycles
Lid Heating	110°C	Until 110°C	
Denaturation	94°C	3 min	1
Denaturation	94°C	30 sec	
Annealing	65°C; -0.7oC/cycle	30 sec	12
Polymerisation	72°C	1 min	
Denaturation	94°C	30 sec	
Annealing	56°C	30 sec	23
Polymerisation	72°C	1 min	
Final Polymerisation	72°C	5 min	1

The amplified products were resolved by PAGE as described in the following section.

4.2.5 Gel Electrophoresis

All amplification products from both microsatellite and AFLP analyses were visualised through the use of a 0.2 mm thick, 25 cm long polyacrylamide gel (based on an 8% Long Ranger Gel Solution) in a Li-COR Gene Readir 4200 DNA sequencer. All products were mixed with a STOP loading buffer at a 1:1 ratio before being denatured at 94°C for 3 min. Reverse primers for microsatellites and +2/+3 EcoRI-primers for AFLP analysis were all fluorescently labelled with IRD800. Through use of a laser emitting a wavelength of 800 nm this dye is excited and fluoresces, allowing the DNA to be detected. A 1x TBE buffer was used for running the gels.

Standard parameters were used for the separation of fragments: 1500 V, 50 W, 35 mA, 48°C. The reagents used for the gel and running of samples are shown in Tab. 13.

Tab. 13: Gel solutions, loading buffer and running buffer composition for PAGE.

Long Ranger gel solution 8%		**5x TBE Buffer**	
Long Ranger 50%	16 mL	TRIS	54 g
Urea	42 g	EDTA (0.5 M, pH 8.0)	20 mL
5x TBE	20	Boric Acid	27.5 g
Milli-Q H_20	add to 100 mL	Milli-Q H_20	add to 1 L
Solution for Polyacrylamide Gel		**STOP loading buffer**	
Long Ranger 8% solution	20 mL	Formamide	95 mL
TEMED	20 µL	EDTA (0.5 M, pH 9.0)	2 mL
10% APS	140 µL	Basic Fuchsin	0.1 g
(**a**mmonium **p**er**s**ulphate)		Milli-Q H_20	add to 100 mL

The size (in base pairs) of bands appearing on the gel were determined by comparison to a 50-350 bp molecular size standard (Li-COR) run on both edges of the gel.

4.3 Statistical Analysis

Unless otherwise stated the population genetics analysis was carried out using the programs The Excel Microsatellite Toolkit (Park 2001) and Arlequin ver. 2000 (Schneider et al. 1997). For Principal Coordinate Analysis (PCA), clustering, and Mantel tests, the software NTSYS-PC (Rohlf 1997) was employed. Testing for bottlenecks was performed with the program BOTTLENECK (Cornuet and Luikart 1996).

Genetic Variability

The within-population genetic diversity was assessed in a variety of ways. For each population the allele frequencies, number of alleles per locus, expected heterozygosity (H_{exp}), and observed heterozygosity (H_{obs}) were calculated using The Excel Microsatellite Toolkit. Expected number of heterozygotes were computed according to Nei's unbiased gene diversity (Nei 1987).

Hardy-Weinberg Expectations

To detect significant deviation of observed genotypic distributions from Hardy-Weinberg equilibrium, a test was carried out using a modified version of the Markov-chain random walk algorithm (Guo and Thompson 1992) with 1000 iterations per batch.

Linkage Disequilibrium

For each population by regime, the genotypic linkage disequilibrium, which refers to the non-random association between pairs of alleles, was tested using an extension of Fisher's exact probability test (Slatkin 1994) using a Markov chain method with 1000 iterations per batch. This test assumes Hardy-Weinberg genotype proportions.

Analysis of Molecular Variance

To elucidate any underlying hierarchical population structure, a Euclidean square distance matrix (pairwise distances) was constructed and used to perform analysis of molecular variance (AMOVA; Excoffier et al. 1992) using Arlequin ver. 2000 (Schneider et al. 1997). Because it is not possible to distinguish genotypes with AFLP data, each sample was considered as a haplotype (Tero et al. 2003).

Bottleneck Detection

Cornuet and Luikart (1996) and Luikart et al. (1998) proposed three tests for detecting recent population bottlenecks from microsatellite frequencies. These tests are based on the fact that when a population experiences a recent reduction in effective size, the reduction of allelic diversity at polymorphic loci occurs quicker than the reduction in heterozygosity (i.e. gene diversity). Thus, in a recently bottlenecked population, a heterozygosity excess is expected when compared to what would be expected from the observed allele frequencies under the assumption of mutation-drift equilibrium. The three tests (the 'standardised test', the 'sign test'

(Cornuet and Luikart 1996), and the 'Wilcoxon signed-ranks test' (Piry et al. 1999)) were carried out using the software package BOTTLENECK[6]. As recommended by Piry et al. (1999), the Two Phase Mutation model (TPM) was set to 95% Stepwise Mutation Model (SMM or single step mutations only) and 5% to multistep mutations.

Another method of detecting bottlenecks was demonstrated by Garza and Williamson (2001). When a population decreases in size both the number of alleles (*k*) and the range of allele sizes (*r*) in the population decreases. However, they noted that *r* decreases at a slower rate than *k* (when an allele is lost in a population *k* will be reduced, but *r* will only decrease when the largest or smallest allele in the population is lost). Therefore, by monitoring the ratio $M = k/r$, changes in population size can be detected.

Binary Data Analysis

The presence (1) or absence (0) of unambiguous AFLP bands was manually scored from the gel images produced by the Li-COR Gene Readir 4200 DNA sequencer. The program RFLP Scan was used to facilitate the scoring; principally due to the programs ability to estimate band size and remove the effects of uneven gels (through the addition of "de-smile" lines). The same procedure was used for microsatellite binary data scoring, except that the data set was doubled with heterozygotes being scored only once for each allele (homozygotes twice). Monomorphic bands were discarded from all binary data sets before analysis.

Principal Coordinate Analysis and Clustering

Pairwise genetic similarities were calculated between individuals using the Jaccard coefficient (Jaccard 1908): $S_{ij} = a / (a + b + c)$ where S_{ij} is the similarity between two individuals *i* and *j*, *a* is the number of bands present in both *i* and *j*, *b* the number present in *i* but absent in *j*, and *c* the number present in *j* but absent in *i*. Principal Coordinate Analysis (PCA; Gower 1966) was performed with the DECENTER and EIGEN procedures.

[6] http://www.ensam.inra.fr/URLB

Dendrograms were constructed using the UPGMA (Unweighted Pair Group Method with Arithmetic Mean) clustering procedure (Michener and Sokal 1957). All analyses were performed using the program NTSYS-PC (Rohlf 1997).

Rates of Migration

The absolute number of migrants, *m*, was calculated according to Slatkin (1991) using Arlequin ver. 2000 (Schneider et al. 1997). This method uses F_{ST} values and assumes that the mutation rate (μ) is negligible compared to the number of migrants exchanged between any given populations.

Mantel Test

The Mantel test (Mantel 1967) was used to estimate the association between genetic and geographic distance. The programme NTSYS-PC (Rohlf 1997) was used to carry out these computations.

4 Results

4.1 Sampling

The total number of seeds which germinated to produce mature adults from the Spring 2000 and 2001 soil sample incubations is shown in Tab. 14 below.

Tab. 14: Number of mature plants extracted from soil samples. The numbers in square brackets relate to the total number of fields sampled per agricultural regime, and the numbers in parentheses represent the average number of plants per field.

	Agricultural Regime			
Species	Intensive [5]	Extensive [6]	Meadow [5]	Fallow [5]
Arabidopsis thaliana	23 (4.6)	166 (27.7)	76 (15.2)	34 (6.8)
Viola arvensis	96 (19.2)	100 (16.7)	40 (8.0)	68 (13.6)
Galeopsis tetrahit	0 (0.0)	14 (2.3)	16 (3.2)	2 (0.4)

As mentioned in Section 3 *Galeopsis tetrahit*, which initially was also to be used for genetic analysis, was not analysed due to the very small number of samples germinating from the soil seed bank.

The difference between the plants extracted from intensive and extensive agricultural regimes is in accordance with previous literature. Although many studied species have declined in abundance on intensive agricultural land over the last few decades, some species, *V. arvensis* included, have shown resistance to herbicide treatments and have successfully survived in such environments (Bachthaler, 1986; Christen et al. 1999; Hyvönen et al. 2003).

If the number of *G. tetrahit* plants from the extensive fields is any reflection of the greater situation, then it would indicate that normal agricultural practices may cause enough environmental stress that this species has difficulty surviving on agricultural land, especially intensively managed fields.

The similarity of the number of *V. arvensis* plants obtained from intensive and extensive land indicates that the differences between these two regimes did not hinder its abundance.

However the is clearly different for *A. thaliana*, which clearly shows a preference for extensive land over intensive land.

Not all samples collected were used for analysis. The extraction of DNA from *V. arvensis* proved complicated due to the viscosity of the solution after incubation in the water bath. Therefore, many samples did not consistently amplify in the subsequent AFLP reactions, presumably due to unclean DNA hindering the restriction digestion.

The number of samples used for analysis is shown by location in Table 15 (*A. thaliana*) and Table 16 (*V. arvensis*), and by regime in Table 17.

Tab. 15: Number of mature *A. thaliana* plants analysed according to location.

Amöneburg			**Erda**			**Eibelshausen**		
Regime Type	Short Label	Number of Plants	Regime Type	Short Label	Number of Plants	Regime Type	Short Label	Number of Plants
Int	AIa	1	Ext	REa	27	Ext	IEa	0
Int	AIb	16*	Ext	REb	33	Meadow	IMa	26
Int	AIc	1	Ext	REc	34	Meadow	IMb	7
Int	AId	0	Ext	REd	39	Meadow	IMc	17
Int	AIe	4	Ext	REe	2	Meadow	IMd	0
Int	AIf	42†	Meadow	RMa	9	Fallow	IFa	0
			Fallow	RFa	11			
			Fallow	RFb	3			
			Fallow	RFc	0			
			Fallow	RFd	14			

* For AFLP analysis only 13 of this population were analysed.

† Plants collected from field Am5 in Spring 2002.

Tab. 16: Number of mature *V. arvensis* plants analysed according to location.

Amöneburg			Erda			Eibelshausen		
Regime Type	Short Label	Number of Plants	Regime Type	Short Label	Number of Plants	Regime Type	Short Label	Number of Plants
Int	AIa	29	Ext	REa	11	Ext	IEa	0
Int	AIb	12	Ext	REb	19	Meadow	IMa	1
Int	AIc	8	Ext	REc	15	Meadow	IMb	0
Int	AId	0	Ext	REd	5	Meadow	IMc	22
Int	AIe	10	Ext	REe	0	Meadow	IMd	2
			Meadow	RMa	5	Fallow	IFa	4
			Fallow	RFa	0			
			Fallow	RFb	0			
			Fallow	RFc	14			
			Fallow	RFd	15			

Tab. 17: Number of mature *A. thaliana* and *V. arvensis* plants analysed according to regime.

	Agricultural Regime				
Species	Intensive	Extensive	Meadow	Fallow	Total
Arabidopsis thaliana	64	135	59	28	286
Viola arvensis	59	50	30	33	172

4.2 Arabidopsis thaliana

4.2.1 Molecular Markers

Microsatellite Markers and Allele Frequencies

A total of 111 alleles were detected at seven microsatellite loci from 286 individuals. All loci examined were highly polymorphic, with a range of 13 to 23 alleles per locus. The average number of alleles per locus was 15.9 (Table 18), which is considerably higher than other reports for *A. thaliana* accessions; Bell and Ecker (1994) recorded 4.1 alleles per locus, Todokoro et al. (1995) 5.3, and Innan et al. (1997) 10.6. All loci deviated significantly from expected allele frequencies under the assumption of Hardy-Weinberg equilibrium ($P < 0.001$).

The distribution of alleles did not follow any particular pattern for all loci (Figure 4a–g), with the allele frequency distributions of all populations varying from locus to locus, e.g. locus nga128 (allele size 134 nearing fixation) and nga139 (very even distribution) for the intensive population. The extensive and fallow populations showed a similar pattern to each other, however the majority of these plants came from the same geographical area and is most likely a representation of the spatial proximity of these populations. No population exhibited fixed alleles for any of the loci. Interestingly, all populations exhibited individuals containing the largest recorded allele (149 bp) for locus nga151, despite the next largest allele (133 bp) being eight repeat units shorter. Large gaps between allele sizes were also seen in all populations for allele nga63.

Tab. 18: Allele numbers and sizes, expected (H_{exp}) and observed heterozygosity (H_{obs}) for the *A. thaliana* microsatellite loci analysed.

Locus	Number of Alleles	Minimum Size (bp)	Maximum Size (bp)	Range (repeat units)	H_{exp}	H_{obs}
nga59	16	109	143	17	0.866	0.106
nga106	14	123	159	18	0.792	0.074
nga111	17	119	167	24	0.849	0.032
nga128	15	161	193	16	0.887	0.092
nga139	23	132	194	31	0.857	0.053
nga151	13	103	149	23	0.762	0.082
nga168	13	129	167	19	0.824	0.092
Average	15.9	125.1	167.4	21.1	0.834	0.076

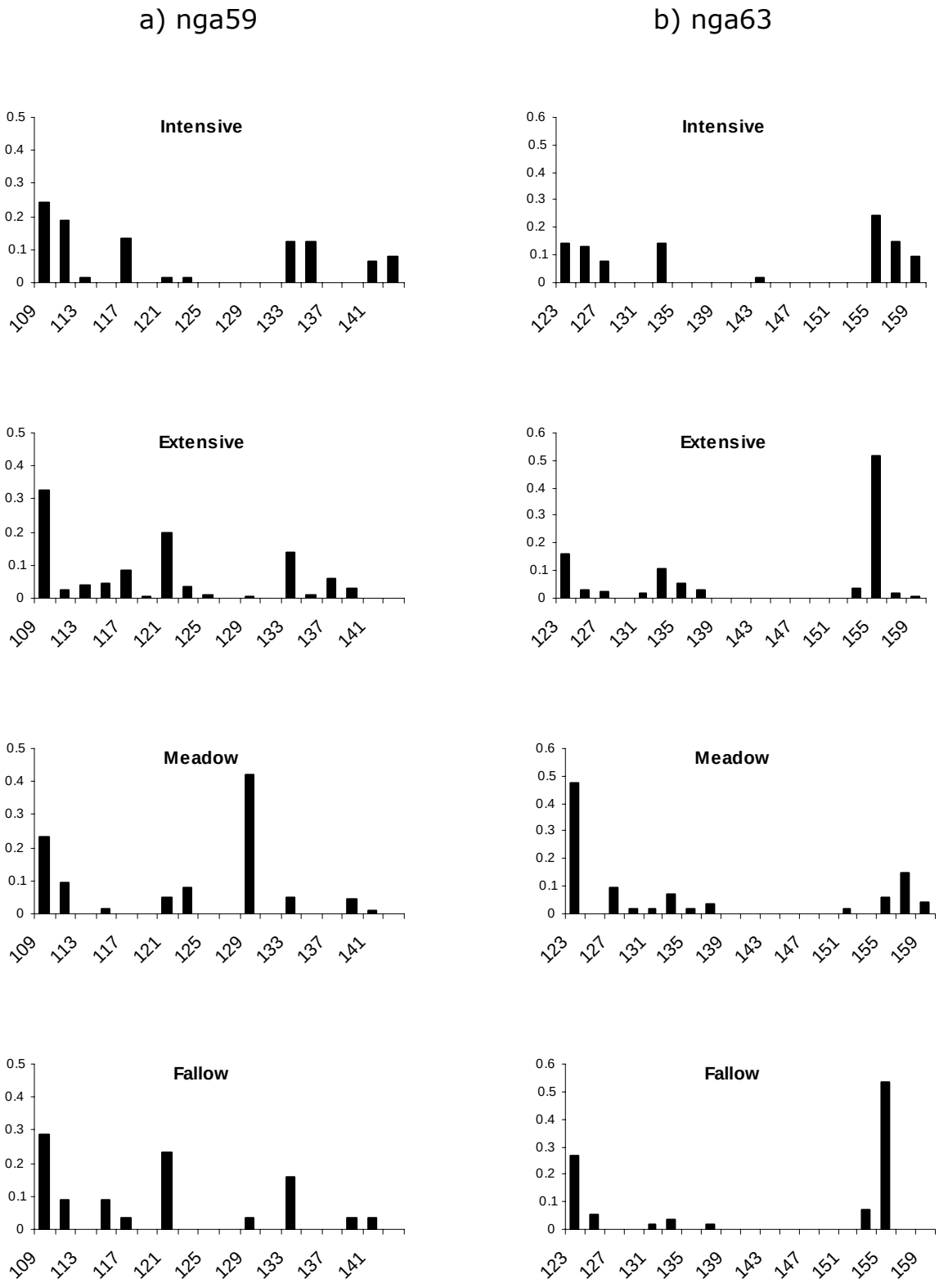


Fig. 4: Allele frequency distributions for *A. thaliana* microsatellite loci nga59 (a) and nga63 (b). The *X axis* represents allele size (in basepairs); the *Y axis* represents allele frequencies.

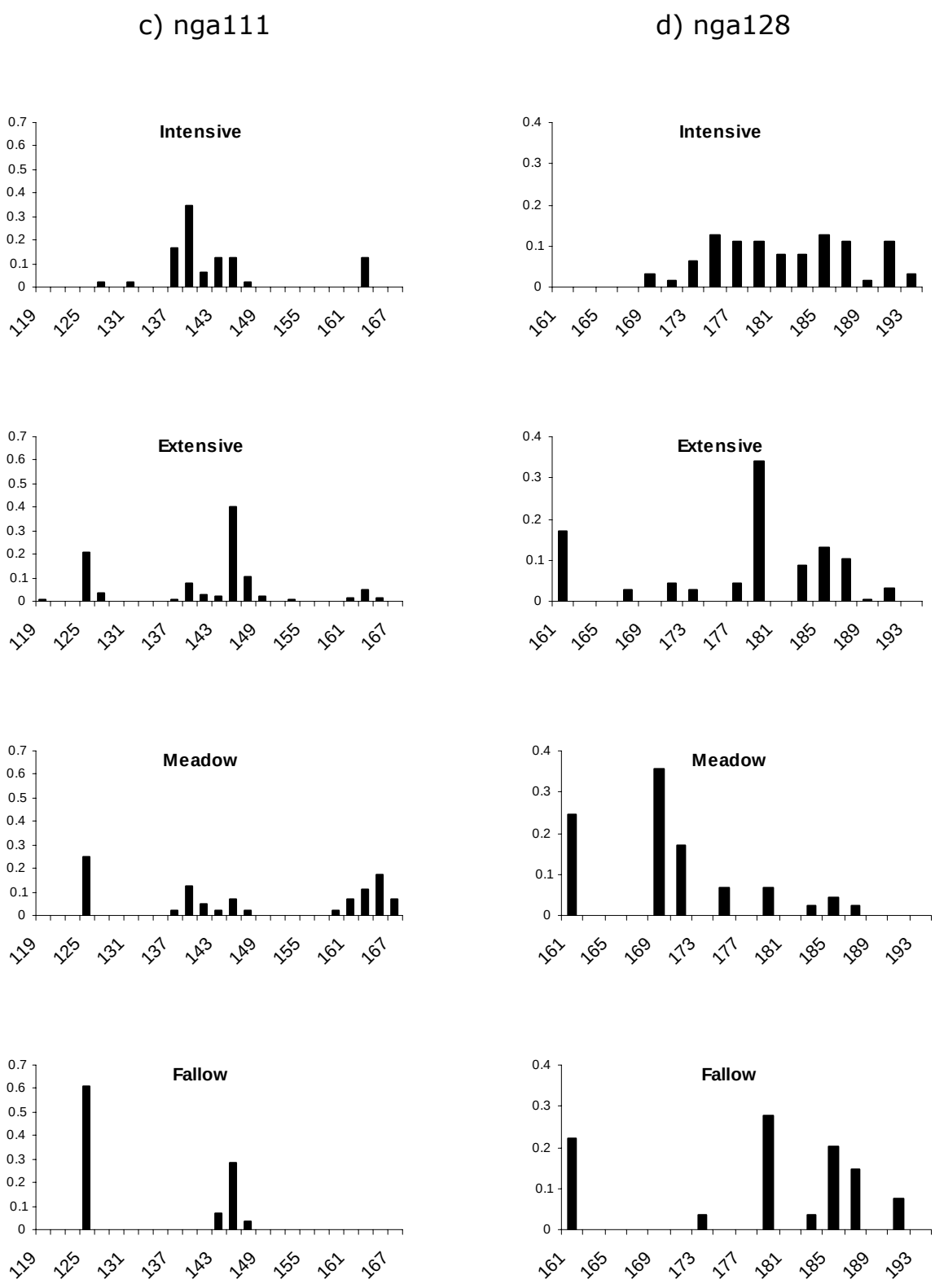


Fig. 4 (cont.): Allele frequency distributions for *A. thaliana* microsatellite loci nga111 (c) and nga128 (d). The *X axis* represents allele size (in base-pairs); the *Y axis* represents allele frequencies.

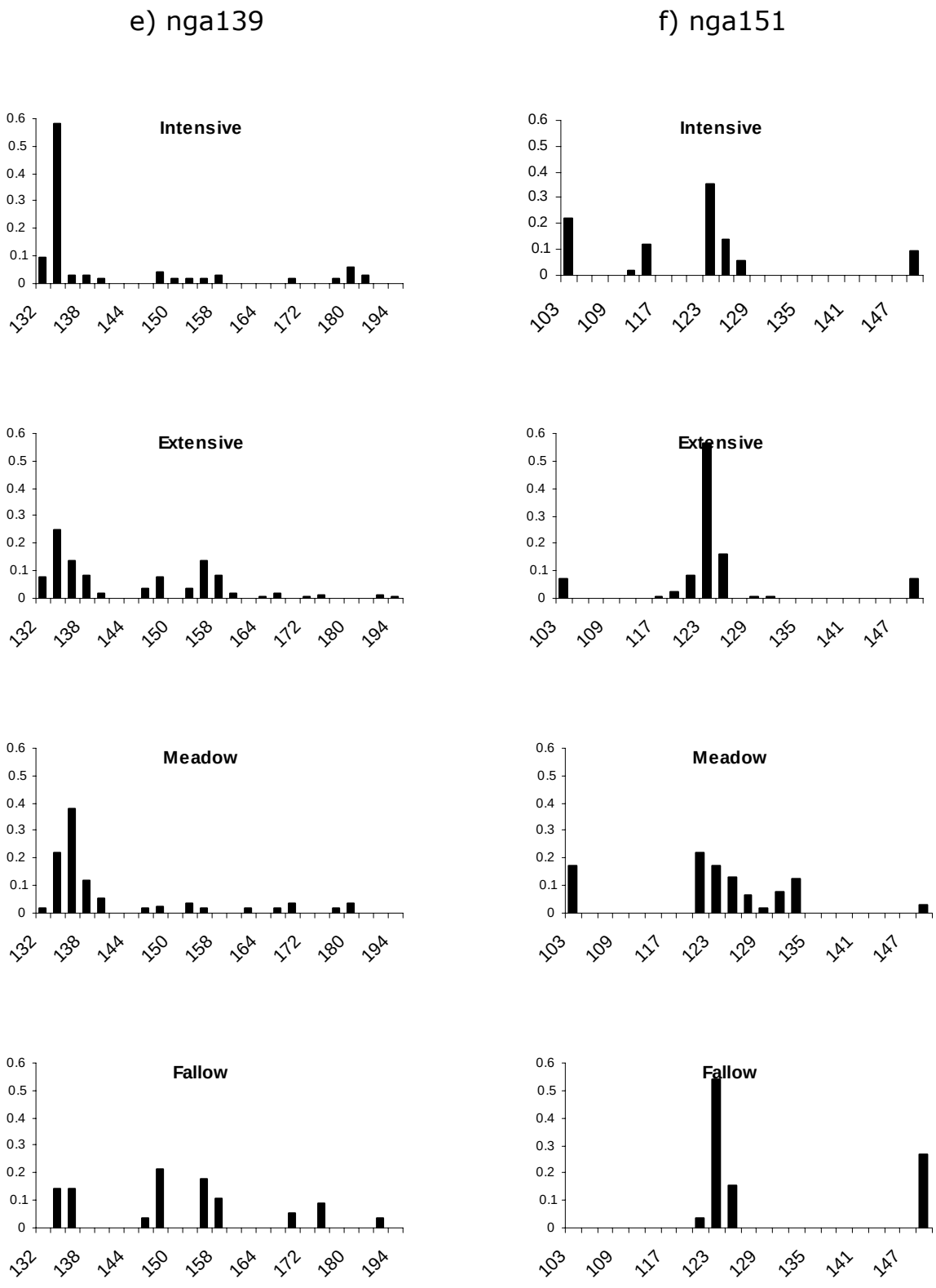


Fig. 4 (cont.): Allele frequency distributions for *A. thaliana* microsatellite loci nga139 (e) and nga151 (f). The *X axis* represents allele size (in base-pairs); the *Y axis* represents allele frequencies.

g) nga168

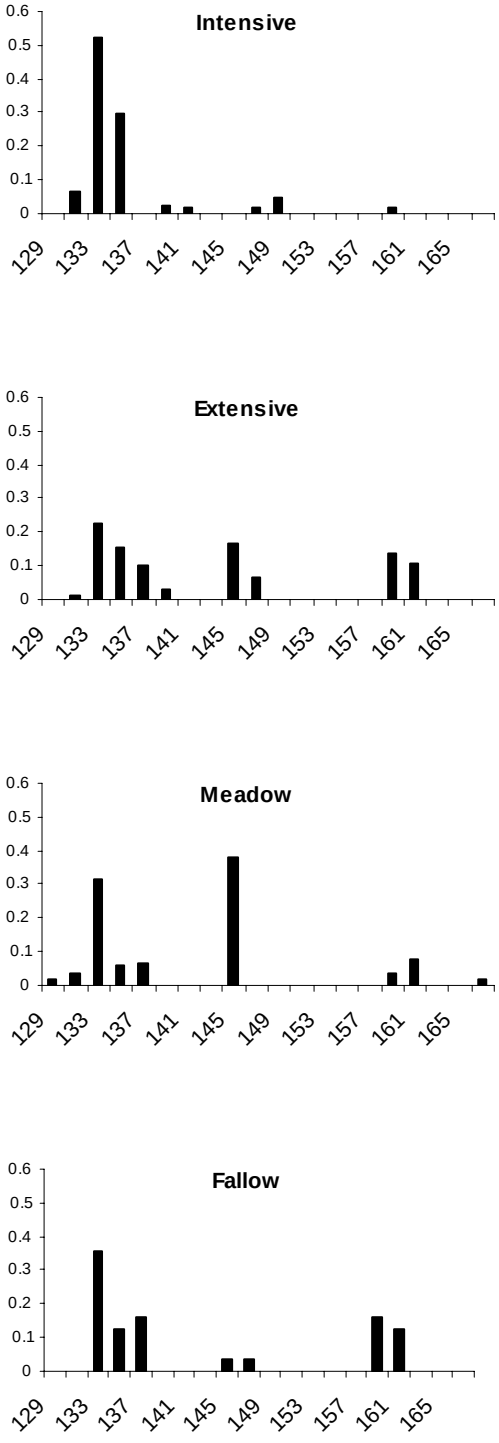


Fig. 4 (cont.): Allele frequency distributions for *A. thaliana* microsatellite locus nga139 (g). The *X axis* represents allele size (in basepairs); the *Y axis* represents allele frequencies.

AFLP Markers

In 283 *A. thaliana* individuals, six primer combinations produced a total of 392 bands, of which 249, or 64%, were polymorphic (Tab. 19). Four of these were singletons (1.6% of the polymorphic bands), however this was reduced to three singletons with the inclusion of the NASC accessions (a singleton is an AFLP band occurring only once from all samples). The number of bands amplified by the different primer pairs varied from 42 to 85, with the degree of polymorphism ranging from 45.5% to 78.8%. The amount of polymorphism is lower than that observed in *A. thaliana* ecotypes using the same restriction enzymes, Miyashita et al. (1999) found 79.2% polymorphism from 38 ecotypes, whereas Breyne et al. (1999) observed 97% polymorphism from 21 ecotypes.

Tab. 19: The six AFLP primer combinations used, the number of bands, and degree of polymorphism in 283 individuals of *A. thaliana*.

Primer Pair	Total Bands	Polymorphic Bands	% Polymorphism
E-CA / M-CC	77	44	57.1
E-CC / M-CA	70	49	70.0
E-CC / M-CG	52	34	65.4
E-CG / M-CC	85	67	78.8
E-CG / M-CT	42	25	59.5
E-CT / M-CA	66	30	45.5
Total	392	249	63.5

4.2.2 Within Population Genetic Diversity

All populations, whether viewed by regime (Tab. 20), or as subpopulations (Tab. 21), showed a significant deviation from HWE for all loci. Only subpopulations with very small sample sizes showed any non-significant results. The subpopulations AIe and IMb were non-significant for one locus ($P > 0.05$), while the subpopulations RFb and Rhön were even less significant ($P > 0.1$ - 0.3) for most loci. Two populations from Amöneburg, AIa and AIc, contained only one individual and so were excluded from analysis in which they would constitute a population by themselves. During analy-

sis by regime they were included into the total intensive population. The populations which do not feature on this list (AId, RFc, IEa, IMd, and IFa), did not produce *A. thaliana* material to be analysed.

The intensive population showed the lowest H_{obs} (0.57) of all the populations by regime (Tab. 20). When only the original intensive samples are considered (removing subpopulation AIf due to the difference in sampling technique, see section 3.2.1), the H_{obs} is reduced to 0.013, a substantially lower H_{obs} than all other populations (and most subpopulations). This result was supported by the H_{exp} of the original intensive population (0.687) being much lower than that of the total intensive population (0.786), indicating that the genetic diversity and the amount of heterozygosity present in the original intensive population is lower than in other populations. However, in contrast to the original intensive population, the intensive subpopulation AIf is relatively heterozygous (0.078), with only two non-fallow subpopulations (REd and IMa) expressing higher heterozygosity (Fig. 21).

The fallow population exhibited more than twice the heterozygosity of any other population (0.160), with the next most heterozygous being the extensive population. Almost all the samples from these two populations originated from the same location, Erda. Therefore, it is possible that the relatively high heterozygosity of the extensive population (0.073) is not because of a higher outbreeding rate amongst individuals over the last 50 years, but due to the legacy of a previously (relatively) highly outbreeding population (the fallow population). This possibility is highlighted by the fact that the meadow population has a lower H_{obs} (0.063) than the extensive population. This result must be viewed in light of the fact that the meadow population incorporates the subpopulation RMa which exhibits no heterozygosity. Without this subpopulation the H_{obs} is slightly higher than that of the extensive population (0.075), while the H_{exp} is lower (0.753). Neither H_{exp} or H_{obs} appear to be closely related to the sample size (n)

(Fig. 5a and 5b), thus the discrepancy in *n* is negligible for the purpose of comparing populations with these two parameters.

Similar trends are seen when examining the other indicators of within-population genetic diversity. The original intensive samples display the lowest *NA* and NA_e of all the populations by regime (6.4 and 3.4, respectively). However, when the subpopulation AIf is incorporated into the intensive population, it exhibits the highest NA_e (5.5). Despite this, the intensive population still displayed the second lowest *NA* (9.9). Care must be taken when interpreting these results as these parameters are closely related to *n* (Fig. 5a, 5b, and 5c). Upon closer examination *NA* appeared to follow a logarithmic function relative to *n*. This observation is in accordance with theories about microsatellites not fitting the IAM, i.e. showing homoplasy. Under an IAM one would expect to see a linear relationship between *n* and *NA* given a constant mutation rate. A similar function, albeit not so clear, was also evident between *n* and NA_e. When the proportion of NA_e to *NA* compared to *n* was calculated, a reverse logarithmic relationship was evident.

The extensive population showed a much lower ratio of *NA* to NA_e in comparison to the intensive population (both the original and total intensive population). This suggests there is comparatively more neutral genetic variability residing in the extensive population, a factor which could become important for the survival of the population during times of environmental change.

Tab. 20: Population statistics by regime for *A. thaliana*.

Population	*n*	H_{exp}	H_{exp} sd	H_{obs}	H_{obs} sd	*NA*	NA_e	NA_e/NA
Intensive	64	0.786	0.040	0.057	0.011	9.9	5.5	0.56
Extensive	135	0.782	0.033	0.073	0.009	12.1	5.1	0.42
Meadow	59	0.791	0.021	0.063	0.012	10.3	4.9	0.48
Fallow	28	0.735	0.047	0.160	0.026	6.7	4.2	0.63
Intensive*	22	0.687	0.042	0.013	0.010	6.4	3.4	0.53
Meadow*	50	0.753	0.032	0.075	0.015	9.0	4.4	0.49

Intensive* - the original intensive population (i.e. excluding subpopulation AIf).

Meadow* - the meadow population from Eibelshausen only (excluding the nine homozygous samples from Erda subpopulation RMa).

Tab. 21: Population statistics by subpopulation for *A. thaliana*.

Population	*n*	H_{exp}	H_{exp} sd	H_{obs}	H_{obs} sd	*NA*	NA_e	NA_e/NA
AIb	16	0.561	0.071	0.009	0.009	4.6	2.6	0.57
AIe	4	0.783	0.025	0.036	0.036	3.4	3.2	0.94
AIf	42	0.757	0.050	0.078	0.016	7.9	4.8	0.61
REa	27	0.702	0.045	0.049	0.016	6.4	3.5	0.55
REb	33	0.715	0.035	0.047	0.014	6.1	3.7	0.61
REc	35	0.738	0.031	0.072	0.017	7.1	4.0	0.56
REd	38	0.786	0.040	0.119	0.020	8.9	5.2	0.60
REe	2	0.000	0.000	0.000	0.000	1.0	1.0	1.00
IMa	26	0.651	0.034	0.099	0.023	6.6	2.9	0.44
IMb	7	0.688	0.051	0.061	0.034	3.7	3.1	0.84
IMc	17	0.674	0.036	0.043	0.019	5.3	3.0	0.57
RMa	9	0.686	0.059	0.000	0.000	4.4	3.5	0.80
RFa	11	0.641	0.105	0.130	0.038	4.3	3.5	0.81
RFb	3	0.533	0.109	0.190	0.088	2.4	2.1	0.88
RFd	14	0.682	0.049	0.176	0.039	5.0	3.2	0.64
Rhön	5	0.483	0.047	0.000	0.000	2.3	1.8	0.78

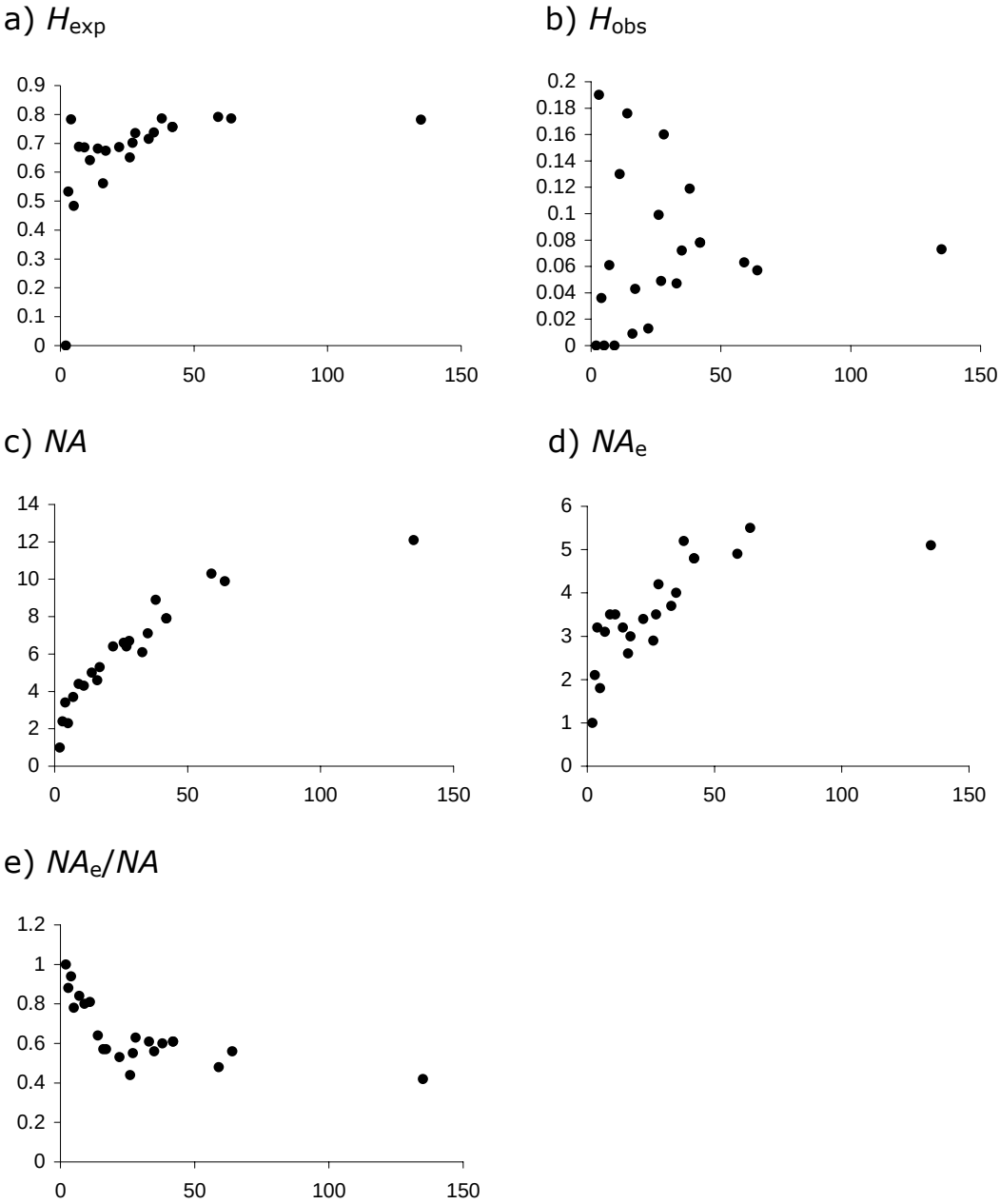


Fig. 5: Relationship between sample size *n* (*x*-axis) and a) H_{exp}, b) H_{obs}, c) NA, d) NA_e, and e) NA_e/NA.

The within-population pairwise differences obtained with AFLP data (Tab. 22) showed similarities to the H_{exp} values obtained with microsatellites: the total meadow and intensive populations exhibited a higher gene diversity (0.227 and 0.231, respectively) than the other two regimes (extensive = 0.220, fallow = 0.209). However, pairwise comparison using microsatellite data (Tab. 22) portrayed a trend more similar to that seen with the observed heterozygosity. The average number of different alleles

between individuals within populations (d-F_{ST}) showed a lower diversity in the intensive population (4.7) compared to the other regimes. Again, this value was reduced considerably by the removal of subpopulation AIf. The average squared difference in allele size between individuals within populations (d-R_{ST}) revealed a comparative lack of diversity in the extensive population (377). This value is equalled or superseded by all other populations. However, from Tab. 23 it can be seen that the subpopulation AIe exhibited a disproportionately high level of within-population allelic diversity. This could be due to the migration of one or more individuals containing very different alleles at the loci studied, or may be due to a non-stepwise mutation event (e.g. unequal crossing over) creating a large difference in allele size between two very closely related individuals. As shown in Fig. 6c the average squared difference in allele size between individuals does not seem to be greatly affected by population size, although as population size decreased the chance of a higher variation increased.

No clear trend was seen between sample size and the other two within-population pairwise differences, d-F_{ST} and gene diversity (Fig. 6a and 6b).

Tab. 22: Gene diversity averaged over all loci for populations by regime.

Population	*n*	Gene Diversity	sd	d-F_{ST}	d-R_{ST}
Intensive	64	0.227	0.110	4.7	389
Extensive	135	0.220	0.106	5.0	377
Meadow	59	0.231	0.112	4.9	403
Fallow	28	0.209	0.104	5.0	485
Intensive*	22	0.223	0.114	4.1	387
Meadow*	50	0.227	0.111	4.6	376

Intensive* - the original intensive population (i.e. excluding subpopulation AIf).

Meadow* - the meadow population from Eibelshausen only (excluding the nine homozygous samples from Erda subpopulation RMa).

Tab. 23: Gene diversity averaged over all loci for subpopulations.

Population	*n*	Gene Diversity	sd	d-F_{ST}	d-R_{ST}
AIb	16	0.212	0.110	3.4	275
AIe	4	0.234	0.154	5.1	700
AIf	42	0.213	0.104	4.6	373
REa	27	0.234	0.116	4.7	304
REb	33	0.199	0.098	4.4	301
REc	35	0.195	0.096	4.6	411
REd	38	0.202	0.099	5.0	397
IMa	26	0.196	0.098	3.8	325
IMb	7	0.188	0.107	4.8	293
IMc	17	0.221	0.113	4.1	403
RMa	9	0.184	0.100	4.7	341
RFa	11	0.200	0.106	4.5	381
RFd	14	0.193	0.100	4.6	470

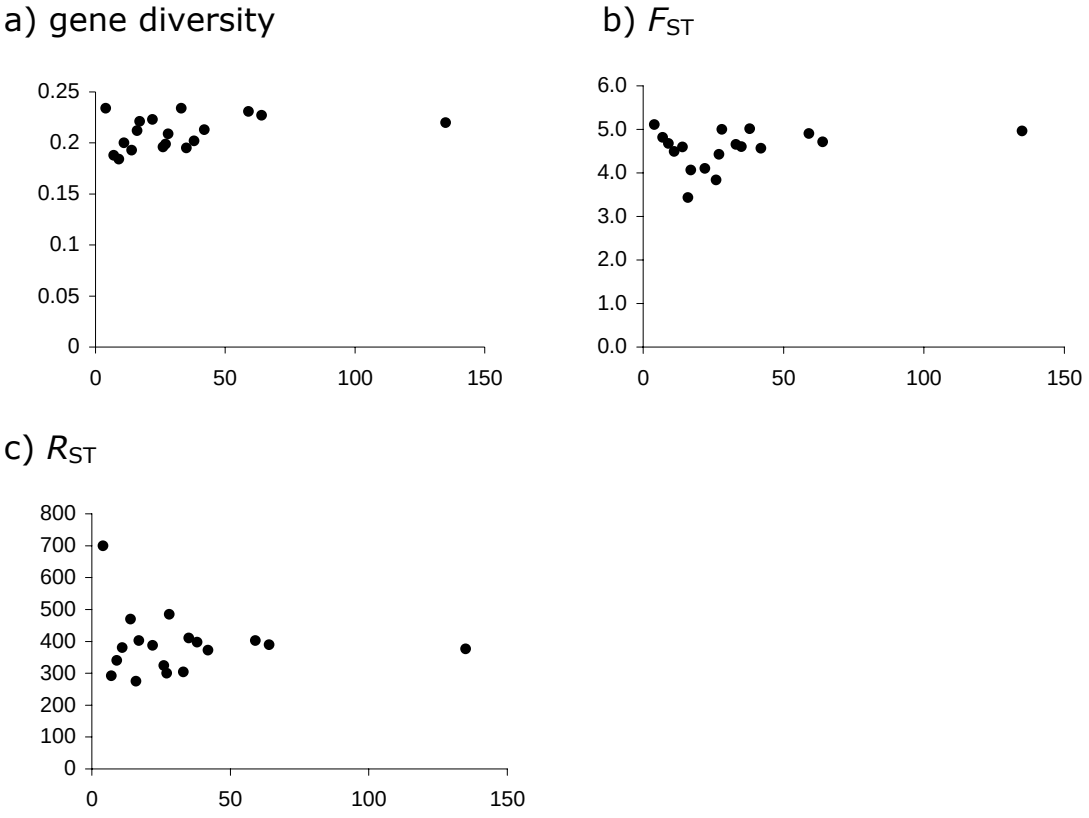


Fig. 6: Relationship between sample size *n* (*x*-axis) and a) gene diversity, b) F_{ST}, and c) R_{ST}.

4.2.3 Population Differentiation and Genetic Distance

AMOVA analysis

The elucidation of population genetic structure in *A. thaliana* was performed using several different techniques. From microsatellite markers, the two distance measures F_{ST} and R_{ST} were used, representing the IAM and SMM, respectively. For the AFLP data, a pairwise distance measure was used.

All measures gave very similar results. The pairwise difference method for both microsatellite and AFLP data gave almost identical results (Tab. 24a, c). Most of the variation, approximately 80%, was contained within populations, suggesting a very localised population structure. Despite this, enough variation at the two higher hierarchical levels was evident to suggest that the subpopulations and populations by regime were significantly different (microsatellite F_{ST} = 0.175, Tab. 24a; AFLP F_{ST} = 0.173, Tab 24c; $P < 0.001$ for both values). Using the sum of squared size difference (R_{ST}; Slatkin, 1995), a slightly higher degree of population differentiation was observed (R_{ST} = 0.189, $P < 0.001$;Tab. 23b).

Under certain conditions, these results would indicate that the populations readily exchange genetic material. While that may be the case (for example seed transfer by birds or agricultural machinery), it must be remembered that many of the populations are separated by 20-30 km and thus such gene flow is very unlikely for such a small, predominantly self-pollinating plant. Therefore, it is possible that these results reflect the recent divergence of the different populations from each other rather than the amount of migration.

Tab. 24: AMOVA results for *A. thaliana* with microsatellite data using (a) pairwise differences, (b) sum of squared size difference, and (c) AFLP data using pairwise distances.

(a)

	d.f.	Sum of Squares	Variance Component	Percentage of Total Variation
Among groups	3	110.25	0.17	6.30
Among populations within groups	8	120.83	0.30	11.15
Within populations	528	1180.16	2.24	82.54
$\boldsymbol{F_{ST}}$ (microsatellite data)	**0.175**			

(b)

	d.f.	Sum of Squares	Variance Component	Percentage of Total Variation
Among groups	3	10555.54	18.54	8.25
Among populations within groups	8	9638.77	23.99	10.67
Within populations	528	96250.55	182.29	81.08
$\boldsymbol{R_{ST}}$ (microsatellite data)	**0.189**			

(c)

	d.f.	Sum of Squares	Variance Component	Percentage of Total Variation
Among groups	3	707.12	2.37	7.61
Among populations within groups	8	711.19	3.02	9.73
Within populations	255	6550.22	25.69	82.66
$\boldsymbol{F_{ST}}$ (AFLP data)	**0.173**			

Population Genetic Distances

Genetic distances between populations were assessed using both pairwise distance (F_{ST}; for both AFLP and microsatellite results) and sum of squared size difference (R_{ST}; microsatellites only).

All the populations were significantly different from each other ($P<0.01$) for both R_{ST} and F_{ST} estimations (Tab. 25). Both distance methods revealed the genetic closeness of the extensive and fallow populations. However, despite these two populations sharing more alleles than with

any other populations (F_{ST} = 0.027), the average difference in allele size between the intensive and extensive samples was the smallest (R_{ST} = 0.059). Both the extensive and fallow populations were made up almost entirely from samples from the same location, Erda, and therefore could be a reflection of their spatial proximity. The similarity between extensive and fallow samples was further supported by the AFLP results (Tab. 26). The genetic distance as measured by F_{ST} was only around half of that separating the next most similar populations (extensive and meadow). However, in contrast to the microsatellite results, the intensive and extensive populations were considered relatively distant from each other (F_{ST} = 0.134), with the only larger genetic distance existing between the intensive and fallow populations (F_{ST} = 0.145). From the microsatellite results, it can be seen that both forms of genetic distance measures find the meadow and fallow populations to be the most distant (F_{ST} = 0.124; R_{ST} =0.213). However, the AFLP results placed the intensive and fallow populations as the most distant (F_{ST} = 0.145)

Tab. 25: Genetic distance by regime as measured by F_{ST} (below diagonal) and R_{ST} (above diagonal) values for *A. thaliana* microsatellites.

	Intensive	Extensive	Meadow	Fallow
Intensive	-	0.059	0.127	0.147
Extensive	0.076	-	0.114	0.076
Meadow	0.117	0.105	-	0.213
Fallow	0.113	0.027	0.124	-

Tab. 26: Genetic distance by regime as measured by F_{ST} values for *A. thaliana* AFLPs.

	Intensive	Extensive	Meadow	Fallow
Intensive	-			
Extensive	0.134	-		
Meadow	0.116	0.083	-	
Fallow	0.145	0.047	0.104	-

Tab. 27: Genetic distance as measured by F_{ST} (below diagonal) and R_{ST} (above diagonal) values for *A. thaliana* microsatellites populations (* = $P > 0.05$). Shaded/non-shaded cells indicate the intersection of the different regimes.

	AIb	AIe	AIf	REa	REb	REc	REd	IMa	IMb	IMc	RMa	RFd	RFa
AIb	-	0.31	0.10	0.10	0.10	0.08	0.14	0.26	0.41	0.26	0.24	0.34	0.23
AIe	0.21	-	0.27	0.34	0.30	0.24	0.19*	0.38	0.45	0.30	0.31	0.34	0.17*
AIf	0.19	0.12*	-	0.10	0.16	0.12	0.07	0.12	0.35	0.21	0.13	0.24	0.26
REa	0.21	0.14	0.14	-	0.09	0.06	0.04*	0.19	0.33	0.29	0.05*	0.23	0.15
REb	0.19	0.17	0.14	0.09	-	0.06	0.10	0.20	0.35	0.27	0.16	0.29	0.15
REc	0.23	0.15	0.12	0.07	0.09	-	0.07	0.15	0.21	0.19	0.08*	0.17	0.16
REd	0.16	0.12	0.12	0.06	0.05	0.08	-	0.16	0.29	0.25	0.03*	0.17	0.10
IMa	0.32	0.27	0.25	0.25	0.27	0.22	0.19	-	0.21	0.08	0.22	0.25	0.38
IMb	0.34	0.20	0.19	0.22	0.21	0.21	0.18	0.22	-	0.13	0.38	0.37	0.48
IMc	0.29	0.23	0.14	0.23	0.22	0.21	0.17	0.16	0.18	-	0.33	0.33	0.43
RMa	0.25	0.18	0.15	0.07*	0.10	0.05*	0.08	0.26	0.23	0.27		0.07*	0.13
RFd	0.29	0.21	0.16	0.13	0.14	0.11	0.10	0.24	0.20	0.19	0.10	-	0.29
RFa	0.23	0.18	0.18	0.07	0.08	0.12	0.07	0.31	0.28	0.28	0.12	0.18	-

Tab. 28: Genetic distance by subpopulation as measured by F_{ST} values for *A. thaliana* AFLPs.

	AIb	AIe	AIf	REa	REb	REc	REd	IMa	IMb	IMc	RMa	RFd	RFa
AIb	-												
AIe	0.08*	-											
AIf	0.14	0.08	-										
REa	0.14	0.10	0.18	-									
REb	0.17	0.15	0.20	0.08	-								
REc	0.18	0.17	0.21	0.09	0.08	-							
REd	0.17	0.14	0.19	0.09	0.07	0.10	-						
IMa	0.18	0.18	0.23	0.21	0.19	0.21	0.16	-					
IMb	0.18	0.17	0.21	0.18	0.14	0.17	0.17	0.18	-				
IMc	0.15	0.11	0.16	0.17	0.17	0.20	0.15	0.16	0.14	-			
RMa	0.22	0.23	0.23	0.13	0.10	0.11	0.09	0.22	0.19	0.17	-		
RFd	0.23	0.22	0.21	0.15	0.12	0.12	0.13	0.24	0.22	0.21	0.16	-	
RFa	0.17	0.14	0.17	0.09	0.05	0.10	0.05	0.18	0.15	0.16	0.11	0.11	-

As expected, the genetic distances between subpopulations followed a similar trend to the population by regime results. Assessing the microsatellite results, the subpopulations AIb and AIf exhibited a much higher similarity to the extensive populations than AIe using R_{ST} values (Tab. 27). However, the F_{ST} values indicate that the subpopulations AIe and AIf are closer to the extensive populations than AIb. It must be remembered that AIe consisted of only four individuals, and it can be seen that some of the results are not significant ($P \geq 0.05$).

The genetic similarity due to proximity can be seen by the values for the Erda subpopulation RMa. Both the F_{ST} and R_{ST} results suggest that this subpopulation is more similar genetically to the fallow and extensive populations (all samples coming from Erda) than to the other meadow populations (which originated from Eibelshausen).

Although not reflecting the general trend, there are numerous examples where subpopulations from different locations or regimes are shown to be more similar than to other subpopulations from the same location or regime. The AIb subpopulation is, for example, as close or closer to most extensive subpopulations than to AIe. This results holds for both F_{ST} and R_{ST} values. A similar situation occurs for subpopulation IMa. Although being very close to IMc ($R_{ST} = 0.08$), it is regarded as being closer to all the extensive subpopulations than to IMb when considering the R_{ST} values. These results are less clear when viewing the F_{ST} values, however, from which it can be seen that the meadow subpopulations are as similar to each other as to the subpopulations from the extensive locations.

A similar situation occurs with the fallow subpopulations. Both these subpopulations are closer to all the other subpopulations from Erda (from both the extensive and meadow regimes) than to each other. The subpopulations which show the most similarity to each other within any of the regimes are those from the extensive regime.

These results are somewhat, but not entirely, supported by the AFLP analysis (Tab. 28). Here we see that the intensive subpopulations are, bar one exception, as close or closer to each other than the extensive sub-

populations. The subpopulation RMa exhibits the same trend of being closer to all of the Erda subpopulations (for both the extensive and meadow regimes) than to the other meadow subpopulations.

The extensive subpopulations are confirmed to be the most similar within a single regime, with no F_{ST} value exceeding 0.10. However, the difference between the fallow subpopulations and the other Erda subpopulations is not so clear with AFLPs as with the microsatellite results. Only RFa exhibits a closer relationship to the other Erda populations with microsatellites, while RFd is as close or closer to RFa than to the other Erda subpopulations.

In general, the subpopulation genetic distances show that the geographical distance between subpopulations is not necessarily indicative of their genetic distances. These results support the AMOVA analysis which showed that, while significant, the genetic differentiation between subpopulations is not large and that most of the genetic differentiation occurs among individuals within subpopulations.

Bottleneck Detection

All tests provided in the BOTTLENECK software were used for bottleneck detection in the four regime populations: Intensive, Extensive, Meadow, and Fallow. The intensive population was also split into the original samples and the subpopulation AIf for analysis. None of the populations showed any significant bottleneck signatures (heterozygosity excess) for the Sign test. The Standardised Difference test was disregarded due to the fact that it requires a minimum of 20 loci. For the Wilcoxon test, and under the assumption of the Infinite Allele Model (IAM), all populations except intensive showed evidence indicating the presence of a bottleneck event ($0.05 > P > 0.01$). However, as microsatellite loci are generally thought to follow the Stepwise Mutation Model (SMM), the results from the Wilcoxon test must be treated with caution before being used as an accurate indication of a bottleneck event. This is especially true for inbreeding populations which do not follow HWE.

Tab. 29: Bottleneck detection for *A. thaliana* populations using *M*, the ratio of number of alleles (*k*) to range in allele size (*r*).

Population	*n*	*M*
Intensive	64	0.29
Intensive*	22	0.19
Extensive	135	0.32
Meadow	59	0.27
Meadow*	50	0.25
Fallow	28	0.21

Intensive* - the original intensive population (i.e. excluding subpopulation AIf).

Meadow* - the meadow population from Eibelshausen only (excluding the nine homozygous samples from Erda subpopulation RMa).

The method of Garza (Garza and Williamson 2001) was also employed to ascertain if bottleneck events have been present. The value *M* was calculated for each population as well as for the subpopulation AIf and the original intensive population (Tab. 29). With the subpopulation AIf removed from the intensive population it can be seen that the original intensive population exhibits a lower *M* (0.19) than all other populations tested. Only the fallow population (0.21) was close to the value of the original intensive samples. It can be seen that the high *M* value for the intensive population is due to the incorporation of the subpopulation AIf. Thus, this method suggests that relative to the other populations, both the original intensive and fallow populations have been through a bottleneck event or events. However, there is a strong relationship between sample size and *M* (Fig. 7), indicating that it would be premature to assume that the populations with a lower *M* value have been through a bottleneck.

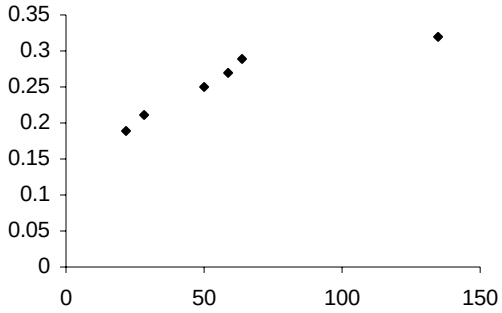


Fig. 7: Relationship between *M* and sample size (r^2 = 0.82).

Linkage Disequilibrium

All four populations showed significant linkage disequilibrium amongst all loci. However, this test is conducted under the assumption that the populations are in Hardy-Weinberg equilibrium. In fact these populations show a very large and significant deviation from HWE. Presumably this is the reason why multiple loci on three different chromosomes were considered linked.

4.2.4 Principle Coordinate Analysis (PCA)

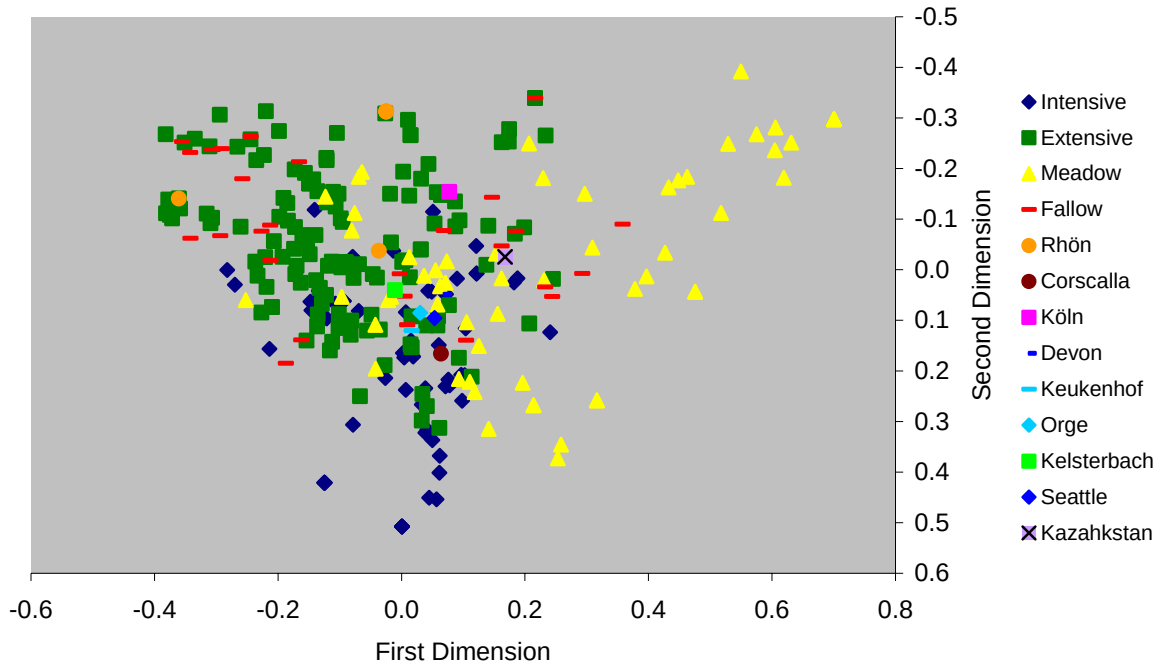
Diagrams of the PCA of pairwise similarities between all the examined *A. thaliana* individuals were carried out assessing the populations by regime and by location for both microsatellite and AFLP data.

A very high dimensionality was observed for the microsatellite data (Fig. 8a and 8b; both figures use the same raw data), with 23 axes being required to explain 50% of the variation. The first two axes represented only 10.5% and 4.4% of the variation, respectively. To account for 95% of the variation, a total of 197 axes were required. The lack of any strong differentiating factor is apparent from the overlapping of many of the individuals from different regimes or locations.

A higher degree of differentiation is seen with the AFLP results (Fig. 9a and 9b; both figures use the same raw data). This is confirmed by the first axis accounting for 59.3% of the total variation. However, the second axis represented only 2.0%, and a total of 121 axes were required to explain 95% of the variation.

It can be seen from the graphs that in both cases individuals are differentiated by location more than by regime. Given the higher level of differentiation from the AFLP data, this would suggest that the first axis is differentiating on the basis of location. That there is a very high dimensionality for the remaining axes indicates that there are no other significant factors separating the individuals. The very high dimensionality of the microsatellite data (and hence the large overlapping) suggests that neither location or regime can be used to accurately differentiate individuals.

(a)



(b)

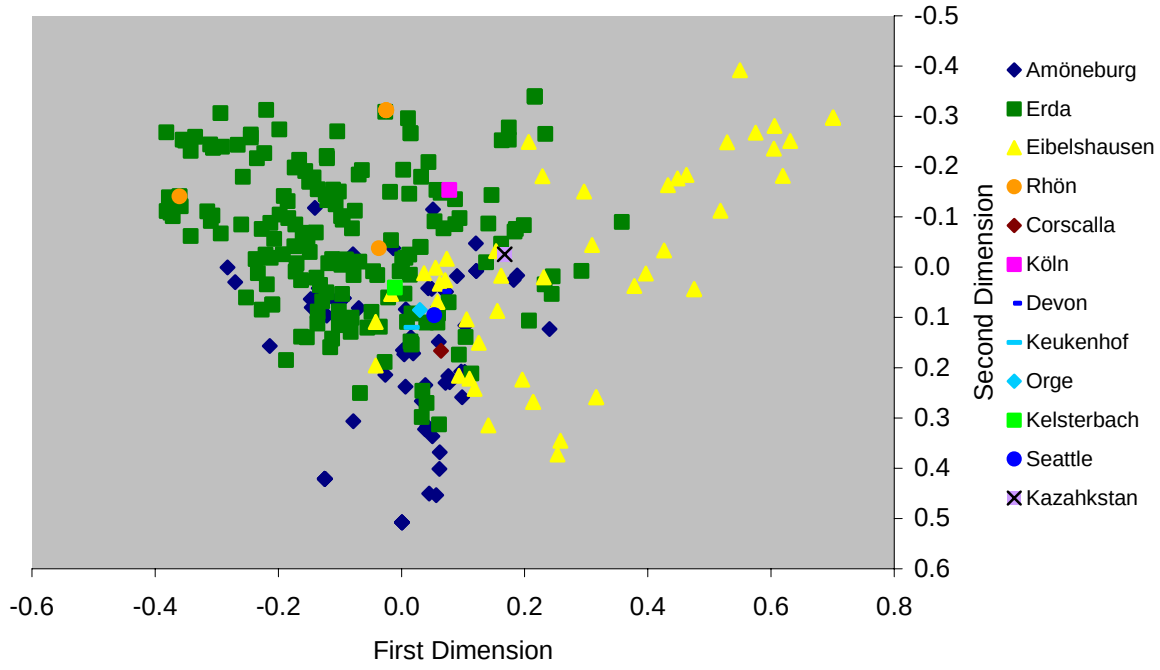
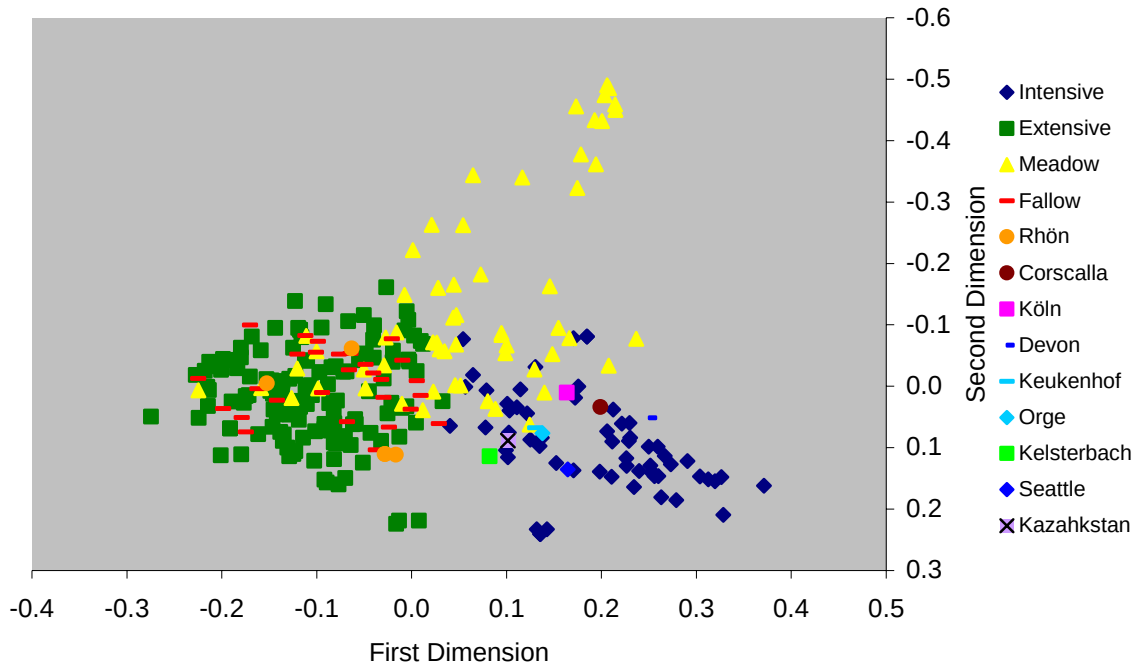


Fig. 8 (a, b): Principle coordinate analysis of *A. thaliana* using microsatellite data by (a) regime, and (b) location.

(a)



(b)

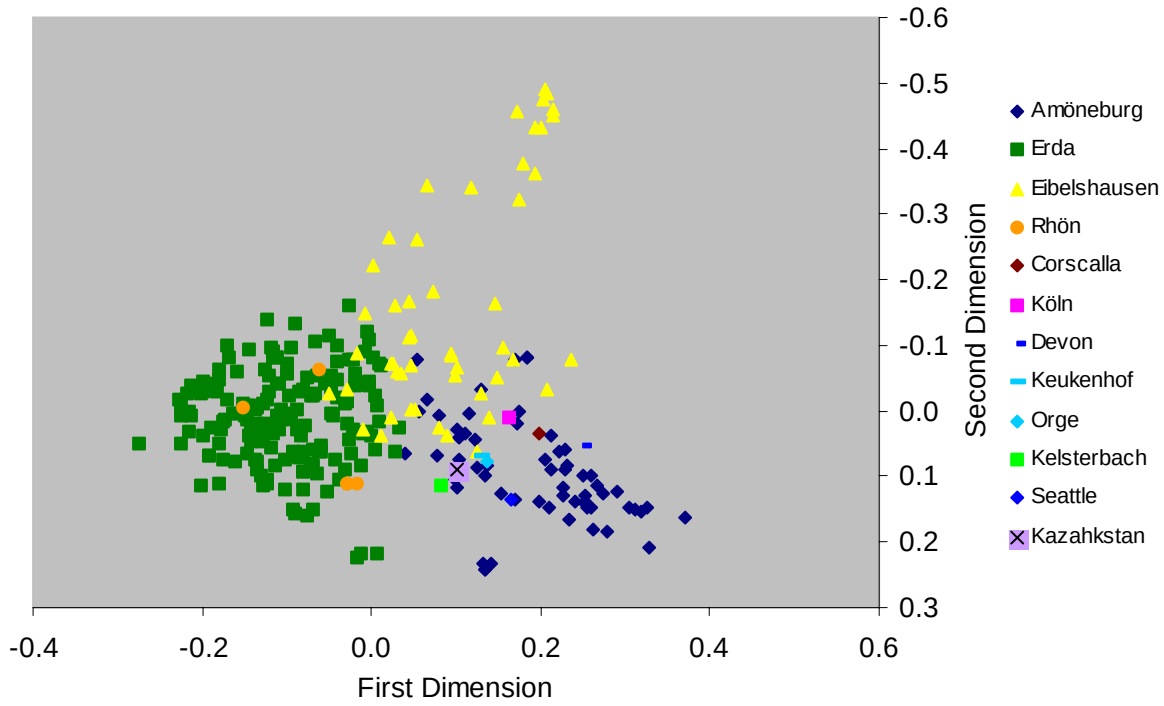


Fig. 9 (a, b): Principle coordinate analysis of *A. thaliana* using AFLP data by (a) regime, and (b) location.

4.2.5 Cluster Analysis

Cluster analysis diagrams of all *A. thaliana* individuals (classified by subpopulation) can be found in Appendix I (AFLP results) and Appendix II (microsatellite results).

In general, the AFLP cluster analysis confirms the trends seen in the PCA diagrams. Most of the individuals from Erda, whether from the extensive, meadow or fallow regimes, cluster together with few exceptions. The three subpopulations from Eibelshausen (IMa, IMb, and IMc) grouped separately. In general, the individuals from each subpopulation clustered together except for IMc, which displayed two separate groups. Most of the intensive samples clustered together, with a few individuals from the subpopulations clustering amongst other regimes and subpopulations. Seven out of the eight accessions obtained from the NASC clustered together, despite representing a global geographic distribution.

Less clarity in the clustering was seen when using microsatellites, a probable effect of homoplasy. The eight NASC accessions were spread throughout the clustering, as were many individuals from various subpopulations. There was little clustering of individuals purely comprising the same subpopulation, as in the AFLP clustering. Many individuals from the subpopulation AIb clustered together, in contrast to subpopulation AIf which had many small groups of individuals distributed throughout the dendrogram, indicating that the samples in this subpopulation have different pedigrees. This is in contrast to the Eibelshausen samples where many individuals from all three subpopulations group clearly, implying that they have all descended from a common ancestor.

In general these results indicate that, for this study, the assignment of a particular *A. thaliana* genotype to any of the subpopulations would be better achieved using AFLPs markers than microsatellites.

4.2.6 Mantel Test

Mantel tests were conducted with AFLP and microsatellite data using F_{ST} values as the measure of genetic distance. For microsatellite data the R_{ST} values were also used.

All measures showed a positive, but small, correlation between genetic and geographic distances. The AFLP results (Fig. 10c) showed the greatest correlation between these two factors (r^2 = 0.45). The F_{ST} values for the microsatellites (Fig. 10a) gave similar results (r^2 = 0.40), with the R_{ST} values (Fig. 10b) being considerably lower (r^2 = 0.22).

In all of the graphs it can be seen that, despite the overall general trend of increased genetic distance with increased geographic distance, a wide range of genetic distances are present at any level of geographic distance. Indeed, several populations which are within a couple of kilometres of each other are genetically more distant than many populations 20-40 kilometres from each other.

(a)

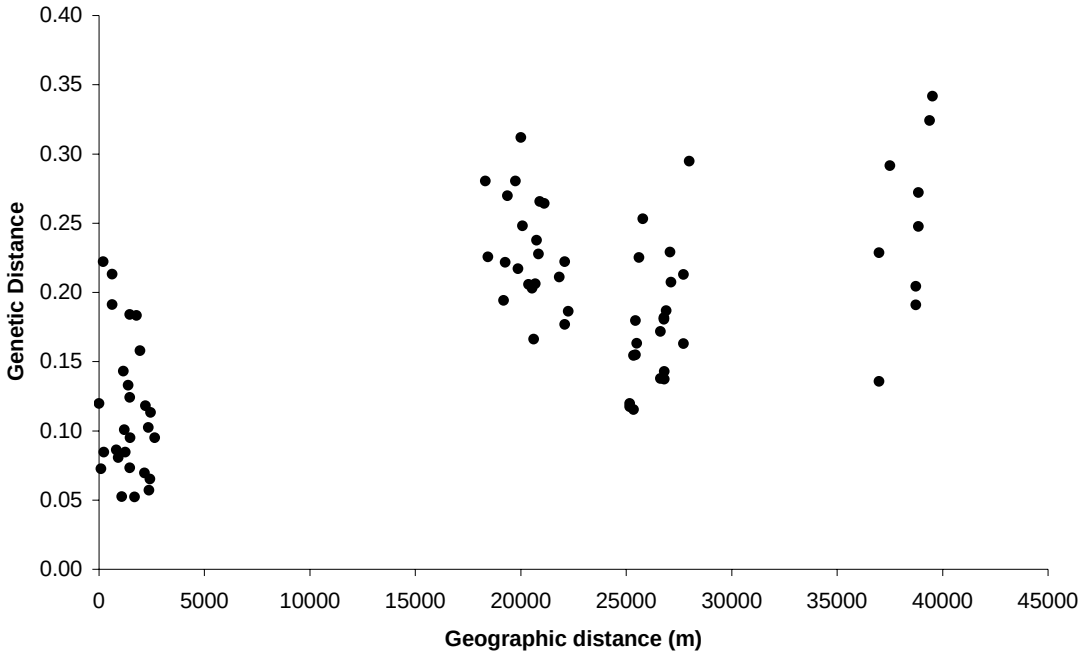
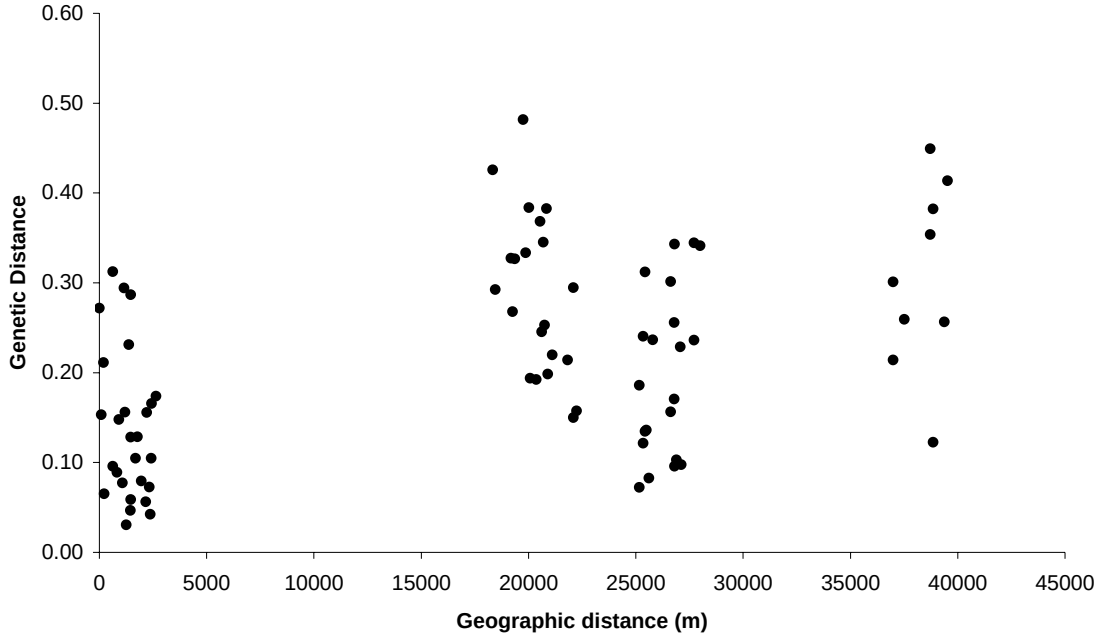


Fig. 10 a: Correlation between genetic and geographical distances for all *A. thaliana* subpopulations using microsatellite data (F_{ST} values, r^2 = 0.40).

(b)



(c)

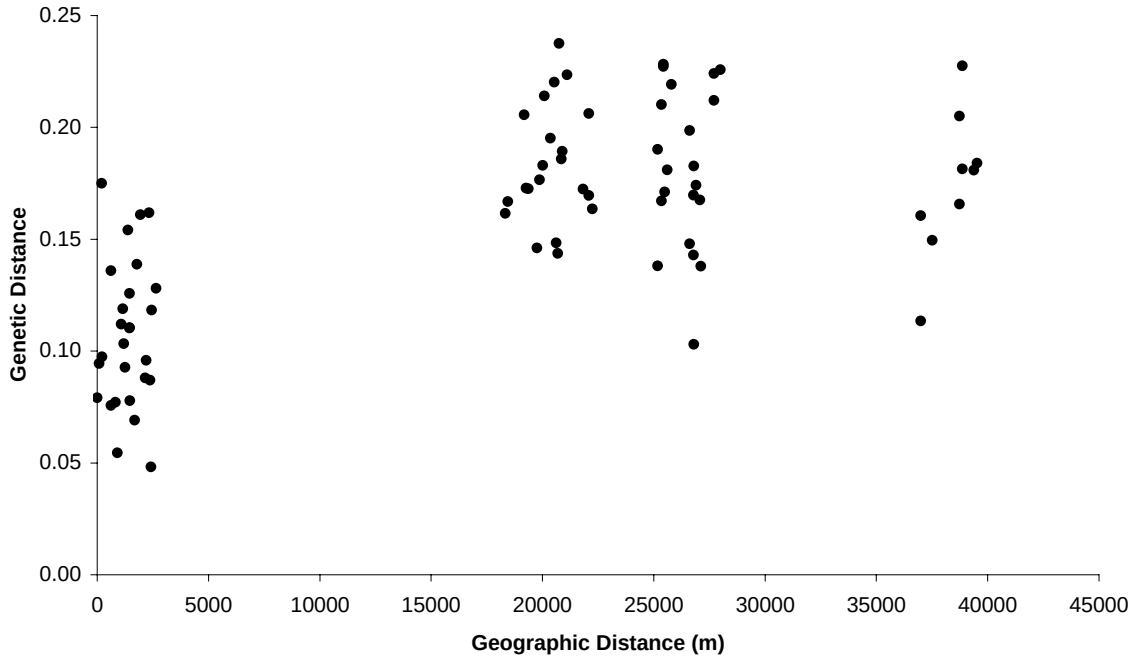


Fig. 10 b and c: Correlation between genetic and geographical distances for all *A. thaliana* subpopulations using (b) microsatellite data (R_{ST} values, $r^2 = 0.22$), and (c) AFLP data (F_{ST} analogue values, $r^2 = 0.45$).

4.2.7 Migration Rates

The number of migrants exchanged between populations per generation, *m*, was calculated by regime (Tab. 30) and between subpopulations (Tab. 31). The results must be treated with caution, however, as the underlying assumptions with the model are that the populations are in migration-drift equilibrium and that the mutation rate is negligible. Given that microsatellite data was used, this last assumption probably does not hold. Because the values of *m* have been calculated using F_{ST} and R_{ST} measurements, they reflect the same pattern observed in the genetic distances calculated using the same measurements.

The migration rates are seemingly very high, with the extensive and fallow populations apparently exchanging nine migrants per generation. This, however, seems unlikely given that the samples obtained from the fallow population are at the youngest 50 years old, compared to the extensive population where the majority of the seeds probably came from the transient seed bank. Of course it is possible that some of the extensive samples came from seeds 50 years or older, however in the extensive fields, which have been tilled regularly and have maintained *A. thaliana* populations for the past 50 years, this should be a rare event.

The geographical distances between the sample sites must also be taken into account. At an average distance of 26 km between the intensive and extensive sample sites, it is difficult to imagine three (F_{ST} value) or four (R_{ST} value) migration events occurring in a species which is predominantly selfing, especially when the sample sizes are considered. However, being agricultural sites it must be considered that pollen and seeds may be carried of between sites by birds or agricultural machinery. This concept is especially relevant to the sample sites at any one particular location where a large amount of soil, fauna and floral transfer is likely to occur through the use of machinery.

Therefore, there is likely to be a higher occurrence of such migration events than in undisturbed habitats. However, subpopulation values of between approximately two and six migrants per generation (extensive

subpopulations) seem unlikely in samples sizes of only around 30 individuals with a very low outcrossing rate. The genetic similarity between subpopulations is therefore likely to be more a legacy of similarity from the founders of these subpopulations than due to gene flow from migration. This is illustrated by the number of migrants exchanged between subpopulations RFa and REd. The former subpopulation is assumed to consist of individuals from seeds which were produced 50 years ago or more, while the latter subpopulation has evolved through at least 50 more generations than RFa. Therefore, the value of 3.58 migrants per generation seems very unlikely and most probably represent the similarity of the populations to each other, either through convergence due to mutational homoplasy, as a representation of the amount of mutation and genetic drift since the subpopulations diverged, or, most likely, a combination of both.

Tab. 30: Estimates of *m* from microsatellite data for population by regime (F_{ST} below diagonal, R_{ST} above diagonal).

	Intensive	Extensive	Meadow	Fallow
Intensive	-	4.00	1.71	1.45
Extensive	3.03	-	1.93	3.03
Meadow	1.89	2.14	-	0.93
Fallow	1.97	9.00	1.77	-

Tab. 31: Estimates of *m* from microsatellite data for subpopulations (F_{ST} below diagonal, R_{ST} above diagonal). Shaded cells indicate the intersection of the different regimes. Shaded/non-shaded cells indicate the intersection of the different regimes.

	AIb	AIe	AIf	REa	REb	REc	REd	IMa	IMb	IMc	RMa	RFd	RFa
AIb	-	0.55	2.36	2.31	2.18	2.77	1.59	0.73	0.35	0.71	0.81	0.48	0.84
AIe	0.92	-	0.67	0.48	0.58	0.79	1.09	0.40	0.31	0.58	0.55	0.48	1.21
AIf	1.06	1.84	-	2.36	1.35	1.81	3.21	1.79	0.46	0.92	1.61	0.81	0.73
REa	0.96	1.50	1.57	-	2.56	4.18	5.64	1.04	0.50	0.60	5.09	0.83	1.38
REb	1.09	1.20	1.56	2.65	-	3.99	2.14	1.01	0.47	0.68	1.35	0.60	1.44
REc	0.86	1.37	1.91	3.34	2.38	-	3.58	1.42	0.92	1.05	2.98	1.26	1.35
REd	1.28	1.84	1.88	4.13	4.54	2.70	-	1.34	0.60	0.77	7.91	1.19	2.13
IMa	0.52	0.67	0.76	0.76	0.69	0.87	1.09	-	0.93	2.90	0.89	0.74	0.40
IMb	0.48	0.97	1.06	0.90	0.96	0.93	1.16	0.87	-	1.69	0.40	0.43	0.27
IMc	0.61	0.84	1.59	0.86	0.88	0.96	1.25	1.33	1.11	-	0.52	0.51	0.34
RMa	0.74	1.14	1.36	3.16	2.23	4.51	2.71	0.70	0.85	0.68	-	3.19	1.70
RFd	0.60	0.92	1.28	1.63	1.50	1.96	2.38	0.80	0.98	1.04	2.19	-	0.62
RFa	0.84	1.14	1.13	3.20	2.85	1.87	3.58	0.55	0.64	0.64	1.76	1.11	-

4.2.8 Estimation of θ

The population parameter θ (equal to $4N\mu$) was calculated according to Chakraborty and Weiss (1991) for both the Infinite Alleles Model and the Single-Step Mutation Model from expected homozygosity values, hence the terminology θ_{Hom}.

The extensive population clearly showed a much higher θ_{Hom} value than the other populations (Tab. 32). This is in part due to the higher sample size, however it can be seen from the meadow population that even with a lower sample size, its θ_{Hom} is substantially greater than that of the intensive population. The fallow population exhibits a θ_{Hom} value similar to that of the intensive population, despite comprising less than half the number of individuals.

The subpopulations show a similar trend, with the extensive subpopulations all exhibiting a substantially higher θ_{Hom} value than any other subpopulation (Tab. 33). The intensive subpopulations, except for AIe (which comprises of only four samples), exhibit higher θ_{Hom} values compared to the meadow and fallow subpopulations, with the exception of subpopulation IMa.

As θ_{Hom} is a product of the relationship between effective population size and mutation rate, if the mutation rate is equal (or similar) for all populations then it can be used as a guide to estimating effective population sizes. As the effective population size is an indicator as to how well any given population can withstand random events (such as genetic drift), it can be seen that the extensive population is in a much more stable position than the intensive population with respect to future survival. This is both due to the abundance of individuals and the genetic make-up of the populations.

Tab. 32: θ_{Hom} values for population by regime.

		Mutation Model	
Population	*n*	IAM	SSM
Intensive	64	60.62	2015.63
Intensive*	22	36.62	776.34
Extensive	135	254.58	33166.10
Meadow	59	76.44	3147.30
Fallow	28	52.14	1512.00

Intensive* - the original intensive population (i.e. excluding subpopulation AIf).

Tab. 33: θ_{Hom} values for all subpopulations.

		Mutation Model	
Population	*n*	IAM	SSM
AIb	16	23.40	340.24
AIe	4	7.08	43.06
AIf	42	30.75	561.27
REa	27	43.33	1064.93
REb	33	60.21	1989.56
REc	35	72.57	2847.26
REd	38	76.27	3133.17
IMa	26	26.10	414.97
IMb	7	15.62	165.12
IMc	17	18.15	215.36
RMa	9	14.44	144.00
RFd	14	20.95	278.57
RFa	11	20.43	266.31

4.3 Viola arvensis

4.3.1 Viola arvensis markers

Seven primer combinations analysed with 172 individuals gave a total of 203 fragments (Tab. 34). From these 132 were polymorphic (65%) and eighteen were singletons (8.9%). The total number of bands observed for each primer pair ranged from 22 to 46. These values were much lower than those for *A. thaliana*, due to the fact that a +3-primer was used for the selective amplification instead of a +2-primer. The degree of polymorphism was very similar (slightly higher) to that observed in *A. thaliana*, ranging from 51.6% to 86.4% with an average of 65.0%.

Tab. 34: Seven AFLP primer combinations used for analysis of *V. arvensis*, the number of bands and degree of polymorphism.

Primer Pair	Total Bands	Polymorphic Bands	% Polymorphism
E-ACT / M-ATC	28	20	71.4
E-ACT / M-ACG	22	19	86.4
E-ACT / M-ACT	31	16	51.6
E-ATC / M-AGG	23	14	60.9
E-ATC / M-ACG	28	16	57.1
E-ATC / M-ATC	25	17	68.0
E-ATC / M-ATT	46	30	65.2
Total	203	132	65.0

4.3.2 Within-Population Genetic Diversity

The within-population genetic diversity, as measured by gene diversity, ranged from 0.171 (extensive) to 0.211 (meadow) (Tab. 35). These values show a similar range of gene diversity within populations as exhibited by *A. thaliana*. Although again the standard deviation for all gene diversity values are high, a trend can be seen in which the extensive subpopulations show a consistently lower average gene diversity than their intensive counterparts (Tab. 36). There appeared to be little or no correlation between sample size and gene diversity (Fig. 11).

Tab. 35: Gene diversity averaged over all loci for populations by regime.

Population	*n*	Gene Diversity	sd
Intensive	59	0.209	0.103
Extensive	50	0.171	0.085
Meadow	30	0.211	0.106
Fallow	33	0.196	0.098

Tab. 36: Gene diversity averaged over all loci for all subpopulations.

Population	*n*	Gene Diversity	sd
AIa	29	0.209	0.105
AIb	12	0.195	0.104
AIc	8	0.193	0.108
AIe	10	0.178	0.097
REa	19	0.163	0.088
REb	11	0.166	0.086
REc	15	0.155	0.081
REd	5	0.164	0.102
IMc	22	0.202	0.102
RMa	5	0.168	0.105
RFc	14	0.174	0.091
RFd	15	0.177	0.093
IFa	4	0.169	0.114

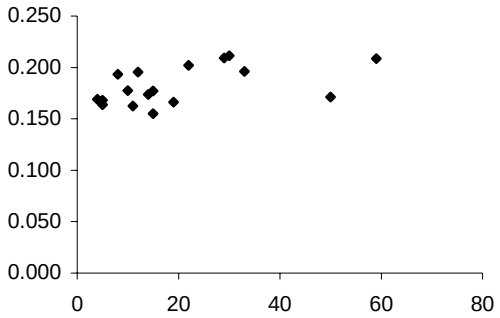


Fig. 11: Relationship between sample size *n* (*x*-axis) and gene diversity.

4.3.3 Population Differentiation and Genetic Distance

AMOVA analysis

The population structuring represented through AMOVA analysis showed that, as with *A. thaliana*, the vast majority of the genetic variation (82.88%) was explained by differences between individuals within populations (Tab. 37). Only a small percentage of the total variation (8.03%) was found between the different agricultural regimes. That only 9.09% of the total variation was explained among subpopulations from the same regime indicates a lack of metapopulation structure.

Tab. 37: AMOVA results for *V. arvensis* with AFLP data.

	d.f.	Sum of Squares	Variance Component	Percentage of Total Variation
Among groups	3	248.07	1.17	8.03
Among populations within groups	9	245.24	1.33	9.09
Within populations	157	1898.69	12.09	82.88
F_{ST}	**0.171**			

Population Genetic Distances

The pairwise genetic distance measures (Tab. 38) between populations showed a similar pattern to that observed for *A. thaliana*: the largest distances were observed between the intensive and extensive populations (0.145), and the intensive and fallow populations (0.135). The extensive and fallow populations were easily the most similar (0.037). However, contrary to the *A. thaliana* results, the meadow population was more similar to fallow for *V. arvensis* than to the extensive population.

Tab. 38: Genetic distance between regimes as measured by pairwise distance.

	Intensive	Extensive	Meadow	Fallow
Intensive	-			
Extensive	0.145	-		
Meadow	0.110	0.132	-	
Fallow	0.135	0.037	0.088	-

The genetic distances between subpopulations revealed that subpopulations from the same location were more closely related to each other than subpopulations from the same regime (Tab. 39). This is exemplified by the distance between RMa and IMc (F_{ST} = 0.20), the two meadow subpopulations from Erda and Eibelshausen, respectively. The fallow subpopulation IFa is very close to IMc (F_{ST} = 0.06), while being relatively distant from the other fallow subpopulations from Erda, RFc and RFd.

Tab. 39: Genetic distance between subpopulations as measured by pairwise distance (* denotes $p > 0.05$). Shaded cells indicate the intersection of the different regimes.

	AIa	AIb	AIc	AIe	REa	REb	REc	REd	IMc	RMa	RFc	RFd	IFa
AIa	-												
AIb	0.06	-											
AIc	0.08	0.02*	-										
AIe	0.08	0.05	0.06	-									
REa	0.17	0.23	0.24	0.25	-								
REb	0.13	0.19	0.20	0.20	0.06	-							
REc	0.16	0.22	0.23	0.22	0.07	0.04	-						
REd	0.17	0.25	0.24	0.26	0.08	0.15	0.15	-					
IMc	0.13	0.16	0.17	0.19	0.18	0.19	0.23	0.18	-				
RMa	0.16	0.22	0.21	0.22	0.02*	0.07	0.04*	0.03*	0.20	-			
RFc	0.21	0.23	0.19	0.22	0.11	0.10	0.12	0.11	0.19	0.07	-		
RFd	0.16	0.20	0.22	0.21	0.08	0.08	0.05	0.17	0.20	0.04*	0.14	-	
IFa	0.17	0.20	0.21	0.23	0.24	0.22	0.28	0.26	0.06	0.26	0.17	0.23	-

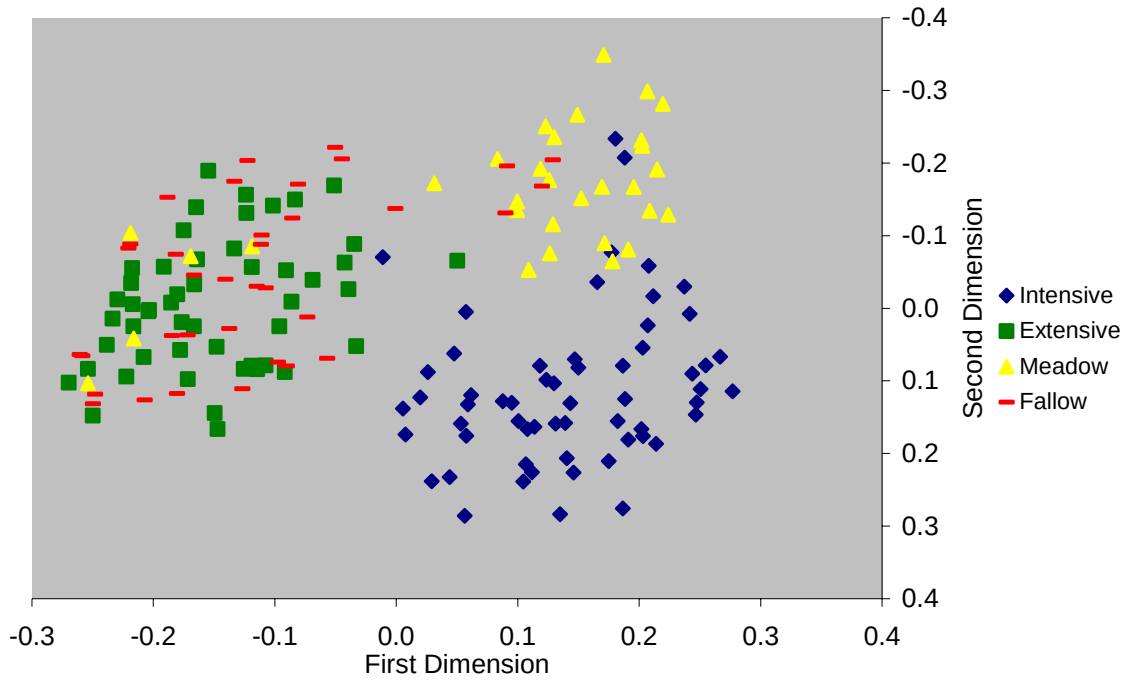
4.3.4 Principle Coordinate Analysis (PCA)

A PCA analysis was carried out for all *V. arvensis* individuals by regime (Fig 12a) and by location (Fig 12b). The results visually confirm the findings from the genetic distances between subpopulations; i.e. the grouping of individuals reflects the geographic location more than the regime.

The first axis accounted for 63.6% of the total variation, and most likely represents geographical location. The second axis represented only 2.6% of the total variation, with 67 axes required to explain 95% of the data.

These findings are very similar to those obtained for *A. thaliana* using AFLP results. After the first axis, the dimensionality is very high, indicating that there is only one major factor which can be used to accurately differentiate individuals.

(a)



(b)

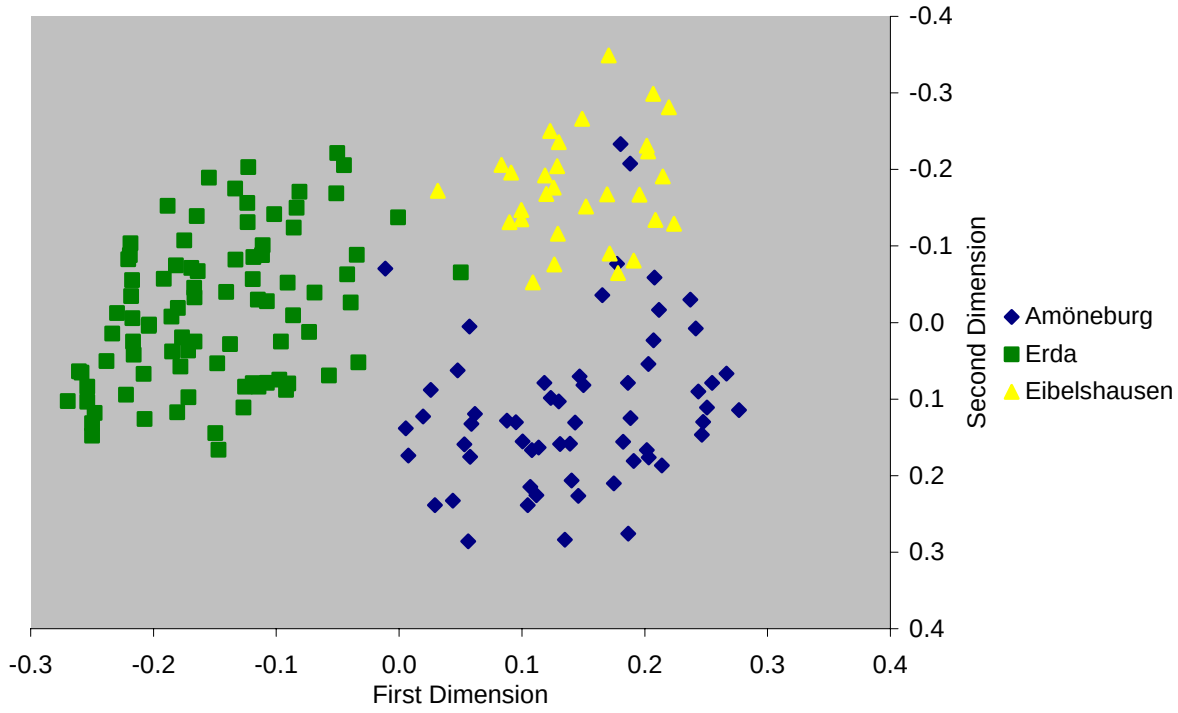


Fig 12 a, b: Principle coordinate analysis of all *V. arvensis* individuals by (a) regime, and (b) location.

4.3.5 Cluster Analysis

Apart from a few individuals, all the intensive samples clustered together, confirming the trend seen in the PCA (for dendrogram see Appendix III). As with *A. thaliana*, individuals from the extensive and fallow subpopulations generally clustered together. The majority of the meadow subpopulation IMc grouped together (including the other Eibelshausen subpopulation IFa), although several outliers were scattered throughout the dendrogram.

As revealed by the PCA, the cluster analysis confirms that the individual plants are, on average, more similar to those from the same location than to those from the same regime.

4.3.6 Mantel Test

As with *A. thaliana*, a Mantel test was conducted comparing the geographical distance with the pairwise genetic distance (F_{ST} analogue values). There was a reasonable overall correlation between the two parameters (r^2 = 0.57; Fig. 13). However, if the group of values whose geographic distances were under five kilometres were removed, it can be seen that the remaining three groups would show a weak negative correlation between genetic distance and geographical distance, i.e. as the geographic distance between subpopulations increases, the genetic distance decreases. This is an extremely unlikely result in a metapopulation in which all individuals could exchange genetic material with one another.

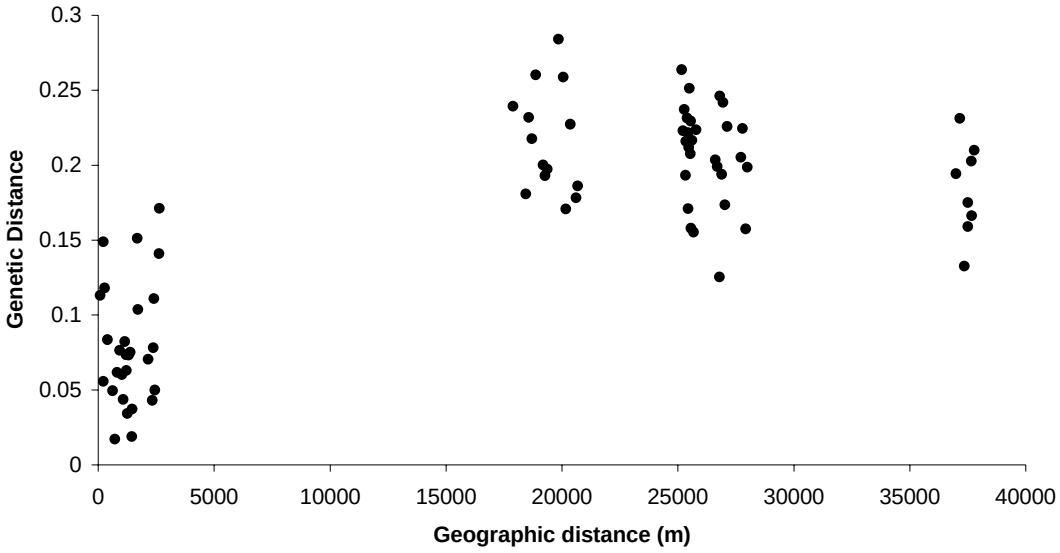


Fig. 13: Mantel test comparing geographic and genetic distance (r^2 = 0.57).

5 Discussion

5.1 Sampling

Before discussing the reasons behind the genetic variation seen in both species in this study, it is necessary to look at the sampling procedures used to collect the samples to assess what influences these procedures may have on the observed results.

The objective of taking soil samples for the fallow and meadow populations was to provide an indication of the genetic structure of populations from these species before (fallow) and shortly after (meadow) the advent of the agricultural intensification. As the fallow sample sites were used for agriculture until 1950 and the meadow sites until 1970, the youngest seeds in the persistent seed bank should be 50 and 30 years old, respectively. Seed decay rates in soil seed banks have been estimated at between 33% (Roberts and Feast 1973) and 50% (Wilson and Lawson 1992) per year in undisturbed habitats. In cultivated habitats, this value has been shown to be lower: after six years buried in the soil, 7 and 38% of ungerminated *V. arvensis* seeds were still viable from cultivated and uncultivated land, respectively (Roberts and Feast 1973). Viable *V. arvensis* seeds from an archaeological site in Denmark were estimated at 463 years old (Odum 1965).

These values suggest that a large variation in ages of seeds is likely when sampling seeds from the persistent seed bank in cultivated land, as the chance of a 30 or 50 year old seed surviving until the present is unlikely to be much greater than that of a 100 year old seed. Therefore, for the fallow and meadow populations, a large temporal bias resulting from the overlapping of generations must be considered when interpreting the results.

Due to the frequent tillage of the soil at the intensive and extensive locations (which generally disturbs the top 20cm of soil), seeds designed for both the persistent and transient seed banks will have invariably been taken. As for the other regimes, the age distribution of the seeds used is an unknown quantity. However, given that the transient seed bank was

also sampled, it is probable that the majority of the seeds obtained from the intensive and extensive regimes exhibit a much smaller variation in age than those from the two older regimes, especially if the seed decay behaviour mentioned above is applicable to these locations.

Therefore, the individuals within subpopulations almost certainly contain samples from different generations which have been selected for seed longevity. The fallow and meadow populations will have been more strongly affected by the seed longevity selection pressure than by intensive and extensive subpopulations.

The subpopulation AIf is an exception to this generalisation as this population was hand picked from immature *A. thaliana* plants growing on the intensive fields, and so represents the only population whereby many (and maybe all) the individuals belong to the same generation.

5.2 Abundance

It is unknown how accurately the number of samples obtained from the sites reflects the actual population sizes for both species. The sample size itself was restricted by the sampling technique; both by the number of soil samples taken, and by the germination rate of the seeds contained within these soil samples. It is interesting to note that, for *A. thaliana*, the meadow population was substantially more abundant than the intensive population, despite the many years of selection pressure on the seeds. Even the fallow population numbered more than the original intensive population, despite the seeds having to survive at least 50 years. More than seven times the number of samples were obtained from the extensive population in comparison to intensive samples originally collected. If these observed numbers are a general reflection of the real population sizes, then it can be seen from the lack of abundance that *A. thaliana* is struggling to survive on intensively managed land in comparison to extensively managed land.

In contrast, *V. arvensis* was slightly more abundant at the intensive sites than at extensive sites. This is not unexpected given that *V. arvensis* has

already been noted for its survival on intensively managed land (Eggers 1978; Bachthaler et al. 1986).

The third species which was initially to be studied, *G. tetrahit*, was considerably less abundant at all sample sites than the other two species, hence it being removed from the population genetic analysis. It is interesting to note, however, that no samples were obtained from the intensive fields, while a moderate number of plants were obtained from the other three sites. This lack of abundance could be representative of the fact that *G. tetrahit* has a decreased fitness on all cultivated land, but especially on intensively cultivated land.

5.3 Arabidopsis thaliana

Within-population genetic diversity

Two common forms of within-population genetic diversity measurements, namely the proportion of polymorphic loci and the number of alleles per locus (*NA*), both showed that the populations from all four regimes exhibited a high degree of allelic diversity.

All seven loci examined were highly polymorphic (*NA* = 15.9). The extensive population was the most polymorphic (*NA* = 12.1), followed by the meadow, intensive and fallow populations (*NA* = 10.3, 9.9, and 6.7, respectively). These values were not adjusted for sample size under the SMM as the function is unknown. Under the IAM this could have been calculated by simply dividing the *NA* by the sample size. However it is accepted that the number of possible allelic states for any given microsatellite locus is restricted by size. The results obtained in this study point to a logarithmic relationship between population size and *NA*. Under this assumption no regime appeared to exhibit a deficiency or excess of *NA*. This was also true even when the intensive population was split into the original samples and the subpopulation AIf. Therefore, the proportion of polymorphic loci and number of alleles per locus were unable to elucidate whether intensive agricultural practices reduce the level of genetic variation in *A. thaliana*.

However, evidence for this was detected from assessment of the observed (H_{obs}) and expected heterozygosity (H_{exp}, also referred to as gene diversity; Nei 1987). The intensive population (with the subpopulation AIf excluded) showed the lowest H_{exp} (0.687) and H_{obs} (0.013). The low H_{obs} indicates a higher rate of inbreeding compared to the other populations. No significant relationship between these heterozygosity measurements and sample size was detected, indicating that the lower sample size of the original intensive population was not a factor in determining its lower heterozygosity values. This was supported by the fact that the fallow population is of a very similar size but exhibited a very high H_{obs} (0.160) and a higher H_{exp} (0.735). The high H_{obs} of the fallow population compared to that of the extensive population (0.073) might be a result of selection for heterozygote seeds over time in the persistent seed bank, given that almost all the samples from these two regimes originated from the same location. However, the H_{obs} of the meadow population (0.063) nominally contradicts this assessment as it is less heterozygous than the extensive population. On the other hand, it is possible that a strong selection for heterozygotes over the past 30 years has also taken place but that the meadow population was initially highly homozygous.

Selection for heterozygotes may also explain the large difference in H_{obs} between the original intensive samples (0.013) and the subpopulation AIf (0.078). If, in this specific situation, heterozygosity confers an increased fitness, an excess of heterozygotes would be expected in all growing populations of *A. thaliana* on intensive fields. The difference in H_{obs} between samples from the seed bank and subpopulation AIf may arise from the action of this selection pressure. The original samples were all grown under optimum conditions in a glasshouse, with individual plants being identified and transferred to a separate pot in which little or no competition occurs. In contrast, the samples from AIf have all been exposed to many biotic and abiotic factors during their lifetime and it is possible that a higher proportion of the homozygous seeds present in that population

(in comparison to heterozygous seeds) either did not germinate or germinated but did not survive.

It is difficult to establish the significance of the high polymorphism found with respect to other studies on *A. thaliana* reported in the literature, as they have almost exclusively focussed on accessions acquired from seed stock centres (Bell and Ecker 1994; Todokoro et al. 1995; Innan et al. 1997; Loridon et al. 1998). The *NA* in this study was considerably higher than that obtained from any of the above mentioned studies. Of these, the highest *NA* value was 10.6 (Innan et al. 1997), observed after studying 42 ecotypes representing a worldwide distribution with 20 microsatellite loci. The same study reported an average gene diversity of 0.79. These values may indicate that the comparison of ecotypes in the literature (most of which are early flowering, selected for ease of laboratory use; Pigliucci 2002) may distort views on how much allelic variation exists in wild populations.

In a study assessing the validity of using microsatellite markers across species, Clauss et al. (2002) examined eighteen *A. thaliana* individuals taken from an isolated population (N > 200) in central Germany with 36 SSR loci. Only 8% of the loci were polymorphic with an observed heterozygosity of 0.2%, resulting in a very low average gene diversity of 0.01. This is in stark contrast to the subpopulation AIf, which displayed a very high level of polymorphism (100% of loci polymorphic), an average gene diversity of 0.757, and 7.8% heterozygosity. The low gene diversity and heterozygosity in the population sampled by Clauss et al. (2002) would arise if the population was recently founded by just one or a few individuals. Therefore, the individuals in subpopulation AIf would be expected to have a comparatively much longer coalescence time.

Analysis by AFLPs showed little difference between each population or subpopulation, especially given the high standard deviation for each value. However, the same order in amount of gene diversity was observed through the AFLP data as that obtained with microsatellite data: meadow (highest), intensive, extensive, and fallow. This order is virtually the op-

posite to the amount of heterozygosity observed in the populations: fallow, extensive, meadow, and intensive.

Another useful parameter for interpreting the state of a population is the θ_{Hom} value. This value can be used as a guide to the evolutionary potential of a population as it combines the parameters of effective population size and mutation (Hartl and Clark 1997). Both the IAM and SSM models of mutation depicted the extensive population and subpopulations as having much more evolutionary potential than all other populations. Even when the total intensive population is used, the extensive population is clearly in a position to be able to withstand environmental changes to a much greater degree than the intensive population.

This is an important measurement with respect to the hypothesis of this study. It is widely accepted that small isolated populations usually exhibit a lower genetic variation (Lammi et al. 1999). However, this is not always the case, and a lack of correlation between sample size and genetic diversity has been explained by unusual histories of bottlenecks and founder events, selection against homozygotes in small populations, varying levels of selfing and outcrossing, and the discrepancy between population size and effective population size.

Hardy-Weinberg Equilibrium (HWE)

All populations and subpopulations showed significant deviations from HWE. This observation was unsurprising given the selfing nature of *A. thaliana*, which has been estimated as being as low as 0.3% (Abbott and Gomes 1989) or as high as 2.2% (Snape and Lawrence 1971). However, the amount of heterozygosity present in the individuals in this study indicates that, in some populations at least, sexual reproduction may occur more frequently than expected. However, given the unknown strength of selection for heterozygosity, both for seeds residing in the seed bank and for germinating individuals, further studies would need to be carried out to confirm this.

Among-Population Genetic Relationships

A significant hierarchical population structure was detected in the populations of *A. thaliana* for all of the measurements used (microsatellites, F_{ST} = 0.175, R_{ST} = 0.189; AFLPs, F_{ST} = 0.173), although diversity was mostly found within populations (approximately 80%). For the interpretation of F_{ST} and R_{ST}, it has been suggested that values between 0-0.05 indicate little genetic differentiation, 0.05-0.15 moderate differentiation, 0.15-0.25 great differentiation, and above 0.25 very great differentiation (Wright 1978; Hartl and Clark 1997). Data in the literature based on isozyme and microsatellite studies suggest that F_{ST} values of 0.2 denote outbreeding species and 0.5 inbreeders (Hamrick and Godt 1989; Bussel 1999).

The results from this study are therefore consistent with many outbreeding species. However, the inbreeding nature of *A. thaliana* is well established and previous studies have not uncovered such a degree of allelic polymorphism or heterozygosity (Bell and Ecker 1994; Todokoro et al. 1995; Innan et al. 1997; Loridon et al. 1998; Clauss et al. 2002). If the individuals sampled in this study were the product of predominantly outbreeding parents then the subpopulations and populations should exhibit genotype frequencies closer to what would be expected under HWE. That this is not the case indicates that other environmental factors, or experimental artefacts, are responsible for the observed phenomena.

Aside from mutation, the only other possibility to maintain genetic variation in populations is from migration, either as the immigration of individuals (seeds) or gametes (pollen) from other populations. Computer simulations have shown that only one migrant per generation is sufficient in order to prevent the negative effects of genetic drift seen in small isolated populations (Lacy 1987). The rates of migration calculated under the assumptions of the IAM and SSM models in this study appear to be unrealistic. The subpopulation AIb consists of only sixteen individuals, but, in order to keep the level of genetic similarity observed between this subpopulations and any of the extensive subpopulations, approximately two individuals per generation would need to be exchanged. Given the dis-

tance between the regimes (approximately 26 km), it seems unlikely that a population of sixteen individuals is exchanging this many migrants per generation (in this case each generation equating to one year) with populations numbering about 30 individuals. It is unlikely that pollen from one population could reach the other, and even less likely to find a recipient given the high inbreeding rate. However, seed dispersal by birds and humans cannot be ruled out. The latter case, with humans, is complicated. The three locations used in this study are all research sites. Therefore, in comparison to normal farmland habitats, a higher rate of movement by humans between these sites probably occurs, allowing the possibility that seeds from one location are carried to another (either on the clothing of humans or machinery which may be used on all locations). Being research areas, the likelihood of other seeds being brought into the environment is probably also higher than on non-research farms. Therefore, one would imagine that research areas are likely to exhibit a slightly higher diversity for many plant species due to the increased movement potentially bringing in new genetic material.

Although these factors are likely to play some role in determining the genetic differentiation observed here, other results from migration indicate that this role may be very minor, and that the most likely reason for the genetic similarities observed (and therefore high migration estimates) is that all the subpopulations were established recently enough so that they all contain alleles which were present in the founding population. This is exemplified by the migration rates estimated for the subpopulation RFa with any of the extensive subpopulations (approximately three individuals per generation). Originating from the fallow regime, the soil from which subpopulation RFa is derived from should not have received any new seed material for approximately 50 years, meaning that any exchange of genetic material between it and subpopulations from the extensive regime must have occurred more than 50 years ago. Because of the sampling procedures used, it would not be expected that many (if any) 50 year old seeds or older would have been extracted from the extensive regime.

Thus there is the real possibility that absolutely no migration between these subpopulations has occurred.

One factor which may have led to an overestimation of similarity between subpopulations from different regimes is the homoplasy associated with microsatellites. However, the AMOVA analysis revealed very little difference in the population genetic structure between data obtained from microsatellite and AFLP analysis. Although homoplasy has probably played a role in overestimating genetic similarity between populations, the similarity due to alleles being identical by descent is exemplified with the locus nga151 (the dimeric repeat $[CT]_n$). Every population exhibits an allele of size 149 bp at this locus. However, the next closest allele in size contains eight less repeat units. It is very unlikely that an unequal crossing over event occurred in all four populations independently to produce the same size allele. Similar differences in allele size were observed in locus nga168, nga111, and nga63. The last of these is an interesting contrast to allele nga151 as it displays a similar difference in allele size (probably resulting from unequal crossing over), but it also appears single-step mutation events have since occurred, giving rise to new alleles one or two repeat units different in size. This has not happened at locus nga151 for the allele of size 149 bp. It is therefore possible that some new alleles mutating under the single-step model from this allele come under selection pressure.

The mutational behaviour of the marker system used and the statistical model employed are also factors to consider when interpreting the results. The high mutation rates of microsatellites are expected to seriously underestimate genetic structure of populations using *F*-statistics when mutation is higher or of the same magnitude as migration (Balloux et al. 2000), and inferences drawn from F_{ST}, such as migration rate, will be biased due to the inability to separate the effects of migration and mutation (Balloux and Lugon-Moulin 2002b). In contrast, *R*-statistics are independent of mutation rate. However, they are very sensitive to variance, and even under a strict SSM they may be outperformed by *F*-statistics (Balloux

and Lugon-Moulin 2002a). In this study, AFLPs and both microsatellite statistical methods for estimating population parameters gave similar results, and thus it cannot be said that one method is more suitable than the other for this study.

Principle Coordinate and Cluster Analysis

The trends revealed through the use of principle coordinate and cluster analysis showed that geographic location was more important than agricultural regime in determining the genetic relationship between individuals. The difference between Figs. 9a and 9b clearly shows how all the fallow samples are within the area of the extensive samples. A similar pattern, although not quite so clear, is seen with the microsatellite data. This observation was supported by the higher dimensionality of the microsatellite data compared to the AFLP data.

It has been shown that when microsatellites and AFLPs have been compared with different marker systems (including each other) microsatellites have revealed the highest level of polymorphism (Morgante et al. 1994; Powell et al. 1996) and AFLPs the lowest (Becker et al. 1995; Russell et al. 1997). The high level of polymorphism present in microsatellite markers is due to the high mutation rate. However, homoplasy is also high and therefore the chance of identity in state is increased. This may cause an overestimation of the similarities between individuals and populations, which in turn makes it more difficult for individuals to be grouped by methods such as PCA and cluster analysis.

Mantel Test

The Mantel test revealed that geographic distance had very little association with all genetic distance measures. There was a slight trend towards higher genetic differentiation with distance, but the variation in genetic similarity between populations approximately equidistant to each other was very large.

Bottleneck Detection

Several tests were carried out in order to detect the presence of bottleneck events. There was, however, little or no evidence for bottlenecks in any of the populations with the tests that were used. The Wilcoxon test detected a slight heterozygote excess in all the regimes except intensive, and under certain conditions this might be valid evidence for recent bottlenecks. However, the tests were conducted under the assumption of the IAM, a model that under some conditions is inapplicable to microsatellites, and therefore these results must be treated with caution.

The method of Garza and Williamson (2001) initially appeared to indicate that bottlenecks had occurred in both the original intensive and fallow populations (in comparison to the other populations). However, a high correlation was detected between *n* and *M*, rendering the findings inconclusive. If *M* is lower for the original intensive and fallow populations due to the relationship between *n* and *M*, then no population appears to have been affected by a bottleneck event (relative to the other populations).

It must also be remembered that all populations consist of individuals that come from a variety of subpopulations (and these themselves are possibly sampled from many small micro-populations). Therefore, evidence for recent bottlenecks may be masked by an exaggerated within-population diversity that might arise from the pooling of individuals from different micro-populations. The same effect may have also occurred through experimental artefacts resulting from sampling over a large temporal scale.

However, the subpopulation AIf, being the only contemporary population sampled, showed no evidence of a recent bottleneck. If this population is indicative of the genetic structure harboured within other subpopulations, it would indicate that the fluctuations in population size are not great enough to subject them to the effects of genetic drift and founding events.

Ecology

Despite all the research carried out on *A. thaliana* over the last half century, little is known about its ecology (Pigliucci 1998). Contrary to other studies, a large amount of within-population variation was uncovered here, in populations which display a genetic structure more usual to that observed in outcrossing species over a range of nearly 40 km. This variation was present in the original intensive population despite the H_{obs} indicating that outcrossing was substantially lower than for other populations.

Different kinds of *A. thaliana* based on flowering time have been previously recognised (Napp-Zinn 1985). Given that *A. thaliana* is considered a poor competitor, preferring to flower as early as possible to minimise inter-specific competition (Pigliucci 2002), it is likely that selection for genotypes conferring early flowering is occurring in agricultural fields. This is likely to be exacerbated for populations occurring on intensive fields due to the addition of fertiliser allowing for more rapid growth of crops (and weeds), providing greater competition for other species. The density of crop plants is also higher on intensively managed fields, and the concomitant increase in shading may also play a role in determining the development of *A. thaliana* populations. Laboratory studies have already demonstrated that *A. thaliana* flowers at a younger developmental stage (i.e. fewer leaves) which sometimes corresponds to earlier flowering in chronological time (Pigliucci 2002). A positive association between plant and seed size and overall fitness has often been established (Grundy et al. 1995; Sletvold 2002). Therefore, if earlier flowering is being selected for due to the harsher abiotic and biotic factors found on intensive fields compared to extensive fields, it is likely to have a negative effect on the fitness of seeds produced and on the plants germinating from these. This may even take the form of selection for individuals possessing loss-of-function FRI alleles, which stem from the dominant FRI gene.

If such selection pressures have been acting on a population of small size, it would be expected that founder events, bottlenecks, reduced genetic

diversity due to genetic drift, and a greater amount of inbreeding would be observable. Loss of heterozygosity through drift alone occurs very slowly (Nei et al. 1975). Therefore it is probable that the lower H_{obs} seen in the original intensive population was due to an increase in inbreeding as opposed to genetic drift.

The amount of genetic diversity seen in all the *A. thaliana* populations, but especially in the intensive population, is surprising in comparison to expectations derived from the literature. This could be explained by a combination of the sampling technique carried out in this study and the behaviour of *A. thaliana* populations on very short spatial scales. A study was carried out by Stratton and Bennington (1996) showed that very localised environmental conditions (less than 50 cm range) may favour one genotype over others, resulting in a seemingly random patchwork of genotypes across the landscape. If this is the situation on the agricultural fields used for sample collection in this study, then it would explain the apparent high variability observed in both populations and subpopulations. However, if genotypic differences between populations can be significant over distances of 50 cm, then it could be that the sampling methods used for this study are too broad, resulting in the pooling of many micro-populations which have their own population histories. Even if these micro-populations exhibit very low within-population genetic diversity, a few individuals from several micro-populations may result in genetically diverse subpopulations as observed in this study.

Therefore, despite probable selection pressure due to the intensity of agriculture (early crop sowing, herbicide application, high crop density), a reduction in genetic diversity due to small population size may be undetectable because the behaviour of *A. thaliana* at the micro-scale makes it very difficult to sample a single population.

5.4 Viola arvensis

Within-Population Genetic Diversity

There appears to be no molecular genetic assessment of *V. arvensis* in the literature. Therefore, unlike *A. thaliana*, the results of this study are difficult to put into perspective. Studies have been conducted on *Viola pubescens* (Culley and Wolfe 2001) and *Viola riviniana* (Auge et al. 2001), but due to the different reproductive strategies of these two species, comparisons of their population genetic structure to *V. arvensis* cannot contribute much information to the present study.

AFLP analysis revealed a very similar pattern of within-population diversity to that seen in *A. thaliana* with AFLP markers. The extensive and fallow populations showed a lower gene diversity than the other two populations, a trend reflected in the subpopulations. As seen with the *A. thaliana* results, however, the gene diversity values may be masking other forms of genetic diversity, such as heterozygosity, which cannot be accurately calculated with a dominant marker system. However, in contrast to the results for *A. thaliana*, in which the intensive population and subpopulations showed a higher gene diversity than the extensive population and subpopulations despite a much lower sample size, the sample size of the intensive population for *V. arvensis* was slightly greater than that of the extensive population. Sample size and gene diversity were shown to be not correlated, as with *A. thaliana*.

Among-Population Genetic Diversity

Analysis with AMOVA revealed significant differences between regimes (F_{ST} = 0.171) despite the majority of the variation (approximately 83%) being within subpopulations. Thus there is a restricted gene flow between the populations, a result not unexpected given the geographic distances separating them.

There is a high level of similarity to the population structure shown by *A. thaliana*, despite the differences in population sample sizes and the supposedly higher outbreeding rate of *V. arvensis*. A more outcrossing spe-

cies would be expected to show less genetic differentiation in a given geographic area due to the higher rate of allelic exchange. There are two possibilities that would explain the similarity of the genetic structure between the two species in this study: either the ratio of selfing to outcrossing in the populations is more similar than has been predicted by the literature, or the temporal and spatial distribution of the samples collected for both species put a bias on the results.

Unfortunately, microsatellite markers are yet to be developed for *V. arvensis*. As seen from the results for *A. thaliana*, the gene diversity values calculated from the AFLP results do not necessarily follow the same trend as diversity values calculated with microsatellites, and thus a study using co-dominant markers would be very helpful for elucidating the amount of genetic variation and differentiation in these *V. arvensis* populations.

Principle Coordinate and Cluster Analysis

Despite the F_{ST} values of *V. arvensis* and *A. thaliana* being almost identical (0.171 and 0.173, respectively), the *V. arvensis* samples grouped more distinctly in the PCA diagrams and clustered tighter in the dendrogram. As with *A. thaliana*, the first axis explained the majority of the variation and this appeared to represent the importance of geographic origin given the difference between Figures 12a and 12b.

Mantel Test

The comparison of geographic and genetic distance for *V. arvensis* showed two separate trends. There is a stronger positive correlation overall between these two parameters than for *A. thaliana*. However, at distances of approximately 20 km to 40 km a slight negative trend is evident. This is in part because of the similarity between the intensive and meadow subpopulations, despite them being the most distant regimes from each other. This is possibly the legacy of a sufficiently recent colonisation event or relatively high gene flow from Eibelshausen to Amöneburg.

Ecology

Unlike *A. thaliana*, *V. arvensis* does not seem to be troubled with growth and survival on intensively-managed agricultural land (in comparison to extensive agriculture). This could be due to two mechanisms. Firstly, *V. arvensis* is thought to resist herbicide treatments to some degree (Christen et al. 1999). Secondly, it is possible that *V. arvensis* is able to utilise higher nitrogen levels, present in the soil from fertiliser, more effectively than some of its competitors.

6 Conclusions and Recommendations

The evidence obtained using microsatellite markers suggests that the less abundant *A. thaliana* population from the intensive regime does not contain the same level of genetic diversity observed in populations from other agricultural regimes. This was particularly true when the subpopulation AIf was removed from the intensive population calculations.

However, not all indicators of genetic diversity supported this result, instead suggesting that the differences between the intensive and other regimes is negligible.

With only AFLP analysis performed on *V. arvensis*, the same level of population genetic structure could not be elucidated. The results showed that, if anything, the intensive population of *V. arvensis* contained more genetic diversity than populations from other regimes.

The similarity of the population structure of the two species, despite their supposedly different reproduction strategies, suggests that either both species from the populations sampled exhibit a similar ratio of asexual to sexual reproduction (with a similar dispersal range), or that the sampling distances relative to the population sizes (geographically) bias the results towards a higher genetic diversity.

If forces such as selection against homozygotes are functioning in peripheral or endangered populations of small size, then genetic diversity cannot always be used as an indicator of the state of a given population. It may well be, given the numbers of *A. thaliana* individuals found at the three locations, that the subpopulations from the intensive regime are indeed declining and will soon be extinct from, in this case, Amöneburg, but that sampling from different areas from multiple generations has depicted this metapopulation to be almost as diverse as that from the extensive regime.

If the biodiversity of the food web surrounding Europe's farmlands is to be protected, then the correct estimators of biodiversity must be in place.

The results from this study make it difficult to reach a conclusion as to the state of the metapopulation of, in particular, *A. thaliana*.

Recommendations

In order to obtain a better understanding of what underlying forces have shaped the situation observed in this study, a number of steps should be undertaken.

The very small spatial scales at which *A. thaliana* (and therefore possibly many arable weeds) can create a patchwork of different genotypes must be taken into account when considering the geographic distances between samples from any one location. In retrospect, taking a larger number of soil samples from a smaller area may have given a better indication as to the structure of the subpopulation diversity.

The fact that only one intensive site was sampled did not allow any general trend from intensive samples to be established, as it remains unknown if the trends observed are due to the populations history or the effects of the regime. Therefore, it is proposed that, if possible, more than one intensive site should be sampled for the future continuation of this study.

The problem of temporal bias (overlapping generations) due to seed bank sampling could also be overcome by taking leaf samples from contemporary individuals growing in the fields.

An increase in number of soil samples taken, combined with an increase in number of sample sites, would obviously produce much more material for analysis, and therefore it is maybe not necessary to sample persistent seed banks.

In analysing species for which no microsatellite markers are available, it would seem appropriate to carry out allozyme analysis in order to obtain estimates of allele frequencies, allowing the estimation of many further population parameters than are possible with dominant markers.

7 Summary

Decreases in population size can lead to a reduction in genetic diversity due to the effects of random genetic drift and inbreeding. It is generally accepted that a higher amount of genetic diversity within a population increases the population's ability to withstand environmental changes, and therefore to survive.

Over the last half century, the agricultural intensification that has occurred throughout Europe has led to the decline in abundance of many flora and fauna associated with farmland habitats.

Two weed species, *Arabidopsis thaliana* and *Viola arvensis*, were selected as indicator species to assess the impact intensive agriculture may have on the genetic diversity of arable weed populations.

Soil samples were taken from a variety of contemporary agricultural sites (both intensive and extensive regimes) and sites previously used for extensive agriculture (designated meadow and fallow regimes) within 30 km of each other in Hesse, central Germany. After incubation, seedlings from either of the two indicator species were isolated.

Molecular marker analysis was carried out on all individuals to elucidate the genetic structure of the populations. For *A. thaliana* both the AFLP and microsatellite techniques were used. However, only the AFLP technique was used for *V. arvensis*.

A similar level of genetic structuring in populations of the predominantly inbreeding *A. thaliana* and the more outcrossing *V. arvensis* was found using AFLPs. Around 80% of the genetic variation of populations from both species lies predominantly among the individuals within subpopulations, indicating that little genetic difference exists between the populations from the different regimes.

However, microsatellite analysis of *A. thaliana* revealed a lower observed heterozygosity in the intensive population in comparison to populations from the other regimes. This result, in combination with other analyses

performed on the data, may be an indication that intensive agriculture reduces the genetic diversity of *A. thaliana*.

However, there are many factors which could combine to produce such an observation, and it is recommended that further studies be carried out on similar agricultural fields to see if the results obtained in this study reflect a general trend on agricultural fields.

8 Zusammenfassung

Populationen mit großer genetischer Variation verfügen über die Fähigkeit, auf veränderte Umweltbedingungen zu reagieren. Verkleinert sich eine Population, so geht damit in der Regel eine Verringerung ihrer genetischen Diversität einher, aufgrund zufällig stattfindender genetischer Drift und zunehmender Inzucht. Als Konsequenz ist eine Reduktion in der Anpassungsfähigkeit der Population zu erwarten.

Ackerwildkräuter stellen Lebens- und Nahrungsräume und Rückzugsgebiete für viele Tiere dar, die selbst Bestandteile des Ackerland-Lebensraum-Nahrungsnetzes sind. Deswegen ist es wahrscheinlich, dass eine Abnahme der Ackerwildkräuter auch im weiteren Verlauf der Nahrungskette zu negativen Effekten führt. So führte z.B. die Beobachtung der deutlichen Abnahme der Anzahl verschiedener Vogelarten zu der Erkenntnis, wie wichtig Ackerwildkräuter für die landwirtschaftlichen Lebensräume sind. Bei vielen im europäischen Raum heimischen Pflanzen und Tieren kam es zu einer Verkleinerung ihrer Populationsgrößen u.a. aufgrund einer Intensivierung der Landwirtschaft, die in ganz Europa während der vergangenen ca. 50 Jahre zu beobachten war. Da Ackerwildkräuter ursprünglich vor allen Dingen als nachteilig auf den Ertrag von Kulturpflanzen angesehen wurden, kam es infolgedessen zur verstärkten Entwicklung und Anwendung von Herbiziden.

Das Verständnis des Einflusses der landwirtschaftlichen Intensivierung auf die genetische Basis der Ackerwildkrautpopulationen kann dabei helfen, Entscheidungen hinsichtlich der Erhaltung der europäischen Unkräuter-Diversität zu treffen. Zu diesem Zweck wurden Populationen zweier Ackerwildkräuter, *Arabidopsis thaliana* und *Viola arvensis*, als Indikator-Arten ausgewählt, um die genetische Struktur von Ackerwildkrautpopulationen aus intensiven und extensiven landwirtschaftlichen Herkünften zu vergleichen.

Die genetische Variation der beiden Arten wurde in Populationen untersucht, die von Standorten mit unterschiedlicher landwirtschaftlicher Nutzungsgeschichte stammen, wobei der Schwerpunkt auf den Vergleich intensiv bzw. extensiv bewirtschafteter Flächen gelegt wurde. Als Hypothese wurde vermutet, dass Populationen der intensiv bearbeiteten Standorte eine geringere genetische Variation zeigen als solche Populationen, die unter natürlichen Bedingungen (einschließlich extensiver Landwirtschaft) kultiviert waren.

Es wurde auf Pflanzenmaterial zurückgegriffen, welches von Flächen mit insgesamt vier verschiedenen landwirtschaftlichen Nutzungsgeschichten stammt, die in dieser Arbeit mit „intensiv", „extensiv", „Weide" und „Brache" bezeichnet wurden. Der intensive Standort, Amöneburg, erhielt sowohl Dünger- als auch Herbizidgaben. Der extensive Standort befand sich in Erda und erhielt keinerlei Herbizidbehandlungen. Die Weiden- bzw. Brachenstandorte befanden sich ebenfalls in Erda, wobei in Eibelshausen ein zusätzlicher Brachen-Standort gelegen war. Die Weiden wurden extensiv bis zu den 1970ern Jahren und seit dem Jahr 2000 als Grünland genutzt. Bei den Bracheflächen erfolgte seit der Unterbrechung der extensiven Landwirtschaft in den 1950er Jahren keine Bewirtschaftung.

Von diesen Flächen wurden im Jahr 1999 an 21 verschiedenen Standorten Bodenproben genommen, wobei jede Standort-Probe als Subpopulation eines Bewirtschaftungssystems angesehen wurde. Während des Frühjahrs und Sommers 2000 und 2001 wurden diese Bodenproben im Gewächshaus kultiviert, um jedes darin vorhandene Saatgut zum Keimen zu bringen. Jedes gekeimte *A. thaliana*- oder *V. arvensis*-Individuum wurde anschließend separat angezogen. Mittels Selbstungstüten wurden Auskreuzungen verhindert. So konnten persistierende und transiente Samen-Banken der intensiven bzw. extensiven Bewirtschaftungssysteme angelegt werden. In den zwei anderen Bewirtschaftungssystemen gelang dies je-

doch nur für das persistierende Saatgut. Die durch Selbstung entstandenen Samen der nächsten Generation wurden nachfolgend wieder ausgesät und ihr frisches Blattmaterial für die anschließende DNA-Analyse verwendet.

Für die genetische Analyse der *A. thaliana*-Populationen wurden sowohl AFLP- als auch Mikrosatelliten-Markertechniken verwendet. Da für *V. arvensis* bisher keine Microsatelliten-Marker entwickelt wurden, fanden für diese Spezies nur AFLP-Marker Anwendung. Alle Markerergebnisse wurden mit Hilfe der Polyacrylamid-Gelelectrophorese (PAGE) in *Li-COR Gene Readir 4200 DNA sequencer*-Geräten sichtbar gemacht und visuell ausgewertet. Dabei wurden AFLP-Gele hinsichtlich der An- bzw. Abwesenheit von Banden, Mikrosatelliten-Banden jedoch auch hinsichtlich ihrer Fragmentlängen ausgewertet. Die Datenauswertung erfolgte mit Hilfe der statistischen Software-Programme *Excel Microsatellite Toolkit*, *Arlequin v. 2000* sowie *NTSYS-PC*.

Von den extensiven Bodenproben konnten ca. sechsmal so viele *A. thaliana*-Individuen generiert werden wie von den intensiven, so dass die Vermutung naheliegt, dass diese Spezies extensiv bewirtschaftete Flächen bevorzugt bzw. besser toleriert. Mit Hilfe der AFLP-Ergebnisse ließ sich nachweisen, dass die *A. thaliana*-Populationen aller vier verschiedener Bewirtschaftungssysteme ähnliche genetischen Diversität zeigten. Mittels der Mikrosatelliten-Analyse konnte jedoch dargestellt werden, dass in der intensiven *A. thaliana*-Population weniger Heterozygote auftraten als in den anderen Populationen. Die Population der Brache zeigte sogar ca. dreimal soviel Heterozygotie wie die der anderen Bewirtschaftungssysteme. Die erwarteten Heterozygotie-Grade in den *A. thaliana*-Populationen aller Bewirtschaftungssysteme waren allerdings sehr ähnlich.

Eine Analyse der *Molecular Variance* (AMOVA), die sowohl AFLP- als auch Mikrosatelliten-Daten nutzte, ergab, dass ca. 80% der gesamten geneti-

schen Variation durch Unterschiede erklärt wurde, die innerhalb der *A. thaliana*-Populationen zwischen den Individuen auftraten. Die *Principle Coordinate Analysis* (PCoA) zeigte eine hohe Dimensionalität der Daten beider Marker-Typen. Es zeigte sich, dass *A. thaliana*-Individuen aufgrund ihrer Herkunft von einem bestimmten Standort eher gemeinsam gruppiert wurden als aufgrund der dortigen Bewirtschaftungssysteme. Die Schätzung des Populationsparameters θ (gleich $4N\mu$) von den erwarteten Heterozygotie-Werten zeigte einen großen Unterschied zwischen den intensiven und extensiven *A. thaliana*-Populationen, und etwas geringer auch zwischen den Weide- und Brache-Populationen. Dies legt die Vermutung nahe, dass die intensive Population ein vergleichsweise niedrigeres Evolutionspotential hat.

Im Gegensatz zu *A. thaliana* wurde für *V. arvensis* die gleiche Anzahl Individuen von den intensiven und extensiven Bodenproben herangezogen. Dies entspricht auch der einschlägigen Literatur, in der ebenfalls beschrieben ist, dass *V. arvensis* eines der wenigen Acker-Unkräuter ist, die auch gut auf intensiv bearbeiteten Flächen überleben kann. Allerdings zeigen die mittels der AFLP-Analyse generierten Ergebnisse, dass die genetische Zusammensetzung der *V. arvensis* Populationen sehr ähnlich der von *A. thaliana* ist. Dies konnte sowohl mittels AMOVA als auch mittels der Gen-Diversitäts-Ergebnisse gezeigt werden. Obwohl die genetische Diversität in der extensiven Population nur geringfügig niedriger war als in der intensiven, waren die Gen-Diversitäts-Werte für alle *V. arvensis*-Populationen ähnlich. Ungefähr 83% der Gesamtvarianz aller Bewirtschaftungssysteme wurde mit Unterschieden erklärt, die innerhalb der *V. arvensis*-Populationen zwischen den Individuen auftraten. Der PCoA zeigte, dass *V. arvensis*–Individuen aufgrund ihrer Standorte besser genetisch unterschieden werden konnten als aufgrund ihrer Bewirtschaftungssysteme.

Zusammenfassend konnte ein ähnliches Struktur der genetischen Diversität in Populationen der überwiegend selbstbefruchtenden Art *A. thaliana*

und der mehr auskreuzenden *V. arvensis* nachgewiesen werden. Die genetische Variation der Populationen beider Arten ist überwiegend zwischen den Individuen innerhalb der Subpopulationen auszumachen. Dies demonstriert dass lediglich ein geringer genetischer Unterschied zwischen den Populationen der verschiedenen Bewirtschaftungssystemen existiert. Der Zeitpunkt der Besiedlung beider Arten in dieser geographischen Umgebung ist unbekannt. Die in dieser Arbeit erzielten Ergebnisse weisen jedoch darauf hin, dass diese erst vor kurzem erfolgte, da es sonst zu einer größeren Differenzierung zwischen den Populationen gekommen wäre. Die Ergebnisse demonstrieren darüber hinaus, dass sich entweder die *A. thaliana*-Populationen in ihrer natürlichen Umgebung häufiger auskreuzen als in der Literatur vermutet, oder es in den *V. arvensis*-Populationen häufiger zu Selbstungen kommt als ursprünglich vorhergesagt.

Die Neigung zu Selbstungen von *A. thaliana* wird durch die Beobachtung unterstützt, dass die hier aufgetretene Heterozygotie deutlich geringer ist als eigentlich nach Hardy-Weinberg-Gleichgewicht erwartet. Leider konnte nur bei *A. thaliana* – anhand der erzeugten Allelfrequenzen der Mikrosatelliten – gezeigt werden, dass die intensive Population eine niedrigere erwartete und beobachtete Heterozygotie zeigte. Diese Ergebnisse unterstützen die Ausgangshypothese, dass Populationen, die unter intensiven Bedingungen heranwachsen, nicht das Niveau an genetischer Diversität besitzen wie solche aus extensiven oder natürlichen Landschaften.

Das experimentelle Design der Versuche dieser Arbeit erlaubt jedoch keinen Vergleich der intensiven Populationen von verschiedenen landwirtschaftlichen Gegenden. Es ist daher möglich, dass das geringere Niveau von Heterozygotie, das bei intensiven Populationen auftritt, nur ein Charakteristikum dieser Population ist. Darüber hinaus könnte unter Berücksichtigung der Raum-Zeit-Achse zum Zeitpunkt der Arten-Sammlung in Kombination mit den sehr geringen räumlichen Distanzen, bei denen es bei *A. thaliana* bereits zu genotypischer Differenzierung kommt - die aktu-

elle genetische Diversität überdeckt werden, die innerhalb der Population auftritt.

Im Rahmen der in dieser Arbeit durchgeführten Untersuchungen konnte gezeigt werden, dass kleine, isolierte oder an Randgebieten auftretende Populationen sowohl niedrige als auch hohe genetische Diversitäten in Abhängigkeit davon zeigen können, welche Evolutionsfakturen dominieren (z.B. genetische Drift oder Selektion gegen die Homozygoten). Die Unterschiede zwischen der erwarteten und beobachteten Heterozygotie der intensiven und extensiven *A. thaliana*-Populationen waren deutlich, aber gering. Deshalb ist es möglich, dass eine stärkere Selektion auf die Heterozygoten bei intensiv bewirtschaftetem Land stattfindet (Heterozygote sollen eine höhere Fitness besitzen). Dieses würde zu einer höheren Frequenz der Heterozygoten führen als normalerweise erwartet. Der Mangel an Mikrosatelliten-Primern für *V. arvensis* führte dazu, dass die genetische Struktur der verschiedenen Populationen nicht so detailliert analysiert werden konnte wie die von *A. thaliana*. Die Häufigkeit, mit der *V. arvensis* auf intensiv bewirtschafteten Ackerflächen auftrat, war der der extensiv bewirtschafteten Flächen sehr ähnlich. Somit kann davon ausgegangen werden, dass diese Art gleich gut unter beiden Bewirtschaftungssystemen wachsen kann. Dies würde auch bedeuten, dass beide Populationen einen ähnlichen Gehalt an genetischer Variation besitzen.

Schlußfolgerungen & Empfehlungen

Mittels dieser Studie konnte die Ausgangshypothese unterstützt werden, dass Populationen der intensiv bearbeiteten Standorte eine geringere genetische Variation zeigen als solche Populationen, die unter natürlichen Bedingungen kultiviert wurden. In *A. thaliana*-Populationen von intensiv bewirtschafteten Flächen wurde ein geringerer Heterozygotiegrad gefunden als von solchen, die von extensiven Flächen stammen. Das Auftreten von Individuen aus den beiden Bewirtschaftungssystemen zeigt ebenso an, dass *A. thaliana* extensivere Flächen bevorzugt.

Ungeachtet dem Mangel an Allel-Daten von *V. arvensis*, legen die ähnlichen Häufigkeiten der intensiven als auch der extensiven Standorte die Vermutung nahe, dass in beiden Pflanzenarten auch ähnliche Gehalte an genetischer Diversität bestehen. Die in dieser Arbeit verwendete Sammelmethode deckt eine große Raum-Zeit-Achse ab. Es ist daher schwierig, die Bedeutungen zu unterscheiden, die die geographische Position bzw. das Bewirtschaftungssystem auf die genetische Populationsstruktur hatte. In zukünftigen Untersuchungen sollten daher mehrere geographische Herkünfte, jeweils mit intensiven und extensiven Ackerflächen, beprobt werden, um den Einfluss der geographischen Position bei der Interpretation der Daten so gering wie möglich zu halten.

9 Bibliography

Abbott RJ, Gomes MF (1989) Population genetic structure and outcrossing rate of *Arabidopsis thaliana* (L.) Heynh. Heredity 62:411-418

Akagi H, Yokozeki Y, Inagaki A, Fujimura T (1996) Microsatellite DNA markers for rice chromosomes. Theor Appl Genet 93:1071-1077

Albrecht H (1995) Changes in arable weed flora of Germany during the last five decades Proceedings 1995 9th EWRS Symposium, Budapest, Hungary, pp 41-48

Alonso-Blanco C, Koornneef M (2000) Naturally occurring variation in Arabidopsis: an underexploited resource for plant genetics. Trends In Plant Science 5:22-26?

Andreasen C, Stryhn H, Streibig JC (1996) Decline of flora in Danish arable fields. Journal of Applied Ecology 33:619-626

Auge H, Neuffer B, Erlinghagen F, Grupe R, Brandl R (2001) Demographic and random amplified polymorphic DNA analyses reveal high levels of genetic diversity in a clonal violet. Molecular Ecology 10:1811-1819

Bachthaler G, Neuner F, Kees H (1986) Entwicklung und Verbreitung des Ackerstiefmütterchens (Viola arvensis Murr.) in Abhängigkeit von den Standorts- und Bewirtschaftungsverhältnissen. Nachr. Bl. dt. Pfl. Schutzd. 38:33-41

Balloux F, Brünner H, Lugon-Moulin N, Hausser J, Goudet J (2000) Microsatellites can be misleading: An empirical and simulation study. Evolution 54:1414-1422

Balloux F, Lugon-Moulin N (2002a) The estimation of population differentiation with microsatellite markers. Molecular Ecology 11:155-165

Balloux F, Lugon-Moulin N (2002b) Statistical properties of population differentiation estimators under stepwise mutation in a finite island model. Molecular Ecology 11:771-783

Barton NH, Hewitt GM (1985) Analysis of hybrid zones. Annual Review of Ecology and Systematics 16:113-148

Baskin JM, Baskin CC (1995) Variation in the annual dormancy cycle in buried seeds of the weedy winter annual *Viola arvensis*. Weed Research 35:353-362

Baumann E, Lewald J, Saedler H, Schulz B, Wisman E (1998) Successful PCR-based reverse genetic screens using an *En-1*-mutagenised Arabidopsis thaliana population generated via single-seed descent. Theor Appl Genet 97:729-734

Becker J, Vos P, Kuiper M, Salamini F, Heun M (1995) Combined mapping of AFLP and RFLP markers in barley. Mol Gen Genet 249:65-73

Bell CJ, Ecker JR (1994) Assignment of 30 Microsatellite Loci to the Linkage Map of *Arabidopsis*. Genomics 19:137-144

Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR (1994) High resolution of human evolutionary trees with polymorphic microsatellites. Nature 368:455-457

Breyne P, Rombaut D, Van Gysel A, Van Montagu M, Gerats T (1999) AFLP analysis of genetic diversity within and between *Arabidopsis thaliana* ecotypes. Mol Gen Genet 261:627-634

Brickle NW, Harper DGC, Aebischer NJ, Cockayne SH (2000) Effects of agricultural intensification on the breeding success of corn buntings *Miliaria calandra*. Journal of Applied Ecology 37:742-755

Brown-Guedira GL, Thompson JA, Nelson RL, Warburton ML (2000) Evaluation of genetic diversity of soybean introductions and North American ancestors using RAPD and SSR markers. Crop Science 40:815-823

Bussel JD (1999) The distribution of random amplified polymorphic DNA (RAPD) diversity amongst populations of *Isotoma petreae* (Lobeliaceae). Molecular Ecology 8:775-789

Castiglioni P, Ajmone-Marsan P, van Wijk R, Motto M (1999) AFLP markers in a molecular linkage map of maize: codominant scoring and linkage group distribution. Theor Appl Genet 99:425-431

Chakraborty R, Weiss KM (1991) Genetic variation of the mitochondrial DNA genome in American Indians is at mutation-drift equilibrium. Am. J. Hum. Genet. 86:497-506

Chamberlain DE, Crick HQP (1999) Population declines and reproductive performance in skylarks *Alauda arvensis* in different regions and habitats of the United Kingdom. Ibis 141:38-51

Chamberlain DE, Fuller RJ, Bunce RGH, Duckworth JC, Shrubb M (2000) Changes in the abundance of farmland birds in relation to the timing of agricultural intensification in England and Wales. J Appl Ecology 37:771-788

Chen, S. X, G. C, Lin XH, Zhang Q (2001) Improving bacterial blight resistance of '6078', an elite restorer line of hybrid rice, by molecular marker-assisted selection. Plant Breeding 120:133-137

Christen O, Evans E, Nielsson C, Haldrup C (1999) Oilseed rape cropping systems in NW Europe Proceddings of the 10th International Rapeseed Congress, Canberra, Australia

Clarke JH, Dean C (1994) Mapping *FRI*, a locus controlling flowering time and vernalization response in *Arabidopsis thaliana*. Mol. Gen. Genet. 242:81-89

Clauss MJ, Cobban H, Mitchell-Olds T (2002) Cross-species microsatellite markers for elcudating population genetic structure in *Arabidopsis* and *Arabis* (Brassicaeae). Molecular Ecology 11:591-601

Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics 1444:2001-2014

Crow JF, Kimura M (1970). In: Harper R (ed) An introduction to Population Genetics Theory, New York, USA, pp 591

Cruzan MB (2001) Population size and fragmentation thresholds for the maintenance of genetic diversity in the herbaceous endemic *Scutellaria montana* (Lamiaceae). Evolution 55:1569-1580

Culley TM, Wolfe AD (2001) Population genetic structure of the cleistogamous plant species *Viola pubescens* Aiton (Violaceae), as indicated by allozyme and ISSR molecular markers. Heredity 86:545-556

Dallas JF (1992) Estimation of microsatellite mutation rates in recombinant inbred strains of mouse. Mamm. Genome 3:452-456

Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M (1994) Mutational processes of simple sequence repeat loci in human populations. Proc. Natl. Acad. Sci. 91:3166-3170

Donald PF, Green RE, Heath MF (2001) Agricultural intensification and the collapse of Europe's farmland bird populations. Proc R Soc Lond B Biol Sci. 268:25-29

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemica. Bulletin 19:11-15

Drake JW, Glickman BW, Ripley SL (1983) Updating the theory of mutation. Am. Sci. 71:621-630

Eggers T (1978) Werden und Wandel der Ackerunkraut-Vegetation. In: Rinteln J (ed) Werden und Vergehen von Pflanzengesellschaften. Cramer 1979, pp 503-527

Environmental Signals 2002. Agriculture. http://reports.eea.eu.int/environmental_assessment_report_2002_9/en/signals2002-chap06.pdf

Erschadi S, Haberer G, Schöniger M, Torres-Ruiz RA (2000) Estimating genetic diversity of *Arabidopsis thaliana* ecotypes with amplified fragment length polymorphisms (AFLP). Theor Appl Genet 100:633-640

Erviö L-R, Salonen J (1987) Changes in the weed population of spring cereals in Finland. Annales Agriculturae Fenniae 26:210-226

Evans AD, Smith KW, Buckingham DL, Evans J (1997) Seasonal variation in breeding performance and nestling diet of cirl buntings *Emberiza cirlus* in England. Bird Study 44:66-79

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

Fuentes JL, Escobar F, Alvarez A, Gallego G, Duque MC, Ferrer M, Deus JE, et al (1999) Analyses of genetic diversity in Cuban rice varieties using isozyme, RAPD and AFLP markers. EUPHYTICA 109:107-115

Fuhr-Boßdorf K, Waldhardt R, Otte A (1999) Auswirkungen der Landnutzungsdynamik auf das Potenzial von Pflanzengemeinschaften und Pflanzenarten einer peripheren Kulturlandschaft. Verh. Ges. Ökol. 29:519-530

Garza JC, Williamson EG (2001) Detection of reduction in population size using data from microsatellite loci. Molecular Ecology 10:305-318

Gaston KJ (1996) Biodiversity. Blackwell Science, Cambridge

Goldstein DB, Linares AR, Cavalli-Sforza LL, Feldman MW (1995) An Evaluation of Genetic Distances for Use With Microsatellite Loci. Genetics 139:463-471

Goodman HM, Eckers JR, Dean C (1995) The Genome of *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. U.S.A. 92:10831-10835

Gower JC (1966) Some distant properties of latent root and vector methods used in multivariate analysis. Biometrika 53:325-338

Green RE (1995) The decline of the corncrake *Crex crex* in Britain continues. Bird Study 42:66-75

Groh S, Zacharias A, Kianian SF, Penner GA, Chong J, Rines HW, Phillips RL (2001) Comparative AFLP mapping in two hexaploid oat populations. Theor Appl Genet 102:876-884

Grundy AC, Froud-Williams RJ, Boatman ND (1995) Maternal effects in progeny of field pansy (viola arvensis) subjected to different herbicide and nitrogen rates. Ann. appl. Biol. 127:343-352

Guo S, Thompson E (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics 48:1933-1940

Hamrick JL, Godt MJ (1989) Allozyme diversity in plants. In: Brown AHD, Clegg MT, Kahler AL, Weir BS (eds) Plant population genetics,

breeding, and genetic resources. Sinauer Association Incorporated, Sunderland, Massachusetts, pp 43-63

Hartl DL, Clark AG (1997) Principles of population genetics. Sinauer Associates, Sunderland, Massachusettes

Henderson ST, Petes TD (1992) Instability of simple sequence DNA in *Saccharomyces cerevisiae*. Mol. Cell. Biol. 12:2749-2757

Hokanson SC, Lamboy WF, Szewc-McFadden AK, McFerson JR (2001) Microsatellite (SSR) variation in a collection of Malus (apple) species and hybrids. Euphytica 188:281-294

Holsinger KE, Lewis PO, Dey DK (2002) A Bayesian approach to inferring population structure from dominant markers. Molecular Ecology 11:1157-1164

Holzmann A (1987) Zur Populationsdynamik und Schadwirkung des Acker-Stiefmütterchens (*Viola arvensis* MURR.) als Grundlage einer wirtschaftlichen Bekämpfung. Hannover

Huston MA (1994) Biological Diversity. Cambridge University Press, Cambridge

Hyvönen T, Ketoja E, Salonen J (2003) Changes in the abundance of weeds in spring cereal fields in Finland. Weed Res 43:348-356

Innan H, Terauchi R, Miyashita NT (1997) Microsatellite Polymorphism in Natural Populations of the Wild Plant *Arabidopsis thaliana*. Genetics 146:1441-1452

Jaccard P (1908) Nouvelles recherches sur la distribution florale. Bulletin de la Société Vaudoise Des Sciences Naturelles 44:223-270

Johanson U, West J, Lister C, Michaels S, Amasino R, Dean C (2000) Molecular analysis of FRIGIDA, a major determinant of natural variation in Arabidopsis flowering time. Science 290:344-347

Jones CJ, Edwards KJ, Castaglione S, Winfield MO, Sala F, van de Wiel C, Bredemeijer G, et al (1997a) Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. Molecular Breeding 3:381-390

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

Katiyar SK, Chandel G, Tan Y, Zhang Y, Huang B, Nugaliyadde L, Fernando K, et al (2000) Biodiversity of Asian rice gall midge (*Orseolia oryzae* Wood Mason) from five countries examined by AFLP analysis. Genome 43:322-332

Kimura M, Crow JF (1964) The number of alleles that can be maintained in a finite population. Genetics 49:725-738

Kimura M, Otha T (1978) Stepwise mutation model and distribution of allelic frequencies in a finite population. Proceedings of the National Academy of Sciences of the USA 75:2868-2872

Koornneef M, Alonso-Blanco C, Peeters AJM (1997) Genetic approaches in plant physiology. New Phytol. 137:1-8

Krauss S (2000) Accurate gene diversity estimates from amplified fragment length polymorphism (AFLP) markers. Molecular Ecology 9:1241-1245

Kudsk P, Streibig JC (2003) Herbicides - a two-edged sword. Weed Res 43:90-102

Lacy RC (1987) Loss of genetic diversity from managed populations: Interacting effects of drift, mutation, immigration, selection, and population subdivision. Conservation Biology 1:143-158

Lahaye T, Hartmann S, Topsch S, Freialdenhoven A, Yano M, Schulze-Lefert P (1998) High-resolution genetic and physical mapping of the Rar1 locus in barley. Theor Appl Genet 97:526-534

Lammi A, Siikamaki P, Mustajarvi K (1999) Genetic Diversity, Population Size, and Fitness in Central and Peripheral Populations of a Rare Plant Lychnis viscaria. Conserv Biol 13:1069-1078

Lewontin RC, Hubby JL (1966) A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. Genetics 54:595-609

Loridon K, Cournoyer B, Goubley C, Depeiges A, Picard G (1998) Length polymorphism and allele structure of trinucleotide microsatellites in natural accessions of *Arabidopsis thaliana*. Theor Appl Genet 97:591-604

Luikart G, Allendorf FW, Cornuet J-M, Sherwin WB (1998) Distortion of allele frequency distributions provides a test for recent population bottlenecks. Journal of Heredity 89:238-247

Lutman PJW, Cussans GW, Wright KJ, Wilson BJ, Wright GM, Lawson HM (2002) The persistence of seeds of 16 weed species over six years in two arable fields. Weed Res 42:231-241

Lynch M (1996) A quantitative-genetic perspective on conservation issues. In: Avise JC, Hamrick JL (eds) Conservation genetics: case histories from nature. Chapman & Hall, New York

Mackill DJ, Nguyen HT, Zhang J (1999) Use of molecular markers in plant improvement programs for rainfed lowland rice. Field Crops Research 64:177-185

Manifesto MM, Schlatter AR, Hopp HE, Suarez EY, Dubcovsky J (2001) Quantitative evaluation of genetic diversity in wheat germplasm using molecular markers. Crop Sci. 41:682-690

Mannion AM (1997) Agriculture and land transformation. 1. Temporal and spatial dimensions. Outlook on Agriculture 26:71-78

Mannion AM (1999) Domestication and the origins of agriculture: an appraisal. Progress in Physical Geography 23:37-56

Mantel N (1967) The detection of disease clustering and a generalized regression approach. Cancer Research 27:209-220

Marshall EJP, Brown VK, Boatman ND, Lutman PJW, Squire GR, Ward LK (2003) The role of weeds in supporting biological diversity within crop fields. Weed Res 43:77-89

Michener CD, Sokal RR (1957) A quantitative approach to a problem in classification. Evolution 11:130-162

Miyashita NT, Kawabe A, Innan H (1999) DNA Variation in the Wild Plant *Aarabidopsis thaliana* Revealed by Amplified Fragment Length Polymorphism Analysis. Genetics 152:1723-1731

Morgante M, Rafalski A, Biddle P, Tingey S, Olivieri AM (1994) Genetic mapping and variability of seven soyabean simple sequence repeat loci. Genome 37:763-769

Nachit MM, Elouafi I, Pagnotta MA, El Saleh A, Iacono E, Labhilili M, Asbati A, et al (2001) Molecular linkage map for an intraspecific recombinant inbred population of durum wheat (*Triticum turgidum* L. var. durum). Theor Appl Genet 102:177-186

Napp-Zinn K (1985) *Arabidopsis thaliana*. In: Halevy AH (ed) CRC Handbook of Flowering. CRC, Boca Raton, FL, pp 492-503

Narvel JM, Fehr WR, Chu WC, Grant D, Shoemaker RC (2000) Simple sequence repeat diversity among soybean plant introductions and elite genotypes. Crop Science 40:1452-1458

Nei M (1977) *F*-statistics and analysis of gene diversity in subdivided populations. Annals of Human Genetics 41:225-233

Nei M (1987) Molecular Evolutionary Genetics. Columbia University Press, New York, NY, USA

Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and genetic variability in populations. Evolution 29:1-10

Nunney L (2002) The effective size of annual plant populations: the interaction of a seed bank with fluctuating population size in maintaining genetic variation. The American Naturalist 160:195-204

Odum S (1965) Germination of ancient seeds. Dansk Botanisk Arkiv 24:1-70

Oostermeijer G (1996) Population viability of the rare Gentiana pneumonanthe: the relative importance of demography, genetics, and reproductive biology [PhD Thesis]. Universiteit van Amsterdam, Amsterdam

Pakniyat H, Powell W, Baird E, Handley LL, Robinson D, Scrimgeour CM, Nevo E, et al (1997) AFLP variation in wild barley (*Hordeum sponta-*

neum C. Koch) with reference to salt tolerance and associated eco-geography. Genome 40:332-341

Panaud O, Chen X, McCouch SR (1996) Development of microsatellite markers and characterization of simple sequence length polymorphism (SSLP) in rice (*Oryza sativa* L.). Mol Gen Genet 252:597-607

Park SDE (2001) Trypanotolerance in West African Cattle and the Population Genetic Effects of Selection [PhD thesis]. University of Dublin

Parsons PA (1996) Conservation strategies: adaption to stress and the preservation of genetic diversity. Biological Journal of the Linnean Society 58:471-482

Pigliucci M (1998) Ecological and evolutionary genetics of *Arabidopsis*. Trends In Plant Science 3:485-489

Pigliucci M (2002) Ecology and Evolutionary Biology of *Arabidopsis*. In: Biologists ASoP (ed) The Arabidopsis Book

Piry S, Luikart G, Cornuet J-M (1999) BOTTLENECK:a computer program for detecting recent reductions in effective population size from allele frequency data. Journal of Heredity 90:502-503

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. Molecular Breeding 2:225-238

Powell WG, Morgante M, Doyle JJ, McNicol JW, Tingey SV, Rafalski AJ (1995) Polymorphic simple sequence repeat regions in chloroplast genomes: applications to the population genetics of pines. Proc. Natl. Acad. Sci. 92:7759-7763

Redei GP (1992) A heuristic glance at the past of Arabidopsis genetics. In: Koncz N-HCCaS, J. (ed) Methods in Arabidopsis Research. World Scientific, Singapore, pp 1-15

Ren N, Timko MP (2001) AFLP analysis of genetic poymorphism and evolutionary relationships among cultivated and wild *Nicotiana* species. Genome 44:559-571

Reusch TBH, Stam WT, Olsen JL (2000) A microsatellite-based estimation of clonal diversity and population subdivision in *Zostera marina*, a marine flowering plant. Molecular Ecology 9:127-140

Roberts HA, Feast PM (1973) Emergence and longevity of seeds of annual weeds in cultivated and undisturbed soil. Journal of Applied Ecology 10:133-143

Rohlf FJ (1997) NTSYS-PC numerical taxonomy and multivariate analysis system, Version 2.01e. Exeter Software, Setauket, NY.

Russell JR, Fuller JD, Macaulay M, Hatz BG, Jahoor A, Powell W, Waugh R (1997) Direct comparison of levels of genetic variation among barley accessions detected by RFLPs, AFLPs, SSRs and RAPDs. Theor Appl Genet 95:714-722

Saghai Maroof MA, Biyashev RM, Yang GP, Zhang Q, Allard RW (1994) Extraordinarily polymorphic microsatellite DNA in barley: Species diversity, chromosomal locations, and population dynamics. Proc. Natl. Acad. Sci. 91:5466-5470

Saiki RK, Scharf S, Faloona F (1985) Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science 230:1350-1354

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. Molecular Ecology 10:1947-1958

Schneider S, Keuffer J-M, Roessli D, Excoffier L (1997) Arlequin ver. 1.1: A software for population genetic data analysis. Genetics and Biometry Laboratory. University of Geneva, Switzerland

Schubert P (2003) Phänotypische Plastizität und Fitness annueller Ackerwildkrautpopulationen von Standorten unterschiedlicher Nutzungsgeschichte [PhD Thesis]. Justs-Liebig Universität, Giessen

Slatkin M (1991) Inbreeding coefficients and coalescence times. Genet. Res. 58:167-175

Slatkin M (1994) Linkage disequilibrium in growing and stable populations. Genetics 137:331-336

Slatkin M (1995) A Measure of Population Subdivision Based on Microsatellite Allele Frequencies. Genetics 139:457-462

Sletvold N (2002) Effects of plant size on reproductive output and offspring performance in the facultative biennial *Digitalis purpurea*. J Ecology 90:958-966

Snape JW, Lawrence MJ (1971) The breeding system of *Arabidopsis thaliana*. Heredity 27:299-302

Snowdon RJ, Friedt W (2004) Molecular markers in *Brassica* oilseed breeding: current status and future possibilities. Plant Breeding 123:1-8

Stoate C, Boatman ND, Borralho RJ, Rio Carvalho C, de Snoo GR, Eden P (2001) Ecological impacts of arable intensification in Europe. Journal of Environmental Management 63:337-365

Stra-Co (2002) Agriculture and Biodiversity in Europe.

Stratton DA, Bennington CC (1996) Measuring spatial variation in natural selection using randomly-sown seeds of *Arabidopsis thaliana*. Journal of Evolutionary Biology 9:215-228

Tero N, Aspi J, Siikamäki P, Jäkäläniemi A, Tuomi J (2003) Genetic structure and gene flow in a metapopulation of an endangered plant species, *Silene tatarica*. Molecular Ecology 12:2073-2085

Todokoro S, Terauchi R, Kawano S (1995) Microsatellite polymorphisms in natural populations of *Arabidopsis thaliana* in Japan. Jpn. J. Genet. 70:543-554

Udupa SM, Baum M (2001) High mutation rate and mutational bias at $(TAA)_n$ microsatellite loci in chickpea (*Cicer arietinum* L.). Mol Genet Genomics 265:1097-1103

Virk PS, Pooni HS, Syed NH, J. KM (1999) Fast and reliable genotype validation using microsatellite markers in *Arabidopsis thaliana*. Theor Appl Genet 98:462-464

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, et al (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407-4414

Waldman M, Shevah Y (2000) Biological Diversity - An Overview. Water, Air, and Soil Pollution 123:299-310

Weber JL, Wong C (1993) Mutation of human short tandem repeats. Hum. Mol. Genet. 2:1123-1128

Wilson BJ, Lawson HM (1992) Seedbank persistence and seedling emergence of seven weed species in autumn-sown crops following a single year's seeding. Annals of applied Biology 120:105-116

Wilson JD, Morris AJ, Arroyo BE, Clark SC, Bradbury RB (1999) A review of the abundance and diversity of invertebrate and plant foods of granivorous birds in northern Europe in relation to agricultural change. Agriculture Ecosystems & Environment 75:13-30

Wolff RK, Plaeke R, Jeffreys AJ, White R (1991) Unequal crossing over between homologous chromosomes is not the major mechanism involved in the generation fo new alleles at VNTR loci. Genomics 5:382-384

Wright S (1921) Systems of mating, I-V. Genetics 6:111-178

Wright S (1931) Evolution in Mendelian populations. Genetics 16:97-159

Wright S (1951) The genetical structure of populations. Annals of Eugenics 15:323-354

Wright S (1965) The interpretation of population structure by F-statistics with special regard to system of mating. Evolution 19:395-420

Wright S (1977) Evolution and the Genetics of Populations, Experimental Results and Evolutionary Deductions. The University of Chicago Press, Chicago

Wright S (1978) Evolution and the Genetics of Population, Variability Within and Among Natural Populations. The University of Chicago Press, Chicago

Zabeau M, Vos P (1993) Selective restriction fragment amplification: a general method for DNA fingerprinting European Patent Application EP 534858A1

Appendices

The appendices are divided into three sections. Each section contains a multi-page dendrogram calculated using NTSYS as described in Section 3: Materials and Methods.

Appendix I: Cluster analysis of Jaccard similarity measurements between *A. thaliana* individuals from the four regimes using AFLP.

Appendix II: Cluster analysis of Jaccard similarity measurements between *A. thaliana* individuals from the four regimes using microsatellites.

Appendix III: Cluster analysis of Jaccard similarity measurements between *V. arvensis* individuals from the four regimes using AFLP.

For all dendrograms the samples have been colour coded to facilitate the identification of groups. The colours used are shown below in Tab. 40.

Tab. 40: Colour code used for sample identification in dendrograms.

	Intensive	Extensive	Meadow	Fallow
Amöneburg				
Erda				
Eibelshausen				

Two other types of samples are shown in the dendrograms, those from Rhön, and those obtained from the NASC. The latter have not been designated any colour, while those from Rhön are labelled in orange.

Appendix I

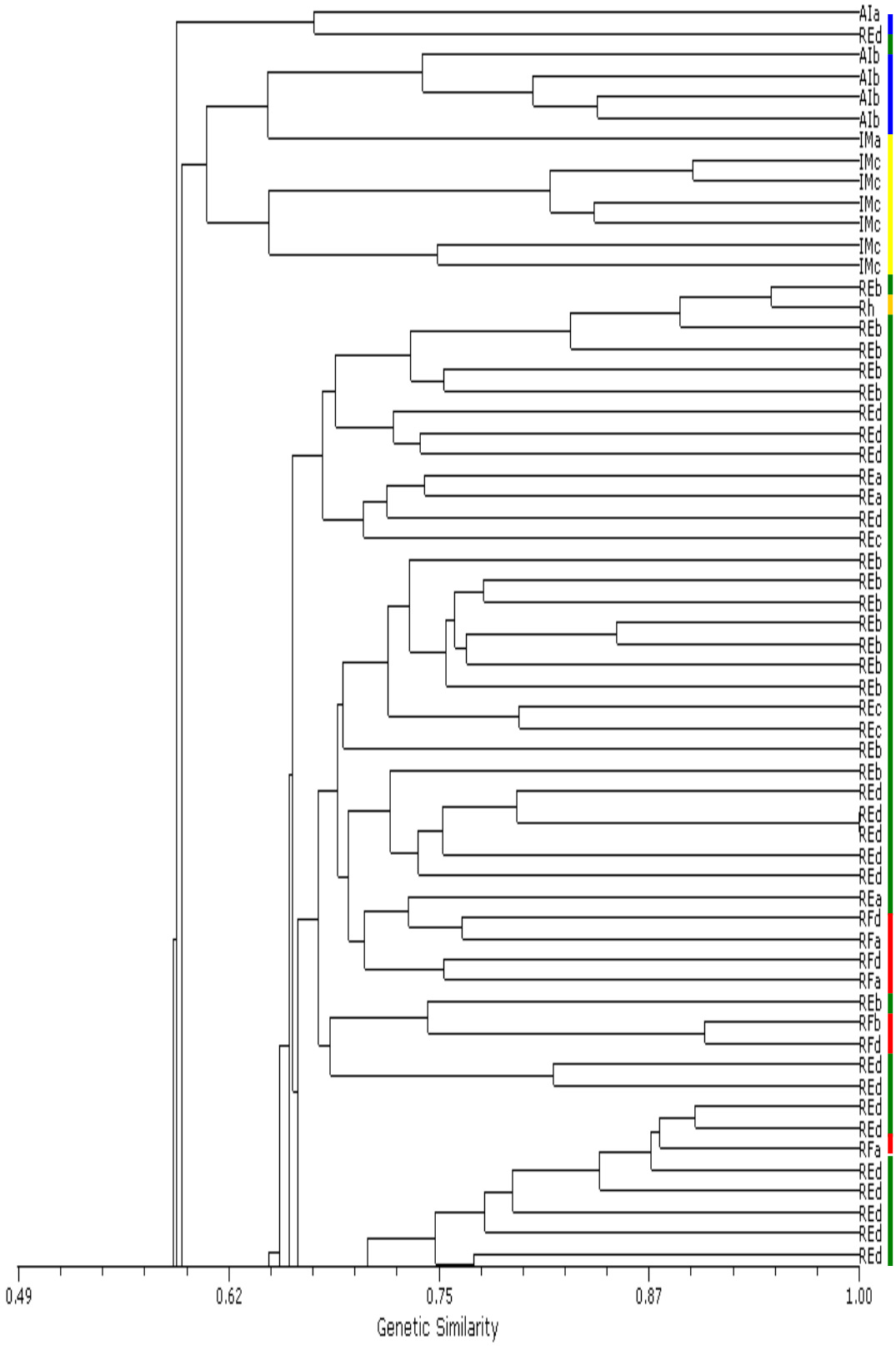


Fig. 14: Cluster analysis of Jaccard similarity measurements between *A. thaliana* individuals from the four regimes using AFLP.

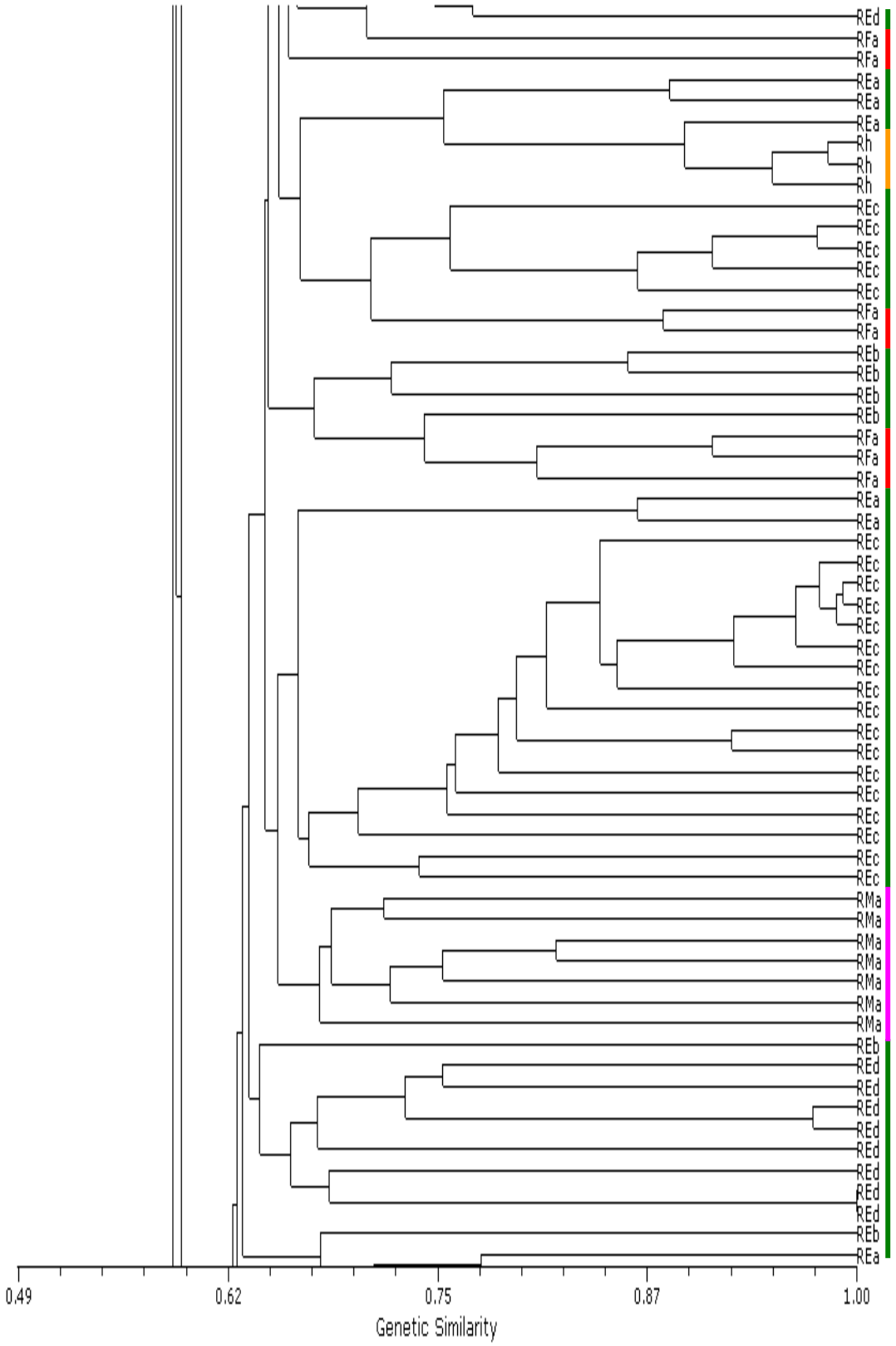


Fig. 14 (cont.): Cluster analysis of Jaccard similarity measurements between *A. thaliana* individuals from the four regimes using AFLP.

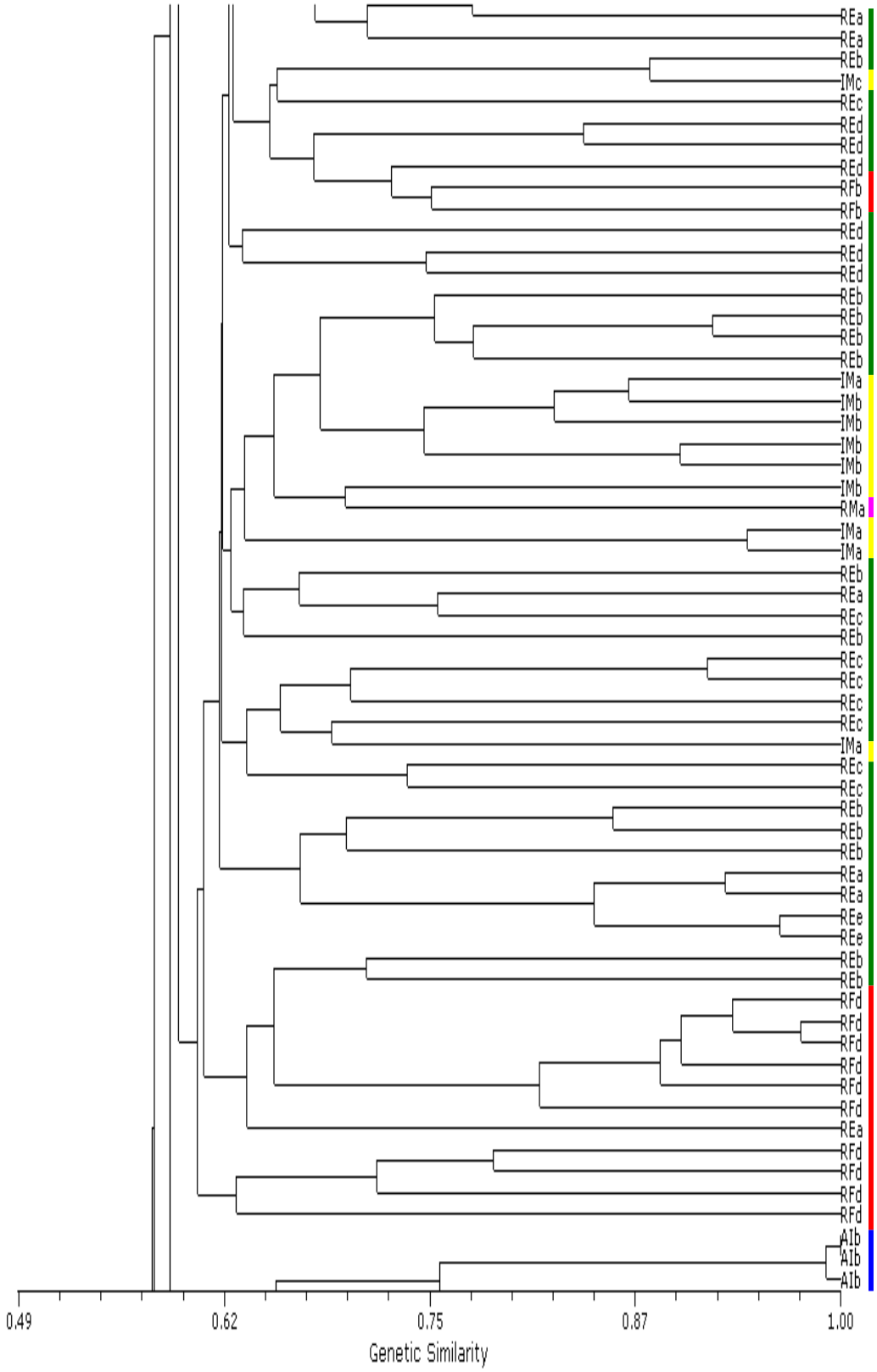


Fig. 14 (cont.): Cluster analysis of Jaccard similarity measurements between *A. thaliana* individuals from the four regimes using AFLP.

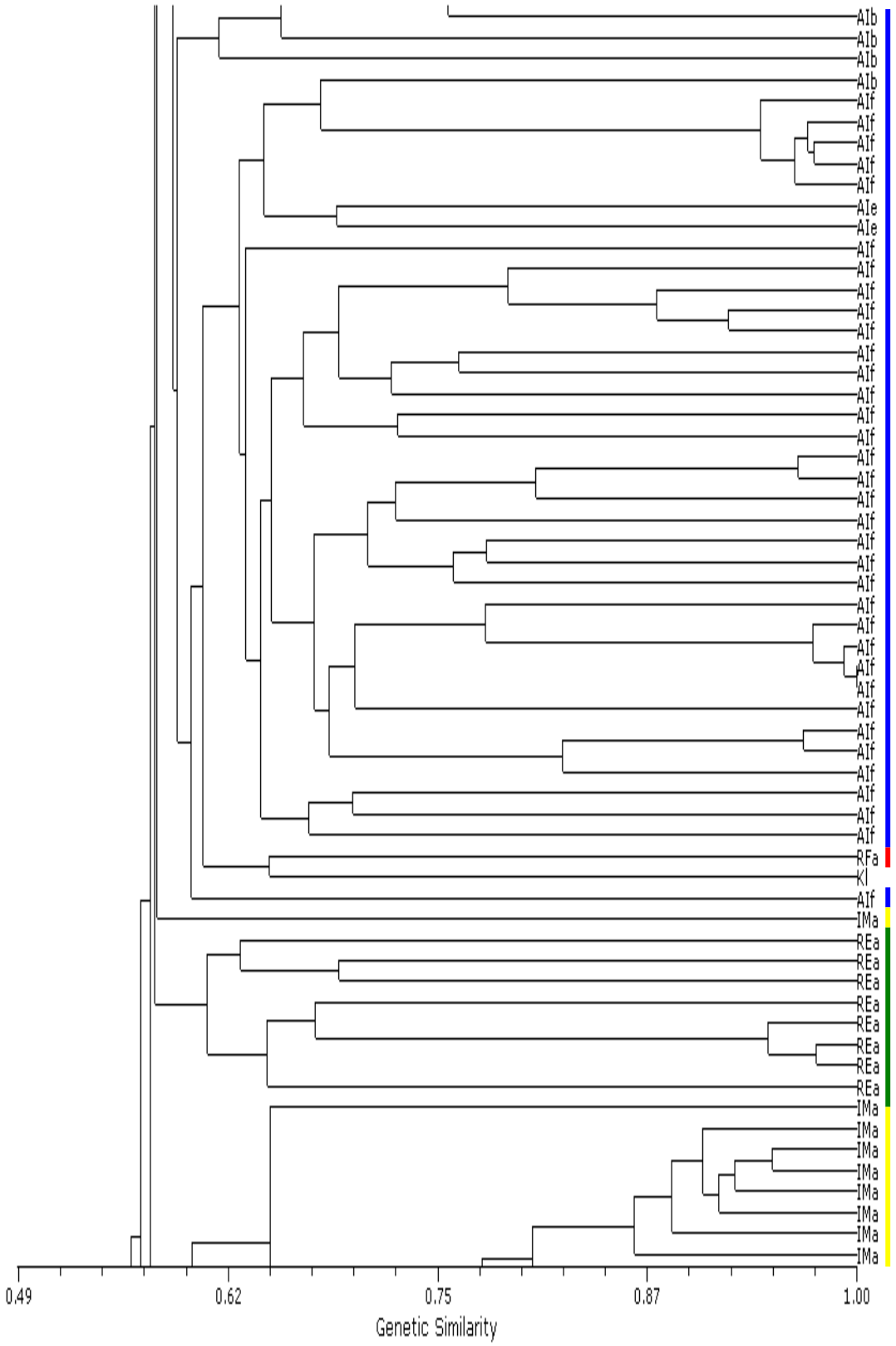


Fig. 14 (cont.): Cluster analysis of Jaccard similarity measurements between *A. thaliana* individuals from the four regimes using AFLP.

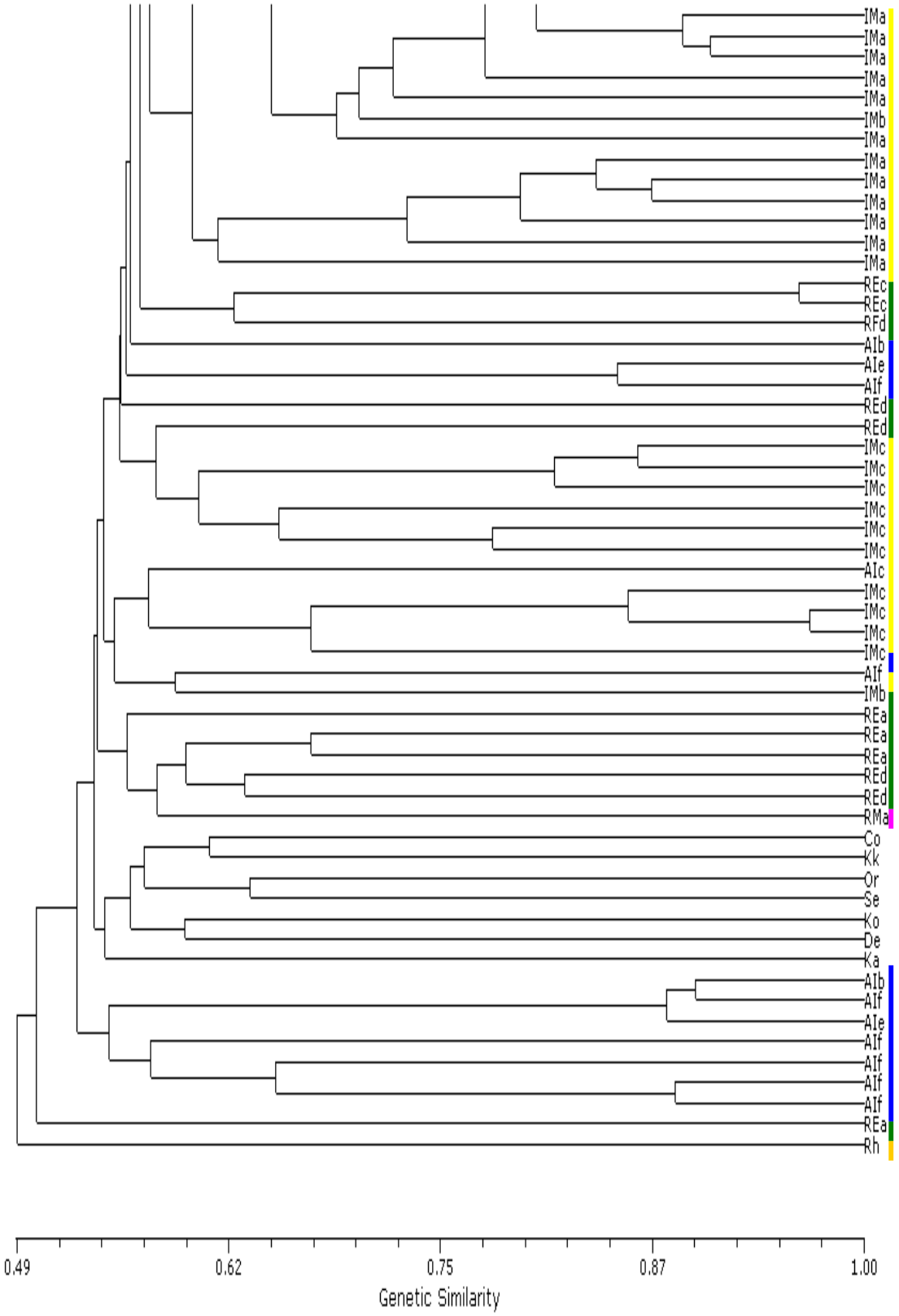


Fig. 14 (cont.): Cluster analysis of Jaccard similarity measurements between *A. thaliana* individuals from the four regimes using AFLP.

Appendix II

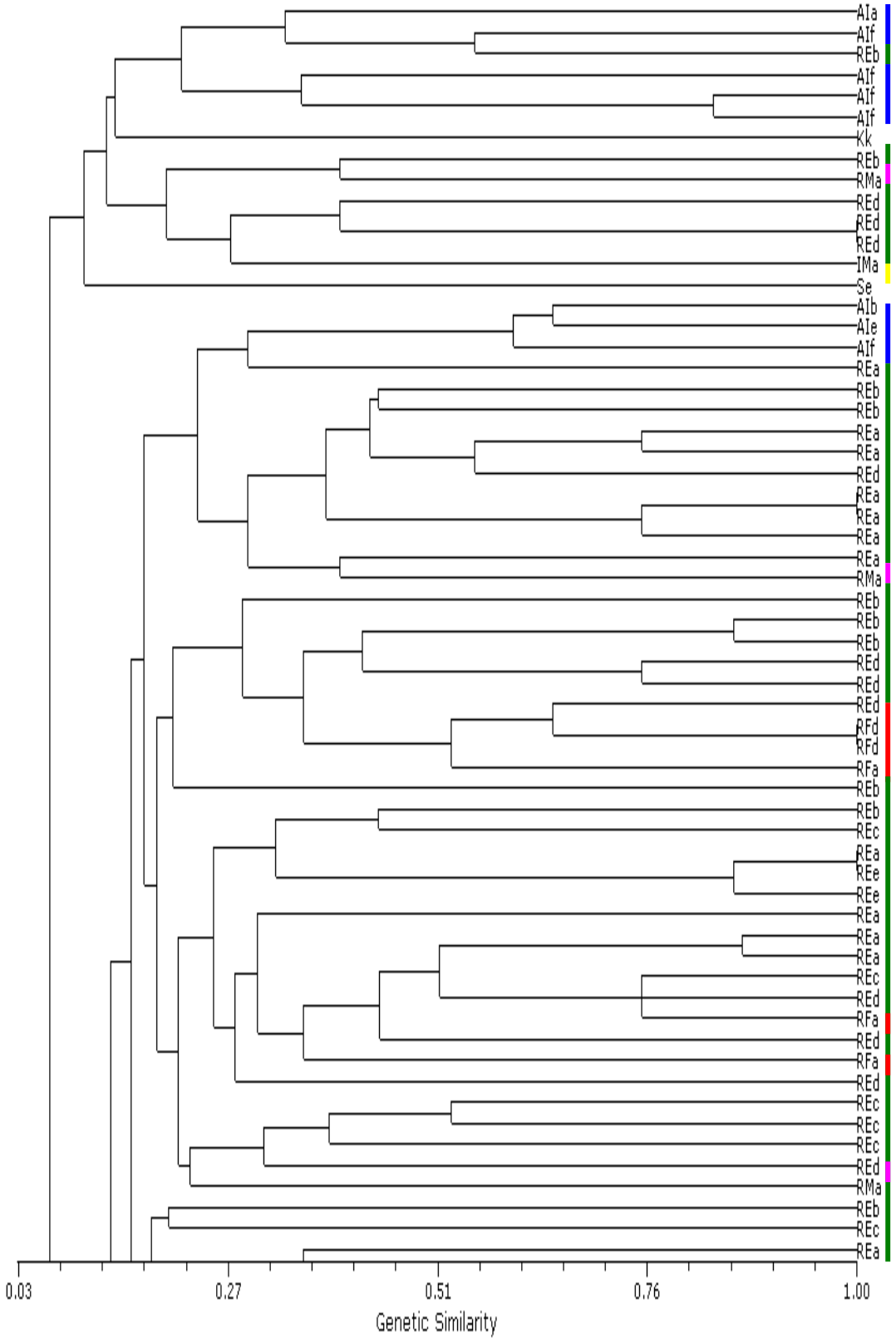


Fig. 15 : Cluster analysis of Jaccard similarity measurements between *A. thaliana* individuals from the four regimes using microsatellites.

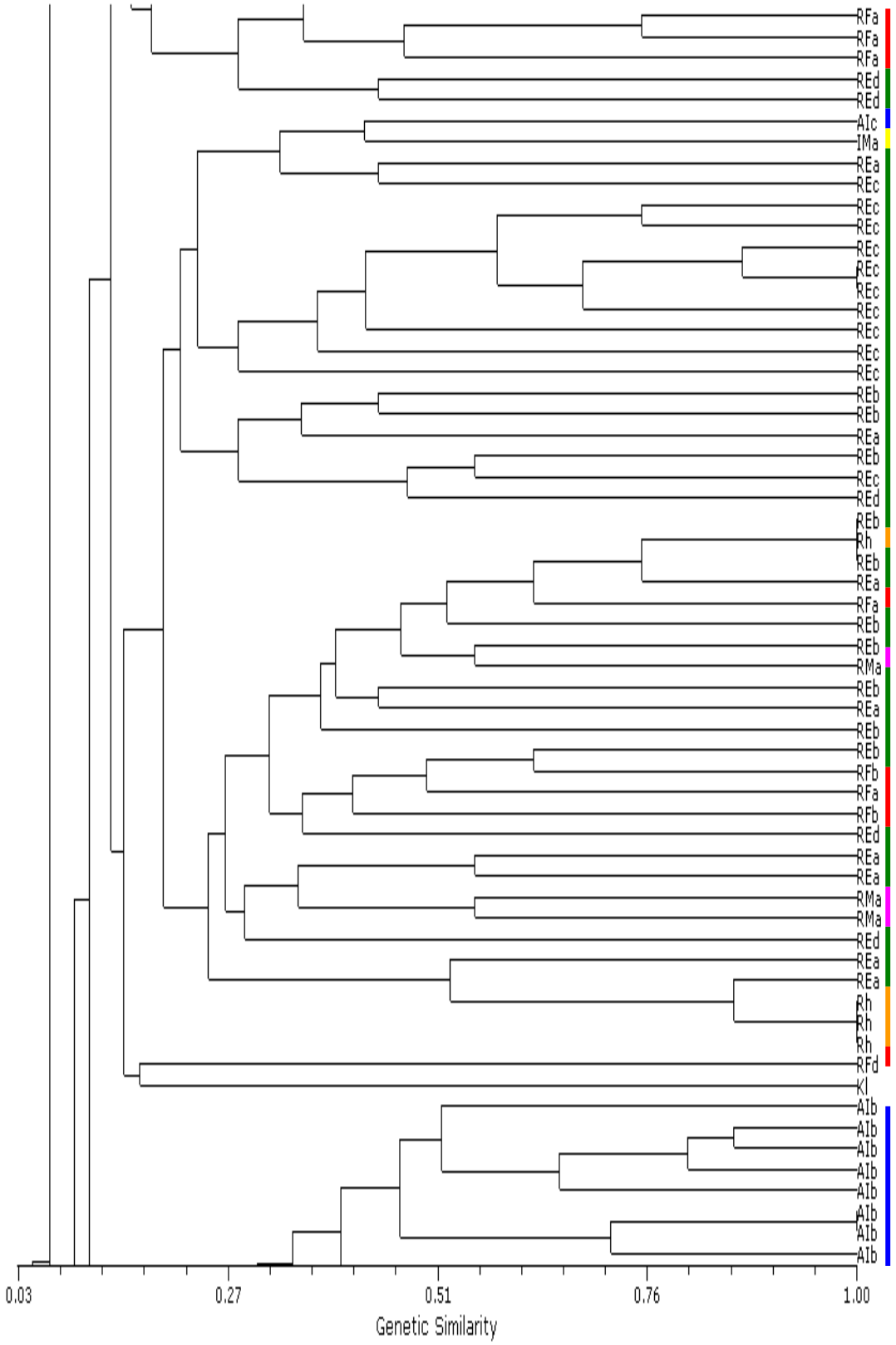


Fig. 15 (cont.): Cluster analysis of Jaccard similarity measurements between *A. thaliana* individuals from the four regimes using microsatellites.

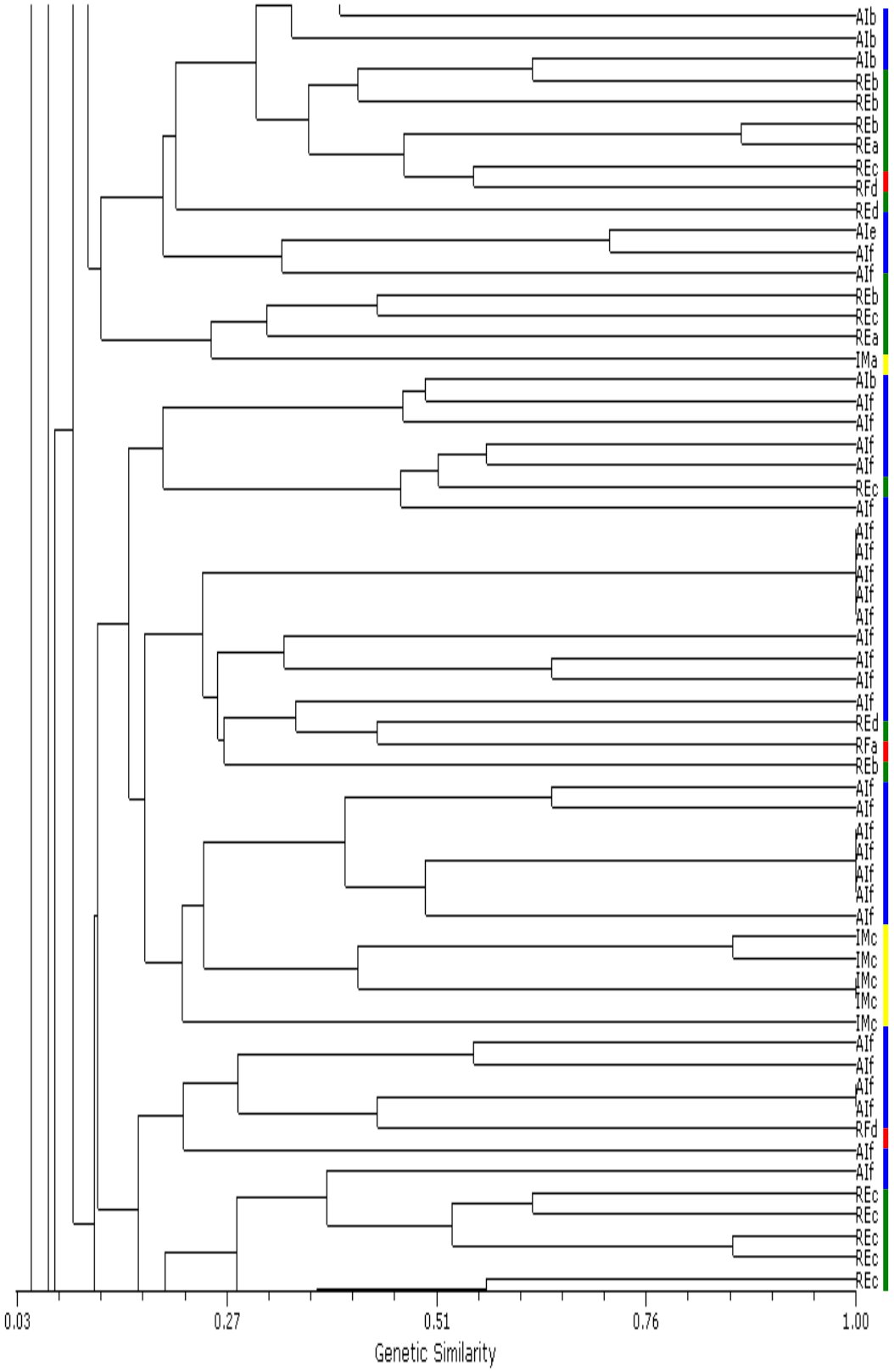


Fig. 15 (cont.): Cluster analysis of Jaccard similarity measurements between *A. thaliana* individuals from the four regimes using microsatellites.

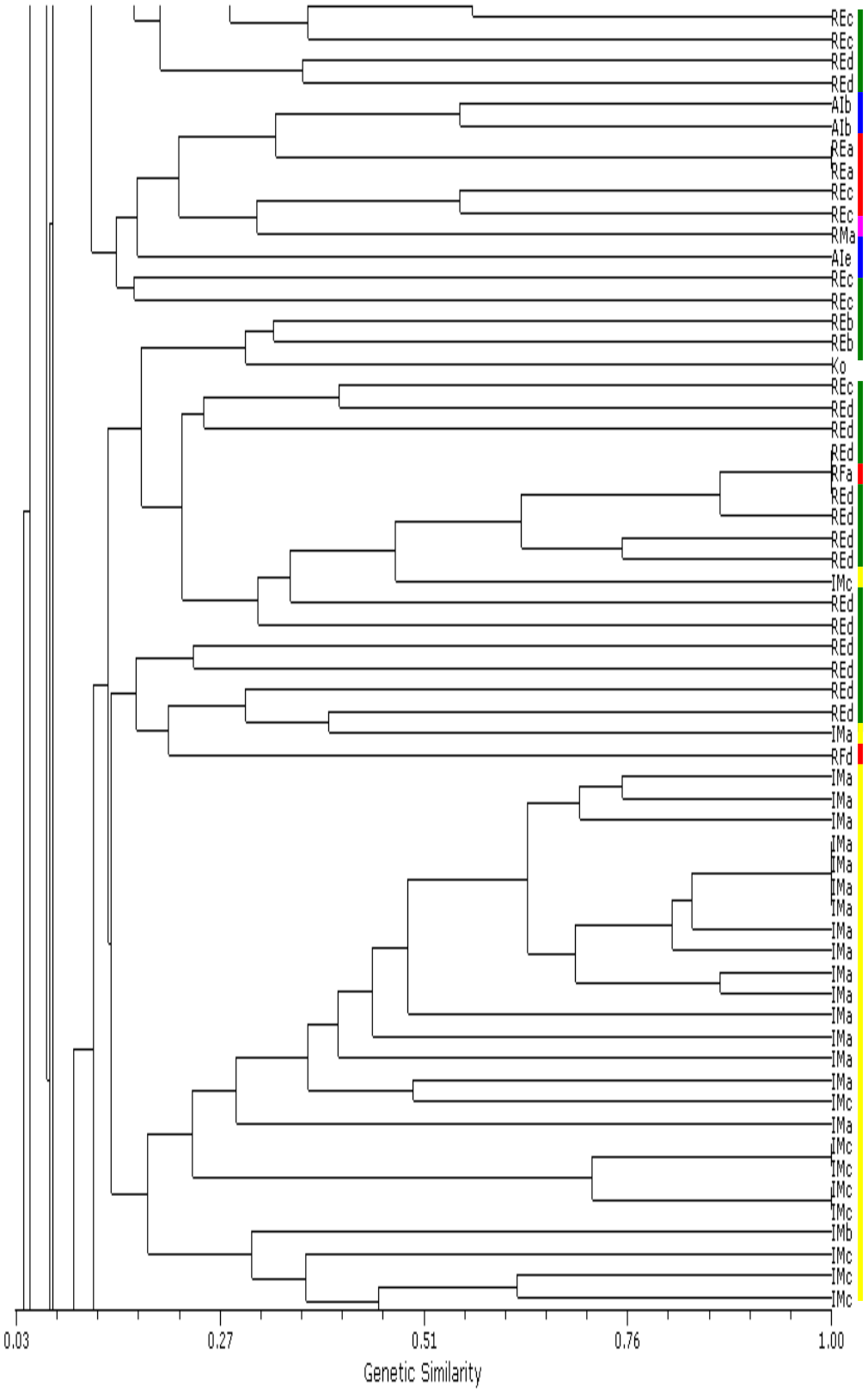


Fig. 15 (cont.): Cluster analysis of Jaccard similarity measurements between *A. thaliana* individuals from the four regimes using microsatellites.

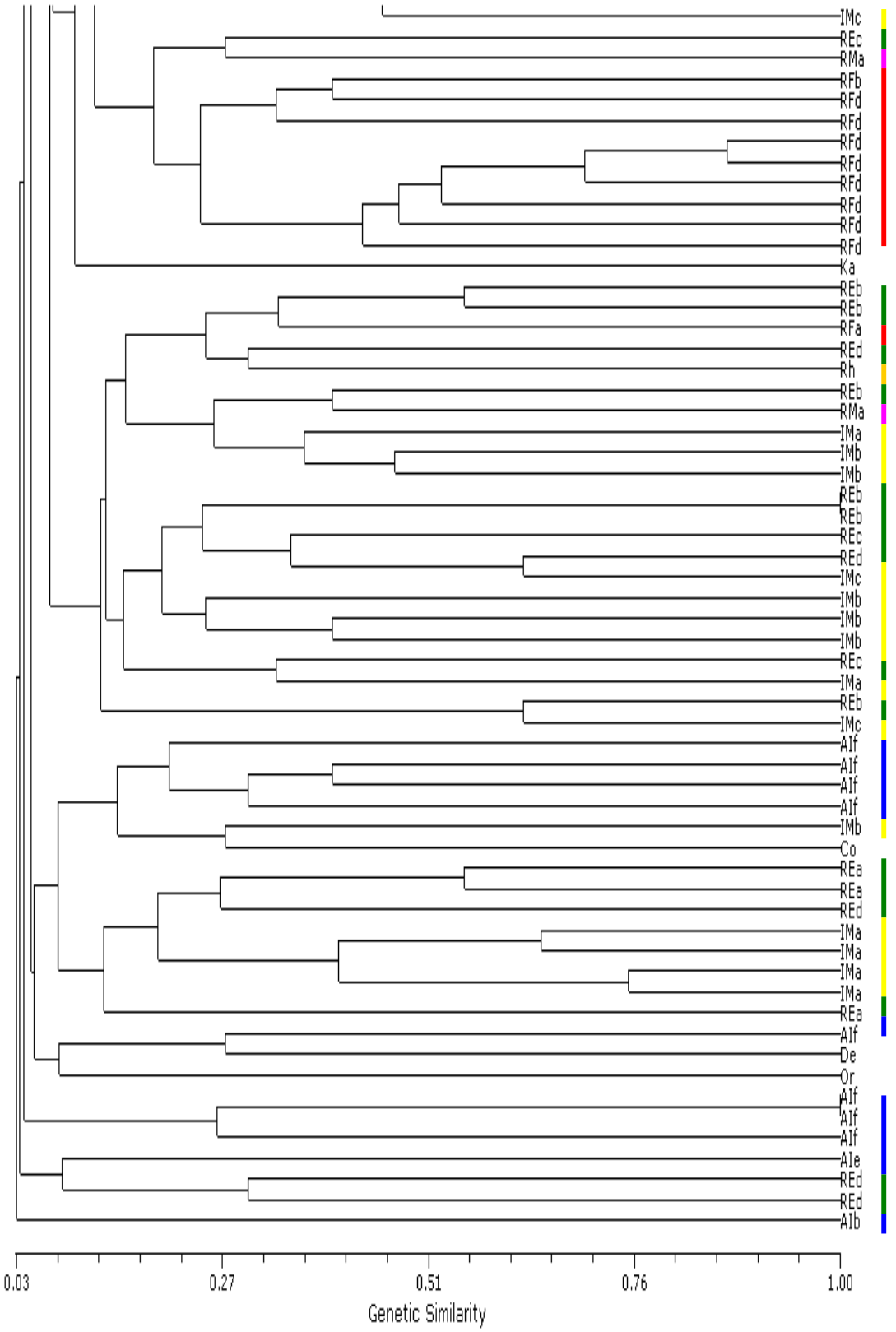


Fig. 15 (cont.): Cluster analysis of Jaccard similarity measurements between *A. thaliana* individuals from the four regimes using microsatellites.

Appendix III

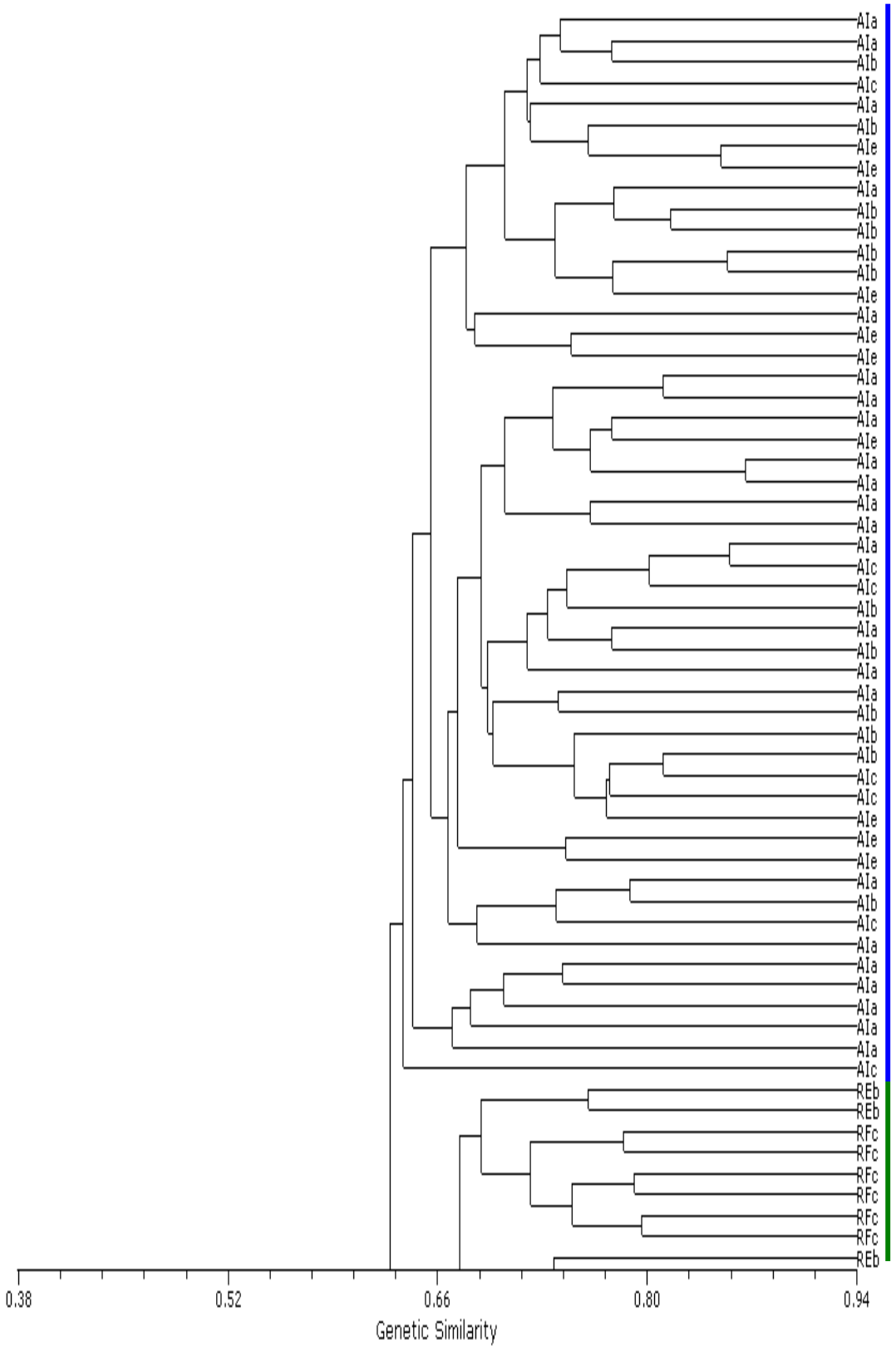


Fig. 16: Cluster analysis of Jaccard similarity measurements between *V. arvensis* individuals from the four regimes using AFLP.

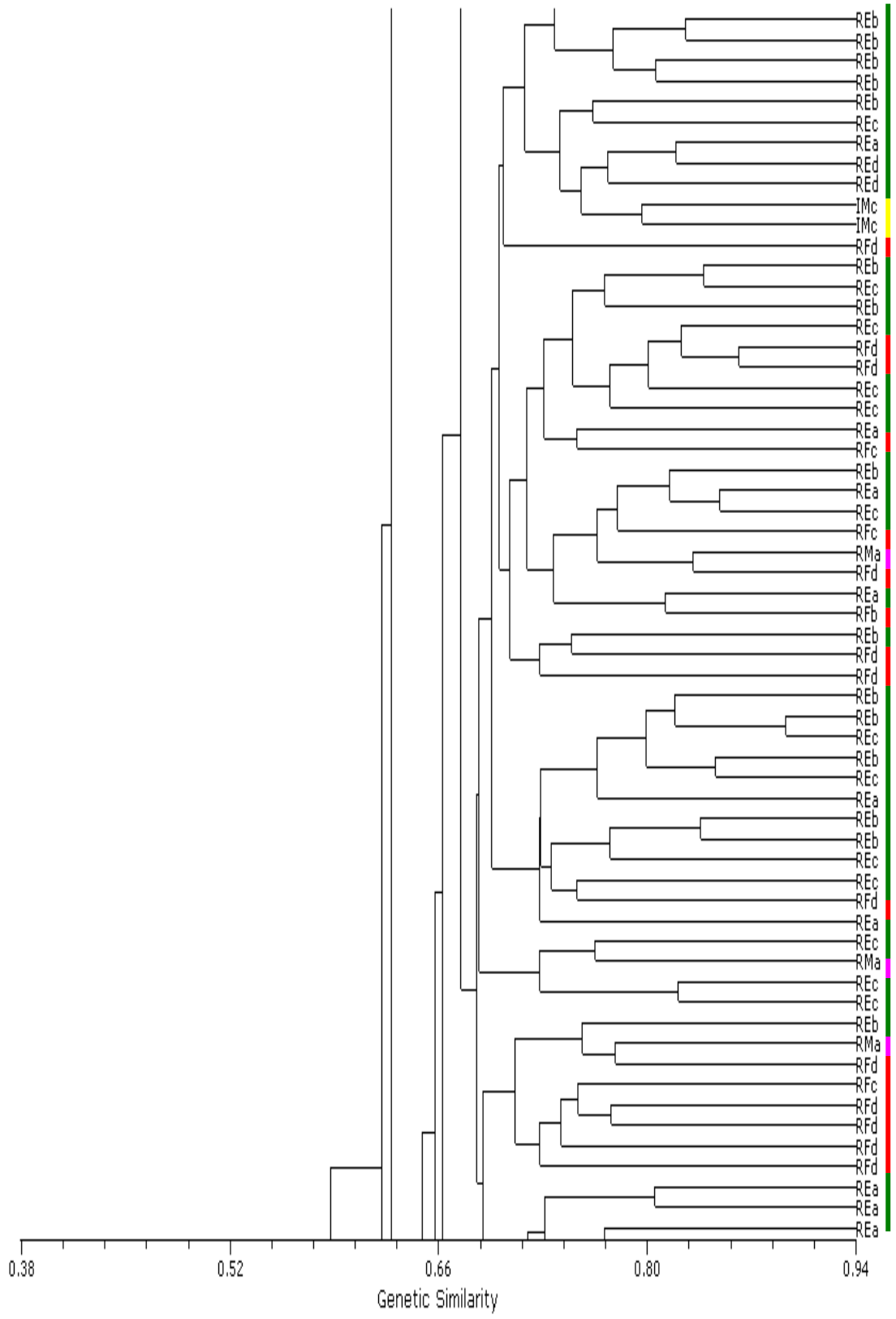


Fig. 16 (cont.): Cluster analysis of Jaccard similarity measurements between *V. arvensis* individuals from the four regimes using AFLP.

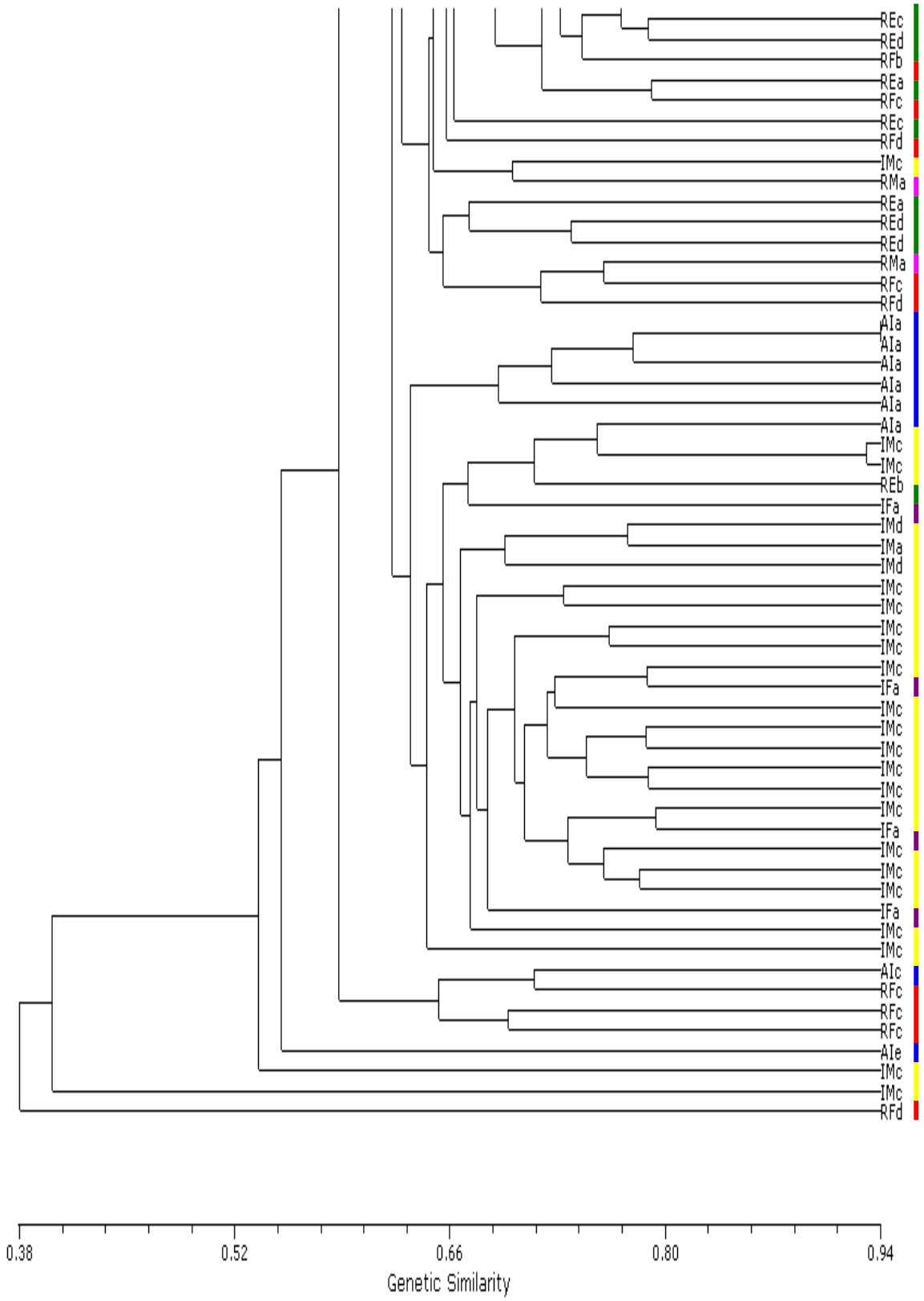


Fig. 16 (cont.): Cluster analysis of Jaccard similarity measurements between *V. arvensis* individuals from the four regimes using AFLP.

Acknowledgements

Many people have been helpful through the process of completing this work.

First and foremost, I would like to thank Prof. Dr. Wolfgang Köhler for allowing me the opportunity to study for a PhD in the Biometry Institute. As this project was also carried out in conjunction with the Land Ecology Institute, I would also like to thank Prof. Dr. Dr. Annette Otte as the second examiner.

Both the DFG and JLU-Giessen have helped fund this project, and without them it could never have been achieved.

I would also like to make a special thanks to Dr. Rod Snowdon, who patiently helped me integrate into the position and culture.

Throughout this project I was helped regularly by Patrick Schubert from the Land Ecology Institute, and his knowledge and insights were invaluable.

There was also much to learn in the laboratory, and I thank Petra Theuer for the time she spent helping me achieve what was needed.

And last, but not least, thanks to all my colleagues in the Biometry Institute for the fun atmosphere that you provided for the duration of my stay here.