Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

Fakulteten för landskapsplanering,
trädgårds– och jordbruksvetenskap

Genetic differences in *Lobesia botrana* populations

– related to host plant or geographic origin?

Sophia Döös

Självständigt arbete vid LTJ-fakulteten, SLU
Examensarbete 30 hp, Alnarp 2013

Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

Fakulteten för landskapsplanering,
trädgårds– och jordbruksvetenskap

Genetic differences in *Lobesia botrana* populations - related to host plant or geographic origin?
Genetiska skillnader mellan *Lobesia botrana* populationer - relaterade till värdväxt eller geografiskt ursprung?

Sophia Döös

Handledare: *Mattias Larsson*, SLU, Institutionen för Växtskyddsbiologi

Btr handledare: *Silvia Vezulli*, Fondazione Edmund Mach di San Michele all'Adige, Italy
Franklin Nyabuga, SLU, Institutionen för Växtskyddsbiologi
Marco Tasin, SLU, Institutionen för Växtskyddsbiologi

Examinator: *Mulatu Geleta*, SLU, Institutionen för växtförädling

Kurstitel: Examensarbete inom hortonomprogrammet
Kurskod: EX0544
Omfattning: 30 hp
Nivå: E
Fördjupning: A2E

Program/utbildning:
Ämne: Biologi

Serienamn: Självständigt arbete vid LTJ-fakulteten, SLU
Utgivningsort: Alnarp
Utgivningsmånad och -år: maj 2013
Elektronisk publicering: http://stud.epsilon.slu.se

Omslagsbild: Andrea Lucchi – Lobesia botrana

Nyckelord: *Lobesia botrana, Grapevine moth, pest, vineyard, AFLP, SSR, population genetics, Daphne gnidium*

Ansvarig institution:
SLU, Sveriges lantbruksuniversitet
Fakulteten för Landskapsplanering, trädgårds- och jordbruksvetenskap
Institutionen för växtförädling

Medansvarig institution:
Institutionen för arbetsvetenskap, ekonomi och miljöpsykologi
Institutionen för biosystem och teknologi
Institutionen för växtskyddsbiologi

Abstract

For the past six decades the model of geographical isolation has been dominating but recently there has been a renewed interest in the possibility of sympatric speciation. Races of phytophagous insects specializing on different species of host plants have often been studied for the purpose of solving the question of sympatric speciation. I have characterized genetic differentiation of populations of a polyphagous moth, *Lobesia botrana* to compare geographic separation and separation based on different host plants as potential causes for population genetic structure. I have used two molecular marker techniques: microsatellites and AFLPs, and the data was analysed using several statistical methods, with the original aim to compare which of these analyses could be most helpful in providing useful information about genetic differentiation. Both microsatellites and AFLPs detected a genetic variance related to Geographic distance and AFLPs were able to detect significant differentiations related to host plant that the microsatellites were not able to detect. Pair wise genetic distance comparisons indicated a higher differentiation according to host plants than the differentiation according to geographic distance. On the other hand, the overall differentiation between samples from different locations was higher than the differentiation between samples from different host plants. This indicates that a host related adaptation can be going on in *L. botrana* populations causing reproductive barriers between groups developing on Daphne and groups developing on grapes, but they do not seem to have formed host races yet. Low heterozygosity values and high inbreeding coefficients in the studied populations need further studies to make strong conclusion.

1. INTRODUCTION

1.1 Speciation

Speciation is the gradual process by which a species is diverged into two or more species. Concerning the mechanisms of speciation, for the past six decades the model of geographical isolation (allopatric speciation) has been dominating (Turelli et al. 2001). However, there has recently been a renewed interest in the possibility of sympatric speciation (Schilthuizen 2000; Mallet 2001; Via 2001), which in contrast to allopatric speciation, does not involve reproductive barriers due to geographic isolation. Instead, natural selection and reinforcement are thought to play the greatest roles in sympatric speciation (Kondrashov 1986; Silvertown et al. 2005). "Host race" is a concept that describes partially reproductively isolated, conspecific populations specializing on alternative hosts (Diehl and Bush 1984). The formation of host races depends strongly on assortative mating (Bush 1994) and it is likely to be an intermediate stage in the continuum between polymorphism and separate species. Differences in behavior between conspecific individuals can lead to the creation of reproductive barriers. Strong host plant association has in some cases been coupled with assortative mating and complete reproductive isolation under sympatric conditions (Craig et al. 1993, 1997, 2001; Itami et al. 1997; Rice 1987)**.** This means that, if host race formation has occurred, gene flow among host-associated populations should be low and the sympatric host-associated populations would likely be genetically differentiated (Feder et al. 1988, 1990; Waring et al. 1990)**.** The plant kingdom produces lots of different chemical substances and the defense mechanisms vary among different plant families and species. Specialization in plant-utilizing insects occurs through adaptation to the specific chemistry of the host (Wennerström et al. 2010).

1.2 Phytophagous insects and sympatric speciation

Races of phytophagous insects specializing on different species of host plants are well documented, and have often been studied for the purpose of solving the question of sympatric speciation (Bush 1969; Diehl & Bush 1984; Via 2001; Drés & Mallet 2002). Many authors have suggested that, for phytophagous insects, a host shift is a significant early step in the diversification process leading towards speciation (Bush 1994; Larsson and Ekblom 1995; Pellmyr and Leebens-Mack 1998; Feder 1998)**.** Recent evidence suggests that host shifts and host race formation is more common among herbivorous insects than was earlier presumed (Bush 1969, 1994; Tauber and Tauber 1989; Craig et al. 1993; Pratt 1994; Crozier and Pamilo 1996; Johnson et al. 1996; Warren et al. 2001)**.** A good example of race formation among insects is the apple race of *Rhagoletis pomonella*, which arose from the ancestral hawtorn-infesting race (Feder 1998). Examples of other known host race forming species are Goldenron Gall Fly *Eurosta solidaginis* (Craig et al. 2001), Pea Aphid *Acyrthosiphon pisum* (Hawthorn & Via 2001), Larch bud moth *Zeiraphera diniana* (Emelianov *et al.* 2003) and Tumbling flower beetle *Mordellistena convicta* (Blair et al. 2005).

It is reasonable to assume that host plant preference and larval performance are genetically controlled traits in herbivorous insects since it has been shown to be the case in several species, for instance the butterfly *Colias eurytheme* (Tabashnik et al. 1981) and the leaf-mining moth *Acrocerus transecta.* The latter species is divided into two host races associated with the hosts; *Juglans ailanthifolia* and *Lyonia ovalifolia*. Transplantation of the larvae has demonstrated that one of these populations completely failed to survive on the hostplant of the other host race. Females preferred to oviposit on their natal host plant (Ohshima 2008). If these two factors; host plant preference and larval performance, are

genetically determined, it might be enough to make sympatric groups diverge into different races and followed by sympatric speciation.

1.2.1 Lobesia botrana

The Grapevine moth *Lobesia botrana* (Denis & Schiffermüller) is a well-known member of the Lepidoptera group. *L. botrana* is a species on which several studies have been made on the female response to different olfactory and contact cues. This is a major pest in vineyards, and has often been studied with the intention to develop new methods of pest control (Gabel & Roehrich 1995; Maher & Thiéry 2006, Maher *et al*. 2006b; Moreau *et al*. 2008; Tasin *et al*. 2010). It is a generalist that is able to complete its development on a wide range of plant families (Stoeva 1982) and its damage in vineyards started to get noticed in the beginning of the 20th century (Gabel & Roehrich 1995). Commercial grape plants are predictable, abundant and easily found; these are qualities that facilitate host-finding in specialized insect species. *L. botrana* could therefore represent a good candidate for specialization on this host plant (Thiéry et al. 2005) but we know that some of them still choose other host plants.

Like many other phytophagous insects, *L. botrana* has low mobility during their developmental stages and the fate of their offspring is greatly affected by the mother's oviposition preference (Ohshima 2008). It has been shown that the quality of larval food affects female reproductive life history traits and that alternative wild hosts provide a greater nutritional value than grapes for *L. botrana* (Moreau et al. 2006 a, b). This could be a reason why *L. botrana*, despite the abundance and predictability of vineyards, has not yet specialized on grapes.

Flax-leaved Daphne, *Daphne gnidium* is one among the popular wild hosts and has been suggested as the presumed native host plant for *L. botrana* (Thiéry et al. 2005). It is an evergreen shrub that grows in the Mediterranean area and has, because of its toxicity, been used by fishermen in Spain for fishing method (Alvarez Arias 2000). It has also been suggested for pest-control due to its pesticidal activities (Pascual-Villalobos et al. 1998).

1.3 My study

Previous studies show that *L. botrana* females often choose to oviposit on the host plant they grew up on. With this in mind, I wanted to investigate if there is a genetic divergence between *L. botrana* collected from grape and *Daphne*. To put this into perspective I also measured the genetic distance between *L. botrana* individuals collected from different provinces to compare which of the two factors (natural host plant and geographic locations) had the bigger influence on genetic divergence.

The aim of this study was to answer whether *L. botrana* form one big connected population or form different subpopulations according to geographic location or their host plant species, as such studies have not been carried out on this species using DNA markers.

By sheding more light on this matter, we will hopefully learn more about the migration of *L. botrana* and whether there are any reproductive barriers between them in the Mediterranean.

I have used DNA sampled from *L. botrana* located at six different locations and from two different host plants, to estimate allele frequencies at a series of loci. I used the estimated allele frequencies to determine the number of clusters, and to compute the likelihood that a given genotype originated from each cluster.

The total sample consisted of DNA from 176 individuals. The questions I tried to answer in this study were: *1)* Is there a significant genetic differentiation between *L. botrana* samples

from different geographical origin? *2)* Is there a significant genetic differentiation between *L. botrana* samples grouped according to their host plant origin?

My hypothesis, based on my previous study of host preferences in *L. botrana* (Döös 2010), was that there would be genetic difference related to geographical origin but not to host plant origin.

2 TECHNIQUES: theoretical background

There are many ways in which one can investigate population genetic structures. I will make a short general presentation of genetic markers, where I explain the differences between some of the methods used within this study.

2.1 Genetic markers

There are three types of genetic markers that have been used in genetic and genomic analysis, which measure variation in the genome at different levels: morphological markers, protein based markers and DNA based markers. To be a genetic marker, the locus has to show experimentally detectable variation among the individuals in the test population. The test population could be natural or the result of a controlled cross. In the latter case, the number and frequencies of alleles may be accurately determined and used to estimate linkage distances between markers, enabling the development of genetic maps. A genetic marker may be operationally defined as a heritable polymorphic marker with clear genetic interpretation and repeatability.

2.1.1 Morphological markers

They are based on discrete traits with simple Mendelian inheritance where traits are mainly controlled by a single dominant gene (see example in Fig. 1). In this case, the morphological characters can be used as reliable indicators for specific genes (shape, color, size) and are useful genetic markers. They often identify functional genes and are easy to observe, but to obtain a reasonable number of polymorphic morphological markers, many mapping populations are needed.

2.1.2 Protein markers

Different alleles of genes may result in proteins with different amino acid compositions, sizes or modifications. Differences in electrical charge or molecular weight/size can easily be detected using gel electrophoresis and be used as genetic markers. Proteins themselves are often not visible in a gel, but they can be visualized by staining, and many enzymes can be visualized by using their activities to create bands where they are present. Isozymes are different forms of enzymes that differ in electrophoretic mobility and are commonly used as protein markers. A drawback with isozymes is that they are limited in number and tissue, and are developmental-stage dependent. And even though many of the detectable protein variations identify allelic sequence variations, some protein variations are due to post-translational modifications.

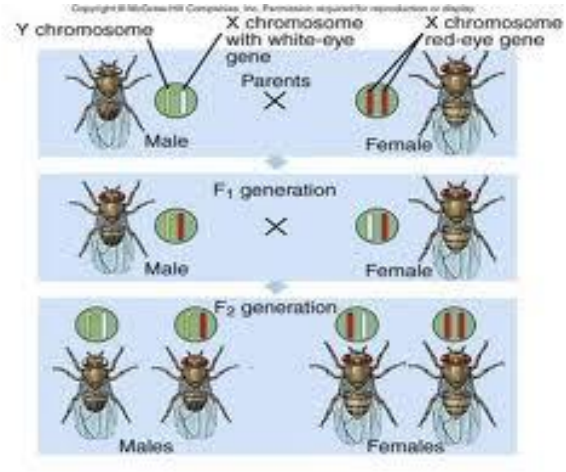


Fig 1 Model showing Mendelian inheritance in Drosophila melanogaster *over three generations(http://www.docstoc.com/docs/517986/Lecture-9-Variations-in-Mendelian-Inheritance)*

2.1.3 DNA markers

These are small regions of DNA that show sequence and size polymorphisms in different individuals within or among species. There are two basic approaches to detect the variations. The first approach is hybridization in which a previously known DNA fragment that shares considerable DNA sequence homology with the fragment of interest is used. This known fragment is labelled and used as a probe. It detects the fragment of interest by complementary base pairing. This is the foundation for Restriction Fragment Length Polymorphism (RFLP) marker technique (developed by Botstein *et al*. 1980). With the help of two-dimensional electrophoresis, the DNA polymorphism in the samples under study can be visualized. The second approach is amplification, in which the target sequence is amplified using Polymerase Chain Reaction (PCR) (Fig. 2). The PCR technique, developed by Mullis in (1986), is based on the knowledge of the segment sequence based on which two flanking primers are designed. Microsatellites and Single Nucleotide Polymorphisms (SNPs) are used as genetic markers based on sequence specific PCR. Short arbitrarily chosen primers have also been used to amplify random polymorphic DNA, these kind of markers include Random Amplified Polymorphic DNAs (RAPDs, a method developed by Williams et al. 1990), and Amplified Fragment Length Polymorphisms (AFLPs , a method developed by Vos *et al.* 1995).

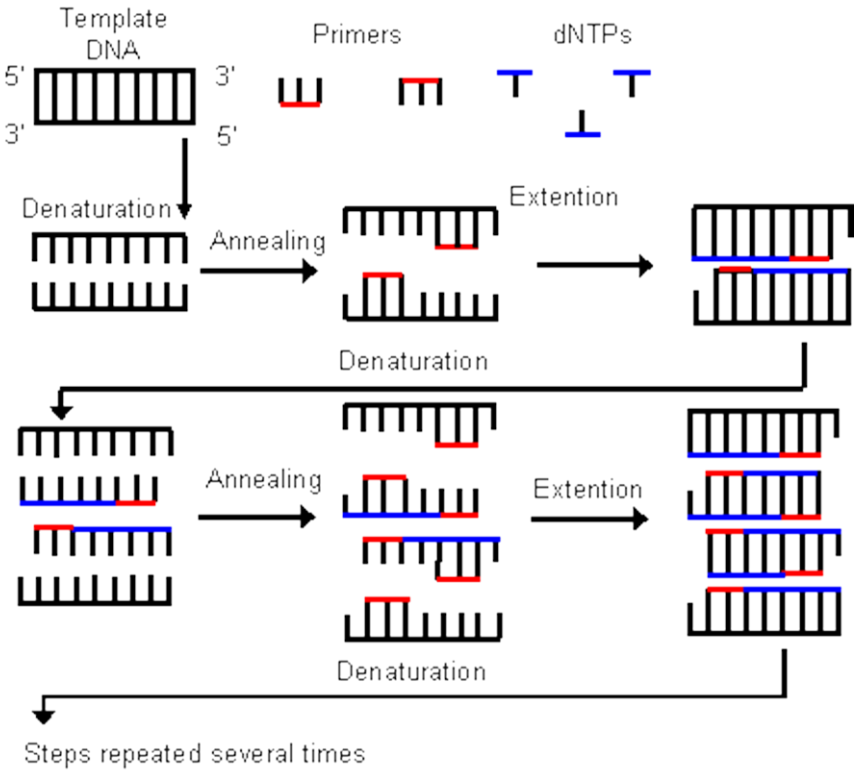


Fig 2 An example of the PCR process, showing two cycles where one fragment of double stranded DNA is turned into four (http://wiki.biomine.skelleftea.se/wiki/index.php/PCR)

2.1.3.1 Microsatellites

Microsatellites, also called Simple Sequence Repeats (SSRs), are tandem repeat sequences of DNA with a repeat motif ranging from 1-6 base pairs. The number of repeats is usually less than 100. Microsatellite variation can be detected by PCR using primers designed on the highly conserved sequences flanking the repeats. Polymorphism in SSRs is due to variation in number of repeats. The tandem arrays are amplified, visualized on an high resolution polyacrylamide gel or run on a capillary electrophoresis system. Microsatellites have high levels of variation in many plant and animal species. The most common forms of repeats are di-nucleotide repeats. Tri- and tetra-nucleotide repeats are also common, but their frequencies are lower than that of di-nucleotide repeats (Hearne et al. 1992). To identify microsatellite loci that are suitable for the development of genetic markers, it mostly takes a large effort using hybridization and sequencing. For a few number of species though, a large amount of DNA sequence data has been accumulated. In these cases microsatellites can be identified by searching through the DNA sequence databases (or even genome databases), and primers designed directly from the sequence data. Microsatellites are widely used in genetic studies because of their co-dominant inheritance and high polymorphism (Jarne & Lagoda 1996; Goldstein and Schlötterer 1999). There are some drawbacks with using microsatellites though. For example, developing microsatellites is both time consuming and expensive.

Fig 3 shows an example of the SSR alleles and what it look like when the locus is homozygous and heterozygous. Fig 4 shows an example of what it look like when the SSR was non-informative.

Fig 3 An example of what it looks like when the SSR is informative. At this single-locus SSR Lobot 6, only one allele-size is present in the first individual (homozygote) and two different sizes in the second (heterozygote).

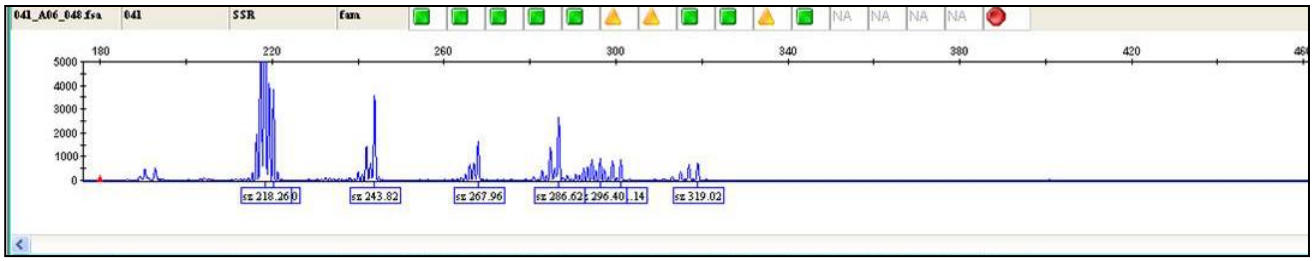


Fig 4 An example of what it look like when an SSR profile is not informative. The fragment analysis generated such peak pattern for the locus Lobot 10 in this study, and hence this locus was not included in the data analysis.

2.1.3.2 AFLPs

Amplified fragment length polymorphism (AFLPs) is another molecular marker technique that is often used in population genetic studies (Vos *et al*. 1995). One difference between AFLPs and microsatellites is that AFLPs are dominant markers. This means that the technique only show if a DNA band is present or absent. In other words, it is not possible to know if an AFLP locus is homozygous or heterozygous. Another difference is that AFLP generate lots of polymorphic bands per primer combination and are relatively easy to develop (Vos *et al*. 1995). However, AFLP is a technique that requires a relatively large amount of high quality DNA (Haig *et al.* 2004).

AFLPs result from restriction enzyme digestion, followed by adapter ligation and two (pre-selective and selective) amplifications. The approach uses PCR to amplify the DNA fragments generated by two restriction enzymes; a rare cutter and a frequent cutter, such as *Eco*RI and *Mse*I.

Adapters are short segments of double stranded DNA with sticky ends complementary to that of the restriction site. Every restriction enzyme needs separate adapters. The DNA fragment created then serves as a template for PCR reactions. Different primers are used to amplify different sequences, and the primers are specific to the combination of adapter sequence, restriction site and extensions. One primer is labeled and only the amplified fragments with this primer will be visualized. By changing the selective extension, different subsets of the restriction fragments will be amplified. The process of AFLP is shown in Fig. 5. Fig 6 shows examples of AFLP profiles among five different individuals.

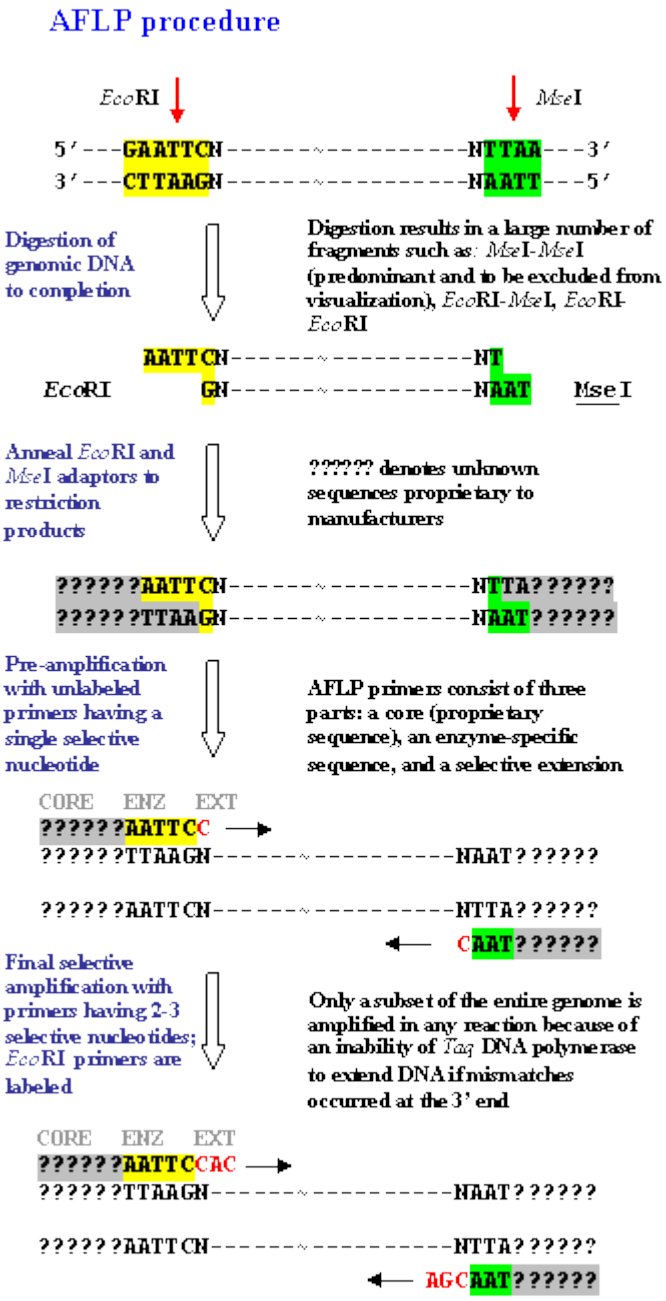


Fig 5 AFLP procedure (http://www.ncbi.nlm.nih.gov/projects/genome/probe/doc/TechAFLP.shtml)

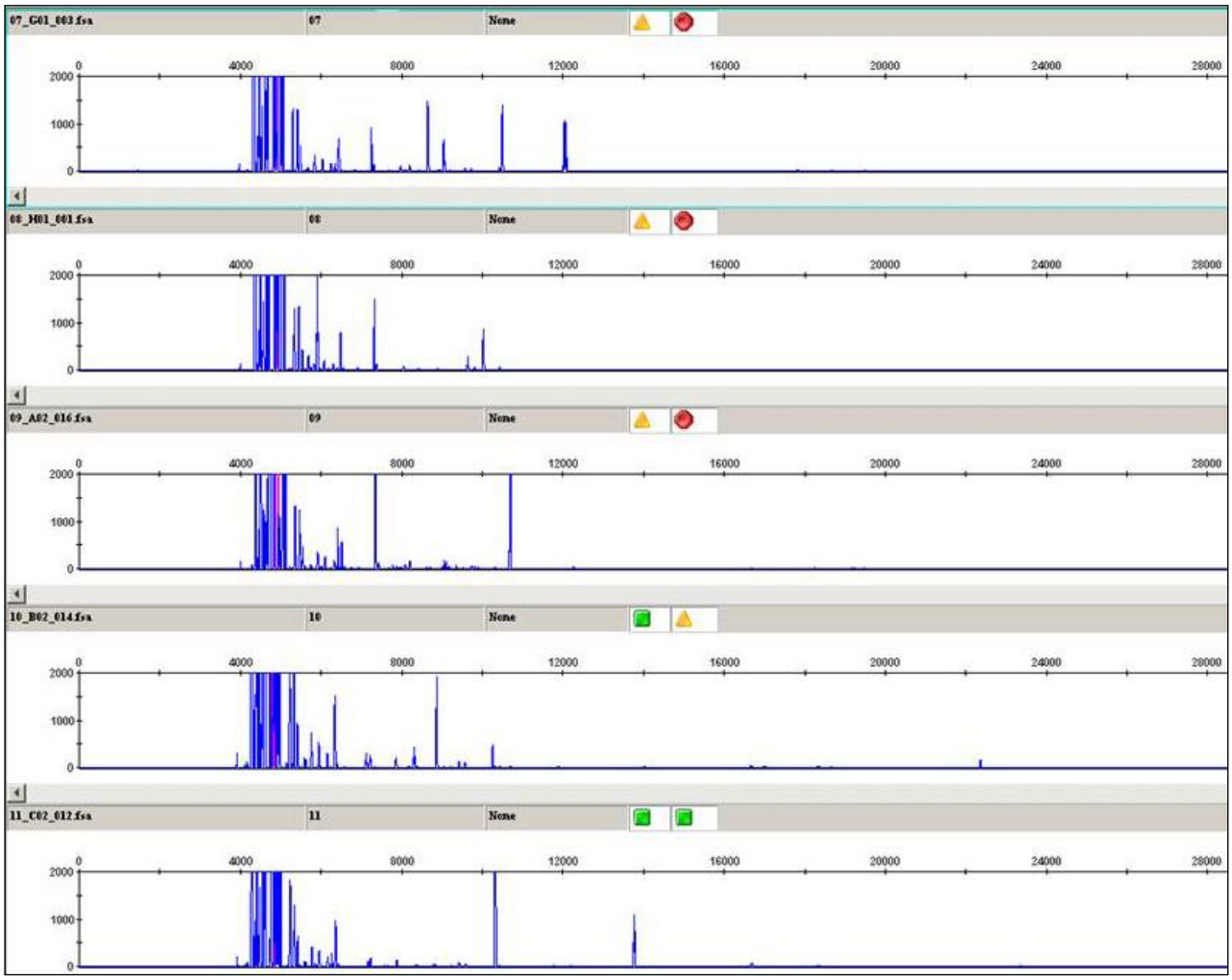


Fig 6 An AFLP profile for five individuals. A peak indicates the presence of a band, where the position on the x-axis is related to the fragment size (the fragment size increases from left to right along the x-axis. The height of the peak has some correlation to the number of copies of amplified product at that locus, but not with enough certainty to assess homo- or heterozygosity. Thus, AFLP is dealing with only the presence or absence of a fragment.

2.1.4 Methods for determining population structure

2.1.4.1 Distance-based methods

Wright's *F* statistics is a widely used method to measure population structure (Wright 1931). By using this method one define individuals into groups and then use their genotypes to compute variance based on allele frequencies. Clusters identified by distance-based methods heavily depend on distance measure and the graphical representation chosen, which makes it difficult to assess how confident one can be that the results obtained by these methods are meaningful. It is also difficult to incorporate additional information in these methods (Pritchard 2000). Even with these shortcomings, distance based methods have continued to be popular among many population genetic studies, and was thus calculated in this study.

2.1.4.2 Model-based approach

Since sometimes it can be hard to predefine populations, a method that does not demanding predefined structure may be needed (Evanno *et al.* 2005). Pritchard *et al*. (2000) has developed a model-based method which is implemented in the software STRUCTURE. This method is the most widely used method among Bayesian clustering

methods. STRUCTURE performs well with both dominant and co-dominant markers (Evanno *et al.* 2005).
By using a Bayesian approach, this new method detects clusters among individuals by comparing their genotypes at multiple loci, and quantifies the likelihood of each individual belonging to each group. With this information it becomes possible to assign the individuals to these clusters to a various degree. The model assumes Hardy–Weinberg equilibrium within clusters and attempts to find population groupings that are not in disequilibrium (Pritchard *et al*. 2000).
An estimation of the most likely number of clusters (K) is among the first steps of analyses using STRUCTURE. The highest of L(K) values have often been used as the true number of clusters (Evanno *et al*. 2005). In 2005, Evanno *et al* made investigations with simulations, to see weather STRUCTURE was able to detect the true number of K. Their results showed that *L*(*K*) actually did not show a clear mode for the true *K*. Instead, the likelihood kept on increasing even after the true K was reached. However, they found that the true value of *K* was correctly identified by an ad hoc quantity, based on the second order rate of change of the likelihood function with respect to *K*(Δ*K*). This is best shown graphically (Evanno *et al.* 2005) as in the materials and methods. Model-based approaches can be applied on various markers, such as microsatellites, AFLPs, RFLPs and SNPs.

In this study, I used both distance and model based approaches

3. MATERIALS AND METHODS

3.1 Insect material and DNA extraction

DNA was isolated from pupae or mature adults of 176 individuals of *L. botrana*. DNA extraction was performed according to the protocol described in Reineke (1998). The samples formed nine groups according to the location and host plant they were sampled from. These groups are presented in Table 1. Fig 7 shows a map with the sampling locations, and Fig 9 shows nests from *L. botrana* on the two host plants.

Table 1 Sample population (N), host plant origin and sample size for microsatellites and AFLP analyses.

N	Location	Host plant/Diet	Samples
1a	Trento	Rearing	8
1b	Trento	Vitis	40
2	Pescara	Vitis	24
3	Verona	Vitis	9
4a	Pisa	Vitis	29
4b	Pisa	Daphne	22
5a	Cagliari	Vitis	12
5b	Cagliari	Daphne	20
6	Israel	Vitis	12

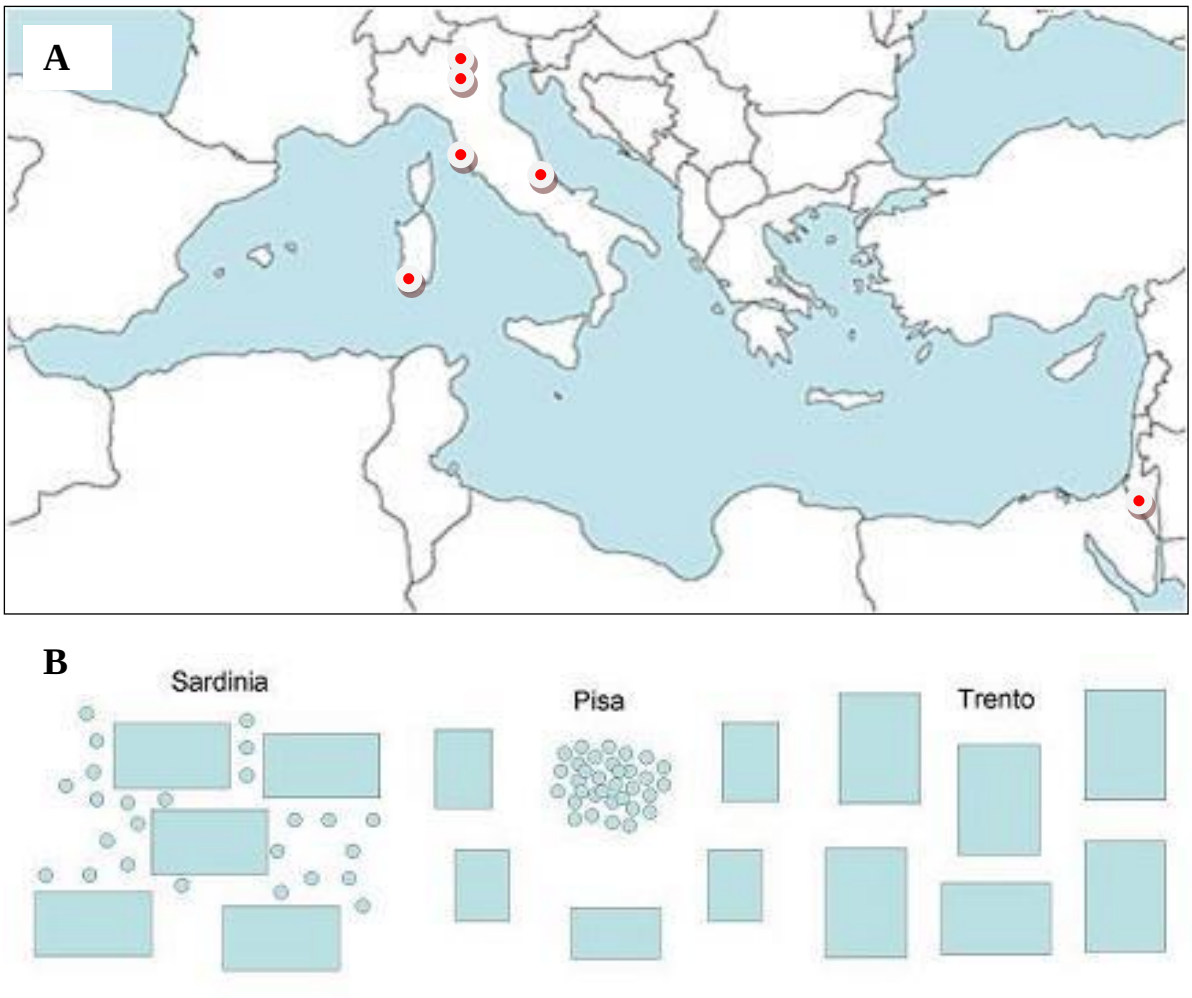


Fig 7 (A) the six locations from which L. botrana samples were collected around the Mediterranean sea (N1-N6);(B) distribution of two host plant species at three sampling locations. Circles represent Daphne plants and squares represent grape fields. In Cagliari-Sardinia, Daphne plants are completely interspersed among grape fields, whereas in Pisa they are mostly found as a separate population inside the national park Parco San Rossore, with a distance of a few kms to the nearest grape fields. In Trento and the remaining sample sites, L. botrana was only sampled from grape fields.

Fig 8 The two host plants (Left: Vitis; Right: Daphne) with nests from L. botrana.

Table 2. Repeat motifs of the SSRs, GenBank accession number of the source sequences and melting and annealing temperatures of the SSR primers for the seven loci studied.

Locus name	Motif	GenBank accession No.	Primer information	Melting T (°C)	Annealing T* (°C)
LOBOT 3	$(CA)^{21}$	AY150994	Amsellem et al. 2003	59.02 59.02	55
LOBOT 6	$(CA)^{17}$	AY150995	Amsellem et al. 2003	58.60 58.65	55
LOBOT 9	$(A)^{13}(CA)^{7}…(CA)^{6}$	AY150996	Amsellem et al. 2003	59.70 59.01	52
LOBOT 10	$(CA)^{42}$	AY150997	Amsellem et al. 2003	59.63 59.07	55
LOBOT 11	$(GT)^{14}(T)^{4}$	AY150998	Amsellem et al. 2003	60.38 60.38	55
LOBOT 14	$(CA)^{4}…(CA)^{9}$	AY150999	Amsellem et al. 2003	59.70 59.15	55
LOBOT 15	$(CA)^{21}$	AY151000	Amsellem et al. 2003	60.24 60.10	55

**Ta: differs between 52-56°C according to the primers, 4-7°C lower than the melting point*

3.2 Microsatellite analysis

All 176 samples were used for SSR analysis. Seven microsatellite loci were tested with the primers developed by Amsellem *et al.* (2003). Repeat motifs, GenBank accession number, melting and annealing temperatures for each of these seven loci are showen in Table 2.

The PCR reaction volumes was 12.5 µl and contains: 10 ng of DNA, 1x PCR reaction buffer (Qiagen) containing $MgCl_2$, 0.2 mM dNTPs, 0.4 µM forward primer, 0.4 µM reverse primer, 0.5 Unit of Taq polymerase and MilliQ water.
PCR were performed according to following conditions:

1) 15 minutes of initial denaturation/activation: this opens the double-stranded DNA and activates the Taq polymerase.

2) 45 seconds of denaturation at 94°C: this starts the cycle and opens up the DNA again.

3) 45 seconds of annealing at Ta: bind primers to the bases of DNA, a higher annealing T gives higher specificity.

4) 1 minute 30 seconds of elongation at 72°C: bases are added to the string.
Step 2-4 were cycled 34 times.

5) 7 min of final elongation at 72°C: the amplicon length is completed.

6) Electrophoresis was performed on a 1.5% agarose gel and a ladder mix (Strategene) was used to determine the fragment sizes of the PCR products. The amplicons were then diluted (1:10 or 1:20) according to how strong the bands (fragments) were on the agarose gel, and mixed with formamide and ROX size standard (Applied Biosystems) at the concentration of 5% diluted PCR product. The samples were loaded on a genetic analyzer based on capillary electrophoresis (3730xl, Applied Biosystems) following the manufacturers´ instructions. Allele size analysis was performed by the GeneMapper v.4.0 software (Applied Biosystems).

3.3 AFLP analysis

All samples were genotyped with AFLP, and was performed based on two steps:
a) Finding the most informative primer combinations. This was done at SLU, Alnarp, Sweden where preliminary experiments were made on a subset of individuals and the primer combinations were graded according to their informativeness.
b) Extending the most informative primer combinations to the entire sample set (176 individuals).

The AFLP procedure was carried out according to the following optimized protocol:

1) 500 ng DNA extract was added to a mixture of 5U hexa-cutter enzyme (*Eco*RI), 5U tetra-cutter enzyme (*Mse*I), 2X restriction enzyme buffer and sterile water up to a final volume of 50 µl which resulted in a restriction mix that was incubated at 37°C for 1 h and then stored at -20°C. Electrophoresis was performed on a 0.8% agarose gel and a ladder mix (Stratagene) was used to assess the quantity of the product.

2) Preparation of two adaptor mixtures each with a final volume of 50 µl; one containing 5µM *Eco*RI adapter and the other containing 5 µM *Mse*I adapter (Table 3). The mixtures were brought to 95°C for 3 minutes and thereafter the temperature was decreased by 1°C/minute until it reached 25°C. The mixtures were then stored at -20°C until used.

3) Preparation of the ligation product. 20µl of the digested DNA, 1µM *Eco*RI adaptor mixture and 10µM *Mse*I adaptor mixture were added to 1X ligation buffer, 1U T4 DNA Ligase and sterile water up to a final volume of 30 µl. This ligation product was diluted 10X and incubated at 16°C for 16 h and then stored at -20°C.

4) The pre-selective PCR product was prepared with AFLP amplification core mix (Applied Biosystems), 0.4 µM pre-selective forward- and reverse primers (Table 3, Fig 6) and 1X ligation product to the final volume of 10 µl. The following PCR program was used: 95°C for 7 min, 95°C for 30 sec, 56°C for 1 min, 72°C for 1 min, step 2, 3 and 4 were repeated 24 times, 72°C for 4 min and at last it was stored at 4°C. Electrophoresis was performed on 0.8% agarose gel and a ladder mix (Stratagene) was used to assess amplification. Then, the pre-selective products of successful amplifications were diluted 10X.

5) For the selective PCR, the product was prepared with 7 µl AFLP amplification core mix, 0.625 µM selective forward primer (labeled), 1.875 µM reverse primers (Table 3) and a 1X diluted pre-selective PCR product to the final volume of 10 µl. The following PCR program was used: 94°C for 7 min, 94°C for 30 sec, 65°C for 1 min, 72°C for 1 min, 94°C for 30 sec, 65°C for 30 sec, 72°C for 1 min, step 5, 6 and 7 were repeated 9 times with the annealing temperature decreasing 0.7°C/cycle, 94°C for 30 sec, 56°C for 30 sec, 72°C for 1 min, step 8, 9 and 10 were repeated 23 times, 72°C for 4 min and at last it was stored at 4°C. Electrophoresis was performed on a 0.8% agarose gel and a ladder mix was used to assess the quantity of the product.

6) For the analysis, a final mixture was prepared where 9.35 µl Formamide (ultra-pure, Applied Biosystems), 0.15 µl size standard (1200 LIZ, ABI) was run on vortex, spun, and added to 0.5 µl of the selective PCR product. The mixture was incubated at 95°C for 2 min, and then put immediately on ice. This final mixture was run in the genetic analyser (3730xl Applied Biosystems) following the manufacturers' instructions. The AFLP fragment analysis was performed using the GeneMapper v.4.0 software (Applied Biosystems).

Table 3 Restriction site, Adapter sequence, pre-selective and selective primer sequences. "_" implicates the restriction cut site or the overhang.

	EcoRI	MseI
Restriction	G_AATTC CTTAA_G	T_TAA AAT_T
Adaptor	F: 5'CTC_GTAGACTGCGTACC R: 5´AATT_GGTACGCAGTCTAC	F: 5'-GACG_ATGAGTCCTGAG R: 5´-TA_CTCAGGACTCAT
Pre-selective primer	F: 5'-GACTGCGTACCAATTC**a**	R: 5'-GATGAGTCCTGAGTAA**c**
Selective primer (Primer+extension)	5´-GACTGCGTACCAATTC**ACG**	5'-GATGAGTCCTGAGTAA **C**AG 5'-GATGAGTCCTGAGTAA **C**CA

3.4 General statistics

Analysis of molecular variance (AMOVA) and pairwise population differentiation estimates (F_{ST}) between groups was calculated for both SSR and AFLP data using ARLEQUIN software V3.0 (Excoffier *et al.* 2005)

3.5 Structure analysis

STRUCTURE v. 2.3 software (http://pritch.bsd.uchicago.edu) was used for Bayesian model-based clustering method of Pritchard et al. (2000). The model with admixture was used with uncorrelated allele frequencies, with the assumed number of populations (K) varying from 1 to 10, 5 replicate runs per K value, a burning period length of 10^6, and a post-burning simulation length of 1.5×10^6. This model assumes that each genotype in the sample comes from more than one differentiated ancestral population (K).

3.5.1 Statistics used to select K

To determine the true number of clusters (K) of the studied sample, I used the STRUCTURE software. This software makes an estimation of the probability that the data belonged to different K´s. This value, called Ln P(D) in STRUCTURE output, is obtained by computing the log likelihood of the data at every step. The averages of these values were then computed and half their variance subtracted from the mean, giving us L(K). To get the true number of K, I calculated an ad hoc quantity based on the second order rate of change of likelihood function with respect to K (ΔK) which, according to Evanno *et al.* 2005, shows a clear peak at the true K value.
This can be carried out in four steps and displayed graphically as in Fig10-13.

1. First I plotted the mean likelihood, L(K), over the five runs for each K (Appendix 3 and 7).
2. Second I plotted L`(K); the mean difference between successive likelihood values of each K, except for K1 since the model assumes more than one K (Appendix 4 and 8).
3. Third I plotted L``(K); the absolute value of the difference between successive L`(K) values (Appendix 5 and 9).
4. Finally I plotted the ΔK; the L´´(K) divided by Standard Deviation of L(K) (Appendix 6 and 10).

4. RESULTS

4.1 Microsatellite analysis

4.1.1 Microsatellite markers

The seven microsatellites loci reported by Amsellem et al (2003) were used in this study. Locus Lobot-9 was not amplified after three trials. Six loci amplified successfully and four, Lobot-6, Lobot-11, Lobot-14 and Lobot-15, were polymorphic. The remaining two, Lobot-3 and Lobot-10, were either non-specific or showed a multi-locus profile.

4.1.2 Microsatellite overall diversity

The observed number of alleles ranged from 6 at locus Lobot-11 to 34 at Lobot-6, mean=15.5 (s.d.=12.97). Overall, observed and expected heterozygosity, respectively, ranged from 0.090 and 0.347 at Lobot-14 to 0.590 and 0.932 at Lobot-6 (Table 4).

Table 4 Fragment size range and number of fragments/alleles yielded by the four polymorphic SSR primers.

SSR loci	Fragment size range (bp)	Number of alleles	Observed heterozygosity	Expected heterozygosity
Lobot-6	199-460	34	0.58960	0.93203
Lobot-11	295-315	6	0.18750	0.60515
Lobot-14	265-283	7	0.09036	0.34729
Lobot-15	180-228	15	0.58235	0.76960

4.1.3 Population differentiation

Out of the total 62 alleles, 14 were population specific alleles (Supplementary appendix 1). There were six such alleles at locus Lobot-6 distributed among four populations and two at locus Lobot-11, one in Cagliari-Daphne and one in Pescara-vitis. Two private alleles at Locus Lobot-14 were found in populations Pescara-vitis and Pisa-vitis. Four population specific alleles at Lobot-15 were obtained in four different populations: Cagliari-vitis, Pescara-vitis, Pisa-Daphne and Trento-vitis (see supplementary Appendix 2 for allele frequencies of each population at each locus).

4.1.3.1 Analysis of molecular variance (AMOVA)

AMOVA was calculated with samples grouped as seven regions (Geographic locality) and two host plants, in total ‘nine populations’. The AMOVA in ARLEQUIN with 10100 permutations showed that among region variance was 5.61% and was significant (p=0.01574, Table 5). The variance among host plants within regions accounted for -

0.49% and was not statistically significant (p=0.6769, Table 5). Pairwise population differentiation (F_{ST}) results (Table 6) also showed non-significant p-values between Cagliari-Daphne and Cagliari- Vitis, and between Pisa-Daphne and Pisa-Vitis. The negative Amova result together with non-significant pairwise differentiation values indicated that the SSR markers could not detect genetic differentiation based on host plants.

Table 5 AMOVA based on the two factors; Host plant and Geographic locality (total= nine populations). The variance component among host plants within regions is not significant.

Source of variation	degrees of Freedom	Sum of squares	Variance components	Percentage of variation	P-Value
Among regions	6	17.977	0.04591 Va	5.61	0.01574
Among host plants within regions	2	1.641	-0.00402 Vb	-0.49	0.67693
Among individuals within populations	167	163.916	0.20526 Vc	25.09	0.0000
among all individuals	176	100.500	0.57102 Vd	69.79	0.0000
Total	351	284.034	0.81817		

Table 6 Pairwise Fst (Lower diagonal) and P-values (Upper diagonal) with nine populations defined by geographic origin and host plant. Genetic distance between Cagliari-Daphne and Cagliari-Vitis, and between Pisa-Daphne and Pisa-Vitis were not significant.

	Cagliari-Daphne	Cagliari-Vitis	Israel-Vitis	Pescara-Vitis	Lab-Rearing	Pisa-Daphne	Pisa-Vitis	Trento-Vitis	Verona-Vitis
Cagliari-Daphne		0.210	0.000	0.185	0.014	0.478	0.343	0.029	0.365
Cagliari-Vitis	0.020		0.000	0.691	0.003	0.748	0.673	0.283	0.082
Israel-Vitis	0.196	0.270		0.000	0.000	0.000	0.000	0.000	0.000
Pescara-Vitis	0.015	0.001	0.225		0.015	0.456	0.278	0.136	0.088
Lab-Rearing	0.075	0.087	0.382	0.060		0.000	0.004	0.019	0.008
Pisa-Daphne	0.003	-0.007	0.235	0.004	0.102		0.961	0.200	0.059
Pisa-Vitis	0.008	-0.001	0.211	0.009	0.093	-0.012		0.022	0.222
Trento-Vitis	0.030	0.017	0.184	0.015	0.069	0.012	0.025		0.037
Verona-Vitis	0.017	0.048	0.202	0.035	0.116	0.034	0.021	0.052	

Since there is no genetic differentiation between Cagliari-Daphne and Cagliari-Vitis, and between Pisa-Daphne and Pisa-Vitis, the pairs were merged together according to their site of collection and made two populations, and hence I proceeded to analyze the

samples as 'seven populations' according to regions. AMOVA as seven populations showed significant differentiations between populations formed based on their region of collection (Variance =5.23%, p<0.001) even though most of the variance was explained by individuals within populations (Table 7). Pairwise F_{ST} values ranged from 0% (between Cagliari and Pisa) to 38.2% (between Lab reared samples and those from Israel, Table 8). Overall F_{ST} was 0.052 (p<0.001).

Table 7 AMOVA based on the two factors; Individual and Geographic Locality (seven populations).

Source of variation	degrees of Freedom	Sum of squares	Variance components	Percentage of variation	P-Value
Among populations	6	17.977	0.04278 Va	5.23	0.00000
Among individuals within populations	169	165.557	0.20430 Vb	24.97	0.0000
within individuals	176	100.500	0.57102 Vc	69.80	0.0000
Total	351	284.034	0.81811		

Table 8 Pairwise F_{ST} (Lower diagonal) and P-values (Upper diagonal) with seven populations defined by geographical location.

	Cagliari	Israel	Lab rearing	Pescara	Pisa	Trento	Verona
Cagliari		0.000	0.006	0.435	0.617	0.031	0.195
Israel	0.204		0.000	0.000	0.000	0.000	0.000
Lab rearing	0.071	0.382		0.014	0.001	0.017	0.009
Peascara	0.005	0.225	0.060		0.154	0.139	0.093
Pisa	0.000	0.215	0.098	0.010		0.006	0.079
Trento	0.021	0.184	0.069	0.015	0.023		0.039
Verona	0.023	0.202	0.116	0.036	0.030	0.052	

4.1.4 Within population genetic diversity (7 populations)

The mean number of alleles per locus ranged from 3.00 (s.d.=1.41, N=8 in the Lab reared population) to 10.5 (s.d.=8.27, N=50 in the Pisa population, Table 9). The correlation between number of samples analyzed and total number of alleles was positive and significant (R=0.876, $F_{1,5}$=16.441, p=0.01 Fig 9). Because of the large variations in number of samples analyzed, allelic richness was calculated. Allelic richness is a relative measure that calculates the alleles using the smallest sample size. In our case, N=8, the Lab reared population. Total allelic richness ranged from 11.73 in lab reared population to 21.45 in

Cagliari (Table 10). The correlation between number of samples and total allelic richness was also positive and marginally significant (R=0.744, $F_{1,5}$=6.217, p=0.0549 Fig 10).
Of the 28 Fisher's exact tests performed (7 populations at 4 loci) to measure deviations from Hardy-Weinberg equilibrium, 18 (64.3%) were significant (p<0.05) (table 9). All the deviations were due to heterozygote deficits. Observed heterozygosity was least in Trento (mean=0.30±0.26 s.d.) and highest in Pisa (mean=0.41±0.32 s.d.). Expected heterozygosity ranged from 0.40±0.16 (s.d.) in Israel to 0.73±0.16 (s.d.) in Pescara. Inbreeding coefficient (F_{IS}) values were all positive except in lab reared population (-0.013, p=0.633), and five of the remaining seven populations had significant P-values (p<0.005). Observed Heterozygosity (HO), Expected heterozygosity (HE) and Inbreeding coefficient values are presented in Table 10.

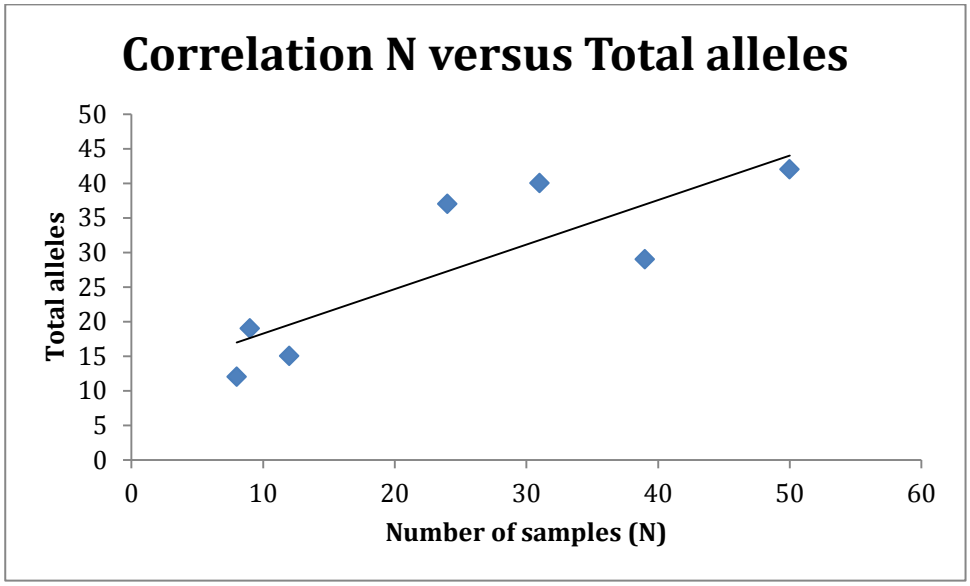


Fig 9 Positive correlation between number of samples analyzed and total number of alleles

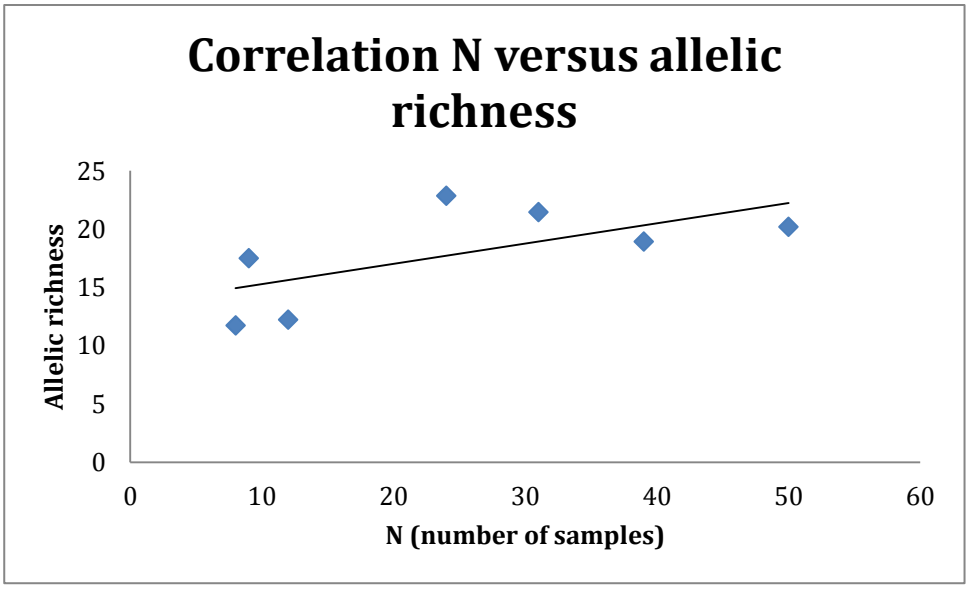


Fig 10 Positive correlation between number of samples analyzed and allelic richness

Table 9 Number of alleles and allelic richness per locus for all 7 populations

		Locus Lobot-6		Locus Lobot-11		Locus Lobot-14		Locus Lobot-15				
Population	N	Number of Alleles	Allelic Richness	Number of Alleles	Allelic Richness	Number of Alleles	Allelic Richness	Number of Alleles	Allelic Richness	Total number of alleles	Mean number of alleles / locus (s.d)	Total allelic richness
Cagliari	31	22	10.33	5	3.26	4	2.45	9	5.44	40	10.00 (8.29)	21.45
Israel	12	7	4.86	3	2.92	2	1.88	3	2.58	15	3.75 (2.22)	12.23
Lab rearing	8	5	4.75	2	1.99	2	2.00	3	2.99	12	3.00 (1.41)	11.73
Pescara	24	18	9.75	4	3.54	5	3.77	10	5.79	37	4.25 (6.40)	22.85
Pisa	50	22	8.91	4	3.03	5	3.10	11	5.16	42	10.5 (8.27)	20.20
Trento	39	12	7.40	4	2.82	3	2.49	10	6.23	29	7.25 (4.43)	18.94
Verona	9	8	7.22	3	3.00	2	1.99	6	5.29	19	4.75 (2.75)	17.50

Table 10 Observed and Expected heterozygosity per locus, mean Observed and Expected heterozygosity per population and Inbreeding coefficient (FIS) for the seven populations. HWE is the number of loci with significant deviations from Hardy-Weinberg equilibrium (Fishers's exact test).

	Locus Lobot-6		Locus Lobot-11		Locus Lobot-14		Locus Lobot-15					
Population	HO	HE	HO	HE	HO	HE	HO	HE	Mean HO	Mean HE	FIS (P-value)	HWE
Cagliari	0.581	0.950	0.219	0.402	0.094	0.279	0.625	0.779	0.38+-0.26	0.60+-0.31	0.292 (<0.001)	3
Israel	0.333	0.554	0.417	0.475	0.000	0.173	0.500	0.409	0.31+-0.22	0.40+-0.16	0.141 (0.219)	1
Lab rearing	0.750	0.750	0.000	0.233	0.000	0.264	0.625	0.608	0.34+-0.40	0.46+-0.26	-0.013 (0.633)	1
Peascara	0.583	0.939	0.190	0.661	0.167	0.552	0.609	0.749	0.38+-0.24	0.73+-0.16	0.287 (<0.001)	3
Pisa	0.760	0.915	0.159	0.403	0.109	0.376	0.609	0.744	0.41+-0.32	0.61+-0.26	0.129 (0.004)	4
Trento	0.462	0.882	0.147	0.575	0.079	0.302	0.500	0.813	0.30+-0.21	0.64+-0.26	0.428 (<0.001)	4
Verona	0.444	0.876	0.222	0.699	0.000	0.233	0.667	0.745	0.33+-0.29	0.64+-0.28	0.328 (0.003)	2

4.2 AFLP analysis

4.2.1 AFLP Markers

The informativeness of the 11 primer combinations was assessed at SLU, Alnarp, Sweden prior to the main study described in this thesis. Out of the 11 selective primer combinations with varying degree of informativeness, the 2 most potentially informative combinations, yielding the highest number of peaks, were extended to the entire sample set of 176 individuals.

4.2.2 Marker Diversity

Out of the starting sample size of 176 individuals, the phenotypes of 158 (89.8%) individuals were successfully scored. With the two chosen primer combinations, 312 fragments were scored, of which primer combination A (E-ACG and M-CAG) produced 123 fragments in the range 50-786bp, and Primer combination B (E-ACC and M-CCA) produced a total 189 fragments in the range 50-940bp. The total number of informative fragments (displaying differences between individuals) was 80 and 152 for primer combinations A and B, respectively. This gave a total of 232 informative AFLP bands among the 158 individuals.

4.2.3 Population structure/ differentiation

Analysis of Molecular Variance in GENEALEX calculated as nine populations showed that among population variance accounted for 11% (PHiPT= 0.114, p=0.002, Table 11), and individuals within populations accounting for 89% (Table 11). I sought to establish whether samples from Cagliari-Daphne and Cagliari_Vitis, and from Pisa-Daphne and Pisa-Vitis were different in terms of molecular variance. Cagliari-Daphne and Cagliari_Vitis showed 11% (PHiPT= 0.107, p=0.002, Table 12), and between Pisa-daphne and Pisa-Vitis 8% (PHiPT= 0.081, p=0.001, Table 13). Due to these result it was not legitimate to merge the samples as was done for microsatellite data. Therefore, results are presented as ‘nine populations’.

Pairwise population genetic differentiation (PHiPT, an analogue of F_{ST} calculated in GENEALEX, Table 14) did show that Cagliari-Daphne and Cagliari_Vitis, and Pisa-daphne and Pisa-Vitis, were indeed genetically differentiated ($P<0.05$). PHiPT values ranged from 0.8% between Trento and Pescara, to 47.3% between Verona and Israel. Cagliari-Vitis and Pisa-Vitis, Pescara and Lab-population, and Pescara and Trento were not genetically differentiated ($p>0.05$).

Table 11 Among and within population molecular variance for the 'nine populations'

Source of variation	Degree of freedom	Sum of squares	Variance components	Percentage of variation	PhiPT	P-value
Among Pops	8	683.416	85.427	11%	0.114	0.002
Within Pops	167	4161.993	24.922	89%		
Total	175	4845.409		100%		

Table 12 Among and within Cagliari-Daphne and Cagliari-Vitis molecular variance

Source of variation	Degree of freedom	Sum of squares	Variance components	Percentage of variation	PhiPT	P-value
Among Pops	1	23,137	23,137	11%	0,107	0,002
Within Pops	30	248,050	8,268	89%		
Total	31	271,187		100%		

Table 13 Among and within Pisa-Daphne and Pisa-Vitis molecular variance

Source of variation	Degree of freedom	Sum of squares	Variance components	Percentage of variation	PhiPT	P-value
Among Pops	1	25,421	25,421	8%	0,081	0,001
Within Pops	49	389,940	7,958	92%		
Total	50	415,361		100%		

4.2.4 Within population Genetic diversity

The number of present allele in each population varied from 22 in the lab reared population to 103 in Trento-vitis (Table 15). The number of polymorphic loci varied from 22 in lab reared population to 41 in Cagliari-Vitis. Expected heterozygosity was always low and ranged from 0.038±0.007 in lab reared population to 0.061±0.009 in Pisa-Daphne (Table 15). The number of final phenotypes was positively correlated to number of bands detected ($R=0.967$, $F_{1,7}=100.86$, $p<0.001$).

Table 14 Pairwise PHiPT values (Lower diagonal) and P-values (upper diagonal) as calculated in GENEALEX with 9999 permutations.

	Cagliari-Daphne	Cagliari-Vitis	Israel	Lab	Pescara	Pisa_Daphne	Pisa_Vitis	Trento	Verona
Cagliari-Daphne		0.002	0.000	0.001	0.000	0.121	0.000	0.000	0.008
Cagliari-Vitis	0.107		0.000	0.000	0.000	0.003	0.097	0.000	0.000
Israel	0.408	0.359		0.000	0.000	0.000	0.000	0.000	0.000
Lab	0.129	0.169	0.385		0.056	0.000	0.001	0.028	0.000
Pescara	0.137	0.133	0.301	0.036		0.000	0.000	0.167	0.000
Pisa_Daphne	0.021	0.089	0.414	0.185	0.169		0.000	0.000	0.039
Pisa_Vitis	0.118	0.026	0.325	0.121	0.096	0.081		0.000	0.004
Trento	0.171	0.168	0.282	0.049	0.008	0.214	0.133		0.000
Verona	0.098	0.183	0.473	0.146	0.149	0.057	0.089	0.189	

Table 15 Sample size, final phenotypes, number of bands, number of private bands, number of polymorphic loci, percentage of polymorphic loci, expected heterozygosity for the nine populations.

N	Sample size	Final phenotypes	No. Bands	No. Private Bands	Number of polymorphic loci	Percentage of polymorphic loci	Expected heterozygosity	S.E. (HE)
Cagliari_D	20	20	65	11	31	13.4	0.05650	0.00817
Cagliari_V	12	9	52	14	41	17.7	0.05776	0.00745
Israel	12	9	39	10	30	12.9	0.04421	0.00758
Lab	8	4	22	0	22	9.5	0.03818	0.00674
Pescara	24	21	77	23	27	11.6	0.04185	0.00617
Pisa_D	22	20	64	8	32	13.8	0.06103	0.00886
Pisa_V	29	26	74	28	28	12.1	0.04512	0.00641
Trento	40	37	103	51	27	11.6	0.04082	0.00581
Verona	9	7	31	2	31	13.4	0.04079	0.00719

4.3 AFLP versus Microsatellite data

A positive and significant correlation was obtained between pairwise genetic distance values (F_{ST} from microsatellite data) and PHiPT (analogue of F_{ST} from AFLP data) (R=0.835, $F_{1, 34}$= 78.51, p<0.001) (Figure 11). This may indicate the consistency of the two molecular markers. The results, however, are to be treated with caution because the number of bands (AFLP data, $F_{1,7}$= 89.033, p<0.001) and number of alleles (microsatellite data, see result above) might have influenced by sample sizes.

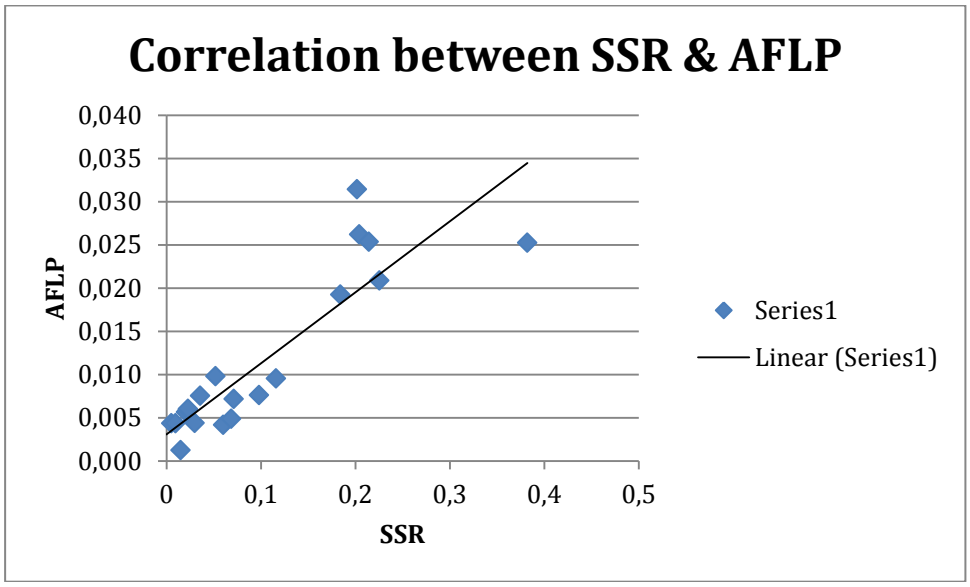


Fig 11 Correlation between pairwise genetic distances between populations, obtained by means of microsatellite data (Fst) and AFLP data (PHiPT)

4.4 Structure analysis

With the results from the AFLP analysis, I tested the probability of the sampled individuals belonging to a given cluster/subpopulation (K). The STRUCTURE analysis according to Evanno *et al.* 2005 suggested that the true K was 3 (Appendix 10). With this in mind, I proceeded with a model-based method developed by Pritchard *et al*. (2000) to assign my populations to the three clusters. Fig 12 shows this according to individual and sampled population, and Fig 13 according to average number of individual per sampled population. The clustering provided by the STRUCTURE analysis revealed no apparent differences in clustering patterns between host plants within the same geographic location. In contrast, the clustering patterns displayed apparent differences between some of the geographic locations. As the STRUCTURE analysis does not provide specific probability estimates, it was not always clear what could be considered a relevant difference between populations based on the observed clustering patterns.

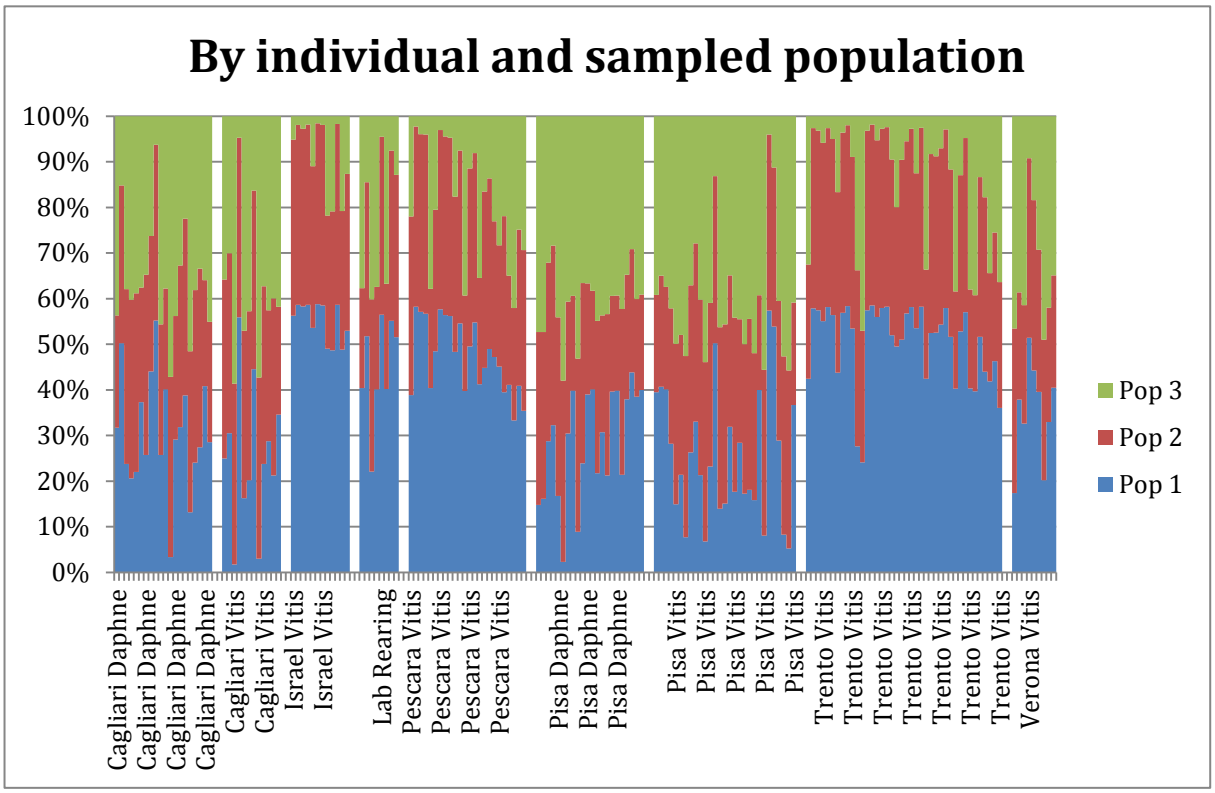


Fig 12 Analysis by Structure shows how Individuals were assigned into three clusters

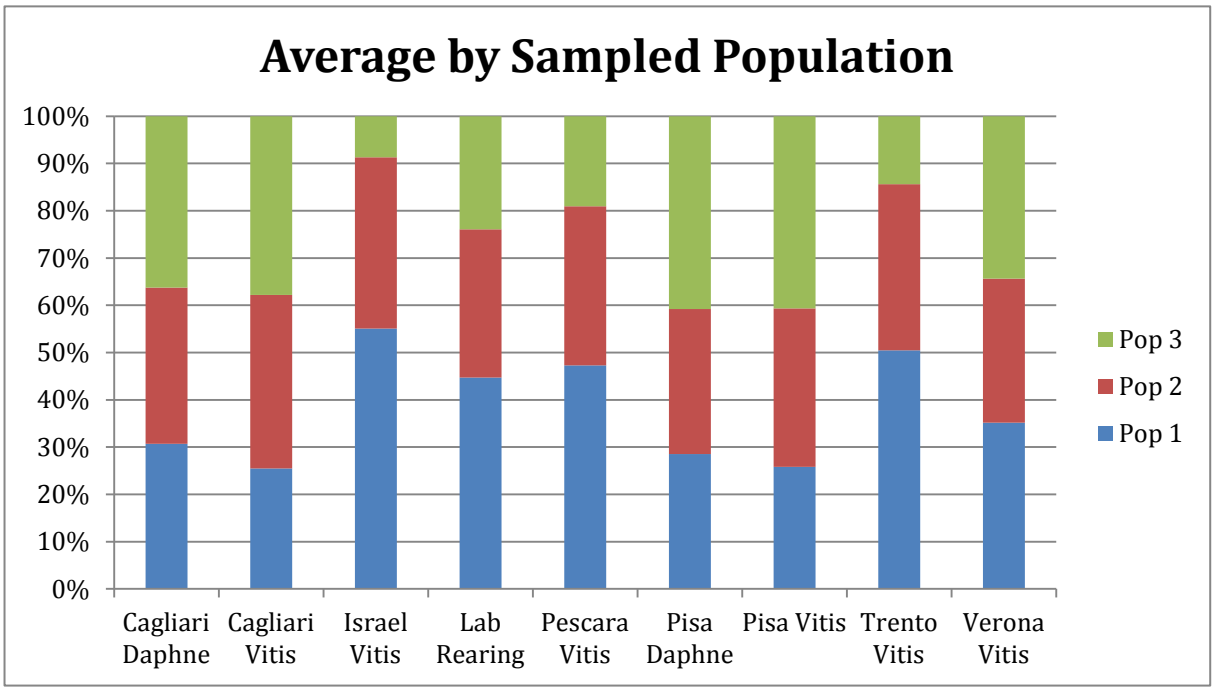


Fig 13 Populations assigned to the three clusters

5. DISCUSSION

In this study, I have characterized genetic differentiation of populations of a polyphagous moth, *Lobesia botrana*. The aim was to compare geographic separation and separation based on different host plants as potential causes for genetic population structure. I used several genetic markers as well as statistical methods in parallel, with the original aim to compare which of these analyses could be most helpful in providing useful information about genetic differentiation.

I have used two genetic markers: microsatellites and AFLPs. With regards to microsatellites, I was limited to the seven available microsatellite loci originally developed for *L. botrana* (Amsellem *et al*. 2003). Of these seven loci, one failed to amplify despite repeated attempts, and another two were either non-specific or showed a multi locus profile. This brought the total number of useful loci, providing information about molecular variation, down to only four. This low number of loci used could be expected to constitute a serious limitation on the power to detect genetic differentiation in the analyses based on microsatellite data. Nevertheless, the two different types of molecular markers (microsatellites and AFLP) generally gave consistent results, although the lower power offered by the microsatellite analysis failed to detect some instances of differentiation that were detected by the AFLP analysis. A linear regression between pair wise genetic distance values of SSR and AFLP indicated general consistency between the two markers (Figure 11).

Two kinds of analyses were made on Microsatellites; Analysis of molecular variance (AMOVA) and Pair wise population differentiation (F_{ST}). Three kinds of analyses were made on AFLP´s; Two analogues to AMOVA and F_{ST} and one Structural analysis performed in STRUCTURE. I detected genetic differentiation based on geographic differences, and also an indication towards differentiation based on host plant.

Variation due to Geographical location: Analysis of molecular variance revealed a significant population differentiation among *L. botrana* populations from different geographic origins (SSR data :5.61%, p=0.016, Table7. AFLP data:11%, p=0.002, Table 11). In pair wise comparisons (see Tables 8 and 14), both SSR and AFLP showed significant differences between Israel and all the other populations, with P values lower than 0.001. Both markers showed significant differences between Trento and all other populations except Pescara, where none of them detected a difference. In most cases of non-consistence between the two markers (like between Pescara-Pisa, Pescara-Verona, Pescara-Cagliari, Cagliari-Pisa, Cagliari-Verona and Verona-Pisa) the AFLPs generated slightly lower P-values than the SSRs, detecting significant differences that the SSRs did not succeed to detect. The only exception here is in the case between Pescara and Lab population, where SSRs detected a significant variance related to geographical location but AFLPs did not show this variance to be significant. I have no explanation to what the reason can be for this. Unlike the other statistical methods, the STRUCTURE analysis provided no specific cutoff limit for when individual populations could be considered statistically different. Based on the clustering patterns obtained from the STRUCTURE analysis, the overall degree of differentiation between populations generally did appear with the results obtained with the other statistical analyses, but no quantitative comparisons could be made with the other methods.

Variation based on host plant: Based on the microsatellite data, neither the AMOVA (0.49%, p=0.677, Table 7) nor the pairwise Fst (Table 8) showed any significant differentiation among host plants within regions. In contrast, both the AMOVA and PHiPT detected significant variation between insects collected from different host plants in Cagliari (PHiPT= 0.107, p=0.002, Table 12) as well as Pisa (PHiPT= 0.081, p=0.001, Table 13) in the case of AFLP markers. Also, in the pair wise comparisons, the AFLPs showed a significant differentiation among host plants within regions (Pisa_Vitis and Pisa_Daphne P=0,000; Cagliari_Daphne and Cagliari_Vitis P=0,002, Table 14) but no significant differentiation between regions within host plants (Pisa_Daphne and Cagliari_Daphne P=0,121; Pisa_Vitis and Cagliari_Vitis P=0,097, Table 14). This indicates a larger differentiation according to host plant that is greater than the differentiation according to geographic distance, which could suggest host plant adaption (Dickey and Medina 2012). The STRUCTURE analysis did not reveal any apparent differentiation in clustering patterns between host plants.

In summary, the microsatellites showed a significant differentiation between regions, but not among host plants within regions. AFLP on the other hand, showed significant variation between regions as well as between host plants within regions, when the data were analyzed with comparable statistical methods, thus succeeding to detect significant variation among groups where SSRs failed to detect. This could suggest that AFLP markers are more effective in detecting differentiation than SSR markers. However, but in my study, the more reasonable explanation to the difference in statistical power between markers is likely be the low number of microsatellite loci, which generated a low amount of data that might not be sufficient to show an existing level of genetic variation (Vos *et al*. 1995). On the other hand, the number of available loci may frequently be a limiting factor in microsatellite studies, unless great effort has been invested in developing microsatellites in advance; a problem that could be bypassed using AFLP. In contrast to the conventional statistical analysis, the AFLP- STRUCTURE analysis showed no clear differences among host plants. This unexpectedly weak discrimination ability of the STRUCTURE analysis is hardly due to lack of data, but might be due to the low number of simulations (5 runs) used in the analysis, resulting in a low resolution (Evanno *et al.* 2005).

It seems, at least according to AFLP data, there is a genetic differentiation among groups of *L. botrana* according to host plant. This is in contrast with my hypothesis, that was based on my previous Degree Project. In the previous project, behavioral studies on the oviposition choice of *L. botrana* were made using grapes and *Daphne gnidium*. No indication on host plant preference, based on host plant origin of females, was found (Döös 2010).

The sampling of *L. botrana* on different host plants was arranged to reflect a situation with limited gene flow between insects from separate host plants in Pisa, as well as a situation where the two host plants were completely interspersed in Sardinia (see Figure 7B). Nevertheless, both situations gave significant genetic differentiation among host plants, suggesting that these differences can arise entirely without physical isolation between host plant patches. The influence of sample size on number of bands and number of alleles tells us to treat these results with caution. Another factor suggesting potential sampling artifacts is the relatively high inbreeding coefficient revealed by microsatellites, where all populations had a significant F_{IS}

value except for Lab rearing. This might sound odd at first taking into the account that laboratory rearing mostly have a lower genetic diversity than wild populations, but it might be explainable by the procedure in which my material was sampled. The lab rearing consisted of individuals sampled from different locations and there was a constant flow of new additions of individuals to keep the rearing healthy and productive. Such results could also be due to different sampling procedure in which, for example, the sampling of populations might have been done in a non-random way where siblings were gathered from the same plant. This is not known since I had no control over how this sampling was done, but if this is the case it could easily have enhanced the genetic differentiation between host plants in the same area too. Dickey and Medina (2012) wrote that one of the genotypic signatures of Host-associated differentiation is that populations exhibit stronger differentiation by host plant species than by geographic isolation. This appears to be partly the case with *L. botrana*. Reciprocal genetic distance comparisons between host plants in Cagliari and Pisa revealed greater genetic distances between host plants within the same site than between sites. On the other hand, the overall differentiation between samples from different locations (PhiPT 0.114 table 11) was higher than the differentiation between host-plants (PhiPT 0.107 table 12 and 0.081 table 13). Comparisons between other known cases of host race formations show that many of them are tightly linked to their host plants in several aspects. For example Host specific mating (Feder 1998) and Host-associated trade-offs (Feder 1998; Ohshima 2008).
Host specific mating, also called host fidelity, means that the adults mate and oviposit on the same species of host plant as they fed on as larvae. This is a system of positive assortative mating acting as a pre mating barrier to gene flow between groups specialized on alternative plants (Feder 1998). No observations have been published on where the mating of *L. botrana* takes place and my previous study on oviposition preference of mated *L. botrana* females showed that they do not necessarily prefer to oviposit on the host-plant they came from (Döös 2010).
Host-associated trade-offs acts as post zygotic barriers to gene flow. Examples of this can be inviability or sterility problems in hybrids between groups specialized on alternative plants. Another example is performance of the offspring. Detoxification of plant secondary compounds can promote host associated adaptation and prevents gene flow between host races (Feder 1998). This is supported by studies of the Leaf mining moth *Acrocercops transecta* and its host races living on *Juglans ailanthifolia* and *Lyonia ovalifolia*. Transplantations of larvae between hosts resulted in larvae from *Juglans* completely failed to survive on *Lyonia* (Ohshima 2008). The anti-insect activities of *Daphne* might be preventing the gene flow between groups of *L. botrana* living on separate hosts, resulting the differentiation I found according to host plant.

The significant differences generated from the AFLP data might, or might not show an actual difference. Considering the great differences in defensive chemistry and other factors between the host plants *Daphne* and grapevine, local adaptations for each host plant could be ongoing, but yet being counter-balanced by gene flow. Despite surrounding conditions that point to a possible host race formation process, looking at the overall results it does not seem like *L. botrana* has formed host races. But to get a more accurate view on whether *L. botrana* is on its way towards dividing into host races, we need more genetic material. To be sure that the data from a location does not consist of mostly siblings, the sampling method should be carefully carried out and documented.

6. REFERENCES

Abrahamson WG, Eubanks MD, Blair CP, Whipple AV (2001) Gall flies, inquilines and goldenrods: a model for host-race formation and sympatric speciation. American Zoologist 41:928-938

Alvarez Arias BT (2000) Ichthyotoxic plants used in Spain, Journal of Ethnopharmacology 73:505-12

Amsellem L, Ristericci AM, Benrey B (2003) Isolation and characterization of polymorphic microsatellite loci in Lobesia botrana Den. & Schiff. (Lepidoptera: Tortricidae). Molecular Ecology Notes 3:117-119

Blair CP, Abrahamson WG, Jackman JA and Tyrrell L (2005) Cryptic speciation and host-race formation in purportedly generalist Tumbling flower beetle. Evolution 304-316

Botstein D, White RL, Skolnick M and Davis RW (1980) Construction of genetic linkage map in man using restriction fragment length polymorphisms. The American journal of human genetics 32:314-331

Bush GL (1969) Sympatric host race formation and speciation in frugivorous flies of the genus Rhagoletis (Diptera, Tephritidae) Evolution 23:237-251.

Bush GL (1994) Sympatric speciation in animals: new wine in old bottles, Trend in ecology & evolution 9:285-288

Craig, TP, Itami K, Abrahamson JW and Horner JD (1993) Behavioral evidence for host-race formation in Eurosta solidaginis. Evolution 47:1696–1710.

Craig TP, Horner JD and Itami JK (1997) Hybridization studies on the host races of *Eurosta solidaginis*: Implications for sympatric speciation. Evolution 51:1552–1560.

Craig TP, Horner JD and Itami JK (2001) Genetics, experience, and host-plant preference in *Eurosta solidaginis*: Implications for host shifts and speciation. Evolution 55:773–782.

Crozier RH. and Pamilo P (1996) One into two will go. Nature 383:574–575.

Dickey AM and Medina RF (2012) Host-associated genetic differentiation in pecan leaf Phylloxera. Entomologia Experimentalis et Applicata 143:127–137

Diehl SR, Bush GL (1984) An evolutionary and applied perspective of insect biotypes, Annual reviews of entomology 29:471-504

Dres M and Mallet J (2002) Host races in plant-feeding insects and their importance in sympatric speciation, Philosophical transactions of the royal society of London B 357:471-492

Döös Sophia (2010) Is the grapevine moth *Lobesia botrana* going towards separate host races? Master´s Degree Project at Lund University

Emelianov I, Simpson F, Narang P and Mallet J (2003) Host choice promotes reproductive isolation between host races of the larch budmoth Zeiraphera diniana. Journal of Evolutional Biology 16:208-218

Evanno G, Regaut S and Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology 14:2611-20

Excoffier L, Laval G and Schneider S (2005) Arlequin (Version 3.0): An integrated software package for population genetics data analysis. Journal of Evolutional Bioinformatics 1:47-50

Feder JL, Chilcote CA and Bush GL (1988) Genetic differentiation between sympatric host races of the apple maggot fly Rhagoletis pomonella. Nature 336:61–64

Feder JL (1998) The apple maggot fly, Rhagoletis pomonella: flies in the face of conventional wisdom about speciation? In: Howard, D. J. and Berlocher, S.H. (eds) Endless forms: Species and speciation pp 130-144 Oxford University Press, New York

Feder JL, Chilcote CA, Bush GL (1990) The geographic pattern of genetic differentiation between host associated populations of Rhagoletis pomonella (Diptera: Tephritidae) in the eastern United states and Canada. Evolution 44:570-595

Gabel B and Roehrich R (1995) Sensitivity of grapevine phenological stages to larvae of European grapevine moth, *Lobesia botrana* Den. et Schiff. (Lep. Tortricidae) Journal of Applied Entomology 119:127-130

Goldstein DB and Schlötterer C (1999) Microsatellites Evolution and applications Oxford University Press

Haig SM, Mullins TD, Forsman ED, Trail PW and Wennerberg L (2004) Genetic identification of spotted owls, barred owls, and their hybrids: legal implications of hybrid identity, Conservation Biology 18:1347-1357

Hawthorn DJ and Via S (2001) Genetic linkage of ecological specialization and reproductive isolation in pea aphids. Nature 412:904-907

Hearne CM, Ghosh S and Todd JA (1992) Microsatellites for linkage analysis of genetic traits. Trend in genetics 8:288–294

Itami, JK., Craig TP, and Horner JD (1997) Factors affecting gene flow between the host races of Eurosta solidaginis. In S. Mopper and S. Y. Strauss (eds.), Genetic structure and local adaptation in natural insect populations: Effects of ecology, life history, and behavior, pp 375–407. Chapman & Hall, New York.

Jarne P, Lagouda PJ (1996) Microsatellites, from molecules to populations and back. Trends in Ecology & Evolution 11:424-9

Johnson PA, Hoppensteadt FC. Smith JJ and Bush GL (1996) Conditions for sympatric speciation: A diploid model incorporating habitat fidelity and non-habitat assortative mating. Evolutionary Ecology 10:187–205.

Kondrashov AS and Mina MV (1986) Sympatric speciation, when is it possible? Biological journal of the Linnean society27: 201-233

Larsson S and Ekbom B (1995) Oviposition mistakes in herbivorous insects: Confusion or a step towards a new host plant? Oikos 72:155–160.

Maher N, Thiéry D (2006a) Daphne gnidium, a possible native host plant of the European grapevine moth Lobesia botrana, stimulates its oviposition. Is a host shift relevant? Chemoecology 16:135-144

Maher N, Thiéry D and Städler E (2006b) Oviposition by Lobesia botrana is stimulated by sugars detected by contact chemoreceptors, Physiological Entomology 31:14–22

Mallet J (2001) The speciation revolution, Journal of evolutional biology 14:887-888

Moreau J, Arruego X, Benrey B and Thiéry D (2006a) Differences in Nutritional quality of parts of Vitis vinifera berries affect fitness of the European grapevine moth, Entomologica Experimentalis et Applicata 119:93-99

Moreau J, Benrey B and Thiéry D (2006b) Assessing larval food quality for phytophagous insects: are the facts as simple as they appear? Functional Ecology 20:592-600

Moreau J, Rahme J, Benrey B and Thiéry D (2008) Larval host plant origin modifies the adult oviposition preference of the female European grapevine moth *Lobesia botrana*, Naturwissenschaften 95:317-324

Mullis K, Faloona F, Scharf S, Saiki R, Horn G and Erlich H (1986) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harbor Symposium in Quantitative Biology 51:263-73

Ohshima I (2008) Host race formation in the Leaf-mining moth Acrocercops transecta (Lepidoptera: Gracillariidae), Biological Journal of the Linnean Society 93:135-145

Pascual-Villalobos MJ and Robledo A (1998) Screening for anti-insect activity in Mediterranean plants, Industrial Corps and Products 8:183-194

Pellmyr O and Leebons-Mack J (1998) Herbivores and molecular clocks as tools in plant biogeography. Biological Journal of the Linnean Society 63:367–378.

Pratt GF (1994) Evolution of Euphilotes (Lepidoptera: Lycaenidae) by seasonal and host shifts. Biological Journal of the Linnean Society 51:387–416.

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

Reineke A, Karlovsky P and Zebits CPW (1998) Preparation and purification of DNA from insects for AFLP analysis. Insect Molocular Biology 7:95-99

Rice WR (1987) Selection via habitat specialization, the evolution of reproductive isolation as a correlated character. Evolutional Ecology 1:301-314

Schilthuizen M (2000) Dualism and conflicts in understanding speciation BioEssays 22:1134-1141

Silvertown J, Servaes C, Biss P and Macleod D (2005) Reinforcement of reproductive isolation between adjacent populations in the Park Grass Experiment, Heredity 95:198–205

Stoeva R (1982) Food-plants of the grape moth (Lobesia botrana Schiff.) in Bulgaria, Gradinarska I Lozarska Nauka 19:83-90

Tabashnik BE, Wheelock H, Rainbolt JD and Watt WB (1981) Individual variation in oviposition preference in the butterfly, Colias eurytheme, Oecologia 50:225-30

Tasin M, Bäckman AC, Anfora G, Carlin S, Ioriatti C and Witzgall P (2010) Attraction of Female Grapevine Moth to Common and Specific Olfactory Cues from 2 Host Plants, Chemical Senses 35:57-64

Tauber CA. and Tauber MJ (1989) Sympatric speciation in insects: Perception and perspective. In Speciation and its consequences eds. Otte D. and Endler J.A. pp 307–345. Sinauer Associates, Sunderland, Massachusetts.

Thiéry D and Moreau J (2005) Relative performance of European grapevine moth (Lobesia botrana) on grapes and other hosts, Oecologia 143:548-556
Turelli M, Barton NH and Coyne JA (2001) Theory and speciation. Trends in Ecology & Evolution 16:330–343
Via S (2001) Sympatric speciation in animals: the ugly duckling grows up, Trends in Ecology & Evolution 16:381-390
Vos P, Hogers R, Bleeker M, Reijans M and Van de Lee T (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research 23:4407-4414
Waring GL, Abrahamson WG, and Howard DJ (1990) Genetic differentiation among host associated populations of the gallmaker Eurosta solidaginis (Diptera: Tephritidae). Evolution 44:1648–1655
Wennerström A, Hjulström L. N, Hjältén J and Julkunen-Tiitto R (2010) Mother really knows best: host choice of adult phytophagous insect females reflects a within-host variation in suitability as larval food. Chemoecology 20:35-42
Williams JG, Kubelik AR, Livak KJ, Rafalski JA and Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Research 18:6531-6535
Wood TK, Tilmon KJ, Shantz AB, Harris CK, and Pesek J (1999) The role of host-plant fidelity in initiating insect race formation. Evolutionary ecology research 1:317–332
Wright S (1931) Evolution in Mendelian Populations. Genetics 16:97-126
Zabeau M and Vos P (1993) Selective restriction amplification: a general method for DNA fingerprinting. European Patent Application, publication no: EP 0534858

Appendix 1: Allele frequencies at each locus for each of the nine populations. Highlighted allele frequencies=Private/ rare alleles

Locus	Cagliari-Daphne	Cagliari-Vitis	Israel-Vitis	Lab-Rearing	Pescara-Vitis	Pisa-Daphne	Pisa-Vitis	Trento-Vitis	Verona-Vitis
Locus: Lobot-6 (Bp)	N=20	N=11	N=12	N=8	N=24	N=23	N=27	N=39	N=9
199	0.000	0.000	0.000	0.000	0.000	0.022	0.000	0.000	0.000
229	0.000	0.000	0.000	0.000	0.021	0.000	0.000	0.000	0.000
260	0.000	0.045	0.000	0.000	0.000	0.000	0.019	0.000	0.000
267	0.000	0.000	0.000	0.000	0.000	0.000	0.019	0.000	0.000
281	0.000	0.000	0.042	0.000	0.042	0.000	0.000	0.026	0.000
284	0.075	0.091	0.125	0.000	0.042	0.087	0.167	0.013	0.056
287	0.050	0.136	0.042	0.063	0.083	0.174	0.167	0.128	0.111
290	0.175	0.000	0.000	0.438	0.125	0.109	0.130	0.051	0.222
293	0.075	0.091	0.000	0.250	0.104	0.043	0.037	0.154	0.000
296	0.000	0.045	0.000	0.000	0.000	0.000	0.056	0.000	0.000
301	0.000	0.000	0.000	0.000	0.042	0.000	0.000	0.000	0.278
310	0.000	0.000	0.000	0.000	0.146	0.000	0.000	0.000	0.000
318	0.025	0.000	0.667	0.188	0.083	0.022	0.019	0.192	0.056
324	0.050	0.045	0.042	0.000	0.042	0.022	0.093	0.026	0.111
327	0.000	0.045	0.000	0.000	0.021	0.022	0.056	0.000	0.000
330	0.000	0.000	0.000	0.000	0.021	0.000	0.000	0.000	0.000
352	0.000	0.000	0.000	0.063	0.000	0.022	0.000	0.051	0.000
357	0.000	0.045	0.000	0.000	0.021	0.000	0.000	0.000	0.000
363	0.000	0.136	0.000	0.000	0.000	0.000	0.000	0.000	0.000
369	0.075	0.000	0.042	0.000	0.063	0.043	0.019	0.000	0.000
372	0.050	0.000	0.042	0.000	0.000	0.022	0.019	0.000	0.000
377	0.050	0.000	0.000	0.000	0.021	0.022	0.037	0.064	0.111
381	0.175	0.000	0.000	0.000	0.000	0.022	0.000	0.000	0.000
383	0.000	0.000	0.000	0.000	0.000	0.022	0.000	0.179	0.000

386	0.025	0.136	0.000	0.000	0.063	0.174	0.111	0.090	0.000
389	0.075	0.000	0.000	0.000	0.042	0.043	0.000	0.000	0.000
392	0.025	0.000	0.000	0.000	0.000	0.043	0.000	0.000	0.056
396	0.025	0.045	0.000	0.000	0.000	0.000	0.000	0.000	0.000
400	0.050	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
403	0.000	0.045	0.000	0.000	0.000	0.043	0.037	0.000	0.000
412	0.000	0.000	0.000	0.000	0.000	0.043	0.019	0.000	0.000
420	0.000	0.000	0.000	0.000	0.021	0.000	0.000	0.000	0.000
440	0.000	0.091	0.000	0.000	0.000	0.000	0.000	0.000	0.000
460	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.026	0.000
Locus: Lobot-11 (Bp)									
295	0.000	0.000	0.000	0.000	0.048	0.000	0.000	0.000	0.000
299	0.050	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
303	0.050	0.083	0.125	0.875	0.214	0.091	0.045	0.441	0.333
307	0.775	0.750	0.167	0.125	0.500	0.705	0.818	0.485	0.389
311	0.075	0.167	0.000	0.000	0.238	0.136	0.114	0.059	0.278
315	0.050	0.000	0.708	0.000	0.000	0.068	0.023	0.015	0.000
Locus: Lobot-14 (Bp)									
265	0.000	0.000	0.000	0.000	0.021	0.000	0.000	0.000	0.000
273	0.100	0.125	0.000	0.000	0.167	0.196	0.022	0.079	0.000
275	0.050	0.000	0.000	0.143	0.000	0.000	0.065	0.092	0.000
277	0.000	0.000	0.091	0.000	0.083	0.000	0.065	0.000	0.000
279	0.825	0.875	0.909	0.857	0.646	0.804	0.761	0.829	0.875
281	0.025	0.000	0.000	0.000	0.083	0.000	0.000	0.000	0.125
283	0.000	0.000	0.000	0.000	0.000	0.000	0.087	0.000	0.000
Locus: Lobot-15 (Bp)									
180	0.025	0.000	0.000	0.000	0.000	0.000	0.037	0.000	0.056

184	0.000	0.000	0.000	0.000	0.022	0.000	0.000	0.000	0.000
186	0.025	0.083	0.000	0.000	0.087	0.026	0.000	0.075	0.000
188	0.050	0.000	0.000	0.000	0.022	0.026	0.019	0.013	0.000
190	0.025	0.000	0.000	0.125	0.000	0.026	0.019	0.050	0.056
192	0.000	0.000	0.000	0.000	0.043	0.000	0.037	0.000	0.056
198	0.300	0.458	0.042	0.563	0.457	0.421	0.389	0.350	0.278
200	0.125	0.042	0.208	0.000	0.087	0.158	0.111	0.100	0.000
202	0.150	0.125	0.000	0.313	0.043	0.026	0.074	0.113	0.111
204	0.300	0.208	0.750	0.000	0.196	0.263	0.296	0.200	0.444
206	0.000	0.083	0.000	0.000	0.000	0.000	0.000	0.000	0.000
210	0.000	0.000	0.000	0.000	0.022	0.026	0.019	0.038	0.000
212	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.025	0.000
214	0.000	0.000	0.000	0.000	0.022	0.000	0.000	0.038	0.000
228	0.000	0.000	0.000	0.000	0.000	0.026	0.000	0.000	0.000

Appendix 2. Allele frequencies for each locus at each population (as 7 populations)

	Cagliari-Daphne-Vitis	Israel-Vitis	Lab-Rearing	Pescara-Vitis	Pisa-Daphne-vitis	Trento-Vitis	Verona-Vitis
Locus: Lobot-6 (Bp)	N=31	N=12	N=8	N=24	N=50	N=39	N=9
199	0.000	0.000	0.000	0.000	0.010	0.000	0.000
229	0.000	0.000	0.000	0.021	0.000	0.000	0.000
260	0.016	0.000	0.000	0.000	0.010	0.000	0.000
267	0.000	0.000	0.000	0.000	0.010	0.000	0.000
281	0.000	0.042	0.000	0.042	0.000	0.026	0.000
284	0.081	0.125	0.000	0.042	0.130	0.013	0.056
287	0.081	0.042	0.063	0.083	0.170	0.128	0.111
290	0.113	0.000	0.438	0.125	0.120	0.051	0.222
293	0.081	0.000	0.250	0.104	0.040	0.154	0.000
296	0.016	0.000	0.000	0.000	0.030	0.000	0.000
301	0.000	0.000	0.000	0.042	0.000	0.000	0.278
310	0.000	0.000	0.000	0.146	0.000	0.000	0.000
318	0.016	0.667	0.188	0.083	0.020	0.192	0.056
324	0.048	0.042	0.000	0.042	0.060	0.026	0.111
327	0.016	0.000	0.000	0.021	0.040	0.000	0.000
330	0.000	0.000	0.000	0.021	0.000	0.000	0.000
352	0.000	0.000	0.063	0.000	0.010	0.051	0.000
357	0.016	0.000	0.000	0.021	0.000	0.000	0.000
363	0.048	0.000	0.000	0.000	0.000	0.000	0.000
369	0.048	0.042	0.000	0.063	0.030	0.000	0.000
372	0.032	0.042	0.000	0.000	0.020	0.000	0.000
377	0.032	0.000	0.000	0.021	0.030	0.064	0.111
381	0.113	0.000	0.000	0.000	0.010	0.000	0.000
383	0.000	0.000	0.000	0.000	0.010	0.179	0.000
386	0.065	0.000	0.000	0.063	0.140	0.090	0.000

389	0.048	0.000	0.000	0.042	0.020	0.000	0.000
392	0.016	0.000	0.000	0.000	0.020	0.000	0.056
396	0.032	0.000	0.000	0.000	0.000	0.000	0.000
400	0.032	0.000	0.000	0.000	0.000	0.000	0.000
403	0.016	0.000	0.000	0.000	0.040	0.000	0.000
412	0.000	0.000	0.000	0.000	0.030	0.000	0.000
420	0.000	0.000	0.000	0.021	0.000	0.000	0.000
440	0.032	0.000	0.000	0.000	0.000	0.000	0.000
460	0.000	0.000	0.000	0.000	0.000	0.026	0.000
Locus: Lobot-11 (Bp)							
295	0.000	0.000	0.000	0.048	0.000	0.000	0.000
299	0.031	0.000	0.000	0.000	0.000	0.000	0.000
303	0.063	0.125	0.875	0.214	0.068	0.441	0.333
307	0.766	0.167	0.125	0.500	0.761	0.485	0.389
311	0.109	0.000	0.000	0.238	0.125	0.059	0.278
315	0.031	0.708	0.000	0.000	0.045	0.015	0.000
Locus: Lobot-14 (Bp)							
265	0.000	0.000	0.000	0.021	0.000	0.000	0.000
273	0.109	0.000	0.000	0.167	0.109	0.079	0.000
275	0.031	0.000	0.143	0.000	0.033	0.092	0.000
277	0.000	0.091	0.000	0.083	0.033	0.000	0.000
279	0.844	0.909	0.857	0.646	0.783	0.829	0.875
281	0.016	0.000	0.000	0.083	0.000	0.000	0.125
283	0.000	0.000	0.000	0.000	0.043	0.000	0.000
Locus: Lobot-15 (Bp)							
180	0.016	0.000	0.000	0.000	0.022	0.000	0.056
184	0.000	0.000	0.000	0.022	0.000	0.000	0.000

186	0.047	0.000	0.000	0.087	0.011	0.075	0.000
188	0.031	0.000	0.000	0.022	0.022	0.013	0.000
190	0.016	0.000	0.125	0.000	0.022	0.050	0.056
192	0.000	0.000	0.000	0.043	0.022	0.000	0.056
198	0.359	0.042	0.563	0.457	0.402	0.350	0.278
200	0.094	0.208	0.000	0.087	0.130	0.100	0.000
202	0.141	0.000	0.313	0.043	0.054	0.113	0.111
204	0.266	0.750	0.000	0.196	0.283	0.200	0.444
206	0.031	0.000	0.000	0.000	0.000	0.000	0.000
210	0.000	0.000	0.000	0.022	0.022	0.038	0.000
212	0.000	0.000	0.000	0.000	0.000	0.025	0.000
214	0.000	0.000	0.000	0.022	0.000	0.038	0.000
228	0.000	0.000	0.000	0.000	0.011	0.000	0.000

Appendix 3 L(K); Mean likelihood over five runs for 10 different K values.

K	L(K) run 1	L(K) run 2	L(K) run 3	L(K) run 4	L(K) run 5	mean L(K)	StDev
1	-4880,1	-4880	-4879,9	-4880,1	-4880,4	-4880,1	0,187082869
2	-4466,1	-4466,5	-4467,8	-4468,1	-4464,9	-4466,68	1,304607221
3	-4123,9	-4124,9	-4128,1	-4123,8	-4120,8	-4124,3	2,620114501
4	-3936,4	-3955,4	-3934	-3930,7	-3928,5	-3937	10,72217329
5	-3742,1	-3951,2	-4012,3	-3752,1	-3884,7	-3868,48	119,6934501
6	-3610,9	-3675,4	-3652,6	-3674,8	-3618,6	-3646,46	30,49209734
7	-3528,5	-3521,2	-3512,4	-3641,5	-3515,6	-3543,84	54,93480682
8	-3502,7	-3448,3	-3454,2	-3457,6	-3452,7	-3463,1	22,38760818
9	-3433,8	-3460	-3435,6	-3451,7	-3429,9	-3442,2	12,96630248
10	-3546,3	-3404,9	-3936,4	-3412,7	-3449,7	-3550	223,2062051

Appendix 4 L´(K); Mean difference between successive likelihood values of K=2 to K=10.

K	L´(K) run 1	L´(K) run 2	L´(K) run 3	L´(K) run 4	L´(K) run 5	Mean L´(K)	StDev
2	414	413,5	412,1	412	415,5	413,42	1,45
3	342,2	341,6	339,7	344,3	344,1	342,38	1,90
4	187,5	169,5	194,1	193,1	192,3	187,3	10,27
5	194,3	4,2	-78,3	178,6	43,8	68,52	116,45
6	131,2	275,8	359,7	77,3	266,1	222,02	115,09
7	82,4	154,2	140,2	33,3	103	102,62	48,18
8	25,8	72,9	58,2	183,9	62,9	80,74	60,31
9	68,9	-11,7	18,6	5,9	22,8	20,9	30,00
10	-112,5	55,1	-500,8	39	-19,8	-107,8	229,29

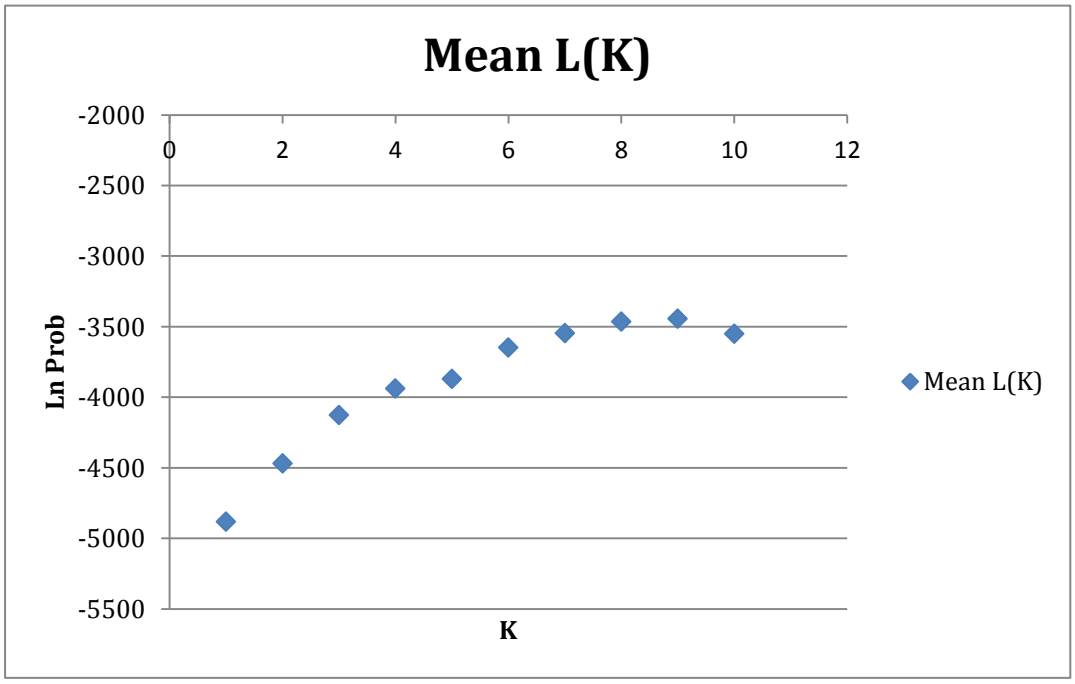
Appendix 5 L"(K); Absolute value of the difference between successive L`(K) values

K	L"(K) run 1	L´´(K) run2	L´´(K) run3	L´´(K) run4	L´´(K) run5	Mean L´´(K)	abs L´´(K)	StDev
2	-71,8	-71,9	-72,4	-67,7	-71,4	-71,04	71,4	1,900789
3	-154,7	-172,1	-145,6	-151,2	-151,8	-155,08	155,08	10,06861
4	6,8	-165,3	-272,4	-14,5	-148,5	-118,78	118,78	115,4244
5	-63,1	271,6	438	-101,3	222,3	153,5	153,5	229,9238
6	-48,8	-121,6	-219,5	-44	-163,1	-119,4	119,4	75,17323
7	-56,6	-81,3	-82	150,6	-40,1	-21,88	21,88	98,02034

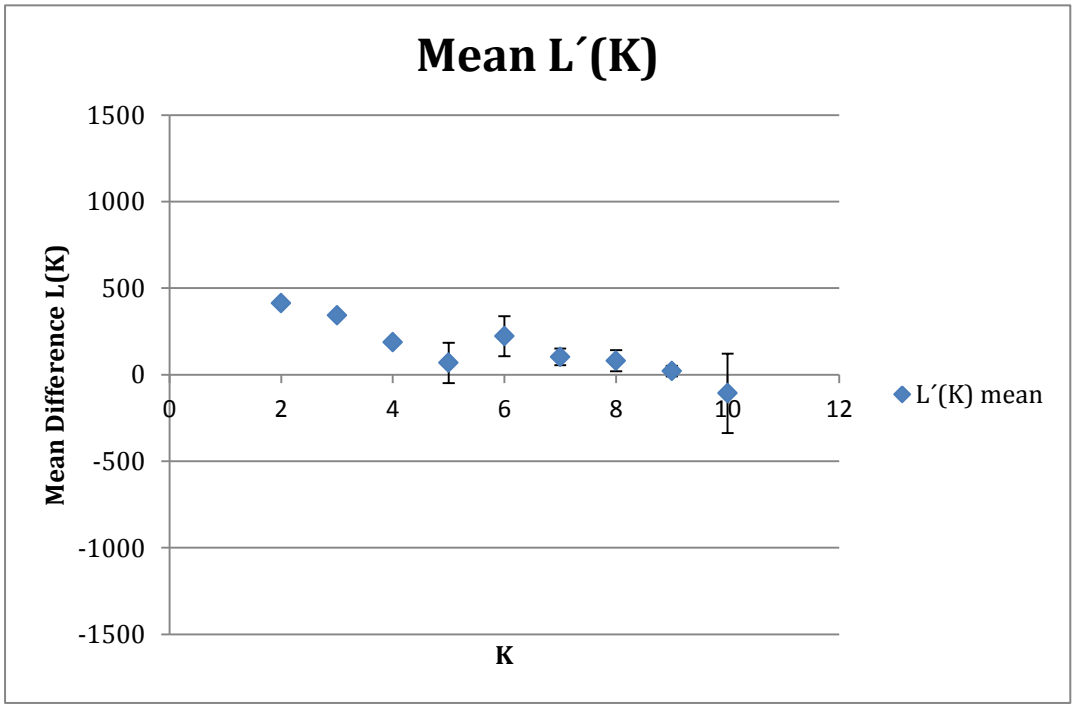
8	43,1	-84,6	-39,6	-178	-40,1	-59,84	59,84	80,57731
9	-181,4	66,8	-519,4	33,1	-42,6	-128,7	128,7	238,3466
10	112,5	-55,1	500,8	-39	19,8	107,8	107,8	229,2933

Appendix 6 ΔK; L´´(K) divided by Standard Deviation of L(K)

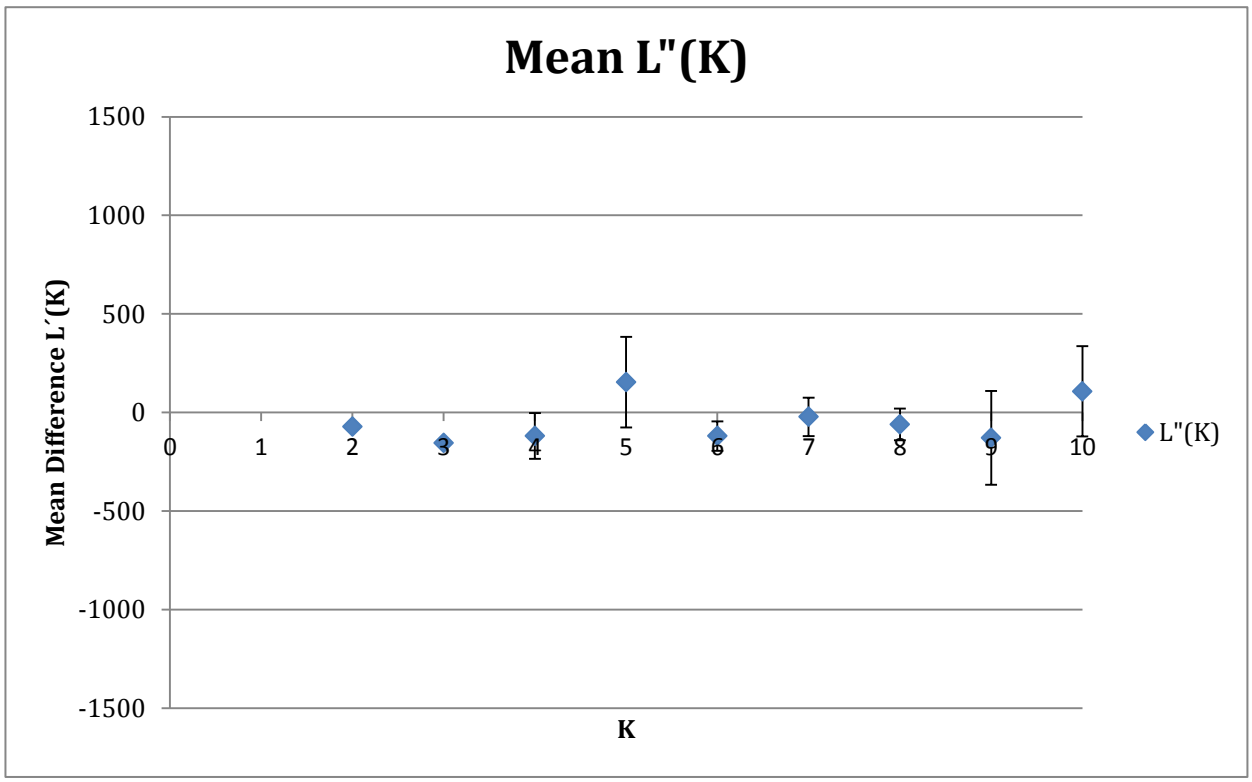
K	StDev[L(K)]	Mean L"(K)	Δ K
2	1,304607221	71,04	54,45317
3	2,620114501	155,08	59,18825
4	10,72217329	118,78	11,07798
5	119,6934501	153,5	1,282443
6	30,49209734	119,4	3,915769
7	54,93480682	21,88	0,39829
8	22,38760818	59,84	2,672907
9	12,96630248	128,7	9,925729
10	223,2062051	107,8	0,482961



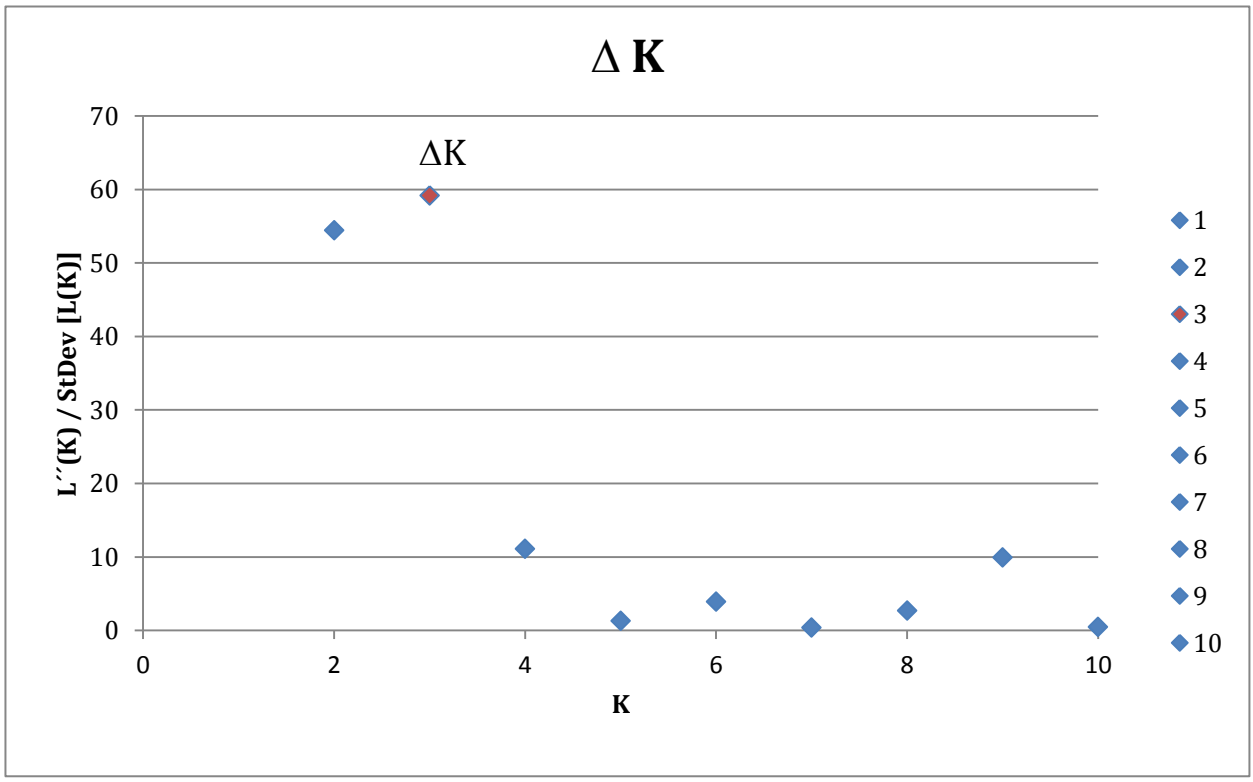
Appendix 7 L(K), mean likelihood over five runs, plotted for 10 different K values.



Appendix 8 L´(K), mean difference between successive likelihood values, plotted for K=2 to K=10



Appendix 9 L``(K), the absolute value of the difference between successive L`(K) values, plotted



Appendix 10 ΔK, the L´´(K) divided by Standard Deviation of L(K), plotted, where the highest value is the true K