

## Evolutionary and Ecological Genetics of *Medicago truncatula*

**Thomas Bataillon** Bioinformatics Research Center & Department of Genetics and Ecology,  
Aarhus University, Hoegh-Guldbergs Gade 10, Building 1090, DK-8000  
Aarhus C. Denmark  
[tbata@daimi.au.dk](mailto:tbata@daimi.au.dk)  
<http://www.daimi.au.dk/~tbata>

**Joelle Ronfort** UMR 1097 DGPC, INRA domaine de Melgueil, F-34130 Mauguio, France.  
[Joelle.Ronfort@ensam.inra.fr](mailto:Joelle.Ronfort@ensam.inra.fr)

### Table of contents

1. Introduction
  2. The mating system of *Medicago truncatula*
  3. Distribution of genetic variation in *Medicago truncatula*
    - 3.1 Within population polymorphism
    - 3.2 Population structure in *Medicago truncatula*
  4. Variation in quantitative traits: Is there ecotypic differentiation in *Medicago truncatula*?
  5. Perspectives: towards the ecological and evolutionary genomics of *Medicago truncatula*
- References  
Figures

### Abstract

This chapter synthesizes our current knowledge on the ecological genetics of the model legume *Medicago truncatula*. After a short introduction to the life cycle of the species, we describe the mating system of *M. truncatula* and how selfing rates in nature can be inferred both by progeny testing and population surveys. We then review what is currently known on the distribution of genetic diversity within and among populations focusing first on the diversity that can be assessed using *a priori* neutral markers (allozyme and SSR). We then consider genetic variation for quantitative traits and discuss whether *Medicago truncatula* exhibits a so called “ecotypic” differentiation. Last we discuss how genomic tools and the availability of sequence information may help shed additional light on the evolutionary forces behind evolution and adaptation in *Medicago truncatula*.

## 1 Introduction

*Medicago truncatula* GAERTN is an annual species which is strongly selfing in nature. It is described as an “omni-Mediterranean” species (Lesins and Lesins 1979). *Medicago truncatula* has also become naturalized in other regions of the world following European migrations (see chapter “Origin and distribution of *Medicago truncatula* world wide” for details about its geographic distribution). It is an opportunistic plant mainly found in open areas and has typically ephemeral populations. *M. truncatula* has been split into three subspecies mainly on the basis of pod characteristics: ssp *truncatula*, ssp *tricycla* and ssp *longeaculata* (see also Chapter “Evolutionary relationships between *M. truncatula* and other legumes and phylogeny of the genus *Medicago*”). Previous molecular analyses demonstrated that individuals from the subspecies *tricycla* are genetically differentiated from the two other subspecies (Baquerizo-Audiot, et al. 2001, Ellwood, et al. 2006).

In Australia, *M truncatula*, among other annual *Medicago* species, is used as a forage crop and soil improver. This has motivated a number of research programs aimed at collecting and characterizing numerous natural populations but primarily in an agronomic context. About 10-15 years ago *Medicago truncatula* was also recognized as a potentially interesting model species for ecological genetics and in particular to study the distribution of genetic diversity at a very fine spatial scale in an annual selfing plant. In many ways, the ecology of *M truncatula* is similar to that of the mouse ear cress (*Arabidopsis thaliana*). Both species are annual and highly selfing and exhibit a colonizing habit although *M truncatula* does not have a geographical range as wide as *A. thaliana*. In that respect it will be illuminating to compare patterns of genetic and phenotypic variation in both species. Finally the choice of *truncatula* as a model legume to dissect the molecular mechanisms enabling the partnership with nitrogen fixing bacteria makes *M truncatula* a species of choice for our evolutionary understanding of plant nitrogen fixing bacteria symbiosis.

## 2 The mating system of *Medicago truncatula*

*M truncatula* is considered to be an autogamous (selfing) species. It bears small inflorescences, racemes, bearing 1 to 5 yellow flowers that are 5-8 mm long. Flowers contain 10 ovules and only open after the auto pollen has fertilized the ovules. Residual outcrossing is however possible after the flower opens. *M truncatula* flowers possess a so called “tripping mechanism”, a floral characteristic shared by the whole *Medicago* genus (Lesins and Lesins 1979).

Various studies report estimates of the selfing rate of *Medicago truncatula* through the use of molecular markers. Self-fertilization is expected to greatly increase levels of homozygosity within genomes. Codominant molecular markers such as allozymes or microsatellites can be used to quantify the level of individual homozygosity within populations. Homozygosity at a locus is quantified through the parameter  $F_{IS}$  and can be simply estimated as  $F_{IS} = 1 - H_{obs}/H_{exp}$ , where  $H_{obs}$  is the observed frequency of heterozygous individuals at that locus and  $H_{exp} = 1 - \sum_i p_i^2$  is the level of heterozygosity expected under random mating at a locus with  $k$  alleles in frequency  $p_1, p_2, \dots, p_k$  (see for instance Hartl and Clark 1989). Under the assumption that increased homozygosity levels are due to selfing at a rate  $s$ , the value of the parameter  $F_{IS}$  is expected to equilibrate at  $F_{IS} = s/(2-s)$  regardless of the effective population or the mutational model of the molecular marker considered (Rousset 1996). This property can in turn be used to estimate

selfing rates from  $F_{IS}$  under the assumption that no within population subdivision creates extra inbreeding.

Figure 1 displays the distribution of selfing rates estimated from mean  $F_{IS}$  values in a set of 77 populations using allozyme data (JM Prospero unpublished data). The distribution of estimated selfing rate is heavily skewed towards high rates of selfing. Selfing rates range from  $s = 0.65$  to  $s = 1$  with a mean selfing rate of  $s = 0.965$ . Selfing rates were also reported in the range of 97.6–99.3% for 192 accessions originating from the SARDI collection (Ellwood, et al. 2006), although the actual procedure used to estimate selfing rates is not clear (a single individual was genotyped per accession).

A detailed study focusing on a single population (Aude, France) and analyzing three sub populations separately also reported very high selfing rates:  $s = 0.95$ ,  $0.97$  and  $0.97$  (Bonnin, et al. 2001). Ultimately robust unbiased estimates of the selfing rate realized over a single generation can be obtained by genotyping of maternal progenies collected in the wild (Ritland 2002). One such study using microsatellites to genotype a large set of maternal plants and progenies, collected as pods still attached to the mother plant, within one population (Salses, France) confirms that selfing rates are indeed potentially very high in natural populations of *Medicago truncatula* ( $s = 0.97$ , Ronfort unpublished data). The distribution of selfing rates reported in Figure 1 suggests that a few populations may have selfing rates of 65 to 70%. Null alleles at marker loci or hidden population substructure within population would create extra homozygosity and thus cannot account for these relatively low selfing rates. Although these selfing rates are indirect estimates and may be also due to the sampling variance associated with the estimation of the  $F_{IS}$  parameter, some populations may indeed experience intermediate levels of outcrossing. The causes of such variation in selfing rates are currently unknown.

### 3 Distribution of genetic variation in *Medicago truncatula*

#### 3.1 Within population polymorphism

Levels of genetic variation within populations of *Medicago truncatula* are highly variable. This is illustrated by a large survey of allozyme polymorphism (Figure 2). Distribution of within population variation among populations, quantified through estimates of the scaled effective size,  $\theta = 4N_e \mu$ , where  $N_e$  is the effective population size and  $\mu$  the mutation rate of the marker used, is very “L shaped”. Although the distribution of  $\theta$  values does not provide estimates of effective size per se, the ratio of  $\theta$  estimates in two populations is akin to the ratio of effective population size (assuming that rates of mutations are conserved across populations for a marker locus). This suggests that realized<sup>1</sup> scaled effective size can vary widely from population to population. This finding is being confirmed by recent studies that monitor temporal variation in allelic frequency, thereby providing direct estimates of the realized effective size over a short time interval within a single population (Siol et al submitted). Numerous populations exhibit little diversity if any, the whole population being dominated by a single homozygous genotypes. Other populations display much higher levels of diversity and comprise numerous multilocus genotypes often dominated by 2 or more multilocus genotypes. This finding was confirmed when studying a sample of 200 individual from 4 populations located in Southern France either using RAPD and microsatellite

---

<sup>1</sup> These scaled effective size estimates assume that each population is at mutation drift equilibrium.

markers (Bonnin, et al. 1996, Bonnin, et al. 2001) When looking closely at the composition of multi locus genotypes in a single population, the genetic makeup of the rarer multilocus genotypes can often be interpreted as the results of rare outcrossing events involving the most frequent multilocus genotypes followed by subsequent selfing and segregation of the F1 hybrids (Bonnin, et al. 2001) – Ronfort unpublished results.

In essence variation at the population level is best viewed as a set of recombinant inbred lines produced *in natura*. Rare outcrossing events are therefore instrumental in shaping within population diversity and these combined with some migration from neighboring populations (see below) may explain the sometimes spectacular levels of within population diversity. The “L shaped” distribution of levels of genetic variation within populations in *M. truncatula* is a characteristic shared by many annual plants which displays high levels of selfing (Schoen and Brown 1991).

The within population polymorphism may be understood in a broader geographical context possibly at the level of a metapopulation comprising numerous ephemeral populations. Each local population prone to extinction and may lose diversity through genetic drift and/or local selection. But periodic migration and re-colonization events may ensure maintenance of substantial levels of diversity at the whole metapopulation level. A great deal of theoretical models make prediction about levels of genetic diversity maintained depending on the amounts of migration and mode of extinction-recolonization (see (Barton and Whitlock 1997) for a review). Documenting in detail the relative roles of residual outcrossing, migration and dynamics of drift and extinction-recolonization is part of an ongoing effort to understand the mechanisms maintaining diversity in natural population of *Medicago truncatula*.

### 3.2 Population structure in *Medicago truncatula*

Population structure can be studied at various scales in both space and time. In this section we focus on the distribution of genetic diversity at various spatial scales ranging from within population structure to the existence of a broad genetic structure found at the whole species level. So far no studies have looked at seed bank dynamics and their impact on genetic diversity in that species.

Analysis of within population structure reveals that strong population subdivision can be found even within population between transects separated by only 10-50 meters (Bonnin, et al. 1996). Within a single sub population isolation by distance can also sometimes be found with significant spatial autocorrelation found among individuals located up to 7 meters apart (Bonnin et al., 2001).

At the regional level, differentiation between populations separated by about 200km as measured by the parameter  $F_{ST}$  (Weir and Cockerham 1984) can be substantial ( $F_{ST}=0.51$  CI 0.36-0.61). But note that levels of differentiation among sub populations within a single population can be as large as that found among populations. Pairwise genetic differentiation between populations can be sometimes correlate with geographic distance between population (Bonnin, et al. 1996) but no relationship between pairwise genetic differentiation and geographic distance was found among 9 Algerian populations ranging in distance from 15 to 700 km (JMP unpublished data).

When examining patterns of genetic diversity at a much broader geographic scale, that of the whole species range, a relatively clear but rather weak stratification emerges. We recently investigated the patterns population structure in a collection of 346 inbred lines sampled all

around the Mediterranean Basin and representing the breadth of naturally occurring diversity in *M. truncatula* using 13 microsatellite loci (Ronfort, et al. 2006). Using a model-based clustering algorithm (Falush, et al. 2003), we identified four main genetic clusters in the set of genotypes analyzed. We also identified a set of “admixed” individuals that do not fit with this population structure scheme (Figure 3a). This stratification of the various genotypes in 4 groups matches broad geographic regions (Figure 3b) and pairwise  $F_{ST}$  values measured *a posteriori* among clusters ranged only from 0.054 to 0.1. Thus the broad scale structure we uncovered although highly statistically significant only accounts for 10% of the whole range of variation found in our sample. This is coherent with the idea that substantial genetic variation and population sub-structure can still be found within clusters. Our finding is somewhat coherent with a recent study of 192 genotypes spanning the same geographic level and genotyped for 6 microsatellites (Ellwood, et al. 2006).

#### 4 Variation in quantitative traits: Is there ecotypic differentiation in *Medicago truncatula* ?

Before discussing heritable variation in quantitative traits and possible ecotypic differentiation in *M. truncatula* it seems worth restating here what ecotype actually means. The term can be traced back to the Swede Göran Turreson (Krebs 2001 and references therein) who in the 1920s was interested in the adaptation of plant to local variations in soil and climate through a species range. The concept of ecotype was then popularized among others by Dobzhansky and Mayr in their contributions to the Neo Darwinian Synthesis.

Ecotypes, also historically referred to as races, are defined genetic entities, members of a population that exhibit a set of (heritable) features signing the adaptation to a particular set of local environmental conditions. In a nutshell the existence of ecotypes within a species requires (i) that the candidate populations or genotypes exhibit heritable differences in morphology, physiology or phenology, (ii) occur in different habitats while still remaining inter-fertile. Last but not least (iii) the genetic differences should be “adaptive” i.e. enhance survival and reproduction in each local habitat. We stress here that this definition implies “much more” than the loose usage of ecotype by a good deal of the *Medicago* community to simply refer to genotypes collected in natural populations, hence “eco”types, instead of artificial mutants obtained in the lab! The selfing nature of *M. truncatula* and the presence of high levels of population structure are *a priori* features that bode well for the development of local adaptation in this species. We review below the empirical evidence currently available.

Various studies of the agronomic characteristics have been conducted on *M. truncatula* and heritable variation for a broad spectrum of phenology and reproductive traits have been reported. But these studies only address tangentially the question of the existence of ecotypic differentiation in *M. truncatula*, In particular the adaptive significance of the variation uncovered in these common garden experiments is at best unclear.

One way to document, albeit indirectly, the existence of locally varying selection on a life history trait across some geographic scale is to compare patterns of population structure on that trait to that revealed over the same set of populations through presumably neutral markers. Patterns of population structure for a quantitative trait can be summarized by the parameter  $Q_{st}$ . Comparison of population differentiation estimates at molecular markers ( $F_{ST}$ ) and quantitative traits ( $Q_{ST}$ ) gives potentially insight on the type of selection acting on a quantitative trait in nature. In a

nutshell, if the quantitative trait is neutral and its determinism is additive, then theory predicts that  $Q_{ST} = F_{ST}$ , while  $Q_{ST} > F_{ST}$  is predicted under directional selection for different local optima, and  $Q_{ST} < F_{ST}$  is predicted under homogenizing selection. Note that a number of statistical caveats arise. For instance departure from additivity in the quantitative trait (for instance through dominance) may complicate the story. The arguments get fairly technical but, in short,  $Q_{ST} > F_{ST}$  can be taken as a good indicator for the presence of selection for different local conditions while  $Q_{ST} < F_{ST}$ , could be the result of several factors other than homogenizing selection. (Goudet and Buchi 2006). One study in *M. truncatula* compared the  $F_{ST}$  revealed by RAPD markers with the  $Q_{ST}$  measured for 24 quantitative traits related to seedling stage, vegetative growth and reproduction (Bonnin, et al. 1996). They found that overall the two populations (Var and Aude) were more differentiated for quantitative traits than for markers. They also documented wide variation between quantitative traits with respect to  $Q_{ST}$ . Interestingly traits a priori are most related to fitness such as pod production or phenology (flowering date) were exhibiting the greatest inter population divergence suggesting that these traits indeed experience locally different selection pressures. The authors could find key heritable differences among populations in the life history of plants (individuals coming from the Aude population grew on average slower, flowered later and produced fewer pods than those sampled in Var); but no obvious ecological variable driving such differences could be identified.

In the future, conducting similar studies over a broader eco-geographic range will be instrumental in documenting rigorously the magnitude of ecotypic differentiation in *M. truncatula* and in understanding which environmental factors shape these differences.

## 5 Perspectives: towards the ecological and evolutionary genomics of *Medicago truncatula*

The sequencing of the gene space of *M. truncatula*, the broad availability of ESTs and the cloning of a growing number of genes involved in a range of biological phenomena opens new research avenues for ecological and evolutionary genetics. This information, combined with QTL mapping by conventional or candidate gene approach, will allow surveying allelic diversity in targeted genomic regions. The analysis of patterns of diversity at loci with known phenotypic effects, instead of anonymous neutral markers, will soon provide a wealth of information on the evolutionary forces shaping variation of a number of ecologically important traits.

To do so, some coordinated efforts are needed within the *Medicago* community to:

- Develop multiple sets of segregating populations, ideally recombinant inbreds. The broad scale stratification we uncovered in populations of *M. truncatula* can serve as a basis for choosing genotypes spanning the range of diversity known to date and complementing the sets of RILs already developed.
- Survey patterns of naturally occurring levels of nucleotide diversity in the *Medicago truncatula* genome by systematic re sequencing numerous short fragments (0.5 -2 kb) in a standardized panel of genotypes. Ideally these genotypes should be chosen in order to represent the breadth of currently sampled genetic diversity and if possible to maximize probability of SNP discovery. The first study published to date (De Mita, et al. 2006), surveying nucleotide polymorphism in a 4 kb fragments of the gene *NORK* (*Mt dmi2*) among 28 genotypes shows levels of nucleotide polymorphism slightly lower but comparable to those found in *A. thaliana* (Nordborg, et al. 2005). This bodes well of the use of naturally occurring nucleotide variation to gain insights into

the demographic and selective forces shaping patterns of polymorphism in *Medicago truncatula*. Such re sequencing will provide a valuable source of SNPs markers for QTLmapping (see Choi, et al. 2004 for a first survey) and comparative genomics.

- Characterize the levels and scale at which linkage disequilibrium (LD) decays in *M. truncatula*. This information will inform us on the physical scale at which LD mapping can be conducted and complement traditional QTL and fine mapping methods (Aranzana, et al. 2005). Ideally LD patterns should be characterized at various physical scales in the genome and in samples representing different levels of population structure (regions, species-wide).

## Acknowledgements

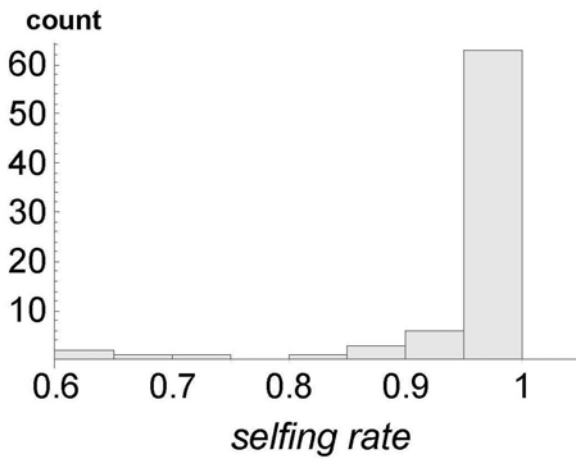
We are grateful to Jean-Marie Prosperi and Sylvie Roussel for making the allozyme data presented in Figure 2 available to us. We thank Bodil K. Ehlers for comments on an earlier draft of this chapter.

## References

- Aranzana MJ, Kim S, Zhao K *et al* (2005) Genome-Wide Association Mapping in Arabidopsis Identifies Previously Known Flowering Time and Pathogen Resistance Genes. *PLoS Genet*: **1**: e60.
- Baquerizo-Audiot E, Desplanque B, Prosperi JM, S. S (2001) Characterization of microsatellite loci in the diploid legume *Medicago truncatula* (barrel medic). *Molec Ecol Notes*: **1**: 1-3.
- Barton NH, Whitlock MC (1997) The evolution of metapopulation; in: Hanski IA, Gilpin ME (eds): Metapopulation Biology - Ecology, Genetics and Evolution.
- Bonnin I, Hugué T, Gherardi M, Prosperi JM, Olivieri I (1996) High levels of polymorphism and spatial structure in a sefling plant species, *Medicago truncatula* (Leguminosae), shown using RAPD markers. *Am J Bot*: **87**: 843-855.
- Bonnin I, Prosperi JM, Olivieri I (1996) Genetic markers and quantitative genetic variation in *Medicago truncatula* (Leguminosae): a comparative analysis of population structure. *Genetics*: **143**: 1795-1805.
- Bonnin I, Ronfort J, Wozniak F, Olivieri I (2001) Spatial effects and rare outcrossing events in *Medicago truncatula* (Fabaceae). *Mol Ecol*: **10**: 1371-1383.
- Choi HK, Kim D, Uhm T *et al* (2004) A sequence-based genetic map of *Medicago truncatula* and comparison of marker colinearity with *M. sativa*. *Genetics*: **166**: 1463-1502.
- De Mita S, Santoni S, Hochu I, Ronfort J, Bataillon T (2006) Molecular Evolution and Positive Selection of the Symbiotic Gene NOR1 in *Medicago truncatula*. *J Mol Evol*: **62**: 234-244.
- Ellwood SR, D'Souza NK, Kamphuis LG, Burgess TI, Nair RM, Oliver RP (2006) SSR analysis of the *Medicago truncatula* SARDI core collection reveals substantial diversity and unusual genotype dispersal throughout the Mediterranean basin. *Theor Appl Genet*: **112**: 977-983.
- Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*: **164**: 1567-1587.
- Goudet J, Buchi L (2006) The effects of dominance, regular inbreeding and sampling design on Q(ST), an estimator of population differentiation for quantitative traits. *Genetics*: **172**: 1337-1347.
- Krebs CJ (2001) Ecology. Benjamin Cummings, 2001.
- Lesins KA, Lesins I (1979) Genus *Medicago* (Leguminosae) - A taxogenetic study. The Hague, Kluwer.
- Nordborg M, Hu TT, Ishino Y *et al* (2005) The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biol*: **3**: e196.
- Ritland K (2002) Extensions of models for the estimation of mating systems using n independent loci. *Heredity*: **88**: 221-228.

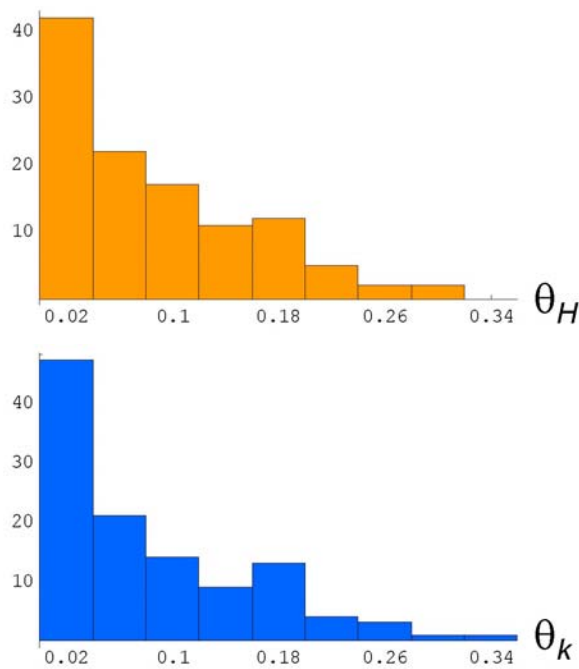
- Ronfort J, Bataillon T, Santoni S, Delalande M, David J, Prosperi JM (2006) Microsatellite diversity and broad scale geographic structure in a model legume: building a set of nested core collections for studying naturally occurring variation in *Medicago truncatula*. *BMC Plant Biology*: in review.
- Rousset F (1996) Equilibrium values of measures of population subdivision for stepwise mutation processes. *Genetics*: **142**: 1357-1362.
- Schoen DJ, Brown AHD (1991) Intraspecific variation in population gene diversity and effective population size correlates with the mating system in plants. *PNAS*: **88**: 4494-4497.
- Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*: **38**: 1358-1370.

**Figures and legends**



**Figure 1** Distribution of selfing rates within populations of *Medicago truncatula*.

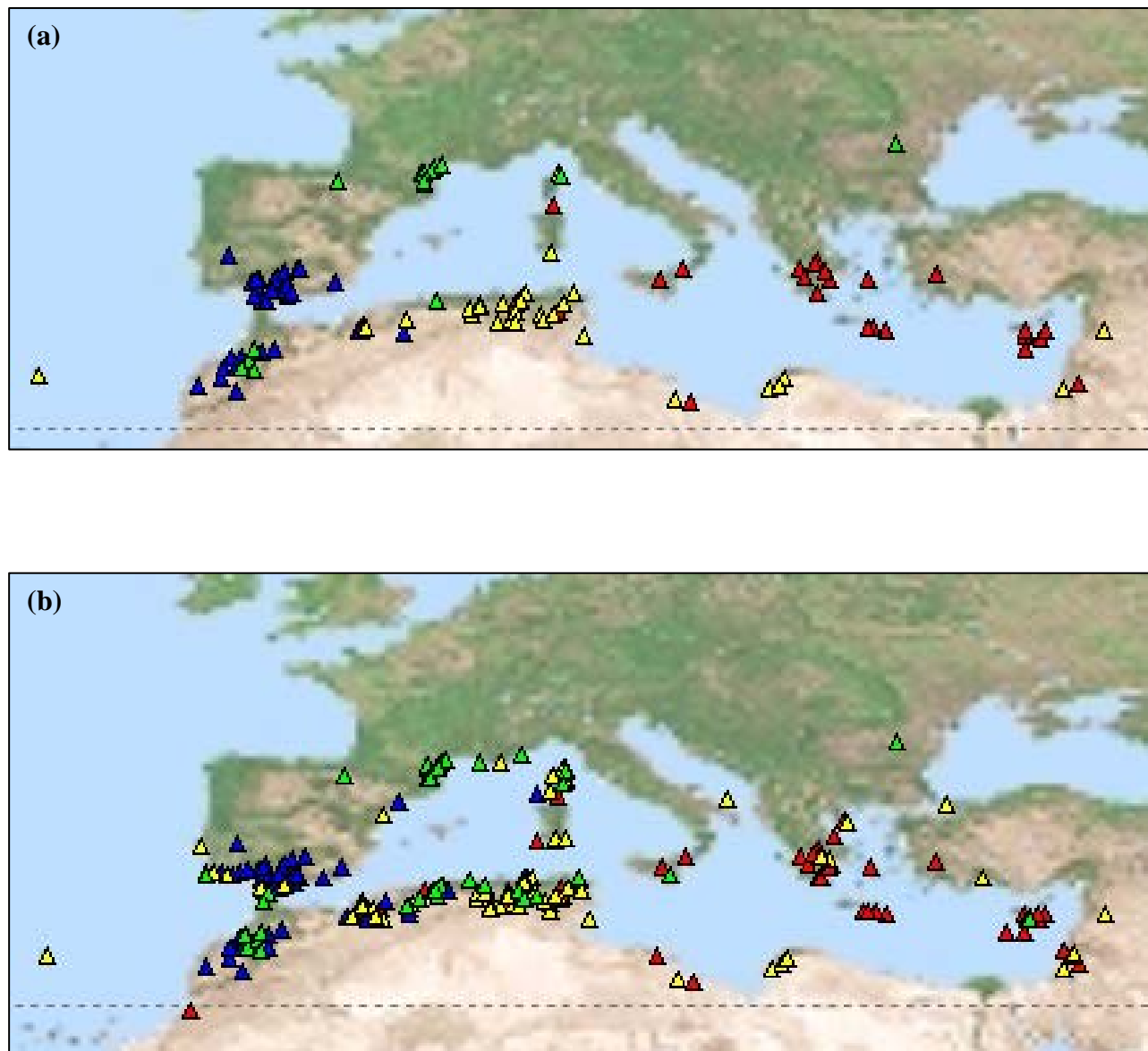
A large number of populations (n=113) were surveyed by sampling 10-30 individuals per population. These individuals were genotyped for 10 enzymatic loci (J-M Prospero unpublished results). Selfing rates were estimated using allozyme polymorphism at 10-15 loci (as in Figure 1). Within population polymorphism allowed estimation of the parameter  $F_{IS}$  for 77 out of the 113 population surveyed. Selfing rates were estimated from the  $F_{IS}$  parameter averaged over loci within each population (see text for details).



**Figure 2: Distribution of levels of allozyme polymorphism within population.**

A large number of populations (n=113) were surveyed by sampling 10-15 individuals per population. These individuals were genotyped for 10 enzymatic loci (J-M Prospero unpublished results).

Variation was measured within each population using two estimators of scaled effective size ( $\theta$ ). At each loci and within each population sample,  $\theta$  is estimated from the number of alleles observed in the sample ( $\theta_k$ ), or from the mean allele heterozygosity in the sample ( $\theta_H$ ), assuming that allozymes follow the so called infinite allele mutation model. Average of  $\theta$  estimates over 10 loci are reported here.



**Figure 3**

The stratification of a species-wide sample of 226 genotypes collection into 4 genetic clusters, plotting only accessions well assigned into a group (a), or plotting all the individuals (b). Clusters were inferred solely from the genotypic information brought by 13 microsatellite loci. Source: (Ronfort, et al. 2006)