GENETIC VARIATION AND SEXUAL SYSTEM EVOLUTION IN THE ANNUAL MERCURIES

DARREN J. OBBARD

THE QUEENS COLLEGE - TRINITY TERM 2004

ABSTRACT

The *Mercurialis annua* L. (Euphorbiaceae) species complex comprises a group of closely related lineages that present a wide range of sexual-systems, making it a valuable model for the study of plant sexual-system evolution. Within this polyploid complex, diploid populations are dioecious, and polyploid populations either monoecious or androdioecious (males coexist with functional hermaphrodites). The primary aim of this thesis was to use patterns of genetic diversity to elucidate the evolutionary origin and maintenance of the sexual-system diversity in *M. annua*.

The phylogeny of the *M. annua* complex was reconstructed using chloroplast and ITS DNA sequence. This, in conjunction with morphometric analysis, showed that both hexaploid *M. annua*, and a novel species from the Canary Islands (newly described here as *Mercurialis canariensis*), were allopolyploid in origin. Such an origin for hexaploid *M. annua* suggests that androdioecy may have been able to arise in this group as a consequence of hybridisation between a monoecious lineage, tetraploid *M. annua*, and a dioecious lineage, *M. huetii.*

Artificial crosses were used to show that hexaploid *M. annua* has disomic marker inheritance, and a statistical approach was developed to quantify genetic diversity and differentiation in polyploids with disomic inheritance. Strong gradients in genetic (allozyme) diversity at a pan-European scale were used to infer the existence of separate glacial refugia for dioecious and monoecious races of *M. annua*, at the eastern and western ends of the Mediterranean basin, respectively.

A metapopulation model had previously been proposed to explain the ecological maintenance of androdioecy in *M. annua*. Here, population-level patterns of genetic diversity were used as an indirect test of this model. The discovery of lower within-population diversity, and of greater genetic differentiation between populations, for monoecious populations than for androdioecious populations was consistent with the metapopulation model, and suggests that androdioecy is maintained by the occurrence of regular local extinction.

ACKNOWLEDGEMENTS

Within evolutionary biology, it seems traditional to start every large piece of text with a quote from Mr. Darwin. Here is mine:

> "*What a very difficult thing it is to write correctly, I am only just beginning to feel my own inaccuracies.*"
>
> (Darwin 1837)

This was written with regard to the editing of a manuscript, and, whilst in my case there are as many inadequacies as inaccuracies, it is a sentiment that resonates.

First and foremost I thank my supervisors, John Pannell and Stephen Harris. They have been unendingly friendly, helpful and enthusiastic (particularly in their criticism of my grammar). They have also put in far more work than I could ever have expected, consistently making themselves available whenever I had a crisis, a query, or merely a new gel photo. Many members of the Pannell and Harris labs have provided valuable help and advice, and within the department of Plant Sciences several people deserve mention for specific contributions, these include Alison Strugnell, Anne Sing, Colin Hughes, Donovan Bailey, John Baker, Richard Buggs, Rosemary Wise and Sarah Eppley. Vast numbers of people have supplied me with *Mercurialis* samples, including Avi Golan, Eirene Williams, Ettore Pacini, Fred Topliffe, Gill Campbell, Luis Fontes, Mark Harris, Pavel Hamouz, Phillip Schlüter, Renate Wesselingh, Tean Mitchell and Tim Rayden. I thank Yong-Ming Yuan, Cecelia Duraes and Salvador Talavera for providing me with particularly hard-to-get samples. Hiroyoshi Iwata and Patrick Meirmans helped me with their software, Johannes Vogel assisted at my initiation into the black art of polyploid isozymes, and Gil McVean helped me with the (marginally less black) art of coalescent simulation. I have been paid for by The Queens College (Oxford) and the Department of Plant Sciences, and The Genetics Society contributed to the cost of fieldwork.

On a personal note, I wish to thank my parents, both generally, for their long-term emotional and financial encouragement over twenty years of full-time education, and specifically, for sacrificing their holiday to the discovery of *Mercurialis canariensis*. Finally, I want to express my everlasting love and gratitude to Dr. Elizabeth Bayne, who now has a psychological allergy to *Mercurialis annua* very nearly as strong as my physical one.

TABLE OF CONTENTS

TABLE OF FIGURES

CHAPTER 2

CHAPTER 4

CHAPTER 5

APPENDICES

1 GENERAL INTRODUCTION

1.1 BACKGROUND

The sexual system, including both breeding system and gender, is one of the most important aspects of a species' biology. It influences evolution in many other traits, and affects patterns of gene flow within and between populations. It thus has implications for the genetic structure of species, and thereby affects the rate and extent of adaptive evolution. Sexual-system evolution in flowering plants is of special interest, because plants display a wide range of sexual system diversity, from self-fertilization through outcrossing hermaphroditism, to complete separation of the sexes (Darwin 1877; Lewontin 1974; Richards 1997; Holsinger 2000; Barrett 2002). This diversity allows an understanding of sexual-system evolution to be reached by making comparisons between groups that differ in the presence of combined and separate sexes, or in the rate of self-fertilisation. However, because differences in sexual-system are often associated with variation in other traits, it can be difficult to attribute associations to sexual-system alone.

The challenge posed by co-variation in other traits can be partly overcome by studying groups in which closely related lineages differ in sexual system, but are otherwise similar in terms of morphology and life history. The *Mercurialis annua* (Euphorbiaceae) species-complex presents both outcrossing dioecious populations, and self-fertile monoecious (functionally hermaphroditic) populations (e.g. Durand and Durand 1985). In addition, populations in which hermaphrodites coexist with pure males are common (Durand 1963; Pannell 1997c). The extreme rarity of this sexual system (androdioecy), combined with the presence of both monoecy and dioecy, has made the *M. annua* complex an important model for the evolution and ecology of plant sexual-systems (Pannell 2002; Pannell *et al.* 2004). For example, studies of sex-allocation and sex ratio have been used to confirm theoretical predictions regarding androdioecy (Pannell 1997c, 1997b). However, many questions remain. In particular, the phylogenetic origin of sexual-system variation within this species-complex is not known, and the maintenance of sexual-system variation is not fully understood. The

primary aim of this study was to address these questions by utilising the information available from patterns of genetic variation.

1.2 SEXUAL-SYSTEM AND POLYPLOID VARIATION IN *MERCURIALIS ANNUA*

Mercurialis annua L. (Euphorbiaceae) is a wind-pollinated annual, almost exclusively confined to anthropogenic habitats (Durand 1963). It is most commonly found as a horticultural weed, at roadsides, and on recently disturbed waste ground (Durand 1963; Kohout and Hamouz 2000). The species' natural range covers northern Europe and the Mediterranean basin, extending southward to the North African coast and eastward as far as Iraq (Hutchinson 1959; Tutin *et al.* 1968); it is also naturalised in North America, South Africa, and parts of Asia (Durand 1963).

Across the majority of this range *M. annua* populations are dioecious, but in Iberia and western North Africa, populations are either monoecious or androdioecious (i.e. males co-occurring with functional hermaphrodites) (Durand 1963; Pannell 1997c). Males are morphologically distinct, as they bear flowers on long pedunculate inflorescences; in females and hermaphrodites the flowers are usually subsessile, being borne on short pedicels in the leaf axils.

The variation in sexual system coincides with variation in polyploid level (Thomas 1958; Durand 1963; Durand and Durand 1992). Dioecious populations of *M. annua* are always diploid, while monoecious populations vary in polyploid level between tetraploid and 12-ploid. These polyploid races are largely allopatric and geographically isolated. Hexaploids are widespread, occurring all across southern Iberia and in North Africa, tetraploids are limited to southern Morocco, octoploids to Tunisia, and higher polyploid levels to Corsica and Sardinia. Androdioecy is limited to hexaploid populations (Durand 1963; Durand and Durand 1992).

Dioecy in *M. annua* has been a point of scientific interest since the existence of gender in plants was first recognised (Camerarius 1694), and dioecious *M. annua* has been widely used for experiments on sex determination and sex expression (Yampolsky 1919, 1930; Gabe 1939; Yampolsky 1957; Durand *et al.* 1987; Hamdi *et al.* 1987; Durand and Durand 1991; Yang *et al.* 1998; Khadka *et al.* 2002). Recently *M. annua* has been proposed as a model for interactions between polyploidy and the

sexual system (Pannell *et al.* 2004, included as Appendix 8.15, page 233). The occurrence of androdioecy in hexaploid *M. annua* makes it particularly valuable as a model for studying the ecological and evolutionary aspects of this rare sexual system (Pannell 1997c, 1997b, 1997a; Pannell 2001; Pannell 2002).

Many specific questions regarding sexual-system and polyploid evolution are posed by the variation seen in *M. annua*. These include questions regarding: (1) the mode of sex determination and expression in different polyploid races; (2) the joint effect of polyploidy and breeding-system on the level of inbreeding depression; (3) the occurrence and significance of hybridisation between polyploid races; (4) the adaptive significance of sexual-system and polyploid variation; and (5) the phylogenetic relationships between the races. Here I focus on evolutionary questions related to the origin and maintenance of sexual-system variation in *M. annua*.

1.3 SEXUAL-SYSTEM EVOLUTION AND GENETIC VARIATION

The first question I address in this thesis is the phylogenetic origin of sexual-system variation in *Mercurialis annua* (Chapter 2). It is well established that monoecy is derived from dioecy in the genus *Mercurialis* (Krahenbuhl *et al.* 2002), but the origin of androdioecy in *M. annua* remains obscure. It is possible that androdioecy evolved from monoecy, as considered by early theoretical models (Charlesworth and Charlesworth 1978; Charlesworth 1984), or that androdioecy evolved directly from dioecy, as appears to be the case in other androdioecious species (reviewed in Pannell 2002).

Shared genetic variation can be used to reconstruct the relationships between lineages, and, in the context of plant sexual-system evolution, molecular phylogenies have been successfully used to identify ecological traits that correlate with the sexual-system (Vamosi *et al.* 2003), and to establish the origin of differences in the sexual system (Weller *et al.* 1995; Weller and Sakai 1999; Renner and Won 2001). By using multiple loci to reconstruct the phylogeny, it is also possible to identify reticulate relationships between lineages, such as those resulting from hybrid speciation or introgression (e.g. Popp and Oxelman 2001; Hughes *et al.* 2002). Additionally, by comparing uniparentally and biparentally inherited markers, it is possible to identify the maternal and paternal parents of a hybrid lineage (e.g. Palmer *et al.* 1983). In

closely related groups, the recognition that relationships can be reticulate is essential, and other data sources such as genome size, chromosome number, and morphology are valuable in doing this. In *M. annua*, this multiple gene-tree approach suggests that hybridisation may have been important in the origin of sexual system diversity (Chapter 2).

The second question I address is the maintenance of sexual system variation across the geographic range of *M. annua*. Both ecological and genetic factors can be important in selecting for one sexual system over another, and key factors include the level of inbreeding depression (Lande and Schemske 1985; Schemske and Lande 1985; Charlesworth and Charlesworth 1987; Cheptou and Mathias 2001), the need for reproductive assurance (Baker 1955; Barrett and Shore 1987; Pannell and Barrett 1998), and the relative gain in fitness received from reallocating resources between genders (e.g. Charnov *et al.* 1976; Bawa 1980; Brunet 1992). In *M. annua*, a metapopulation model has been proposed to explain the selective maintenance of androdioecy (Pannell 2001), and this model can be extended to explain the full range of sexual-system variation seen (Chapter 5). The model makes predictions regarding the rates of local extinction and migration under different sexual systems, allowing patterns of genetic diversity to be used as an indirect test (Chapter 5). However, the way in which genetic markers are inherited is altered by polyploidy, and this can affect the level and distribution of genetic diversity. Thus, before inferences regarding population structure can be made, the mode of marker inheritance needs to be established (Bever and Felber 1992; Ronfort *et al.* 1998).

Establishing the inheritance pattern shown by nuclear genes in hexaploid *M. annua* is the primary aim of Chapter 3. The mode of inheritance in polyploids depends upon the degree of differentiation between the duplicated genomes. If there is no differentiation between genomes, as when a single genome has been duplicated (autopolyploidy), inheritance will be different to cases in which there are genetically distinct genomes, either through hybridisation (allopolyploidy) or subsequent evolution (diploidisation) (reviewed in Ramsey and Schemske 2002; Jenczewski *et al.* 2004). The standard approach to inferring the mode of inheritance is to use Chi-squared tests on progeny arrays to exclude some models of inheritance (e.g. Soltis and Soltis 1988; Krebs and Hancock 1989; Rieseberg and Doyle 1989; Wolf *et al.* 1989). However, this is statistically weak, and an alternative approach has recently been proposed (Olson 1997; Ridout *et al.* 2001). Unfortunately, this approach requires many

potential parental genotypes to be enumerated, making it difficult to apply by hand to highly polyploid systems (Ridout *et al*. 2001). In Chapter 3, I introduce a computer program to apply both statistical approaches to the analysis of self-fertilised progeny, and I use this to establish the mode of marker inheritance in hexaploid *M. annua*.

Polyploidy can in itself be an impediment to quantifying patterns of genetic diversity (reviewed in Chapter 4). For example, whilst there are many computer programs available to calculate summary statistics in diploids, there are very few which can analyse polyploid data. The problem in analysing polyploid data stems primarily from the presence of more than two alleles at each locus, making it difficult to interpret banding patterns in terms of genotypes. Thus, although software is available to analyse genotypic polyploid data (Hardy and Vekemans 2002), there is no standard approach to summarising genetic diversity using non-genotypic data. In Chapter 4, I review the various *ad hoc* approaches that have been used, and propose a new measure of diversity. I compare this summary statistic to earlier ones in terms of its stability to different levels of polyploidy, and the amount of useful information that is lost because genotypes cannot be scored.

In Chapter 5, I use patterns of genetic diversity to test the metapopulation model of sexual-system maintenance in *M. annua* (see below). Patterns of genetic diversity depend strongly on factors such as migration between spatially structured populations (reviewed by Charlesworth *et al.* 2003), the rate of self-fertilization (reviewed by Charlesworth and Pannell 2001; Charlesworth 2003), and past fluctuations in population size (reviewed by Pannell and Charlesworth 2000). This allows diversity patterns to act as an indirect test of demographic models. However, patterns of genetic diversity result not only from ongoing population processes, but also from historic ones (Charlesworth *et al*. 2003). In particular, climate change has repeatedly confined temperate species to small refugia, from which they later spread. For many species this range-expansion has had a major impact on the level and large-scale distribution of genetic diversity (reviewed by Hewitt 1999). Therefore, in testing demographic models with genetic diversity data, this post-glacial range-expansion must be accounted for.

Briefly, the model of sexual system dynamics in *M. annua*, suggests that regular recolonisation in a metapopulation context selects for reproductive assurance, maintaining self-fertile hermaphrodites, but that males are able to invade established populations that have lower rates of self fertilization (see Chapter 5). According to this model, monoecious populations result from more recent recolonisation than

androdioecious populations, making the testable prediction that monoecious populations will have lower genetic diversity, and higher genetic differentiation than androdioecious ones. In Chapter 5, the background information available from chapters 3 and 4 is combined with genetic diversity data and a survey of population size and abundance (data provided by S. M. Eppley and J. R. Pannell) to infer post-glacial range expansion, and to provide an indirect test of the metapopulation model of sexual-system variation in *M. annua.*

In this thesis, my approach to understanding sexual-system evolution in the *M. annua* complex uses genetic variation across a range of geographic and temporal scales. First, genetic variation between species and between sub-specific races is used to infer the phylogenetic origin of sexual-system variation. Second, pan-European gradients in genetic diversity are used to infer the occurrence of post-glacial range expansion and the location of separate refugia for the two sexual-system races. Third, population-level diversity and inter-population genetic differentiation is used to support the hypothesis that sexual-system variation in *M. annua* is maintained by population-turnover in a metapopulation context. This integrated phylogenetic, phylogeographic and population-based genetic approach, helps to provide an evolutionary explanation of the origin and maintenance of sexual-system variation in the *M. annua* species complex.

2 THE PHYLOGENETIC ORIGIN OF SEXUAL-SYSTEM VARIATION

2.1 INTRODUCTION

An evolutionary transition from one sexual system to another is one of the most profound biological changes a lineage can undergo, as the result will affect the context, rate, and direction of future adaptation in many other traits. Our understanding of such transitions, e.g. from combined to separate sexes, comes primarily from theoretical (reviewed in Charlesworth 1999) and population-level studies (reviewed in Webb 1999). However, as robust and highly resolved gene trees become increasingly easy to obtain, more information, such as the frequency and direction of the sexual-system transitions over evolutionary time, is available from phylogenetic analysis (reviewed in Weller and Sakai 1999; Renner and Won 2001; Vamosi *et al.* 2003). This combined approach is helpful in understanding the processes that give rise to sexual-system transitions, and in explaining the correlation between sexual systems and other ecological traits.

In the flowering plants, which are particularly rich in sexual-system diversity (Richards 1997; Barrett 2002), the mapping of polymorphic breeding systems onto phylogenetic trees has been successfully used to infer the direction of changes in sexual-system within families and genera, as well as at the specific level (reviewed in Weller and Sakai 1999). It has also been successfully employed to find the frequency of transitions between combined and separate sexes at higher taxonomic levels, such as in the monocotyledons (Weiblen *et al*. 2000), and to test correlations between ecological traits and combined or separate sexes in major angiosperm clades (Vamosi *et al*. 2003). However, the sexual system can be a highly labile trait, and a lack of resolution in higher-level phylogenetic studies may lose fine detail of sexual-system variation within taxa. Consequently, studies within groups that display variation in gender dimorphism at, or below, the species level may provide the most information about sexual-system evolution.

Although sub-specific groups that differ in gender dimorphism are likely to be highly informative for research on the evolution of combined and separate sexes, there are relatively few of them. Some of those that do display both monoecy and dioecy have been studied extensively, e.g., *Cotula* spp. (Lloyd 1975a, 1975b), *Ecballium elaterium* (Costich and Meagher 1992), *Sagittaria latifolia* (Dorken *et al.* 2002; Dorken and Barrett 2004), *Mercurialis annua* (Durand and Durand 1992; Pannell 1997c), as have other groups with gender variation such as sub-dioecy and gynodioecy, *Wurmbea dioica* (Case and Barrett 2001, 2004), *Schiedea globosa* and *S. salicaria* (Sakai *et al*. 1989; Sakai and Weller 1991; Weller *et al*. 1998), *Hebe subalpina* and *H. strictissima* (Delph 1990a, 1990b). However, amongst these model taxa, most studies have investigated ecology or population genetics; very few have been explicitly phylogenetic (but see Weller *et al.* 1995; Krahenbuhl *et al.* 2002).

Of the species displaying variation in gender dimorphism, *Mercurialis annua* is particularly worthy of detailed phylogenetic study because it not only presents both monoecious and dioecious populations (Durand 1963; Durand and Durand 1985), but also a high proportion of populations containing both functional hermaphrodites (monoecious individuals) and pure males (Pannell 1997c). This sexual system, described as androdioecy, is exceptionally rare, probably because of the very high siring success needed by males if they are to coexist with hermaphrodites (Charlesworth 1984; Pannell 2002). The only plant in which androdioecy has been the subject of explicit phylogenetic investigation is *Datisca glomerata*, where androdioecy appears to have evolved from dioecy (Rieseberg *et al*. 1992; but see also Swensen *et al*. 1998). However, there is good circumstantial evidence that androdioecy is derived from dioecy in other androdioecious plants; e.g., *Schizopepon bryoniaefolius* (Akimoto *et al*. 1999), *Castilla elastica* (Sakai 2001) and *Spinifex littoreus* (Connor 1996). The preponderance of dioecy in the genus *Mercurialis* similarly suggests a dioecious origin for androdioecy in *M. annua* (Pannell 2001; Krahenbuhl *et al.* 2002). However, phylogenetic analysis of sexual-system evolution in *Mercurialis* has so far only been applied to samples from dioecious and monoecious races of *M. annua*; while the origin of monoecy from dioecy at the level of the genus is now well-established, the analysis has not yet been extended to include samples from an androdioecious lineage (Krahenbuhl *et al*. 2002).

Models for the evolution of androdioecy from both monoecy (Charlesworth and Charlesworth 1978; Vassiliadis *et al*. 2000) and dioecy (Pannell 2001; Wolf and

Takebayashi 2004) have been proposed, but because there are very few androdioecious species available, opportunities to test between the alternative pathways have been rare (but see Rieseberg *et al*. 1992). The *M. annua* complex displays all three sexual systems, so that a sufficiently well resolved phylogeny of the annual mercuries should allow us to distinguish between the evolution of androdioecy from monoecy, and the evolution of androdioecy from dioecy, in this group.

Unfortunately, well-supported and well-resolved phylogenetic trees, which are a prerequisite of this phylogenetic approach, are most difficult to obtain in closely related lineages. This is essentially because character differentiation (e.g. DNA sequence) is reduced, and lineage relationships are more likely to be genuinely reticulate. In particular, the occurrence of polyploidy, which is widespread in the genus *Mercurialis* and particularly in the *M. annua* complex, may complicate the inference of phylogeny (e.g. Popp and Oxelman 2001; Hughes *et al.* 2002; Raymond *et al.* 2002). To overcome this, the presence of intra-individual variation in biparentally inherited (nuclear) loci can be used to identify reticulate relationships (e.g., Palmer *et al.* 1983; Soltis *et al.* 1991; Soltis *et al.* 1995), and targeted sequencing can be used to identify the presence of intra-individual variation in DNA sequences such as ITS, which may also be indicative of hybridisation (e.g. Popp and Oxelman 2001; Hughes *et al.* 2002; Rauscher *et al.* 2002). Additional evidence for hybridisation and lineage relationships in closely related species complexes can be gained from genome size comparisons (e.g. Ohri 1998; Horandl and Greilhuber 2002; Levin 2002), morphometric analysis (reviewed in Rieseberg and Ellstrand 1993), and allozyme frequencies (e.g. Crawford 1989; Raybould *et al.* 1991; Brochmann *et al.* 1992; Wendel *et al.* 1992).

Here I resolve the evolutionary relationships between the annual mercuries. Through the combined use of nuclear ITS (5.8S ribosomal DNA with the flanking regions ITS1 and ITS2) and plastid DNA sequence I am able to provide a robust phylogeny for the inference of sexual-system evolution within the *M. annua* complex. This phylogenetic approach, in conjunction with allozyme data, genome size measurements, direct chromosome counts, and morphometric analysis strongly suggests that allopolyploid hybridisation has played an important role in the polyploid and sexual-system diversification of the annual mercuries, and is implicated in the origin of androdioecious hexaploid *M. annua* and the dioecious tetraploid Tenerife mercury.

2.2 MATERIALS AND METHODS

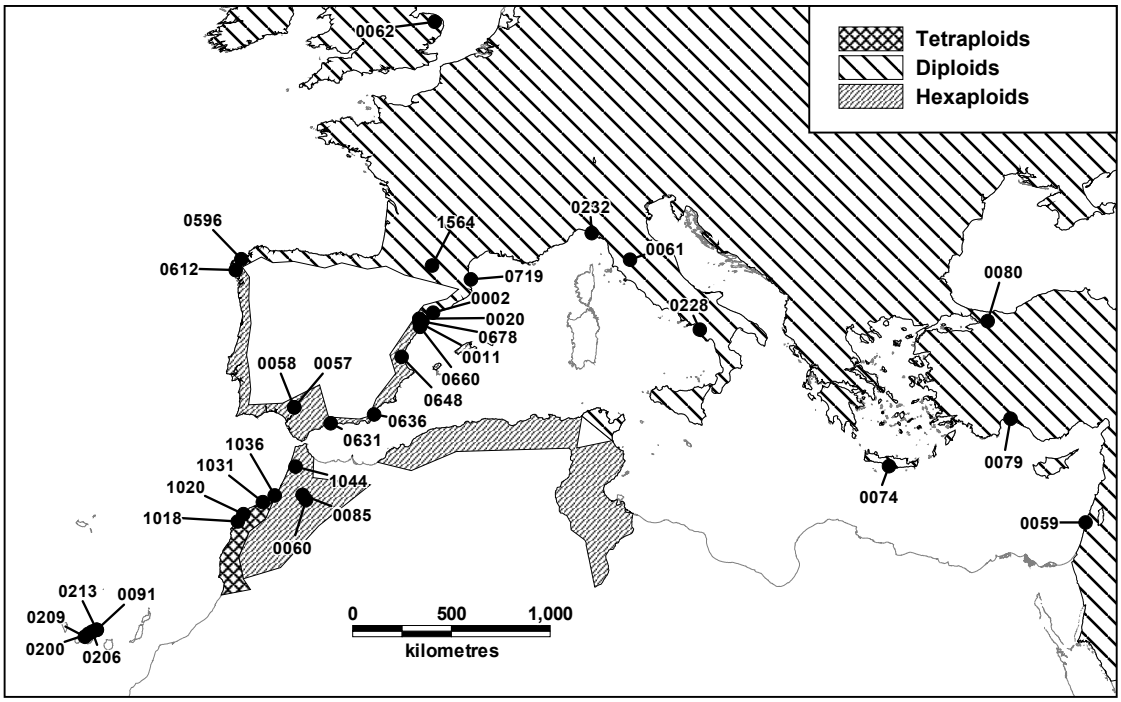
2.2.1 Study Species

All species of *Mercurialis* are wind pollinated, and the majority are dioecious (Table 2.1); all except *M. leiocarpa* are native to Europe and the Mediterranean basin (Tutin *et al.* 1968; Krahenbuhl *et al.* 2002). There are several weedy annual lineages, two of which are widely recognised as distinct species: *M. annua* and *M. huetii* (Table 2.1). *Mercurialis huetii* is morphologically similar to *M. annua*, but smaller, and restricted in distribution (Durand 1963; Durand and Durand 1985). *Mercurialis annua* comprises a polyploid complex (2x-12x) of ruderals (Durand 1963). In the past this complex has been spilt into as many as three taxa, according to polyploid level and sexual system (e.g. Durand 1963): *M. annua sensu stricto*, *M. monoica* (Moris) Durand and *M. ambigua* L. *sensu* Durand (1963). However, although dioecy in the *M. annua* complex is limited to the diploids, and monoecy is limited to the polyploids, vegetative differentiation is slight (Durand and Durand 1985, and this chapter). Because it is difficult to distinguish morphologically between the polyploid levels some authors choose to lump these polyploid lineages into one taxon (Tutin et al. 1968); *M. annua sensu lato*. Here I follow the latter example, and refer to all members of the complex as *M. annua*, noting polyploid level where necessary. In addition to *M. huetii*, and the *M. annua* complex, in which dioecy is limited to diploids and monoecy to polyploids, a previously unknown dioecious polyploid annual species has been recently identified from Tenerife, hereafter denoted "Tenerife mercury" (Appendix 8.4, page 172).

2.2.2 Sampling

Annual mercury samples were selected to cover the largest possible geographic and morphological range (Figure 2.1, Sample details in Appendix 8.8). More than 4,400 individuals, from *ca.*100 different populations across the entire natural range were screened for isozyme variation (Appendix 8.8, page 195). For sequence-based phylogenetic analysis DNAs from a subset of these populations were

Figure 2.1: Distribution of the annual mercuries in Europe, with the origin of samples used in this chapter



Modified from Durand (1963); the north African distribution is approximate, and depends strongly on sampling bias. Hatched areas show the distribution of *M. annua* diploids, tetraploids and hexaploids; a complex polyploid swarm up to 12x occurs in Corsica and Sardinia (not shown). *M. huetii* occurs in the southeast of France and north east of Spain. Numbers refer to populations used in the morphometric, DNA content, and phylogenetic studies (Appendix 8.8).

Table 2.1: *Mercurialis* species

M. annua L.	2x *(x=8)* [i]	Dioecious	Annual
	4x—12x *(x=8)* [i]	Monoecious / Androdioecious	Annual
M. spec. nov. (Tenerife)	4x *(x=8)*	Dioecious	Annual
M. huetii Hanry.	2x *(x=8)* [i]	Dioecious	Annual
M. perennis L.	6x—12x *(x=8)* [ii]	Dioecious	Rhizomatous perennial
M. ovata Sternb.& Hoppe.	2x—4x *(x=8)* [ii]	Dioecious	Rhizomatous perennial
M. leiocarpa Sieb.& Zucc.	2x *(x=8)* [ii]	Monoecious	Rhizomatous perennial
	6x *(x=8)* [ii]	*(Not reported)*	Rhizomatous perennial
M. elliptica Lam.	*2n=42, 2n=220* [iii]	Dioecious	Woody Perennial
M. corsica Cosson.	*2n=66* [iii]	Dioecious	Woody Perennial
M. tomentosa L.	*2n=26* [iii]	Dioecious	Woody Perennial
M. reverchonii Rouy.	*2n=26* [iii]	Dioecious	Woody Perennial

[i](Durand 1963) [ii](Krahenbuhl and Kupfer 1995) [iii](Krahenbuhl *et al*. 2002)

extracted from glasshouse-grown plants: (i) diploid *M. annua* (populations 0002, 0059, 0061, 0080), (ii) tetraploid *M. annua* (1018, 1020, 1031), (iii) hexaploid *M. annua* (0011, 0020, 0058, 0060), (iv) Tenerife mercury (0091, 0200, 0209) and (v) *M. huetii* (0678, 0719). Where the same population appears in both the ITS and cpDNA analyses, the same individual was used. Silica-dried leaf material was used for DNA extractions from woody perennial *M. elliptica*, *M. perennis* and *M. tomentosa,* whilst *M. reverchonii* DNA was extracted from herbarium specimens. Based on morphology, the most likely sister groups to the genus *Mercurialis* are *Leidesia* and *Seidelia* (Pax 1914; Webster and Rupert 1973). However, as these were unavailable I followed Krahenbuhl *et al*. (2002) in using *Ricinus communis* as an outgroup.

The morphological analysis used 26 population samples in the following five groups: (i) diploid *M. annua* (1564, 0002, 0596, 0062, 0074, 0079, 0228, 0232), (ii) tetraploid *M. annua* (1018, 1020, 1031) (iii) hexaploid *M. annua* (0012, 0620, 0631, 0636, 0648, 0660, 0085, 1036, 1044), (iv) *Tenerife mercury* (0200, 0206, 0209, 0213) and (v) *M. huetii* (0678, 0719) (Appendix.8.8). The individuals were not those in the DNA analyses, but were grown from the same site-collected seed bulks. With the exception of population 0232, genome-size measurements were made on all the populations used in the morphometric analysis.

2.2.3 DNA content

DNA content was measured by flow cytometry (FC) on three plants from each population. Approximately 15mg of leaf material from each plant was used, along with 5 mg *Lycopersicon esculentum* cv. 'Gardener's Delight' leaf as an internal standard. Leaf material was chopped with a razor blade in 1 ml ice-cold 'LB01' lysis buffer and staining solution, modified from Dolezel *et al*. (1989): 15 mM Tris base, 2 mM Na_2EDTA, 0.5 mM spermine tetrahydrochloride, 80 mM KCl, 20 mM NaCl, 0.1% (v/v) Triton X-100, 15mM ß-mercaptoethanol, 50 µg ml^{-1} propidium iodide and 50 µg ml^{-1} RNAse. The resulting suspension was filtered through 30 µm mesh 'CellTric' disposable filter (Partec GmbH, Münster, Germany) and analysed with a Becton Dickinson FACScan flow cytometer. For each sample, 5000 events were recorded in each of five runs and mean peak values were evaluated using CellQuest software (Becton Dickinson). The coefficient of variation for each peak used was above 2.00 and below 5.00. For each run, the mean DNA content of the sample peak was

calculated with reference to *L. esculentum*, which has a DNA content of 4.10 pg (Bennett and Leitch 2003). Linearity of the flow cytometer was confirmed using a suspension of 2.0 x 10^7 chicken erythrocyte nuclei per millilitre fixed in ethanol-PBS (Biosure®, Grass Valley, California) diluted 1:10 in a staining solution consisting of calcium- and magnesium-free Dulbecco's PBS, 0.05 mg ml^{-1} propidium iodide and 0.6% (v/v) Nonidet P40. All C-values will be submitted to the RBG Kew C-value database on publication. [DNA content measurements were made by RJA Buggs.]

2.2.4 Chromosome counts

Where possible, population polyploid level was inferred from FC measurement of DNA content in conjunction with mating-system and geographic location (using the extensive survey of Durand 1963). When direct chromosome counts were used in addition to FC (e.g. for tetraploids and the Tenerife samples) 2-5 mitotic cells were used. Seeds were germinated in the dark on moist filter paper at room temperature. Root tips, 1-3cm long, were harvested, and treated with 0.2% (w/v) colchicine for 3 hours, then fixed in ethanol:acetic acid (3:1) for a minimum of 24 hours. The fixed root tips were incubated at 60°C for 10 mins in 1M HCl, and then stained with Schiff's reagent, and examined using standard procedures. [Counts were made by SA Harris]

2.2.5 DNA isolation, polymerase chain reaction and sequencing

The nuclear ITS region (the 5.8S ribosomal RNA with flanking transcribed spacers ITS 1 and ITS2) and two chloroplast sequences; *trn*L-*trn*F (the non-coding intergenic spacer between *trn*L and *trn*F) and *mat*K-*trn*K (protein-coding sequence at the 3' end of the *mat*K gene, and the non-coding intronic region between *mat*K and the 3' exon of *trn*K) were sequenced for phylogenetic analysis. These sequences were selected for their utility in phylogeny reconstruction at the species level, and the availability of near-universal primer sequences (Taberlet et al. 1991; Gielly and Taberlet 1994; Hsiao et al. 1994).

DNA was extracted from fresh and dried leaf material according to the modified CTAB procedure of Doyle and Doyle (1987). Leaves were ground under liquid nitrogen in microcentrifuge tubes, and after incubation in 2X CTAB at 65°C the samples were purified with two chloroform:isoamyl alcohol (24:1) extractions. Following propan-2-ol precipitation at -20°C, samples were washed in 76% ethanol,

Table 2.2: PCR primer sequences and reaction conditions

*trn*L-*trn*F non-coding region		35 cycles of 1min at 94°C, 1min at 53°C and 2mins at 72°C.
e[i]:	5'-GGT TCA AGT CCC CTC TAT CCC-3'	
f[i]:	5'-ATT TGA ACT GGT GAC ACG AG-3'	
*mat*K-*trn*K 5' intron		3min at 95°C; 35 cycles of 1min at 95°C, 1min at 50°C and 1min at 72°C; 7mins at 72°C
trnK-2R[ii]:	5'-CCC GGA ACT AGT CGG ATC-3'	
1908F[ii]:	5'-GGC ATC CCA TTA GTA AGC-3'	
ITS 'universal'		40 cycles of 1min at 97°C, 1min at 48°C and 2mins at 72°C; 7mins at 72°C.
ITSL[iii]:	5'-TCG TAA CAA GGT TTC CGT AGG TG-3'	
ABI102[iv]:	5'-TAG AAT TCC CCG GTT CGC TCG CCG TTA C-3'	
ITS *M. huetii* specific (1a)		2min at 97°C; 30 cycles of 30s at 97°C then 60s at 74°C; 3min at 74°C.
Huet60MMF[v]:	5'-TCC GCG CCC CTC ATT CTC **CTG** **A**CG a**G**-3'	
ABI102[iv]:	5'-TAG AAT TCC CCG GTT CGC TCG CCG TTA C-3'	
ITS *M. huetii* specific (1b)		2min at 97°C; 30 cycles of 30s at 97°C, 35s at 69°C and 35s at 72°C; 3min at 72°C
ITSL[iii]:	5'-TCG TAA CAA GGT TTC CGT AGG TG-3'	
Huet468MMR[v]:	5'-AAC ATA **A**AT TTT GGG CCA ACC ACA TGa **A**-3'	
ITS Tenerife exclusion (2a)		2min at 97°C; 30 cycles of 30s at 97°C then 60s at 74°C; 3min at 74°C
Ann47MMF[v]:	5'-**T**AG T**C**G GG**T** GAA TTT GTG GCT **CC**a C-3'	
ABI102i[v]:	5'-TAG AAT TCC CCG GTT CGC TCG CCG TTA C-3'	
ITS Tenerife exclusion (2b)		2min at 97°C; 30 cycles of 30s at 97°C, 35s at 66°C and 35s at 72°C; 3min at 72°C.
ITSL[iii]:	5'-TCG TAA CAA GGT TTC CGT AGG TG-3'	
Ann459MMR[v]:	5'-**C**MG ACG GCT **A**AG AAC AGC GCA CGt **C**-3'	

[i]Taberlet *et al.* (1991), [ii]Lavin *et al.* (2000), [iii]Hsiao *et al.* (1994), [iv]Lola Lledo (Royal Botanic Gardens Kew), [v]Designed by D.J.O. These primers are positioned internally to the 'universal' ITS primers, and amplify ITS from *M. huetii* but not *M. annua* (1 a and b) and *M. annua* but not Tenerife mercury (2a and b), (see main text for details). The bases marked in bold are mismatches to the excluded template (*M. annua* and Tenerife mercury, respectively), and bases in lower case are mismatches to both the excluded template and the target templates (see Appendix 8.2 page 163 for the effect of technique).

dried, and re-suspended in water. Samples were stored at -20°C until needed.

For all PCRs the reagents were as follows: 2.5 μM each of dATP, dTTP, dGTP and dCTP, 0.8 μM of each primer, 1Unit of *Taq* DNA polymerase and 10-100 ng DNA. Reaction volumes were 25 μl or 50 μl. All primer sequences and reaction conditions are shown in Table 2.2. Following amplification PCR products were checked for homogeneity on an agarose gel and then purified using 'QIAquick' purification spin columns (Qiagen Ltd) according to the manufacturer's instructions. Sequencing reactions were performed in both directions for each PCR product using BigDye^{TM}version 3.1 (Perkin Elmer), according to the manufacturer's protocol. In all cases the sequencing reactions utilised the same primers as the amplification PCRs. The sequencing products were analysed using an ABI Prism DNA Sequencer 3730.

2.2.6 Identification of heterogenous ITS types

The presence of two divergent ITS types within an individual can be indicative of a hybrid origin (Popp and Oxelman 2001; Hughes et al. 2002; Rauscher et al. 2002). However, direct sequencing of ITS using 'universal' primers, may fail to identify all the sequences present, e.g. because concerted evolution has reduced the copy number of one type (Wendel et al. 1995). A particularly sensitive method of identifying such 'hidden' sequences, is the use of specific primers designed to amplify the ITS region of a putative parent, but exclude the ITS sequence initially found using universal primers (for a discussion of this approach see Rauscher *et al*. 2002). To test the hypotheses that *M. huetii* might be a parent of polyploid *M. annua*, and that *M. annua* might be a parent of the Tenerife mercury, specific forward and reverse primers covering the 3' end of ITS1, 5.8S RNA, and 5' end of ITS2 were designed for two purposes; 1a and b (Table 2.2) to amplify *M. huetii* ITS DNA, but exclude *Mercurialis annua* sequences, and 2a and b (Table 2.2) to amplify diploid *Mercurialis annua* ITS DNA, but exclude the Tenerife tetraploid sequences initially amplified by the universal primers. The specific primers were positioned internally to the universal primers, providing two overlapping sequences that covered the full length of the ITS1-ITS2 region. Following Cha *et al.* (1992), specific primers were designed for maximum stringency by ensuring mismatches with the excluded sequences were present at the most 3' position in the primer sequence, and incorporating an additional mismatch in

the penultimate 3' base. All primer sequences and reaction conditions are shown in Table 2.2.

2.2.7 Sequence editing and analysis

Primary sequence editing was done using Sequencher (Gene Codes Corporation). Preliminary sequence alignments were made using clustalX 1.81 (Thompson 1997), and these were refined by hand using BioEdit (Hall 1999). Indel characters were coded following the "simple gap coding" method of Simmons and Ochoterena (2000) using the program "gap-coder" (Young and Healy). The aligned sequences will be submitted as a PopSet to GenBank on publication. Parsimony analyses were done using PAUP* (Swofford 2002), using a heuristic search with tree bisection and reconnection. Bootstrap replicates had a limited search time imposed (500 seconds). Bayesian analyses were done using MrBayes 3.0 (Ronquist and Huelsenbeck 2003). For all analyses, chloroplast sequences *trn*L-*trn*F and *mat*K-*trn*K were concatenated to form a single combined matrix, whilst the ITS region was analysed separately. Bayesian analysis used a General Time Reversible (GTR) model with gamma distributed rate variation. Starting trees were random, with all other substitution-model and sampling parameters set to the MrBayes 3.0 default values. Combined chloroplast data were divided into five regions with unlinked model parameters: *trn*L-*trn*F, non-coding *mat*K-*trn*K, and each codon position within the *mat*K coding region. ITS and 5.8S data were analysed as a single set, allowing a proportion of sites to be invariable. For chloroplast data, the Markov chain was observed to converge after *ca*. 10,000 steps, and samples were taken every hundredth step for 450,000 steps after the first 50,000 steps. For ITS data, the Markov chain was observed to converge after *ca*. 50,000 steps, and samples were taken every hundredth step for 500,000 after the first 100,000 steps. Multiple runs converged on similar consensus trees.

2.2.8 Isozymes

Only four isozyme systems were informative in polyploid *M. annua*: AAT (aspartate aminotransferase E.C. 2.6.1.1), PGI (glucose-6-phosphate isomerase E.C. 5.3.1.9), 6-PGD (two loci, phosphogluconate dehydrogenase E.C. 1.1.1.44), and IDH (isocitrate dehydrogenase E.C.1.1.1.42). Protocols and solutions were adapted from Wendel and Weeden (1990). For each individual, approximately 2 cm^2 of fresh young

leaf tissue was ground with extraction buffer (1.21 g Tris-HCl, 0.04 g EDTA, 0.076 g KCl, 0.2g $MgCl_2.6H_20$, 4 g PVP, 0.5 g PVPP in 100 ml stock, made up with 90:10:0.5 stock:DMSO:ß-mercaptoethanol) on a pre-chilled ceramic block. The resulting paste was soaked into 3mm by 8 mm Whatmann No. 3 paper wicks for loading into starch gels, made from 14% w/v hydrolysed potato starch (StarchArt Corporation, Texas) and 0.25% w/v sucrose. Three buffer systems were used. Lithium-borate gels (electrode buffer: 0.039 M LiOH, 0.263 M boric acid; gel buffer: 33 mM Tris-base, 5 mM citric acid, 4 mM LiOH, 30 mM boric acid, pH 7.6, lithium borate), stained for AAT and PGI. Morpholine-citrate gels (electrode buffer: 0.04 M citric acid, 0.068 M N-(3-aminopropyl)-morpholine pH 6.4; gel buffer: 1:14 dilution of the electrode buffer), were stained for 6-PGD. Tris-citrate gels (electrode buffer: 0.135 M Tris-base, 0.03 M citric acid, pH 8.3; gel buffer: 1:19 dilution of the electrode buffer), stained for IDH. All gels were run at 4 ºC for approximately 6 hours. Lithium-borate gels were run at 300 V, morpholine-citrate gels at 250 V, and Tris-citrate at 150 V. Gels were cut into 1 mm thick slices and stained at 40 ºC. Staining solutions were adapted from Wendel and Weeden (1990). Reactions were stopped after staining, and gels were stabilised by removing stain solution and adding 30 ml 25% v/v glycerol. Gels were scored either from photographs or at the time of staining.

2.2.9 Morphometrics

Plants from 26 populations were grown in a glasshouse in Oxford from the 1st August 2003 to approximately the 1st October 2003. Seeds were sown in single-population seed trays, and within a few days of germination, seedlings were selected at random and planted into individual pots. Plants were grown in a climate-controlled glasshouse at 25 °C with 16h daylight, watered as necessary. Pots were positioned in eight randomised blocks, each population being represented once in each block. Blocks were re-randomised at intervals of four to five days. Disease reduced the total sample size to 179 plants.

The following morphological characters were measured for each plant (indices refer to leaf nodes on the main stem, with the cotyledons denoted as node zero): plant height, dry-mass, length of each of the first five internodes, length of petioles two to four, length of leaf three, width of leaf three, length of branches number zero to three, length of internodes zero and one for branches zero and one, diameter of internode

one, length of stipules at node five, length of peduncles (if present), flower-bearing peduncle length, proportion of bi- tri- and tetra-capsulate female flowers. Areas and perimeters were calculated from the scanned images of five leaves per plant (from nodes one to five).

The software package "Shape" (Iwata and Ukai 2002) was used for all leaf-shape analysis. Elliptic Fourier descriptors were calculated from chain-codes of leaf perimeters and normalised by the longest axis, with some correction by eye. As there is no reason to expect a heritable consistent left-right asymmetry in *Mercurialis* leaves, asymmetric components of leaf shape were not used in the analysis (Iwata et al. 1998). The symmetric coefficients were used in a Principal Component Analysis (PCA), and Principal Component Scores (PCS) for the first five components (those that each explained > 0.5% of the variation, 98.4% in total) were calculated for each leaf. These scores were then used as measures of leaf shape in all later analyses. This process is a highly efficient way of capturing subtle information about leaf shape, and thereby making it available to standard statistical analysis (Iwata and Ukai 2002).

Multivariate analysis was in two parts: (1) a clustering analysis to identify 'natural' groups, and (2) a Discriminant Function Analysis (DFA) to find whether the 55 vegetative characters above can be used to support an *a priori* distinction between the five groups. In analysis (1), z-score scaled population-mean morphological data were used in a clustering analysis using within-group linkage of squared Euclidian distances. For (2) all vegetative characters were used in a principal component analysis (PCA), and overall morphological principal component scores (PCS) on each of the first 21 principal components were calculated (90.4% of the variation). These PCS were then used in the DFA, specifying the following five groups: *M. huetii*, Tenerife mercury, diploid *M. annua*, tetraploid *M. annua*, and hexaploid *M. annua*. PCS were used instead of the raw data because DFA assumes normality, equal variance, and low-correlation between measures (Manly 1986), all of which were better satisfied by the PCS (data not shown). All statistical analyses were performed using SPSS (SPSS for Windows, release 11.0.0 © SPSS Inc.) and Minitab (Release 12.1 © Minitab Inc.)

2.3 RESULTS

2.3.1 DNA content and ploidy

In the annual mercuries DNA content was highly correlated with ploidy: diploid, tetraploid, and hexaploid samples of *M. annua* formed an almost linear series (Figure 2.2). Diploid *M. annua* had a 4C DNA content between 2.62 pg and 2.65 pg (99% confidence interval of the grand mean), i.e. about 3.8 times larger than *Arabidopsis thaliana*. Tetraploid *M. annua* had a 4C value between 5.14 pg and 5.19 pg, whilst the hexaploid *M. annua* 4C value was between 7.73 pg and 7.80 pg. *Mercurialis huetii* had a 4C DNA content between 2.82 pg and 2.86 pg, significantly larger than diploid *M. annua* (Figure 2.2). Tenerife mercury is tetraploid (2n=32), but had a significantly larger 4C value (6.43pg to 6.46pg, Figure 2.2) than tetraploid *M. annua*. [These measurements are unpublished data provided by RJA Buggs.]

2.3.2 Heterogenous ITS types

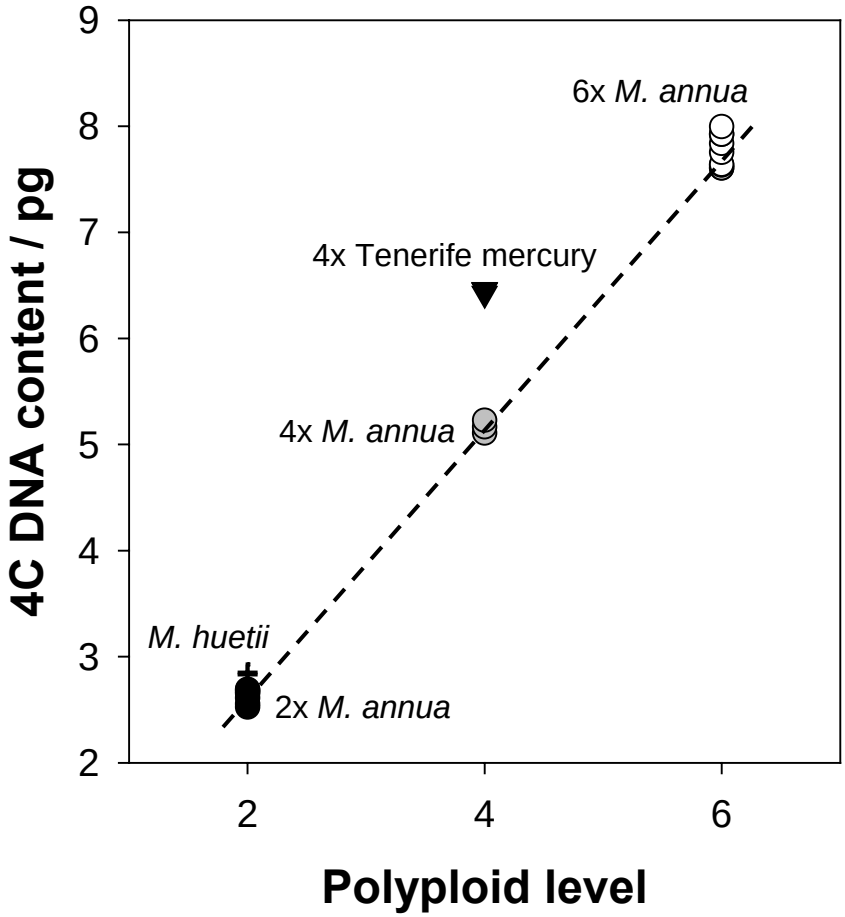
Direct sequencing from the PCR products obtained using universal primers did not find any evidence of multiple ITS sequences, e.g. consistent double-peaks in both sequencing directions. However, the use of specific targeted primers to amplify putative parental sequences identified heterogeneous ITS types in two of the annual mercury groups, hexaploid *M. annua* and the Tenerife mercury (hereafter denoted 'secondary', purely to distinguish them from those sequences obtained using universal primers). Under the chosen reaction conditions, neither set of taxon-specific primers amplified a detectable product from controls containing only the alternative template (Appendix 8.2, page 163). *M. annua*-specific primers amplified a divergent ITS sequence from dioecious polyploid Tenerife accessions, that are nested within a strongly-supported group comprising all of the *M. annua* accessions (Polyploid *M. annua* in Figure 2.3). *Mercurialis huetii* -specific primers amplified an additional ITS sequence identical to that found in *M. huetii* from hexaploid *M. annua* (*M. huetii* in Figure 2.3). No *M. huetii*-like ITS sequence was amplified from tetraploid *M. annua* (for figures, see Appendix 8.2 page 163).

2.3.3 ITS phylogeny

The total aligned sequence, including the 5.8S subunit of nrRNA, was 774 bp. There were no indels in the amplified 5.8S nrDNA sequences, and the only substitution in 5.8S subunit of nrDNA was shared by all the woody perennial species in the analysis (woody perennial species, Figure 2.3). The absence of indels in the 5.8S region indicates that none of the sequences are evolving as pseudogenes. In the alignment of ITS1 and 2 there are 30 indels of 1-3 bp, one of 9 bp, and one of 13 bp. Of these, nine are parsimony-informative. The exclusion of these gap characters from the analysis does not qualitatively alter the results of the analysis, but does slightly alter support for some nodes (see Appendix 8.3.2, page 170). Although it does not affect any conclusions regarding sexual system and polyploid evolution within the annual group, the alignment between *Mercurialis* and the outgroup *Ricinus* was problematic in the spacer regions and ought to be treated with some caution.

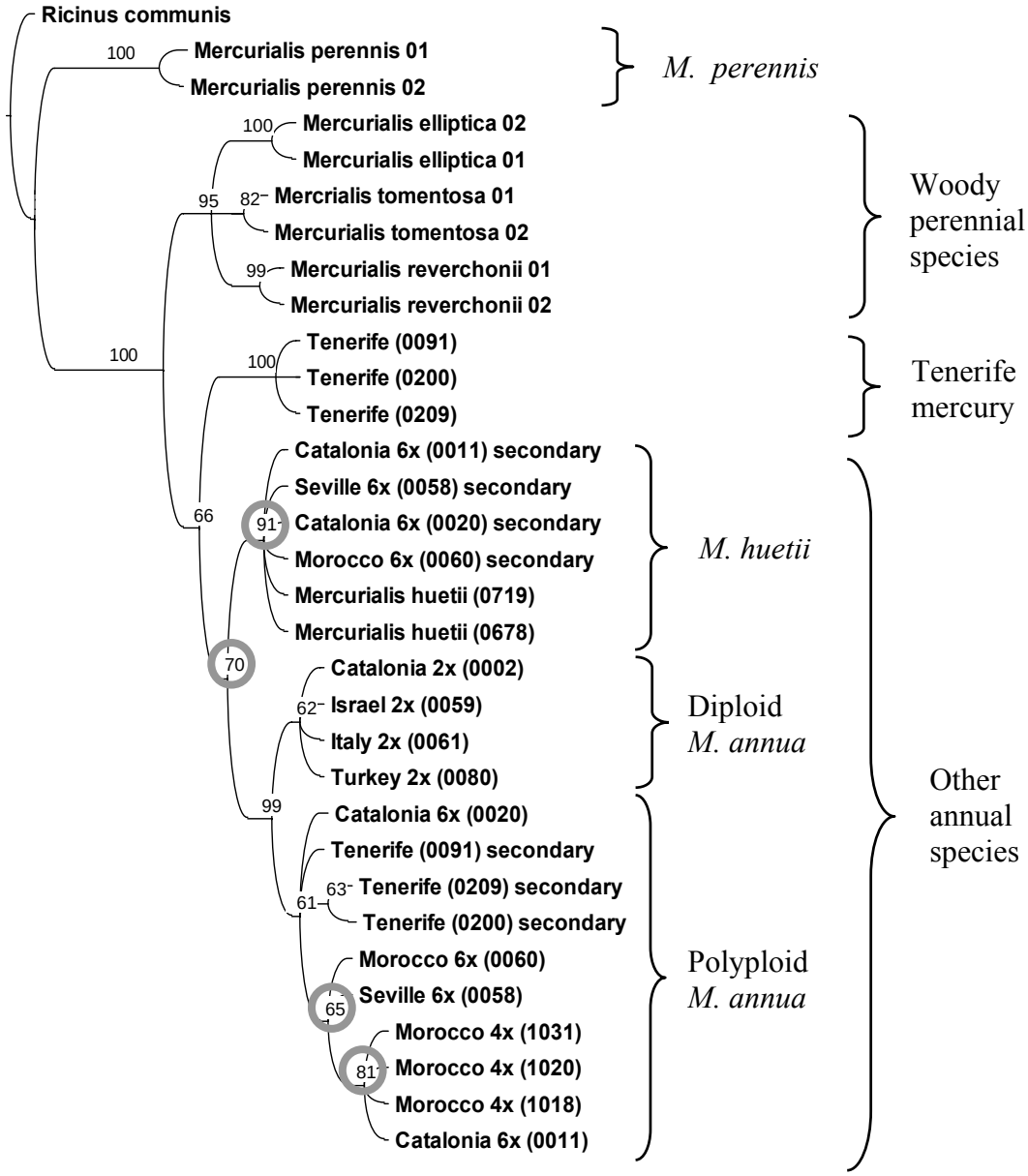
Parsimony and Bayesian analyses of ITS data resulted in gene trees that were identical in all important respects (Figure 2.3, and Appendix 8.3.2, page 170). All clades that received high bootstrap support also had a high posterior probability, although some with a relatively low bootstrap support in the parsimony analysis were strongly supported by the Bayesian analysis, e.g. the grouping of *M. annua* and *M. huetii* had a posterior probability of 0.96 but only 70% bootstrap support. The gene tree inferred from ITS data showed the annual and woody-perennial species of *Mercurialis* form a monophyletic group, sister to the *Mercurialis perennis* group (Figure 2.3). Within the annual and woody-perennial clade, there were three major clades whose relationships to each other were not strongly supported (woody perennial species, Tenerife mercury, and the other annual species in Figure 2.3). Within the annual species, the group consisting of *M. huetii* and identical hexaploid *M. annua* ITS sequences amplified by *M. huetii*-specific primers, was sister to *M. annua* (Figure 2.3). Within the *M. annua* clade, the diploids and polyploids (including the Tenerife mercury 'secondary sequences') formed distinct groups, and although bootstrap support was low for these groups, some had Bayesian support, e.g. the *M. annua* diploids (Figure 2.3, posterior probability = 0.91).

Figure 2.2: Genome sizes of the annual mercuries



Mean genome size (4C-value/pg) for the populations used in the morphometric analysis. Standard errors are too small to show. All five labelled groups are significantly different from each other in DNA content (Tukey test). Although diploid, tetraploid, and hexaploid *M. annua* appear to form a linear series, both polyploids have smaller genomes than would be expected based exclusively on summations of diploid *M. annua*, and/or *M. huetii*. Tenerife mercury has a much larger genome size than would be expected if it resulted exclusively from recent autopolyploidisation within *M. annua*. [This figure is drawn from unpublished data provided by RJA Buggs].

Figure 2.3: ITS gene tree for *Mercurialis*



[See next page for legend]

Figure 2.3: ITS gene tree for Mercurialis

[See previous page for figure]

The strict consensus of 12 equally most parsimonious trees based on the 111 parsimony-informative characters (after indel-coding) from the combined ITS1-5.8s-ITS2 dataset (Consistency index = 88, Retention index = 94). All samples identified by source location and ploidy are *M. annua s.l.* Sequences denoted 'secondary' are those divergent sequences only amplified by specific primers. Numbers above nodes indicate bootstrap support (as a percentage) based on 585 replicates. Terminal polytomies are sets of identical sequences [identical sets are as follows: (1) *M perennis* samples, (2) all Tenerife 'primary' sequences, (3) Tenerife 'secondary' sequences 0091 and 0209, (4) *M. reverchonii* 01 & 02, (5) *M. elliptica* 01 & 02, (6) *M. huetii* and hexaploid *M. annua* secondary ITS, (7) diploid *M. annua*, (8) hexaploids 0060 & 0058, (9) polyploids 1031, 1020, 1018 and 0011]. Nodes circled in grey are those that have low bootstrap support in the parsimony analysis, but in the Bayesian analysis are found to have greater than 95% posterior probability (for the Bayesian tree see Figure 8.12, page 171).

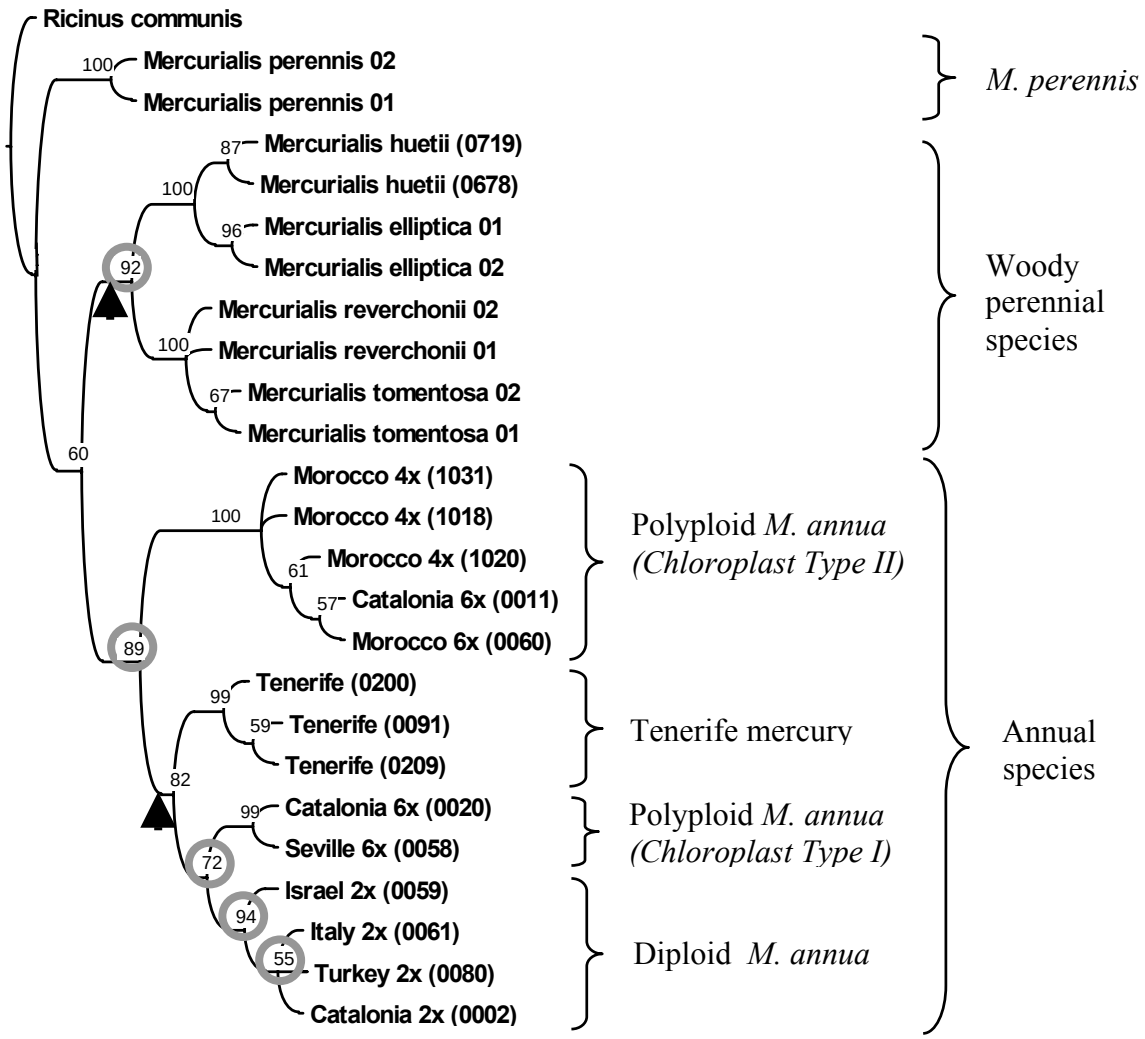
2.3.4 Chloroplast phylogeny

The total aligned length of *trn*L-*trn*F was 474bp, and that of *mat*K-*trn*K was 664bp. There were 26 indels in *trn*L-*trn*F dataset and 16 in the non-coding part of *mat*K-*trn*K; 13 of the cpDNA indels were associated with poly-T regions. Despite the large number of indels, alignment was straightforward. The amplification of *mat*K-*trn*K from *Mercurialis reverchonii* and *M. tomentosa* failed, and the sequences were treated as missing data for all analyses. Parsimony and Bayesian analyses of combined *trn*L-*trn*F and *mat*K-*trn*K chloroplast data resulted in gene trees that were identical in all important respects (Figure 2.4 and Appendix 8.3.1, page 169). All clades that received high bootstrap support also had a high posterior probability, although some clades with a relatively low bootstrap were still strongly supported by the Bayesian analysis (Figure 2.4). The exclusion of the 32 parsimony-informative gap characters from the analysis did not qualitatively alter the results, but it did decrease support for some nodes (see Appendix 8.3.1, page 170). The phylogenetic positions of the two largest indels (91 bp and 106 bp) are marked by black arrows on Figure 2.4. Although cpDNA data resolved the *M. perennis* clade as being sister to the other included *Mercurialis* species, support for the monophyly of the other species in the analysis was low (Figure 2.4, 52% bootstrap 0.91 posterior probability). Each of the three main clades (*M. perennis*, woody perennial species, annual species, Figure 2.4) was individually well supported. Importantly, the most parsimonious tree inferred from chloroplast data differed from the ITS tree in the position of hexaploid *M. annua* and Tenerife mercury (compare Figure 2.3 and Figure 2.4), and *M. huetii*, which appeared as sister to *M. elliptica* within the woody perennial clade (Figure 2.4).

2.3.5 Isozymes

Isozyme banding patterns could be interpreted in terms of the alleles present or absent, but allele-dosage could not be assessed (Appendix 8.9, page 197 for example gels and interpretation). The Tenerife mercury was distinct in terms of the allozymes present, with four alleles not seen in any of the other groups. Tetraploid and hexaploid *M. annua* shared several alleles with each other that were not present in diploid *M. annua* (e.g. *Pgi*-1^{050}, *Pgi*-1^{136}, 6*Pgd*-1^{073}, 6*Pgd*-2^{114}, Figure 2.5) and some that were additionally shared with *M. huetii* (e.g. *Aat*-1^{075} *Pgi*-2^{122} 6*Pgd*-1^{080}, Figure 2.5).

Figure 2.4: Chloroplast gene tree for *Mercurialis* based on combined *trn*L-*trn*F and *mat*K-*trn*K data



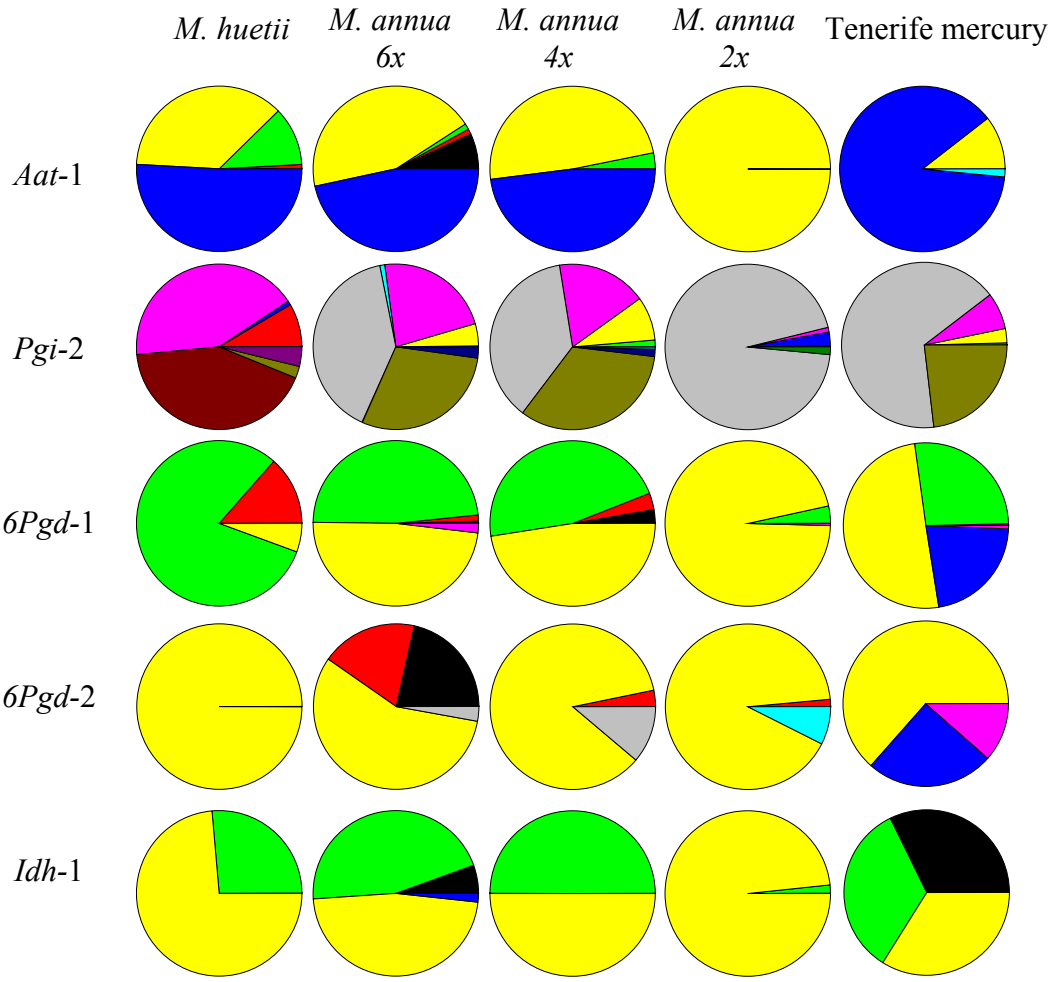
[See next page for legend]

Figure 2.4: Chloroplast gene tree for Mercurialis *based* on combined trnL-trnF *and* matK-trnK

[See previous page for Figure]

The strict consensus of 4 equally most parsimonious trees based on the 125 parsimony-informative characters (after indel-coding) from the combined chloroplast dataset (Consistency index = 84, Retention index = 94). All of the terminals identified by source location and polyploid level are *M. annua s.l.* The tree includes several pairs of identical sequences, as follows: *M. perennis* 01&02, *M. reverchonii* 01& 02, *M. annua* 1031&1018, 0011&0060, 0020&0058. Numbers above nodes indicate bootstrap support (as a percentage) based on 1000 replicates. Nodes circled in grey are those that have low bootstrap support, but in the Bayesian analysis are found to have greater than 95% posterior probability. Black arrows mark the positions of two large deletions in *trn*L-*trn*F, one of 91 bp, and one of 106 bp.

Figure 2.5: The distribution of isozyme alleles in the annual mercuries



Pie charts show which alleles are carried at five isozyme loci by each of the annual mercury groups. As the allele dosage could not be scored in polyploid *Mercurialis* genotypes, the allele frequencies are unavailable. Instead, pie charts show the proportion of visible alleles that were of each type, i.e. if 100 individuals were seen carrying allele only *a*, 50 carrying alleles *a* and *b*, and 5 carrying only allele *b* then 205 alleles were seen, of which 73% were *a* and 27% *b*. If all individuals display fixed heterozygosity for alleles *abc* then 33.3% of alleles seen are of each type.

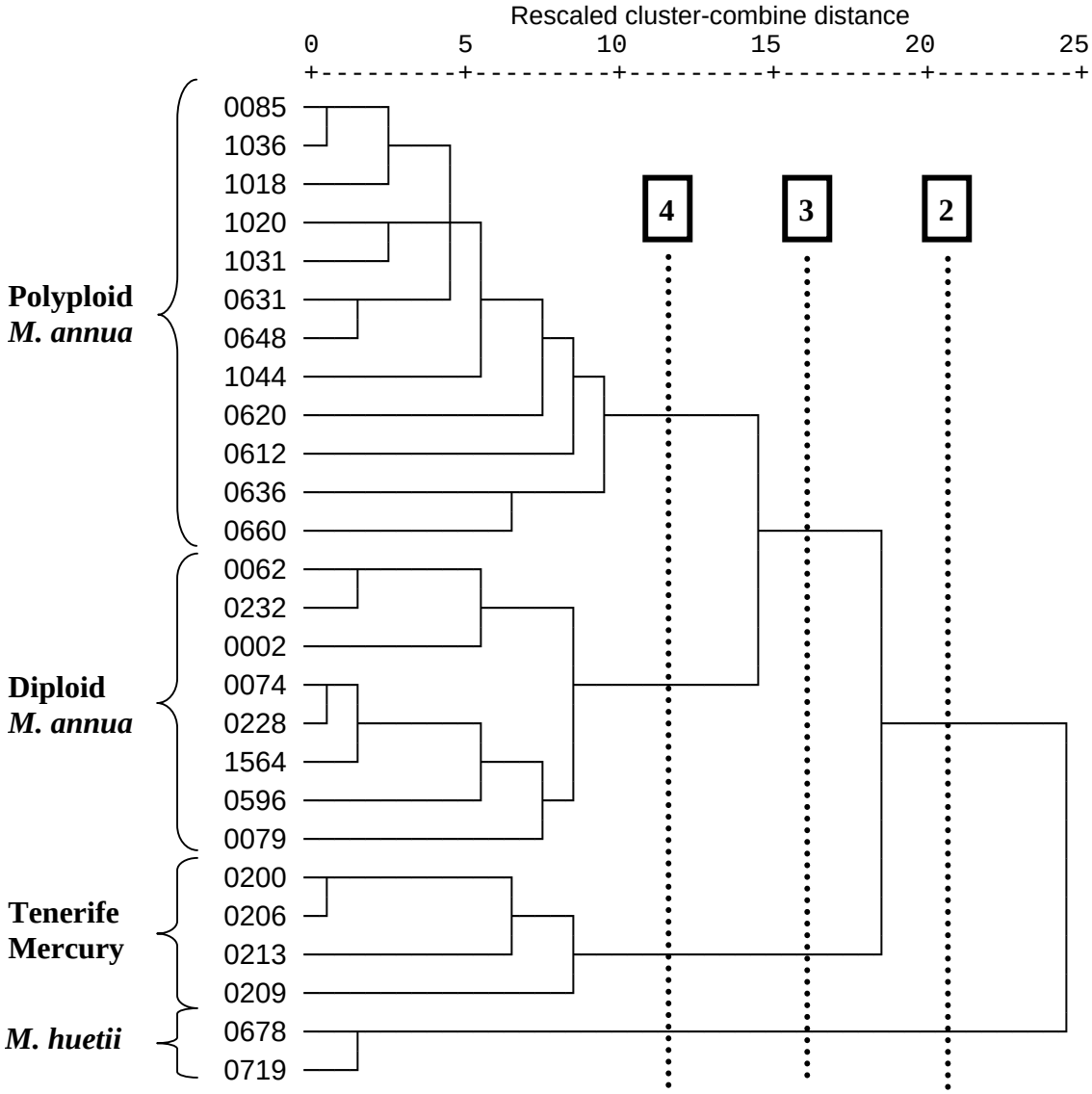
2.3.6 Morphology

Despite the absence of discrete diagnostic characters, *M. huetii* is morphologically distinct from *M. annua* (Durand and Durand 1985). Clustering analysis of glasshouse-grown material grouped the two *M. huetii* populations together (Figure 2.6), and the discriminant functions (DFs) correctly assigned 100% of *M. huetii* individuals. (Figure 2.7 for a plot of the first two DFs). *M. huetii* is found in less anthropogenic environments, and is generally smaller than *M. annua*, with relatively longer branches and smaller leaves (Durand and Durand 1985). These field-based morphological observations were also true for glasshouse-grown material ($p < 0.001$ for each).

The Tenerife mercury is morphologically distinct. In the clustering analysis of glasshouse data, populations clustered together (Figure 2.6) and all individuals were correctly assigned by the canonical discriminant functions (Figure 2.7 for a plot of the first two DFs). Morphological characters associated with Tenerife mercury included; large size, large stipules, male flower bracts, and a high frequency of tricapsulate female flowers (see Appendix 8.4). Of these characters, the presence of male flower bracts and stipules > 4 mm long are diagnostic. For illustrations of plant morphology see Figure 8.13 (page 177), Figure 8.14 (page 179) in Appendix 8.4, plus photographs in Appendix 8.5.

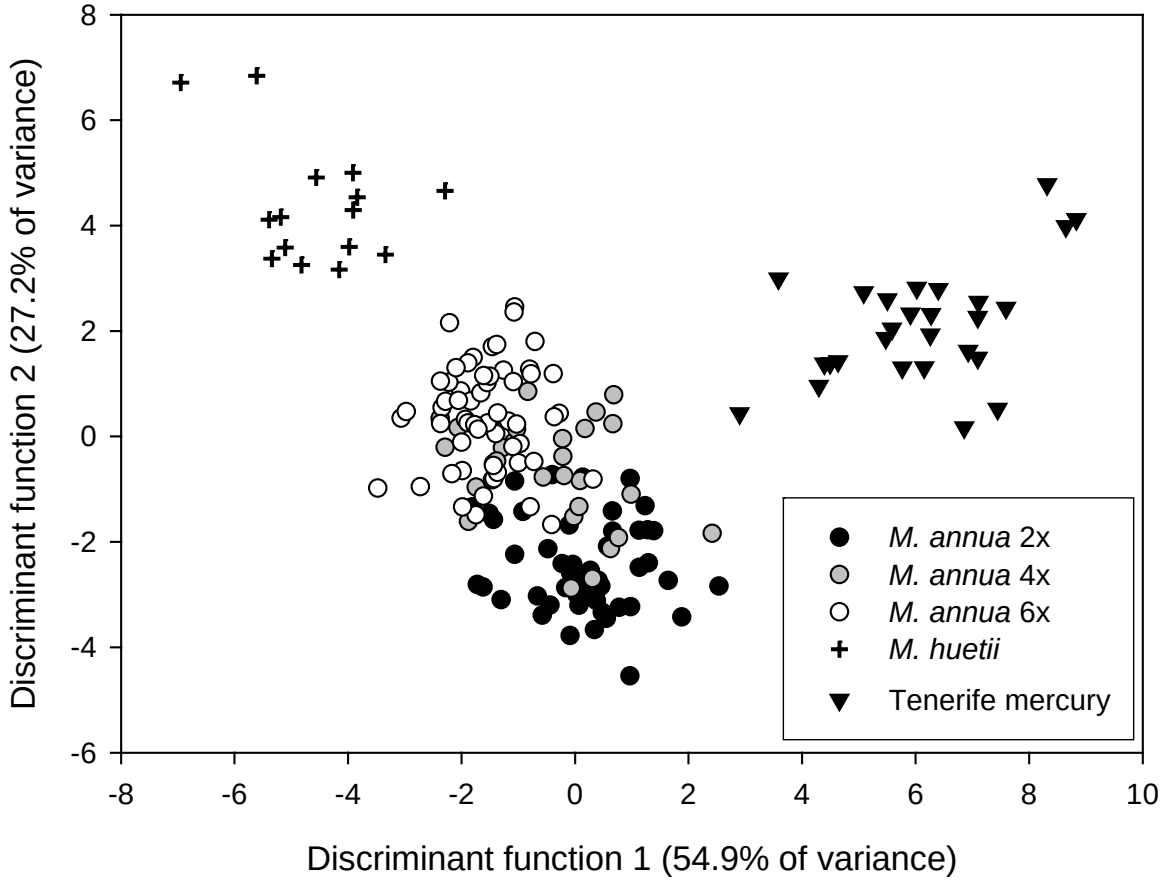
Diploid and polyploid annual mercuries are most easily distinguished by their sexual-system: diploids are always dioecious, and (with the exception of the Tenerife mercury populations) polyploid populations always contain cosexual individuals (Durand 1963; Durand and Durand 1985). In the controlled environment of the glasshouse some subtle but statistically significant vegetative differences were found between diploid and polyploid *M. annua*, e.g. polyploids had rounder leaves (t-test on the first PCS of leaf shape: $p < 0.001$, see Appendix 8.6 for details of leaf-shape) and were about 10% taller ($p < 0.001$). It does not seem to be possible to distinguish between tetraploids and hexaploids purely on grounds of vegetative morphology; direct chromosome counts or DNA content is also required. The clustering analysis did not separate tetraploids from hexaploids, and the discriminant function analysis misclassified 13% of tetraploids as hexaploids, and 10% of hexaploids as tetraploids, despite both having been grown in common, controlled glasshouse conditions.

Figure 2.6: Morphological clustering analysis



The dendrogram shows hierarchical clustering of scaled population-mean morphological data. Clustering is based on average within-group linkage, using squared Euclidian distances calculated from 55 z-score transformed vegetative characters (sample sizes of 5-8 individuals per population). The three most informative classifications divide the dataset into two, three, or four distinct groups respectively (marked with dotted lines). The most divisive classification identifies *Mercurialis huetii*, Tenerife mercury, diploid *M. annua*, and polyploid *M. annua* populations as distinct clusters. No clear morphological distinction between tetraploid and hexaploid *M. annua* is identifiable.

Figure 2.7: Discriminant function analysis of morphological data



Plot of glasshouse grown individuals along the first two discriminant functions (explaining 82.1 % of the variation). Four canonical discriminant functions make a significant contribution. Overall, 92.7% of individuals are correctly assigned using the four discriminant functions, including 100% of individuals from *M. huetii* and Tenerife mercury. The least morphologically distinct pair of groups is tetraploid and hexaploid *M. annua*, between which more than 10% of individuals are misclassified in each direction.

2.4 DISCUSSION

The primary results of this chapter can be summarised as follows: (1) There was intra-individual variation in ITS sequences (Figure 2.3), and ITS and cpDNA gene trees were incongruent for the phylogenetic positions of *M. huetii*, hexaploid *M. annua*, and the Tenerife mercury (Figure 2.3 and Figure 2.4), (2) genome sizes in the *M. annua* polyploid complex formed a linear series increasing with ploidy, but *M. huetii* had a slightly larger genome than diploid *M. annua*, and the tetraploid Tenerife mercury had a very much larger genome than tetraploid *M. annua* (Figure 2.2), and (3) Tenerife mercury and *M. huetii* were morphologically distinct groups, with no significant overlap with the *M. annua* polyploid complex (Figure 2.6 and Figure 2.7), but diploid, tetraploid, and hexaploid *M. annua* were vegetatively very similar to each other (Figure 2.7). Below, I discuss how these results provide evidence for hybridisation and allopolyploidy within the annual mercuries, and what the implications of this are for sexual-system evolution in the genus.

2.4.1 Evidence for hybridisation and allopolyploidy in the annual mercuries

Despite the potential problems of paralogy and intra-individual variation, the ITS regions of nrDNA have been very successfully used for species-level phylogeny reconstruction in plants (Baldwin *et al.* 1995; Buckler *et al.* 1997; Alvarez and Wendel 2003; Bailey *et al.* 2003). Furthermore, differentiated sequences resulting from hybridisation can provide valuable information regarding the parentage of hybrid lineages (e.g. Popp and Oxelman 2001; Hughes *et al.* 2002; Rauscher *et al.* 2002). In the annual mercuries, the identification of two divergent ITS types within hexaploid *M. annua* individuals and tetraploid individuals from Tenerife strongly supports a hybrid origin for each. The absence of indels and very low variation within the 5.8S subunit of nrDNA indicates that none of the sequences included in this study were evolving as pseudogenes (Bailey *et al*. 2003). In contrast to hexaploid *M. annua*, only one ITS sequence was amplified from tetraploid *M. annua* individuals. This may be because (1) divergent sequences resulting from hybridisation have been wholly homogenised by

concerted evolution (Wendel et al. 1995) (2) one sequence fails to amplify because of divergence at the primer sites, or (3) because only one of the sequences was present originally , i.e. tetraploid *M. annua* has an autopolyploid origin.

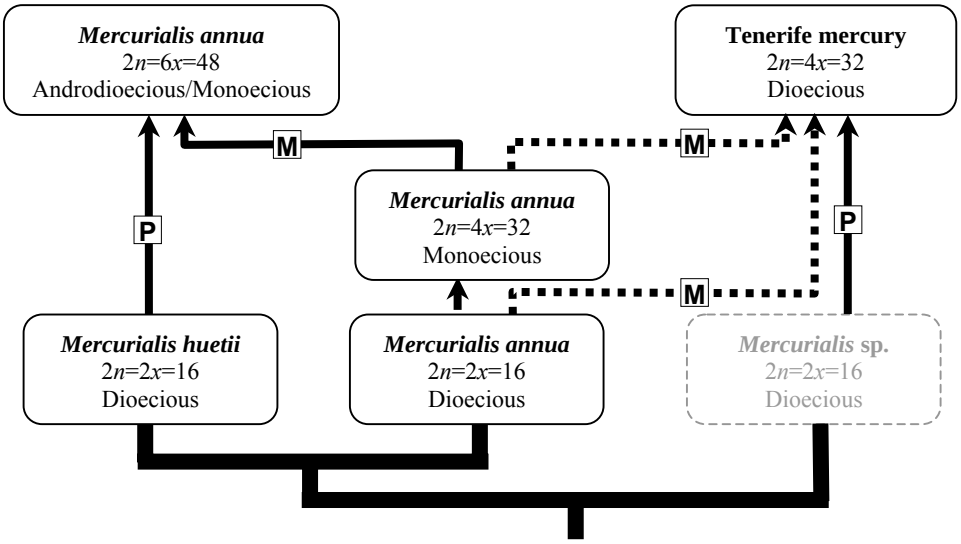
The ITS gene tree (Figure 2.3), including both of the ITS sequence types found in hexaploid *M. annua* individuals, suggests that hexaploid *M. annua* is a hybrid between *M. huetii* and tetraploid *M. annua* (Figure 2.8a). The ITS data allow the possibility that the *M. annua*-like parent was actually a diploid ancestor of tetraploid *M. annua* (though not extant diploid *M. annua*). However, the polyploid level (6x) is more consistent with a diploid-tetraploid cross, followed by chromosome doubling (Figure 2.8a).

The two ITS sequences isolated from each Tenerife mercury individual are phylogenetically distant from each other; one sequence is nested within the polyploid *M. annua* clade and the other is even more distantly related to *M. annua* than is *M. huetii* (Figure 2.3). This second sequence does not resemble any known species of *Mercurialis* (i.e. those examined here or in Krahenbuhl *et al*. 2002), suggesting that one parental lineage of Tenerife mercury is either extinct, or remains undiscovered. The phylogenetic position of the *M. annua*-like ITS isolated from the Tenerife mercury, within the polyploid clade of *M. annua*, suggests that the *M. annua* parent was a polyploid. However, this would require a considerable reduction in chromosome number, as Tenerife mercury is tetraploid. It is more plausible that the *M. annua*-like parent of Tenerife mercury was the diploid progenitor of extant tetraploid *M. annua* and was differentiated from extant *M. annua* (Figure 2.8b).

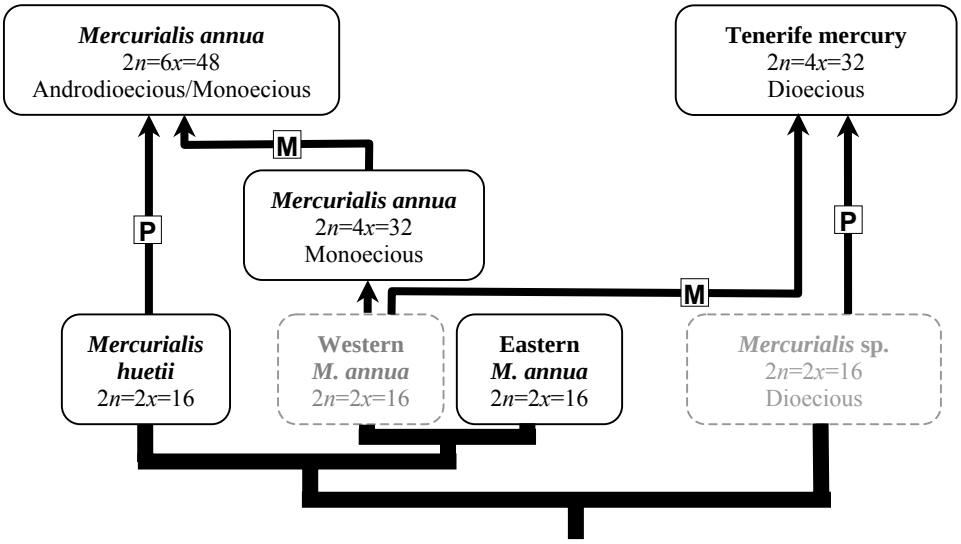
In the most parsimonious cpDNA gene tree, *M. huetii* chloroplast sequences are nested within the woody perennial chloroplast sequences (Figure 2.4). This implies they are only distantly related to the chloroplasts of all the other annual mercuries. Given the similarity in morphology between *M. annua* and *M. huetii* (Durand 1963), and the relative positions of their ITS sequences in the ITS gene tree (Figure 2.3), the best explanation for this incongruence seems to be past chloroplast capture of *M. elliptica*-like chloroplasts by *M. huetii*. Such introgression of cpDNA, not necessarily accompanied by the evident transfer of other markers or morphological traits, is a widely recognised phenomenon in many species (Soltis *et al.* 1996; Cottrell *et al.* 2002; Tsitrone *et al.* 2003), and chloroplast capture between *M. huetii* and *M. elliptica* may have been facilitated by sympatry and the low specificity of wind-pollination.

Figure 2.8: Model for the origin of the annual mercuries

a) Extant diploid *M. annua* as parent of polyploid *M. annua***.**



b) Divergent diploid *M. annua* as parent of polyploid *M. annua.*



[See next page for legend]

Figure 2.8: Model for the origin of the annual mercuries

[See previous page for Figure]

Heavy lines indicate the phylogenetic relationships between diploid species. Thin arrows show polyploidisation and / or hybridisation events; 'M' indicates proposed maternal parentage, and 'P' paternal parentage.

a) Extant diploid *M. annua* as parent of polyploid *M. annua.* —The heterogenous ITS types present in hexaploid *M. annua* show it has an allopolyploid origin between *M. annua* and *M. huetii*, and the hexaploid chromosome complement is consistent with hybridisation between a tetraploid and a diploid, followed by chromosome doubling. ITS data also show *Tenerife mercury* to be allopolyploid in origin, probably a hybrid between *M. annua* and an unknown taxon (dashed grey box). Chloroplast sequence similarity to *M. annua* suggests that *M. annua* was the maternal parent, and chromosome numbers are consistent with both parents being diploid.

b) Divergent diploid *M. annua* as parent of polyploid *M. annua.*— As Figure 2.8a, except for the addition of a second (hypothetical) diploid *M. annua* lineage, slightly divergent from extant diploid *M. annua*. Such a lineage could explain the similarity in allozyme frequencies and ITS sequence between tetraploid and hexaploid *M. annua*, and the Tenerife mercury. In this scenario, the two diploid *M. annua* lineages would represent opposite ends of the species geographic range; the western end giving rise to the polyploid complex, but failing to survive the last glaciation, and the eastern end surviving, giving rise to all extant diploid *M. annua*. Such a hypothesis is consistent with the observed post-glacial migration patterns (Chapter 5).

Chloroplast inheritance in *M. annua* is maternal (Appendix 8.1.1) allowing comparison of the ITS gene tree and the cpDNA gene tree to identify maternal and paternal parents of the hybrid lineages (e.g. Palmer *et al.* 1983; Soltis *et al.* 1991; Soltis *et al.* 1995). The Tenerife mercury cpDNA sequences are nested within the *M. annua* clade (Figure 2.4), suggesting that *M. annua* was its maternal parent (Figure 2.8). Similarly, extant *M. huetii* is unlikely to have been the maternal parent of hexaploid *M. annua* (Figure 2.8). However, as the capture of *M. elliptica*-like chloroplasts by *M. huetii* would have obscured its previous chloroplast type, there is no power to distinguish between the maternal and paternal parents of hexaploid *M. annua*, if the origin of hexaploid *M. annua* predates the chloroplast capture event. It is possible that *M. elliptica*-like chloroplasts are not shared by all extant *M. huetii*, and wider geographic sampling of *M. huetii* might prove informative.

There are two divergent chloroplast types found in polyploid *M. annua* populations (Type I and Type II in Figure 2.4); Type I is similar to chloroplasts from diploid *M. annua*, while Type II is basal to other *M. annua* plastids (with the exception of *M. huetii*). In a survey of cpDNA types (Appendix 8.1.2), Type I was found to be restricted to Iberian hexaploids, and though rarer (26% of surveyed individuals) than Type II, it is widely distributed across Iberia. The more common Type II chloroplasts occur in tetraploids and hexaploids all over their Iberian and Moroccan ranges. The similarity between Type I chloroplasts, and the chloroplasts found in diploid, *M. annua* allows the speculation that Type I may represent a relatively recent chloroplast capture from diploid *M. annua* by hexaploid *M. annua*.

Genome size, morphology, and isozyme data, though certainly not conclusive, support the conclusions from the DNA sequence analysis. *Mercurialis huetii* and diploid *M. annua* have very similar DNA content (Figure 2.2). This means there is little power to distinguish between an auto- and allopolyploid origin of hexaploid *M. annua* based on genome size. However, Tenerife mercury has a much larger genome than would be expected from a sum of *M. annua* and/or *M. huetii* genomes, providing further evidence for an allopolyploid origin for Tenerife mercury, with an unknown taxon as the second parent.

Fixed heterozygosity in some hexaploid *M. annua* populations, and experimental evidence of disomic inheritance in variable populations (see Chapter 3) is consistent with an allopolyploid origin, but is not conclusive evidence as diploidisation generally leads to disomic inheritance, even in autopolyploids (reviewed in Ramsey and

Schemske 2002). Hexaploid *M. annua* shares alleles with both diploid *M. annua* and *M. huetii*, superficially supporting an allopolyploid origin. However, tetraploid *M. annua* contains many alleles seen in hexaploid *M. annua*, but not in diploid *M. annua*, and for some loci (e.g. Pgi-1 in Figure 2.5) frequencies are similar. This suggests an allopolyploid origin for tetraploid *M. annua* as well as hexaploid *M. annua*. However, it may also be the result of on-going gene flow between the polyploids, or, given the ITS sequence data, evidence that the putative diploid ancestor of tetraploid *M. annua* was divergent from extant diploid *M. annua* (Figure 2.8b). This is consistent with the suggestion that extant diploid *M. annua* and polyploid *M. annua* had different glacial refugia (see Chapter 5).

Morphologically, the Tenerife mercury falls well outside the *M. annua* complex, being as distinct as *M. huetii* (Figure 2.6, Figure 2.7 and Appendix 8.4, page 172). There is little vegetative differentiation between diploid, tetraploid, and hexaploid *M. annua* (Figure 2.7), but it is notable that a plot of the first two discriminant functions finds hexaploid *M. annua* to be intermediate between its putative parents: *M. huetii* and tetraploid, or diploid, *M. annua* (Figure 2.7). See Figure 8.18 (page 187) in Appendix 8.5 for a photograph of hexaploid *M. annua* with its putative parents.

It was previously thought that the *M. annua* polyploid complex was exclusively autopolyploid in origin (Durand and Durand 1985, 1992; Krahenbuhl *et al.* 2002). This conclusion was based on extensive artificial hybridisation and colchicine-induced polyploid studies (Durand 1963). Autotetraploids and autohexaploids derived by colchicine treatment of dioecious diploid *M. annua* are partially fertile, and more than 10% of individuals are monoecious. In contrast, allotetraploids created by colchicine treatment of hybrids between diploid *M. annua* and *M. huetii* have very low fertility and segregation of the sexes persists (Durand 1963). However, since some disturbance to sex expression was reported in the artificial allotetraploids (Durand 1963), it is possible that monoecy could have been selected for in either artificial polyploid. A previous ITS-based phylogenetic study of Mercurialis (Krahenbuhl *et al*. 2002) found no evidence of a hybrid origin for hexaploid *M. annua*, despite finding mixed ITS types in putative hybrid individuals between the woody perennial taxa. This can be explained by the low power available to identify multiple ITS types when sequencing using universal primers (Rauscher *et al.* 2002; Alvarez and Wendel 2003).

2.4.2 Implications for sexual-system evolution in *Mercurialis*

The vast majority of *Mercurialis* species are dioecious (Table 2.1) and within the *M. annua* complex, only polyploids are monoecious (Durand and Durand 1992). This provides strong circumstantial evidence that monoecy has been derived from dioecy in this genus (Durand and Durand 1992), a conclusion that has been confirmed by explicit phylogenetic analysis (Krahenbuhl *et al*. 2002). However, many hexaploid *M. annua* populations display the rare sexual system androdioecy (Durand and Durand 1985; Pannell 1997c), thus an alternative scenario would be that monoecy evolved indirectly from dioecy, via androdioecy.

Early theory assumed the opposite pathway; that androdioecy evolves from hermaphroditism, possibly as a stepping-stone on the pathway to dioecy (Charlesworth and Charlesworth 1978; Charlesworth 1984). These models suggested that the invasion of a hermaphroditic population by a male (female-sterile mutant) could only occur under a highly restrictive set of conditions, making the monoecy-androdioecy route unlikely. More recent models invoking pollen limitation, e.g. during colonisation, suggest that the evolution of androdioecy from dioecy through the modification of females may be less restrictive (Pannell 2001; Wolf and Takebayashi 2004). In such a dioecy-androdioecy pathway, males would not have to evolve adaptive outcrossing traits *de novo* (Pannell 2001), and this pathway seems to have been prevalent in nature; *Datisca glomerata* (Swensen *et al*. 1998), *Schizopepon bryoniaefolius* (Akimoto *et al*. 1999), *Castilla elastica* (Sakai 2001), *Spinifex littoreus* (Connor 1996), and *Eulimnadia texana* (Weeks *et al*. 2000).

Mapping characters onto phylogenetic trees to infer the rate, direction, and ecological correlates of evolution in sexual systems (or other traits), assumes that the underlying tree is bifurcating. However, whilst bifurcating trees probably represent the true phylogenetic relationship for the majority of species, there are a large number of plant lineages, such as hexaploid *M. annua*, for which the tree is genuinely reticulate. An allopolyploid origin of hexaploid *M. annua*, between monoecious tetraploid *M. annua* and dioecious diploid *M. huetii*, raises the interesting possibility that androdioecy in *M. annua* initially arose directly as a result of hybridisation; i.e. male-determining factors were inherited from the *M. huetii* ancestor. This sidesteps a need for the re-evolution of specialised male inflorescence morphology, likely to be a considerable barrier since the high siring-success required of males is the favoured

explanation for the extreme rarity of androdioecy (Pannell 2002). An alternative to the hybrid origin of androdioecy, which would also avoid the re-evolution of specialist male traits, would be the re-activation in androdioecious hexaploid *M. annua* of male-inflorescence genes that are unexpressed in the tetraploid, octoploid, and other polyploid lineages. However, such a scenario requires the preservation of unexpressed male sex-determination and sex-expression genes in monoecious races.

Although a hybrid origin for androdioecy remains highly speculative, given the extreme rarity of androdioecy, and how little is known about its origin, such a novel pathway is worthy of more attention. In particular, the model presented here makes predictions that could be used to distinguish between the alternative origins for androdioecy outlined above. A hybrid origin would be favoured if the genes involved in sex determination and male morphology in hexaploid *M. annua* were more closely related to those in *M. huetii* than those in *M. annua*, while a 'dormant males' origin would be favoured if both tetraploid and hexaploid *M. annua* carried close relatives of the diploid *M. annua* sex-determination genes, but they were only expressed in the hexaploid.

Hybrids might be expected to be intermediate between their progenitors, and this is true for many traits (e.g., Rieseberg and Ellstrand 1993), perhaps suggesting adaptive niches for hybrids that are intermediate between those of the parents (e.g. Cruzan and Arnold 1993). However, many species are believed to be hybrid in origin, either homoploid hybrids (Rieseberg 1997) or allopolyploids (Otto and Whitton 2000; Levin 2002), and hybridisation has been implicated as a contributive factor in adaptive radiation (Seehausen 2004). It is now known that interactions between parental genomes can allow hybrid lineages to show more extreme traits than either parent (Lexer *et al*. 2003), so that hybridisation can act as an important source of new variation (Rieseberg *et al*. 2003). As in hexaploid *M. annua*, which apparently displays the rare combination of male and hermaphrodite genotypes because of hybridisation, it appears that hybrid origins are a major source of evolutionary novelty.

3 DISTINGUISHING BETWEEN POLYPLOID GENOTYPES USING SELFED PROGENY

3.1 INTRODUCTION

Polyploidy is a widespread phenomenon, occurring in almost all groups of multicellular organisms. It provides new variation on which selection can act, and may even have been a major source of evolutionary innovation (Otto and Whitton 2000). Amongst plants, it seems to be a significant causal factor in speciation, and it has been implicated in the evolution of many biologically important traits (reviewed in Levin 2002), including ecological range (e.g. Brochmann and Elven 1992), herbivore tolerance (e.g. Nuismer and Thompson 2001) and changes in sexual-system (reviewed in Pannell *et al*. 2004, included as Appendix 8.15, page 233)

Polyploidisation (genome duplication) commonly occurs through the union of unreduced gametes or through somatic doubling in cells that later go on to form the germ-line (Bretagnolle and Thompson 1995; Ramsey and Schemske 1998). It is often associated with hybridisation, as genome duplication offers an escape from the failure of meiosis that occurs when chromosomes lack pairing partners (Sybenga 1975; Ramsey and Schemske 1998). Polyploids formed by the union of divergent genomes (e.g. interspecies hybrids) are described as 'allopolyploid', whilst those formed by genome duplication within a single species are 'autopolyploid'. These different origins are often reflected by differences in the mode of inheritance in the neopolyploid (Ramsey and Schemske 2002).

Autopolyploids have multiple copies of each chromosome (described as homologous chromosomes), and if all pairings of these chromosomes are equally frequent during meiosis then inheritance is described as 'polysomic'. Although homologous chromosomes may exclusively form pairs (bivalents), in many species they also have the potential to form larger groups (multivalents), caused by the arms from a single chromosome pairing with different partners (e.g. Sybenga 1975). If only bivalents are formed, then inheritance is a simple extension of that seen in diploids (i.e.

Mendelian inheritance); a polysomic tetraploid individual carrying alleles *abcd* will form gametes *ab*, *ac*, *ad*, *bc*, *bd*, *cd*, with equal frequency (Bever and Felber 1992; Ramsey and Schemske 2002). If multivalents are formed, inheritance may be complicated by recombination occurring between the centromere and the locus in question (described as "double reduction", reviewed in Bever and Felber 1992).

In allopolyploids it is likely that the (homologous) chromosomes inherited in pairs from each parent will associate at meiosis, but that (homeologous) chromosomes inherited from different parents will not pair (Ramsey and Schemske 2002). Thus alleles are restricted to different duplicate loci ('isoloci', Waples 1988). Such inheritance is described as 'disomic' (Ramsey and Schemske 2002) and is identical to Mendelian inheritance in a diploid organism; a disomic tetraploid individual carrying alleles (*ab*|*cd*) will form gametes (*a*|*c*), (*a*|*d*), (*b*|*c*), (*b*|*d*) with equal frequency (where the symbol '|' is used to separate differentiated genomes). If a population is monomorphic for both isoloci (e.g. *aa*|*bb*) all gametes will be of the same type, (*a*|*b*), and the situation is described as 'fixed heterozygosity'.

The correspondence between origin and inheritance is far from complete (Ramsey and Schemske 2002). This is for two reasons. Firstly, hybridising lineages are often closely related, and some homeologous chromosomes may retain enough similarity to allow pairing (reviewed in Ramsey and Schemske 2002). Secondly, both genetic and epigenetic changes to the genome gradually 'diploidise' polyploids until they are indistinguishable from diploids; in autopolyploids, a major part of this process is the onset of disomic inheritance (Wolfe 2001; Jenczewski *et al.* 2004). This probably occurs because homologous chromosomes gradually diverge to the extent that they no longer 'recognise' each other at meiosis. In some organisms, inheritance is best described by a probabilistic model of chromosome pairing in which chromosomes display a 'pairing preference' (e.g. Wu *et al*. 2001), ranging from an exclusive partner (disomic inheritance) to random pairing between all potential partners (polysomic inheritance) (but see Sybenga 1996). This may lead to different loci displaying different modes of inheritance, within the same organism (Marsden *et al*. 1987).

The mode of inheritance has important implications for evolution. For example, polysomic inheritance may mitigate the effect of inbreeding depression (Lande and Schemske 1985; Ronfort 1999) or increase fitness through increased genetic diversity (Soltis and Soltis 2000). On the other hand, the duplication of loci that results from disomic inheritance permits differentiation and specialisation (Mazet and Shimeld

2002; Prince and Pickett 2002; Wagner 2002), e.g. to different developmental stages, environmental conditions, or tissues (Osborn et al. 2003) . Duplicate loci resulting from hybridisation have even been hypothesised to "fix" combinations of useful traits from the parental lineages (Brochmann and Elven 1992; but see also Ramsey and Schemske 2002). Knowledge of the mode of inheritance is also essential to our understanding of a species' biology. Without it, genes cannot be mapped by recombination analysis (Luo *et al*. 2004), nor can quantitative trait loci be identified (Wu *et al*. 2004). It is also essential when making inferences about population processes or history using observations of genetic diversity or differentiation (Bever and Felber 1992).

Because disomic and polysomic inheritance usually differ in terms of predicted gamete frequencies (see above), artificial cross- or self-fertilisation followed by analysis of genotype frequencies amongst the progeny can be used to distinguish between them (e.g. Marsden *et al.* 1987; Soltis and Soltis 1988; Krebs and Hancock 1989; Soltis and Soltis 1989; Wolf *et al.* 1989; Shore 1991; Murawski *et al.* 1994; Laushman *et al.* 1996; Maki *et al.* 1996; Olson 1997; Hardy *et al.* 2001). Most such studies select heterozygous parents, cross (or self) them, and compare the observed genotype frequencies to those expected under disomic and polysomic inheritance. Alternatively, if the parental genotype is unknown, it can be inferred along with the mode of inheritance (e.g. Ridout *et al*. 2001). In the latter case, there may be many competing hypotheses for the parental genotype. For example, if the progeny of a self-fertilised tetraploid carry alleles *a* and *b*, the parent may have had any one of seven possible genotypes, three polysomic genotypes *abbb*, *aabb*, *aaab*, and four distinct disomic genotypes *ab*|*bb*, *aa*|*bb*, *ab*|*ab*, *aa*|*ab.*

A further complication arises when the copy-number of each allele cannot be inferred from gel banding patterns; i.e. an individual of genotype *aabb* cannot be distinguished from one of genotype *aaab*. This occurs either because band-intensity is not directly proportional to allele copy-number, or because differences in band-intensity are too subtle to judge consistently (e.g. 4:4 versus 5:3 in an octoploid). If this is the case, 'allelic phenotypes' (banding patterns) can be recorded instead of genotypes; e.g., genotypes *aaab*, *aaaa*, and *abcc* scored as allelic phenotype "*ab*", "*a*" and "*abc*" respectively (e.g. Krebs and Hancock 1989; Rieseberg and Doyle 1989; Huang *et al.* 1997; Hardy *et al.* 2001; Ridout *et al.* 2001).

In most studies of marker inheritance (e.g. isozymes, microsatellites), goodness-of-fit test statistics (e.g. Chi-squared) are computed to distinguish between the competing hypotheses, and parental genotypes that give *p*-values below an arbitrary threshold (e.g. $\alpha = 0.05$) are discounted. Although this approach has been used to identify the mode of inheritance in many different polyploid species (e.g. Soltis and Soltis 1988; Krebs and Hancock 1989; Rieseberg and Doyle 1989; Soltis and Soltis 1989; Wolf *et al.* 1989; Laushman *et al.* 1996; Huang *et al.* 1997; Hardy *et al.* 2001), it has statistical limitations (Olson 1997). Firstly, it does not favour one parental genotype over the others; it merely allows those that are not consistent with the data to be discarded. Thus, if more than one of the possible parental genotypes is consistent with the data, there is no way to distinguish between them (Olson 1997). Secondly, there may be large numbers of competing hypotheses, so that the multiple tests involved will inflate the chance of a Type I error (Olson 1997; Ridout *et al*. 2001). Thirdly, unless sample sizes are very large, expected phenotype frequencies will often fall below the minimum class size of five required to fulfil assumptions underlying the use of the Chi-Squared distribution. To counter these shortcomings, a Bayesian approach has been proposed (Olson 1997). This has the advantage of assessing the support for one hypothesis relative to the others (Olson 1997; Ridout *et al*. 2001). Moreover, it can be used to include *a priori* support that may be available for each hypothesis (Olson 1997; Shoemaker *et al*. 1999), such as that which might be gained by earlier inheritance studies, e.g. by karyotype analysis.

Despite its inherent advantages, the Bayesian approach has not been widely used to distinguish between disomic and polysomic inheritance. This may be due to a lack of appropriate software, since making the calculations by hand for polyploid levels greater than hexaploid would be impractical. A major practical challenge to be overcome is the enumeration of all potential parental genotypes based only on an allelic phenotype (Ridout *et al*. 2001); e.g., an octoploid with five distinct alleles has 600 possible disomic genotypes and 35 possible polysomic genotypes. To overcome this, I have written a computer program ("PolySelf") that reads a list of observed allelic phenotypes (i.e. the progeny of a self-fertilisation), and, given the polyploid level (tetraploid, hexaploid, octoploid etc), lists all the possible parental genotypes, and calculates Chi-squared and Bayesian posterior statistics for each.

Here I use the Bayesian and Chi-squared approaches to distinguishing between modes of inheritance and putative parental genotypes in hexaploid *Mercurialis annua*

L. (Euphorbiaceae). This serves as an example of my implementation ("PolySelf") of the method recently proposed by Olson (1997) and establishes the mode of inheritance in hexaploid *M. annua*. Obtaining an inheritance model for this species is particularly important because it is a valuable system for studying the ecology and evolution of combined versus separate sexes, and in particular for the study of androdioecy, an otherwise exceptionally rare sexual system (Pannell 2002; Pannell *et al*. 2004). *Mercurialis annua* is a small, pan-European, wind-pollinated annual, comprising a polyploid complex (2x to 12x). Until recently hexaploid *M. annua* was thought to be autopolyploid in origin, a conclusion reached on the basis of morphology and colchicine-induced polyploidisation (Durand 1963; Durand and Durand 1985, 1992). However, recent molecular data suggest it is an allopolyploid hybrid between *M. annua* and *M. huetii* (Chapter 2). A few hexaploid *M. annua* populations display fixed heterozygosity (Chapter 4, and see Appendix 8.9 page 197), a feature that is diagnostic of disomic inheritance and is consistent with an allopolyploid origin. However, other populations present extreme inter-individual variation, with as many as five alleles seen in some individuals. Such diversity is a common feature of polysomic inheritance (e.g. Rieseberg and Doyle 1989), and has even been used to distinguish between modes of inheritance in the absence of artificial crosses (Machon *et al*. 1995). Thus, a primary aim in analysing *M. annua* was to discount the possibility that there was geographic or inter-locus variation in the mode of inheritance.

3.2 METHODS

3.2.1 Sampling and artificial crosses

Monoecious *Mercurialis annua* plants are self-compatible, and can easily be induced to self fertilize by isolation in pollen-proof boxes (Pannell 1997a). Three plants, known to be heterozygous for some loci, were selected from a population presenting high inter-individual variation (Seville, Spain 37'31" N, 006'16" W). The variation seen suggested that this population was amongst those most likely to have polysomic inheritance. Plants were isolated for between 6 and 10 weeks, until enough seed had been set to ensure adequate sample sizes. Whole plants were harvested, and seeds were collected as they were released upon drying.

3.2.2 Isozymes

Four isozyme systems (five loci) were analysed: AAT (aspartate aminotransferase E.C. 2.6.1.1), PGI (Glucose-6-phosphate isomerase E.C. 5.3.1.9), 6-PGD (two loci, phosphogluconate dehydrogenase E.C. 1.1.1.44), and IDH (isocitrate dehydrogenase E.C.1.1.1.42). Protocols and solutions were adapted from Wendel and Weeden (1990) and are described in detail elsewhere (Chapter 5, page 97). All gels were run at 4 ºC for approximately 6 hours. Lithium-borate gels were run at 300 V, morpholine-citrate gels at 250 V, and Tris-citrate at 150 V. Gels were cut into 1 mm thick slices and stained at 40 ºC. Reactions were stopped after staining and gels were stabilised by removing stain solution and adding 30 ml 25% v/v glycerol. Gels were scored either from photographs or at the time of staining.

3.2.3 Computational analysis

3.2.3.1 Generating alternative hypotheses

Given the polyploid level, and a list of distinct alleles present in the parent, the computer program PolySelf generates all the possible disomic and polysomic parental genotypes, and calculates the expected frequencies of offspring allelic phenotypes for each one. This process is divided into three stages. Stage (1): Given the alleles present, all of the distinct allele-sets are listed; i.e. for a tetraploid carrying alleles *abc* the distinct allele sets are *aabc*, *abbc* and *abcc*. This requires all the integer partitions of the polyploid level ($2n$) that have as many members (m) as there are alleles (Ridout *et al*. 2001); e.g. in a hexaploid with three alleles, $2n = 6$, $m = 3$, and thus integer partitions are (1,1,4), (1,2,3) and (2,2,2). These are used, along with all the permutations of the list of alleles present, to generate all the distinct allele sets (those that are identical are discarded). Because the alleles in a polysomic genotype are equivalent (i.e. order independent), allele sets correspond to polysomic genotypes. Stage (2): The allele sets are used to generate potential disomic genotypes. This is done by taking all permutations of the set, dividing each one into pairs (corresponding to isoloci), and discarding those that are identical; e.g., permutations and divisions of *abbc* lead to identical disomic genotypes *ac*|*bb* and *ca*|*bb*, which, because isoloci are equivalent during selfing, are also identical to *bb*|*ac* and *bb*|*ca*. In the case of the allele set *abbc*, the only distinct disomic genotypes are *ac*|*bb* and *ab*|*bc*. There appears to be

no simple way of generating all possible distinct disomic genotypes without an exhaustive search (Ridout *et al*. 2001). Stage (3): these disomic and polysomic genotypes are used to calculate the expected phenotype frequencies amongst selfed offspring.

3.2.3.2 Statistical analysis

Following Olson (1997) and Ridout *et al* (2001), two statistical tests are used, a goodness-of-fit test (Chi-squared), and the Bayesian posterior probability. These approaches are complementary (Ridout *et al*. 2001). In the event that the correct genotype and mode of inheritance are not amongst those considered, the most-likely parental genotype will not be the correct one, thus (although its posterior probability may be high), the goodness-of-fit test will exclude it.

The Chi-squared test statistic is calculated using observed and expected allelic phenotype frequencies, o_i and e_i:

$$\chi^2 = \sum \frac{(o_i - e_i)^2}{e_i} \tag{3.1}$$

(Sokal and Rohlf 1995). The number of degrees of freedom for each test corresponds to the number of different allelic phenotypes that can occur under that parental genotype (or one less if the parental phenotype is calculated from the offspring phenotypes) (Olson 1997). The probability of observing a particular Chi-squared test-statistic can be found using standard statistical tables.

The Bayesian posterior probability for each hypothetical parental genotype is calculated using Bayes' theorem for discrete distributions:

$$P\left(H_i \mid data\right) = \frac{P\left(data \mid H_i\right) P(H_i)}{\sum P\left(data \mid H_i\right) P(H_i)} \tag{3.2}$$

(Olson 1997) where $P(H_i \mid data)$ is the posterior probability of hypothesis (parental genotype) H_i being the true one, given the data. $P(data \mid H_i)$ is the probability of observing the data under that hypothesis, and $P(H_i)$ is the prior probability associated that hypothesis. The probability of observing the dataset, given a particular hypothetical parental genotype, $P(data \mid H_i)$, is calculated using the multinomial distribution, such that:

$$P\left(data \mid H_j\right) = \frac{n!}{\prod n_i!} \prod p_i(H_j)^{n_i} \quad (3.3)$$

(Olson 1997) where n is the total number of progeny, n_i is the observed number of progeny of phenotype i, and $p_i(H_j)$ is the expected proportion of progeny that are of phenotype i under hypothesis H_j. If the prior probability is set to be equal across all hypotheses, as here, this approach is effectively equivalent to maximum likelihood. The posterior probability can be interpreted as the support for a particular parental genotype in light of the observed data and the other possible genotypes considered (Olson 1997; Shoemaker *et al*. 1999).

3.3 RESULTS

In no case was a polysomic genotype found to be more likely than a disomic genotype. Of the five loci in three individuals tested, one locus was homozygous in one individual, and thus did not represent a test of inheritance (6Pgd-2 Individual II, Table 3.1). Of the 14 heterozygous tests, six apparently showed "fixed heterozygosity", i.e. all offspring displayed the same heterozygous banding pattern (Table 3.1), broadly supporting the hypothesis of disomic inheritance. The remaining eight tests displayed variation amongst progeny phenotypes. In one case, the disomic and polysomic genotypes were necessarily indistinguishable; in the other seven cases, Bayesian analyses supported disomic inheritance over polysomic inheritance. However, in three cases, Chi-squared tests discounted all of the parental genotypes considered.

3.3.1 Fixed heterozygous loci

Despite the presence of more than one allele, offspring showed no variation in allelic phenotype for the following loci: 6*Pgd*-1 (all three progenies, Appendix 8.12), *Aat*-1 (selfed progeny of individuals I and II, Appendix 8.12) and *Idh*-1 (selfed progeny of individual III only, Appendix 8.12). Chi-squared tests could not reject a

Table 3.1: Differences between Bayesian and Chi-squared analysis in *M. annua*

		Individual I	Individual II	Individual III
Bayesian Posterior	Aat-1	Disomic (fixed)	Disomic (fixed)	Disomic
	Pgi-2	Disomic	Disomic	Disomic
	6Pgd-1	Disomic (fixed)	Disomic (fixed)	Disomic (fixed)
	6Pgd-2	Indistinguishable	(no test)	Disomic
	Idh	Disomic	Disomic	Disomic (fixed)
Chi-Squared	Aat-1	Undecided	Undecided	(None)
	Pgi-2	Disomic	Undecided	Undecided
	6Pgd-1	Undecided	Undecided	Disomic
	6Pgd-2	Indistinguishable	(no test)	(None)
	Idh	(None)	Disomic	Undecided

Support for disomic or polysomic parental genotypes in *M. annua*, using (1) the largest Bayesian posterior, or (2) the exclusion of all other possible parents with Chi-squared tests. Grey cells under Chi-squared indicate differences from the Bayesian result.

genotype in any of these cases. However, because some classes of offspring phenotypes had small expected frequencies, Chi-squared tests could not be used to exclude the competing polysomic genotypes in all but Progeny III 6*Pgd*-1. Thus in five of the six "fixed heterozygous" progenies Chi-squared tests did not distinguish between disomic and polysomic inheritance (Table 3.1).

By contrast, in all six cases the parental genotype with the highest posterior probability was one with disomic inheritance (Table 3.1). When there are only two alleles present, many of the disomic genotypes are indistinguishable from each other. Because of this, where sample sizes were smallest (e.g. progeny II, $n = 49$) the single most likely disomic genotype was only 1.28 times more likely than the most likely polysomic genotype. Even for the largest sample size (e.g. progeny III, $n = 196$), the single most likely disomic genotype was only *ca.* 2.75 times more likely than the most likely polysomic genotype, below the factor of three recommended as "firm evidence" in favour of a hypothesis (Olson 1997). However, even in the worst case ($n = 49$) the probability of *any* disomic genotype being the correct one was 3.16 times greater than the probability of *any* polysomic genotype being the correct one (found by summing the posterior probability across all disomic and polysomic genotypes, respectively). This supports a disomic mode of inheritance, though without identifying a particular parental genotype.

3.3.2 Variable loci

Offspring were variable in eight cases (Table 3.1, Table 3.2, Appendix 8.12): *Aat*-1 (progeny III), *Pgi*-1 (all three progenies), 6*Pgd*-2 (progenies I and III), *Idh*-1 (progenies I and II). For an illustration of an isozyme gel showing variable progeny from a self-fertilised hexaploid, see Appendix 8.11 page 203. In the case of 6*Pgd*-2 from progeny I, the most likely disomic parent and the most likely polysomic parent give the same predicted offspring phenotype frequencies, and were therefore necessarily indistinguishable; disomic (*bb*|*bb*|*bd*) versus polysomic (*bbbbbd*) both predict 75% '*b*' and 25% '*bd*' offspring. In the other seven cases (Table 3.2, Appendix 8.12), the parental genotype (or genotypes, where different disomic parents give the same predicted offspring ratios) with the largest posterior probability was a disomic one.

Table 3.2: Offspring frequencies and most probable parents for crosses that were variable

Progeny	Locus	Observed Offspring Phenotype frequencies	Most strongly supported parental genotype(s)	Notes
I	*Pgi-2*	bc (33) ab (28) abc (53)	(*ac*|*bb*|*bb*)	
I	*6Pgd-2*	b (29) bd (85)	(*bbbbbd*) or (*bb*|*bb*|*bd*)	*Indistinguishable*
I	*Idh-1*	be (49) bef (50)	(*bb*|*bf*|*ee*) or (*bb*|*ee*|*ef*)	*excluded by Chi-Square*
II	*Pgi-2*	a (4) ac (7) ab (11) abc (27)	(*aa*|*ab*|*ac*)	
II	*Idh-1*	ae (11) aef (36)	(*aa*|*af*|*ee*) or (*aa*|*ee*|*ef*)	
III	*Pgi-2*	bc (43) abc (153)	(*ab*|*bb*|*cc*) or (*ac*|*bb*|*cc*)	
III	*6Pgd-2*	b (19) bf (33) be (48) bef (73)	(*bb*|*be*|*bf*)	*excluded by Chi-Square*
III	*Aat-1*	b (22) ab (172)	(*ab*|*ab*|*bb*)	*excluded by Chi-Square*

Observed phenotype numbers and the most strongly supported parental genotype (more than one if predictions are identical) for each of the loci that displayed variation amongst the offspring. In three cases, the most strongly supported (Bayesian posterior) parental genotypes were excluded ($\alpha = 0.05$) by Chi-squared tests.

In three cases, even though the most likely parental genotype was a disomic one, (Table 3.1, Table 3.2, Appendix 8.12) the Chi-squared analysis excluded ($\alpha = 0.05$) the possibility of the most likely parental genotype being true. This suggests that in these three cases none of the hypothetical parental genotypes actually represented the true mode of inheritance.

3.4 DISCUSSION

In all the crosses that had the potential to differentiate between disomic and polysomic inheritance, the genotype with the highest posterior probability was a disomic one. This, in conjunction with the presence of fixed heterozygosity in a minority of natural populations (Chapter 5), strongly supports disomic inheritance in hexaploid *M. annua*. There is no evidence of inter-individual or inter-locus variation in the mode of inheritance, as has been proposed for some species (e.g. Marsden *et al.* 1987; Machon *et al.* 1995). However, in three cases none of the parental genotypes considered satisfactorily explained the offspring frequencies. This suggests the presence of an additional factor (discussed below), e.g. the existence of null alleles, or a non-zero rate of homeologous chromosome pairing.

Disomic inheritance is consistent with the newly proposed allopolyploid origin for hexaploid *M. annua* (Chapter 2), and has important implications for future studies of population structure (Chapter 5). Additionally, the analysis presented here raises important points with regard to the analysis of progeny ratios in hexaploids more generally; especially (1) many genotypes may be necessarily indistinguishable, and (2) sample sizes needed to distinguish between disomic and polysomic inheritance using traditional goodness-of-fit tests may be prohibitively large. Below, I discuss each in turn.

3.4.1 Implications of disomic inheritance in *Mercurialis annua*

Patterns of isozyme allele inheritance are often used to distinguish between autopolyploid and allopolyploid origin. However, although polysomic inheritance is more likely to occur in autopolyploids (Ramsey and Schemske 2002), diploidisation means that many older autopolyploids will display disomic inheritance (Wolfe 2001; Ramsey and Schemske 2002; Jenczewski *et al.* 2004). Thus, while disomic inheritance

is consistent with the allopolyploid origin recently proposed in *M. annua*, and with the absences of multivalents at meiosis (Durand 1963), it would be equally consistent with an autopolyploid origin followed by diploidisation.

The existence of alleles that occur at more than one isolocus (e.g. allele *a* in Pgi-2, Table 3.2) could be used to argue for autopolyploidy followed by diploidisation. However, this assumes that every mutation leads to a new allele (i.e. an "infinite-alleles" model), when in fact some mutations may lead to convergence. Additionally, there is the possibility that gene-conversion between isoloci could lead to a sharing of alleles. Finally, in the case of *M. annua*, the putative parents are close relatives, and it is possible that the copies of allele *a* occurring on different isoloci were inherited from their common ancestor. Sequence-level studies of isozyme-encoding genes could distinguish between these three scenarios.

Regardless of the origin, the distinction between disomic and polysomic inheritance is essential for studies of (1) population structure (e.g. Bever and Felber 1992), (2) breeding system (e.g. Murawski *et al*. 1994) and (3) inbreeding depression (Lande and Schemske 1985; Ronfort 1999). This is particularly important in hexaploid *M. annua* where the sexual system is variable (Durand and Durand 1992; Pannell 1997c), and the coexistence of males and hermaphrodites makes the species a valuable model for the evolution of androdioecy and sexual-systems in general (Pannell 2002; Pannell *et al*. 2004).

3.4.2 Unconsidered genotypes and models of inheritance

Of the 14 loci that represented a test, three resulted in offspring phenotype frequencies that were not consistent with any of the parental genotypes considered (Chi-squared tests, $\alpha = 0.05$, Table 3.1). Although I considered all possible disomic genotypes and all possible polysomic genotypes without double reduction, I did not consider (1) double reduction in polysomic inheritance, (2) the presence of null alleles, or (3) modes of inheritance intermediate between disomic and polysomic.

Low rates of double reduction, which result from recombination between the locus in question and the centromere, lead to expected genotype frequencies similar to that of simple polysomic inheritance, but with increased frequencies of homozygous genotypes and additional (homozygous) genotypes (Bever and Felber 1992; Ramsey and Schemske 2002). However, this is not consistent with the allelic phenotype

frequencies seen (Table 3.2), or with the support for disomic (rather than polysomic) inheritance at all other loci. The presence of a null allele, i.e. an allele that does not function under the assay conditions (Wendel and Weeden 1990), seems particularly likely for Idh-1 progeny II, where a disomic genotype of (*bb*|*ee*|*f*-) would give the observed 50:50 ratio (observed frequencies are given in Appendix 8.12).

Inheritance that is intermediate between disomic and polysomic could lead to the offspring phenotype ratios seen, but would be difficult to analyse because several additional parameters would have to be considered. In tetraploids, if only bivalents are formed, intermediate forms of inheritance can be characterised by one parameter, the relative frequency with which homologous and homeologous chromosomes pair (the "pairing preference", Wu *et al*. 2001). However, if quadrivalents form, parameters for the pairing preference, frequency of quadrivalent formation, and double reduction are needed (Wu *et al*. 2001). While this may be possible in tetraploids, in hexaploids, there could be up to three distinct pairs of homologous chromosomes (under fully disomic inheritance). Thus, pairing preferences would need to be specified by several (non-independent) parameters, and parameters for the relative frequencies of the three classes of multivalents (one hexavalent, one quadrivalent and one bivalent, or three bivalents) and relative double reduction rates in the difference multivalents, would have to be estimated (Wu *et al*. 2001). Therefore, in hexaploids and above, it seems likely that the sample sizes required to make meaningful estimates of all the parameters would be prohibitively large.

3.4.3 Necessarily indistinguishable genotypes

In tetraploids, some disomic genotypes are necessarily indistinguishable from polysomic ones. This only occurs when there are two alleles present, but not when there are three or four; the two cases are (*aaab*) = (*aa*|*ab*) and (*abbb*) = (*ab*|*bb*). In hexaploids the analogous problem occurs when there are two alleles (Table 3.3): (*aaaaab*) = (*aa*|*aa*|*ab*) and (*abbbbb*) = (*ab*|*bb*|*bb*). In addition, because alleles are 'hidden' by the fixed heterozygosity, there are also disomic genotypes that are indistinguishable: (*aa*|*bb*|*bb*) = (*aa*|*ab*|*bb*) = (*aa*|*aa*|*bb*). The same problem occurs with three alleles: (*aa*|*ac*|*bb*) = (*aa*|*bb*|*bc*), (*aa*|*ab*|*cc*) = (*aa*|*bc*|*cc*), and (*ab*|*bb*|*bc*) = (*ab*|*bc*|*cc*). However, with four or more alleles present all possible disomic and

Table 3.3: Sample size and power for two alleles in a hexaploid

				Sample Size					
				50		100		1000	
Phenotypes:	*b*	*a*	*ab*	**Posterior**	**Chi^2 *p***	**Posterior**	**Chi^2 *p***	**Posterior**	**Chi^2 *p***
(aa aa bb)			100.000	0.231	1.000	0.271	1.000	0.333	1.000
(aa bb bb)			100.000	0.231	1.000	0.271	1.000	0.333	1.000
(aa ab bb)			100.000	0.231	1.000	0.271	1.000	0.333	1.000
(aaabbb)	0.250	0.250	99.500	0.180	0.969	0.164	0.918	0.002	0.170
(ab ab ab)	1.562	1.562	96.875	0.047	0.656	0.011	0.358	0.000	0.000
(aaaabb)		4.000	96.000	0.030	0.353	0.005	0.125	0.000	0.000
(aabbbb)	4.000		96.000	0.030	0.353	0.005	0.125	0.000	0.000
(aa ab ab)		6.250	93.750	0.009	0.189	0.000	0.036	0.000	0.000
(ab ab bb)	6.250		93.750	0.009	0.189	0.000	0.036	0.000	0.000
(aaaaab)		25.000	75.000	0.000	0.000	0.000	0.000	0.000	0.000
(abbbbb)	25.000		75.000	0.000	0.000	0.000	0.000	0.000	0.000
(aa aa ab)		25.000	75.000	0.000	0.000	0.000	0.000	0.000	0.000
(ab bb bb)	25.000		75.000	0.000	0.000	0.000	0.000	0.000	0.000

Illustration of necessarily indistinguishable parental genotypes, and the difficulty in applying Chi-squared tests to exclude the possibility of polysomic inheritance when 50, 100, or even 1000 heterozygous phenotypes are seen. In this table, disomic genotypes are indicated by the separation of alleles into pairs.

polysomic genotypes are essentially distinguishable using allelic phenotype frequencies of selfed offspring. The number of indistinguishable disomic genotypes increases with polyploid level; e.g. an octoploid with three alleles has four sets of six equivalent disomic genotypes, three sets of three, and one set of two. This apparent trend suggests that large numbers of alleles might be needed to avoid the problem of equivalent genotypes for tests of inheritance in highly polyploid organisms.

3.4.4 Low power to distinguish between genotypes

Even when genotypes are essentially distinguishable using offspring phenotype frequencies, the sample size needed to do so may be prohibitively large. This is exemplified by the difficulty of distinguishing between disomic and polysomic inheritance in some *M. annua* crosses above. Genotypes (*aa*|*bb*) and (*aabb*) are most easily distinguished by the presence of homozygotes (allelic phenotype "*a*" or "*b*"). However, for the genotype with polysomic inheritance (*aabb*), their combined frequency is expected to be only 5.5% of the progeny; i.e. a sample size of approximately 100 is needed before Chi-squared tests can be used to exclude the possibility of polysomic inheritance. In a hexaploid, it becomes more difficult (Table 3.3). It is far harder to distinguish between a fixed heterozygous genotype (e.g. *aa*|*ab*|*bb*) and genotype with polysomic inheritance (*aaabbb*) by observing a homozygote amongst the progeny, as the expected frequency of homozygotes generated by the genotype with polysomic inheritance is 0.5%. In such a case, even a sample size of 1000 offspring gives an expected homozygous class too small to use Chi-squared tests. This difficulty in excluding a polysomic parent because of small sample size, even when all the offspring are off the same heterozygous phenotype, is not always appreciated. Some studies have pragmatically reached the conclusion of disomic inheritance without testing the possibility of polysomic inheritance at all (e.g. Widen and Widen 2000).

For few alleles and high polyploid level, Bayesian or maximum likelihood approaches are easier to apply, as they do not suffer from the minimum class size of five required to use the Chi-squared distribution (Olson 1997; Ridout *et al*. 2001). However, power to distinguish between parental genotypes that differ only in terms of rare offspring classes is still likely to be low. For example, to distinguish between octoploid genotypes (*abbbcccc*) and (*ac*|*bc*|*bc*|*bc*) requires a sample large enough to

distinguish between phenotype percentages (0.02, 24.98, 0.49, 0.02, 74.49) and (0.39, 24.61, 1.17, 0.39, 73.45). That power must decrease with increasing polyploid level is evident from the large numbers of different genotypes possible for higher polyploid levels; as there are more possible parental genotypes, the differences in expected offspring frequencies must be increasingly small. It seems that a progeny-phenotype approach to identifying polyploid parental genotype is unlikely to be viable for polyploid levels greater than octoploid, unless very many alleles are available. Certainly, approaches relying on the Chi-squared distribution, which are the most commonly used, and which are based on the exclusion of all possible competing hypotheses, are unlikely to be successful.

The examples presented above show the difficulty of distinguishing between potential parental genotypes using offspring allelic phenotype frequencies depends jointly on the polyploid level and the number of distinct alleles present, as well as on the precise parental genotype, which is probably out of the experimenters control. Power to distinguish between parental genotypes can be improved by selecting individuals likely to display large numbers of distinct alleles. This is because some genotypes are necessarily indistinguishable when there are less than n distinct alleles in a $2n$-ploid parent, and even when parental genotypes are essentially distinguishable, small differences in expected frequencies reduce power and make Chi-squared tests difficult to apply.

4 POPULATION STRUCTURE IN POLYPLOIDS WITH DISOMIC INHERITANCE

4.1 INTRODUCTION

The spatial structuring of populations is of great importance in biology, affecting many different aspects of ecology and evolution. As well as necessarily spatial traits, such as dispersal or the breeding system, the isolation of populations affects fundamental biological processes such as local adaptation, speciation and local extinction. The effects of spatial structure are reflected in the level of genetic diversity, and the degree of genetic differentiation between populations (reviewed by Charlesworth *et al*. 2003). Thus, measuring and accounting for the level and distribution of genetic diversity is a primary aim for population geneticists. In particular, patterns of genetic diversity and differentiation can provide information regarding population processes such as migration (e.g. Cockerham and Weir 1993; Ennos 1994; Neigel 1997), and local extinction (e.g. Pannell and Charlesworth 2000; Wakeley and Aliacar 2001).

Patterns of genetic diversity can be quantified in many different ways, depending on the biological questions being asked, and the molecular techniques being used to address them. For molecular markers with a clear genetic interpretation, such as microsatellites, isozymes and DNA sequences, commonly used measures of diversity include: allelic richness; the proportion of polymorphic loci (A_p); Nei's gene diversity (H_e); and (for DNA) the proportion pairwise site differences (π). Genetic differentiation between populations is most often quantified by measures analogous to F_{ST}, which can be interpreted as the proportion of variation that is due to differences between populations (Hartl and Clark 1997). Alternatively, genetic similarity between populations may be summarised by a genetic distance (see Weir 1996; Hartl and Clark 1997).

All of these summary statistics are commonly employed to quantify patterns of genetic diversity in diploid organisms. However, many organisms are polyploid, and

although polyploidy is particularly common amongst plants, fish, and amphibians, it is also found amongst birds, mammals and many invertebrates (Leitch and Bennett 1997; Otto and Whitton 2000; Legatt and Iwama 2003). Thus, exactly the same questions regarding population structure and genetic diversity are regularly posed by polyploid organisms, and polyploidy itself may be a focal point for population genetic studies (e.g. Mahy *et al.* 2000; Hardy and Vekemans 2001). Unfortunately, it is often more difficult to calculate and interpret genetic diversity summary statistics in polyploid organisms, and this is possibly what has lead some authors to avoiding completely the use of genetic summary statistics in polyploids (e.g. Glover and Abbott 1995; Rumsey *et al*. 1999; Vogel *et al*. 1999).

This chapter has four aims. (1) To outline the challenges to the calculation and interpretation of genetic summary statistics in polyploids, and refer the reader to the literature on genotype-based polyploid statistics. (2) To briefly review some of the various *ad hoc* genetic summary statistics that have been quoted for polyploid populations in which genotype data could not be interpreted. (3) To introduce new statistics to quantify diversity and differentiation, and provide a computer program to calculate this (and the other *ad hoc* statistics) from allelic phenotypes. (4) To evaluate the behaviour of these statistics under an island model of population structure, using a simulation approach.

4.2 GENETICS SUMMARY STATISTICS IN POLYPLOIDS

At least two aspects of polyploidy give rise to difficulties in the calculation and interpretation of genetic summary statistics. First, in polyploids it is often difficult to distinguish between different genotypes, making estimates of genotype or allele frequency impossible (e.g. Kahler *et al.* 1980; Krebs and Hancock 1989; Brochmann *et al.* 1992). This is essentially because more than two alleles are present at any one locus. For example, in a diploid heterozygote, any two distinct alleles must each be present as a single copy, whilst in a polyploid heterozygote, the number of copies of each allele must also be determined. Allele copy-number (or dosage) can be difficult to estimate from electrophoretic banding patterns, because it need not be proportional to band intensity (or peak height for automated sequencers). This problem affects both protein-based (isozyme) and DNA-based (e.g. microsatellite) markers, and may occur

if, for example, allozymes differ in activity under assay conditions, or PCR amplification is affected by allele-specific primer binding.

If these effects can be ignored, such that band intensity (peak height) can be assumed to be proportional to copy-number, then allele dosage may be estimated reliably in tetraploids (e.g. Arft and Ranker 1998; Prober *et al.* 1998; Young *et al.* 1999; Hardy and Vekemans 2001; Nassar *et al.* 2003). However, this is only feasible in tetraploids because there are large differences between 'balanced' (2:2) and 'unbalanced' (3:1) heterozygotes. It becomes more difficult to estimate allele dosage in higher polyploid levels, where band intensity ratios such as 4:2 and 5:1 (hexaploids) or 5:3 and 6:2 (octoploids) need to be consistently distinguished. In such cases, data can only be summarised as banding patterns (e.g. Jain and Singh 1979; Gaur *et al.* 1980; Chung *et al.* 1991; Brochmann *et al.* 1992), or 'allelic phenotypes' (e.g. Murdy and Carter 1985; Bayer and Crawford 1986; Rogers 2000; Berglund and Westerbergh 2001), which list only the distinct alleles carried by each individual for each locus. That is to say, for example, tetraploid genotypes *aabc* and *abbc* would both be recorded as allelic phenotype *abc*.

The second difficulty in the calculation and interpretation of polyploid summary statistics stems from the mode of polyploid inheritance, i.e. whether it is polysomic or disomic. This can affect the apparent level and distribution of genetic diversity, and the interpretation of diversity patterns in terms of population processes (Bever and Felber 1992; Ronfort *et al*. 1998). In an autotetraploid with polysomic inheritance, all four chromosomes are homologous and either pair at random or form multivalents during meiosis (reviewed in Bever and Felber 1992). If multivalents are never formed, or if the locus in question is tightly linked to the centromere, polysomic inheritance is an extension of Mendelian inheritance seen in diploids, e.g. a tetraploid locus with alleles *abcd* generates gametes *ab*, *ac*, *ad*, *bc*, *bd*, *cd* with equal frequency (reviewed in Bever and Felber 1992). If there is multivalent formation, and recombination occurs between the centromere and the locus, then a proportion of the gametes will be homozygous (described as 'double reduction', Bever and Felber 1992). Although double reduction may complicate multi-locus estimates of some statistics, gene diversity (H_e) and estimates of F_{ST} are in principle calculable in a way analogous to the diploid approach (Ronfort *et al*. 1998; Thrall and Young 2000). Polysomic inheritance is most commonly displayed by neo-autopolyploids, as there is initially no differentiation between chromosomes (Ramsey and Schemske 2002).

The quantification and interpretation of genetic diversity in polyploids with disomic inheritance can be more problematic. If chromosome pairs are strongly differentiated from each other, as in allopolyploids or diploidised autopolyploids, chromosomes will consistently form the same pairings at meiosis (Ramsey and Schemske 2002; Jenczewski *et al.* 2004). Thus, an *ab*|*cd* individual (the symbol '|' denotes separation of the chromosome pairs) with disomic inheritance will generate gametes *a*|*c*, *a*|*d*, *b*|*c* and *b*|*d*. In effect, the loci have been duplicated and, in principle, the genetic system is identical to that of diploids. However, the duplicate loci will often share alleles, particularly if they result from diploidisation or from hybridisation between closely related species. If this is the case, duplicate loci may co-migrate during electrophoresis, and must be treated together for interpretation and analysis. These genetically independent, but experimentally confounded, duplicate loci are described as 'isoloci', and the term 'locus' is usually used to refer to the set of isoloci that are derived from the ancestral locus (Waples 1988; Hedrick *et al.* 1991; Prober *et al.* 1998).

If inheritance is polysomic and allele dosage can be consistently scored, extensions of standard diploid summary statistics such as H_e and F_{ST} can be used to quantify genetic diversity and population differentiation in polyploids (Ronfort *et al*. 1998; but also see Thrall and Young 2000 for references). Moreover, computer programs are freely available to do so (AUTOTET - Thrall and Young 2000; SPAGEDi - Hardy and Vekemans 2002). If inheritance is disomic and allele dosage can be scored, it may be possible to estimate the underlying allele frequencies for different isoloci using the superficial genotypes (Waples 1988). These estimates can then be used to calculate genetic diversity statistics (Waples 1988; Bouza *et al.* 2001). However, this approach requires the mating-system to be known, or for artificial crosses to be used (Hedrick *et al*. 1991). For markers with disomic inheritance where allele dosage *cannot* be scored, phenotype-based statistics need to be used.

4.3 PHENOTYPE-BASED STATISTICS

When allele-dosage cannot be inferred from banding patterns, a range of different *ad hoc* approaches have been advocated that quantify diversity, differentiation, and genetic distance using the frequencies of different banding patterns or allelic

phenotypes. Although polyploid genetic data may be entered into some packages as pseudohaplotypes, until recently no dedicated software has been available to calculate allelic-phenotype statistics (but see GENOTYPE in Meirmans and van Tienderen 2004). This dearth of software for the analysis of allelic phenotype data has limited analysis in some systems (e.g. Lack and Kay 1988; Schierenbeck *et al.* 1995; Ainouche *et al.* 1999). It has also lead some authors to attempt analysis as if the data were diploid, by assuming that one or more isoloci are monomorphic (e.g. Watson *et al.* 1991; Prober *et al.* 1998).

4.3.1 Genetic diversity

Simple measures of genetic diversity, such as allelic richness (a count of the alleles present) and the proportion of loci that are polymorphic, are often reported for polyploids (e.g., Gaur *et al.* 1980; Murdy and Carter 1985; Chung *et al.* 1991; Garcia *et al.* 1991; Rogers 2000; Taylor and Foighil 2000; Berglund *et al.* 2001). These statistics can be calculated from allelic phenotype data in exactly the same way as they are calculated from genotypic data, because they only require each allele's presence to be identified, not its frequency. Additionally, when allelic phenotypes or banding phenotypes are scored, the total number of different phenotypes per population is often reported (Gaur *et al.* 1980; Chung *et al.* 1991; e.g. Rogers 2000; Berglund *et al.* 2001). However, because rare and common phenotypes contribute equally to these statistics, they are less informative about genetic diversity than measures like Nei's (1987) gene diversity, H_e (expected heterozygosity in diploids), which take frequency information into account.

When genotypic data are unavailable, diversity statistics similar to H_e can be calculated from phenotype frequencies instead of allele frequencies. In particular, a statistic directly analogous to Nei's gene diversity has been used (Yunus *et al.* 1991; Meerts *et al.* 1998), such that:

$$H^{Phen} = \sum_{i=1}^{n} p_i(1-p_i) = 1 - \sum_{i=1}^{n} p_i^{\,2} \qquad (4.1)$$

where p_i is the frequency of the ith phenotype and n is the number of distinct phenotypes. This is the diversity statistic calculated when data are entered into standard population genetic packages as pseudohaplotypes, and can be viewed as the probability that two randomly drawn phenotypes are different. Similarly, a Shannon-

Weaver diversity index of phenotypes is often reported (e.g. Jain and Singh 1979; Gaur *et al.* 1980; Chung *et al.* 1991), such that:

$$H^{SW} = \sum_{i=1}^{n} p_i \log(1 - p_i) \tag{4.2}$$

Both of these measures of diversity treat phenotypes as being equally distinct; they do not account for the fact that phenotypes which share many bands (i.e. alleles) are more genetically similar than those which share few bands.

4.3.2 Genetic differentiation

Genetic differentiation between populations is most commonly quantified by statistics similar to F_{ST}, which can be defined as

$$F_{ST} = \frac{H_T - H_S}{H_T} \tag{4.3}$$

where H_T is the total diversity, and H_S is the average within-population diversity (e.g. Hartl and Clark 1997). The most widely used differentiation statistics that can be regarded as estimators of F_{ST} are: (1) G_{ST}, which can be calculated directly from equation (4.3) by substituting Nei's gene diversity, H_e, calculated over all populations and within populations as appropriate (e.g. Hartl and Clark 1997); and (2) θ, which is calculated using the variance in genotype frequencies (Weir and Cockerham 1984; Weir 1996).

If allele-dosage can be inferred from band intensity, F_{ST} can be estimated for polyploids that have polysomic inheritance by either of these two methods (Nei 1987; Ronfort *et al*. 1998). When such genotypic data are unavailable, a measure of genetic differentiation between populations can be calculated by substituting allelic-phenotype diversity into equation (4.3), in place of the allele-based gene diversity H_e. This may be done with either of the two phenotypic diversity measures outlined above, H^{Phen} (e.g. Meerts *et al*. 1998) or H^{SW} (e.g. Chung *et al*. 1991).

4.3.3 Genetic similarity

Several statistics have been used to quantify genetic similarity between polyploid populations or individuals. These can be used for the creation of dendrograms that link them according to similarity (e.g. Chung *et al.* 1991; Brochmann *et al.* 1998), but they

can also be used to identify whether genetic similarity varies with spatial separation. In (partially) asexual organisms, the genetic similarity between individuals can be used to define clonal groups of individuals (e.g. Meirmans and van Tienderen 2004).

For phenotypic polyploid data, the genetic similarity between populations has often been quantified using Hedrick's (1970) genotypic identity (Gaur *et al.* 1980; Chung *et al.* 1991; Meerts *et al.* 1998; Rogers 2000; Berglund *et al.* 2001; Berglund and Westerbergh 2001). This identity measure, I_{xy}, was originally intended for use with diploid genotypic data (Hedrick 1970), but because it is phrased in terms of genotype frequencies rather than allele frequencies, phenotype frequencies can easily be substituted instead. This genotypic (or phenotypic) identity can be interpreted as the probability of randomly drawing the same genotype (or phenotype) from both populations, relative to the average chance of drawing that genotype twice from either one of them, and is calculated as:

$$I_{xy} = \frac{\sum_{j=1}^{n} p_{jx} p_{jy}}{\frac{1}{2}\left(\sum_{j=i}^{n} {p_{jx}}^2 + \sum_{j=i}^{n} {p_{jy}}^2 \right)} \qquad (4.4)$$

where p_{jx} is the frequency of the jth genotype (or phenotype) in population x.

Allelic phenotype-based similarity between individuals has been quantified in several ways, according to the mutational relationship assumed between alleles. If it is reasonable to treat alleles as being equally different from each other (i.e. an infinite alleles model) then one possible genetic distance between a pair of individuals is the number steps needed to convert one phenotype into the other (Meirmans and van Tienderen 2004). Thus, individuals with allelic phenotypes *abc* and *cdef* have a distance of three, because it would take three transformations (e.g. *a* → *d*, *b*→ *e*, and ? → *f*) to convert one to the other. Another possible inter-individual distance measure is a Dice similarity index (see Meirmans 2004), which is based on the proportion of alleles by which a pair of individuals differ, i.e. *abc* and *cdef* have six distinct alleles between them, and they differ by five, giving a dice similarity of 5/6. Other infinite-alleles measures of similarity are discussed by Meirmans and van Tienderen (2004) and Meirmans (2004). If a stepwise mutation model can be assumed (e.g. for microsatellites), information may also be available from differences in allele length or repeat number. Bruvo *et al.* (2004) have recently proposed a genetic distance measure

for polyploid individuals that takes account of the number of stepwise mutations between genotypes (but see also Meirmans and van Tienderen 2004). This measure can be calculated between allelic phenotypes (described as 'partial heterozygotes' by Bruvo *et al.*) by averaging the genetic distance across all possible underlying genotypes (Bruvo *et al*. 2004).

4.4 A NEW DIVERSITY STATISTIC BASED ON ALLELE DIFFERENCES

Genetic diversity statistics based on allelic phenotype frequencies, such as H^{Phen} and H^{SW}, record the same level of genetic diversity whether genotypes vary by a few alleles or by many alleles. Clearly, a measure of phenotype diversity that accounts for the degree of differentiation between phenotypes would be preferable. The number of alleles by which a pair of individuals differ is a simple measure of the differentiation between them, i.e. *abc* differs from *cdef* by *a*, *b*, *d*, *e* and $f = 5$, (compare the Dice similarity index described above). Therefore, I suggest the use of a phenotype-based genetic diversity statistic, H', defined as the average number of allele differences between pairs of individuals:

$$H' = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j>i}^{n} \sum_{k=1}^{alleles} x_{ijk} \qquad (4.5)$$

where n is the total number of individuals and x_{ijk} is an indicator variable that takes the value 1 when individual i carries allele k but individual j does not, or when individual j carries allele k but individual i does not, and otherwise takes the value 0. This measure of genetic diversity, calculated within populations and over all populations as appropriate, is then substituted into equation (4.3), to give a measure of genetic differentiation analogous to F_{ST}, which I denote F'_{ST}. I have written a computer program, 'FDASH', to calculate H'_S, H'_T, and F'_{ST}, along with H^{Phen}, H^{SW}, and their associated differentiation statistics using allelic phenotype data (Appendix 8.13 page 209).

4.5 THE PROPERTIES OF PHENOTYPE-BASED STATISTICS

Before inferences can be drawn from statistics calculated from allelic phenotype data, it is important to identify how they respond to differences in polyploid level and migration; i.e. whether F'_{ST} is comparable between diploid and polyploids, and whether it is qualitatively similar to F_{ST} in its response to migration. It is also interesting to compare allelic phenotype statistics to those calculated from genotypic data, since statistics based on allelic phenotypes clearly contain less information than those based on allele frequencies, as the allele copy-number is 'hidden'. They are therefore likely to be poorer descriptors of population structure than statistics based on allele or genotype frequencies, such as θ or G_{ST}. Additionally, since they are calculated simultaneously from several independent isoloci, it seems likely that they will be affected by the number of isoloci present (i.e. the polyploid level) and the degree of differentiation between the isoloci.

One way to address the differences between genotype and phenotype-based statistics, in terms of their response to polyploid level, their deviation from the expectation of F_{ST}, and their evolutionary and sampling variance, is to simulate populations with different levels of polyploidy and different rates of migration. Here, two main effects are of interest: (1) the effect of disomic polyploidy on allelic phenotype-based measures of genetic diversity and differentiation; and (2) the relative quality of the phenotype-based differentiation statistic F'_{ST} as an estimator of F_{ST}, compared to genotype-based estimators such as θ.

In assessing the F'_{ST} statistic I use Wright's Island Model of population structure, which has the benefit of being simple, and being a model that is often implicitly used for inference in diploids (discussed in Whitlock and McCauley 1999). This model divides the total population into discrete demes (of size N), between which migration occurs, such that each generation a proportion (m) of individuals in each deme are drawn randomly from the other demes. Computer simulations of such models are usually done using a forward-time individual-based approach (e.g. Balloux and Goudet 2002), but here I choose to use a coalescent framework (see below).

The coalescent is an explicitly genealogical approach to population genetics, which considers genetic variation in terms of the common ancestors that link sampled genes. Moving backward through time, coalescences (corresponding to the occurrence of common ancestors) occur between pairs of lineages at a rate that depends on the

number of remaining lineages, and population parameters such as the effective population size. This approach is widely used to obtain results regarding evolutionary and ecological genetics (Hudson 1990; Nordborg and Donnelly 1997; Nordborg 2001; Rousset 2003). As a framework for computer simulation of population genetics, it has the advantage of being more efficient than individual-based simulation (Hudson 1990).

4.5.1 The coalescent process on two timescales

For an introduction to the coalescent, see Hudson (1990). To summarise, for the simple case of a single locus in a structured (diploid) population, the process starts with a sample of individuals from each of several different demes; each (diploid) individual consisting of a pair of gene lineages. A genealogy for this sample is constructed, moving back in time, according to the rate of migration, and the effective population size of demes. Lineages coalesce (i.e. share a common ancestor) within demes, or migrate to different demes, with the time between migrations and coalescences being exponentially distributed according to the migration rate and the number of lineages remaining. Mutations are distributed randomly on the tree, with the mean number of mutations per branch being proportional to its length.

Migration is usually a very fast process compared to mutation, and, if the number of sampled demes is small compared to the total number of demes (such that migration events carry lineages into demes that were not sampled), it is likely that all the lineages being followed during the simulation will be in different demes before any mutations occur (e.g. Wakeley and Aliacar 2001). These assumptions allow the coalescent process to be separated into two parts; a rapid phase in which lineages coalesce within demes or migrate out of them, called the ‘scattering’ phase; and a slow phase in which lineages from different demes coalesce and mutations occur, called the ‘collecting’ phase (e.g. Wakeley and Aliacar 2001). The collecting phase can be treated as a coalescent in a single (unstructured) population, with the effective population size modified to account for population structure (Wakeley and Aliacar 2001; Rousset 2003). This approach to the coalescent is shown in Figure 4.1.

4.5.2 Including polyploidy with disomic inheritance

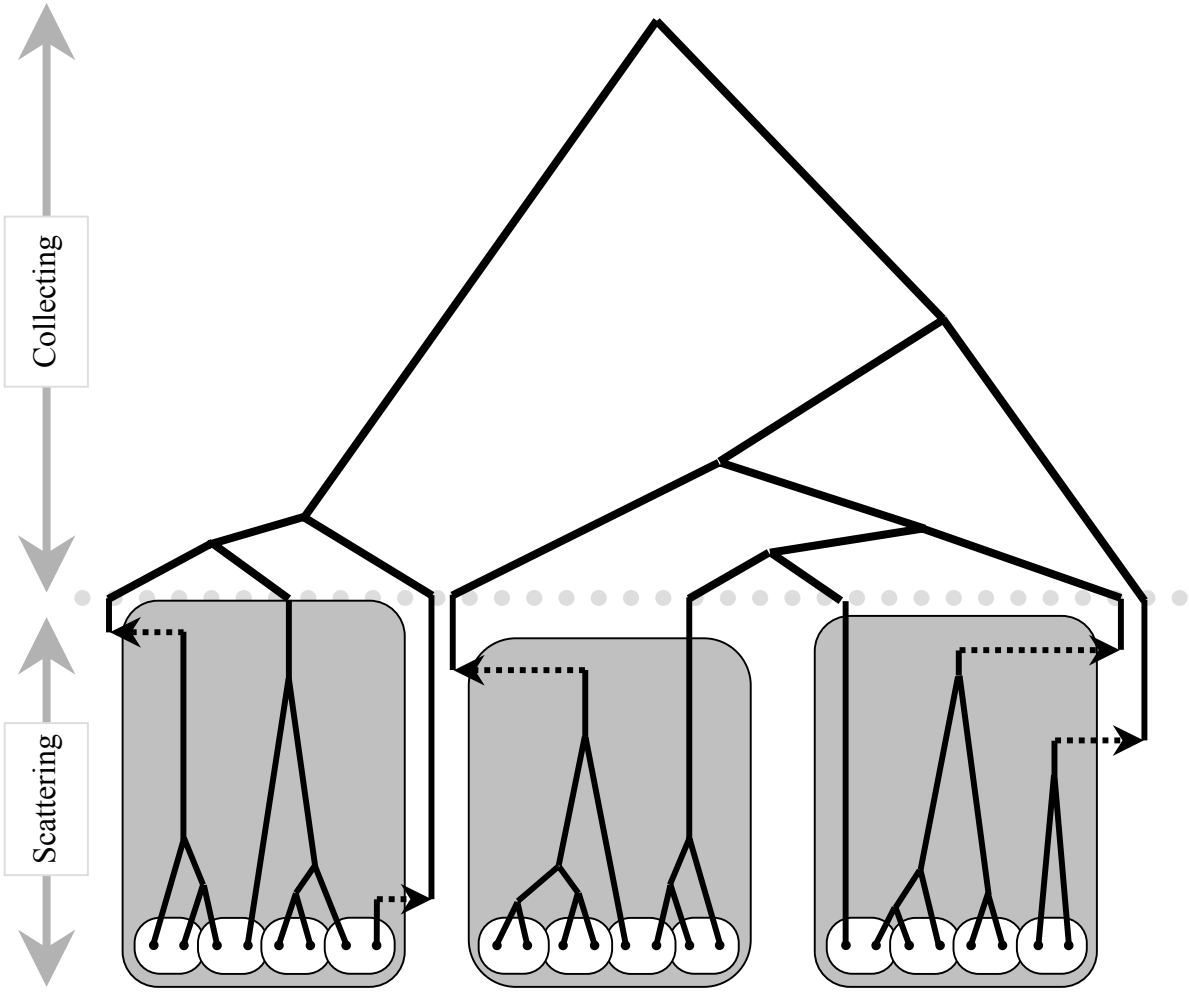
For polyploids with disomic inheritance, an important extension to the standard approach outlined above needs to be made. This is to allow for multiple isoloci, which

are independent at the time of sampling, but share a common ancestral locus in the distant past. The aim is to allow isoloci to share alleles through common ancestry. In this modified coalescent model, the initial sample of lineages consists of 2*x*-ploid individuals rather than diploid individuals (i.e. $x > 1$). Thus, the scattering phase is simulated as before, but with a simultaneous coalescent process for each of the *x* different isoloci in each deme. The collecting phase starts as before (though with *x* simultaneous coalescent processes), but after some threshold time coalescences can occur between lineages from different isoloci. In forward time, this is equivalent to polysomic inheritance having become disomic, as through diploidisation, at some point in the past. The threshold time, at which the inheritance model changes, controls the extent to which isoloci share alleles by descent. If it is in the very distant past, they will share no alleles and the markers will be effectively diploid (paleopolyploid), but, if it is in the recent past then isoloci will share alleles, and banding patterns may look superficially like polysomic inheritance, i.e. with the appearance of frequent homozygotes.

4.5.3 Computer simulation of the structured coalescent for disomic polyploids

To examine the statistical properties of F'_{ST}, I used a computer to simulate the structured coalescent for polyploids with disomic inheritance. In each run of the simulation, a sample of several 2*x*-ploid individuals was taken from each of several demes. The scattering phase for each of the *x* isoloci was constructed for each deme independently. Coalescences occurred within the deme and migrations carried lineages out of the deme (into an un-sampled deme). The scattering phase ended when all lineages were in different demes. Parallel collecting-phase coalescents (one for each of the *x* isoloci) were then constructed for all of the lineages that remained at the end of

Figure 4.1: A structured coalescent in two timescales



A single coalescent tree for a sample of four diploid individuals from each of three demes. The sample is depicted at the bottom of the figure (oval individuals in grey demes), with their common ancestor at the top. Moving back in time, coalescences happen within demes, and migration events (dotted arrows) move lineages out of the deme in which they were sampled. Migration is a much faster process than mutation, and if only a small proportion of demes are sampled, migration events will carry all the lineages into 'un-sampled' demes before any mutations occur. This allows the coalescent process to be separated into two independent phases. In scattering phase lineages only coalesce within demes, or migrate out of them. Once all lineages are in different demes the collecting phase begins. This is like the coalescent process in a single population, with the effect of migration accounted for by a change in the effective population size.

the scattering phase. After the inheritance transition-time was reached, and inheritance was no longer disomic, coalescences could occur between the different isoloci. Once the tree was complete, a Poisson-distributed number of mutations were applied to each branch, according to the branch length and the mutation rate. Using an infinite-alleles model of mutation (as appropriate to isozymes) the allelic state of each of the sampled alleles was identified, and diversity and differentiation statistics calculated for the sample. This process was repeated 20,000 times to obtain an estimate of the mean and variance of each summary statistic.

4.5.4 Calculation of statistics

Differentiation statistics based on phenotype frequencies were calculated from equations (4.1) to (4.3). F'_{ST} was calculated using equations (4.3) and (4.5). A genotype-based estimate of F_{ST} (θ) was calculated as described by Weir (1996), with multi-locus (i.e. multi-isolocus) estimates calculated as a ratio of averages (Weir 1996). The same genotype-based statistic was also calculated as if the polyploid had polysomic inheritance, i.e. a single locus with four alleles rather than two isoloci with two alleles each (Ronfort *et al*. 1998). The expectation of F_{ST} was calculated for the island model as $E[F_{ST}]=1/(1+4Nm)$. As a measure of the quality of F_{ST} estimators, I follow Balloux and Goudet (2002) in using the mean square error of estimates (MSE), calculated as the sum of squared bias and the variance ($bias^2 + var$).

4.5.5 Model parameters

To examine how polyploid level and differentiation between isoloci affects H'_T and F'_{ST}, I simulated a structured population according to Wrights Island Model, with 500 demes each of 250 (polyploid) individuals. Migration was assumed to be haploid (as by pollen). This was done for diploids, tetraploids, and hexaploids, with three different levels of divergence between isoloci. The divergence times between isoloci were selected such that when divergence time was low (divergence time = 0.01 x $2N_e$ generations), most alleles in the sample occurred at all isoloci, and when divergence was high (divergence time = 100 x $2N_e$ generations), alleles almost never occurred at more than one of the isoloci. Illustrative migration and mutation rates were selected so that diploid estimates of allele number and differentiation were similar to those actually seen in plants. To this end, ten demes were sampled, and the migration rate

was set to $m = 0.0062$ and the mutation rate to $\mu = 5.7\text{x}10^{-5}$ (found by trial and error), so that the number of distinct alleles observed in the sample ($A = 1.99$) and the differentiation between demes ($F_{ST} = 0.197$) matched those reported as an average for isozymes in outcrossing plants by Hamrick and Godt (1990).

To examine the relative utility of differentiation statistics based on genotype and allelic phenotype data, I again simulated a structured population according to Wright's Island Model (500 demes each of 250 individuals), but for tetraploids with one (intermediate) level of divergence between isoloci. The simulation covered a range of migration rates (expected F_{ST} values between 0.02 and 0.99), with the mutation rates as given above.

For all simulations, the sample from which statistics were calculated consisted of 25 individuals drawn from each of 10 demes. For each parameter combination, the simulation was repeated 20,000 times to obtain the mean and variance of the statistics in question.

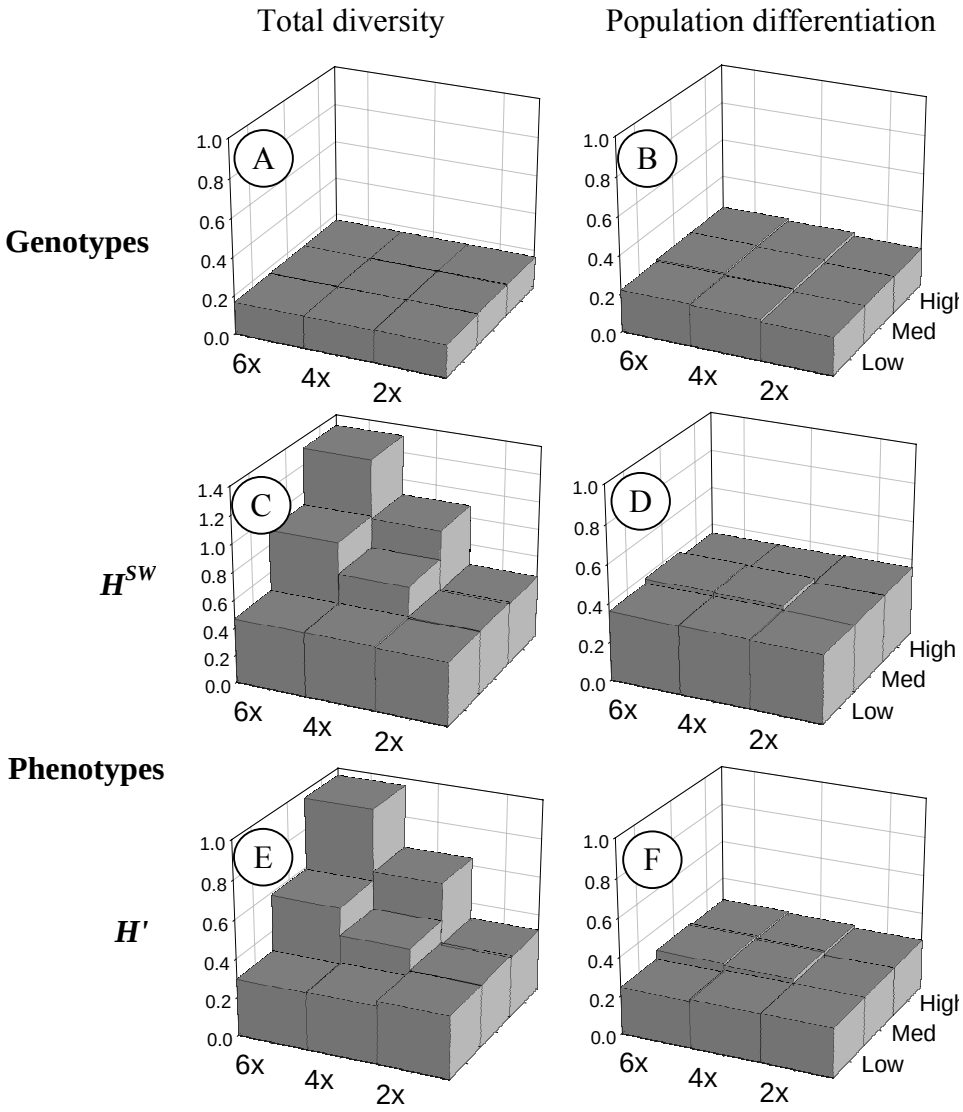
4.5.6 Simulation results

4.5.6.1 The effect of polyploidy

As expected, the genotype-based genetic diversity statistic, H_T, calculated as an average across isoloci, did not vary with increasing polyploid level or increasing differentiation between isoloci (Figure 4.2A). The genetic diversity statistics based on phenotype data (i.e. the joint diversity across multiple isoloci), increased with the level of polyploidy, (i.e. the number of isoloci) and the degree of differentiation between isoloci. This was true of Shannon-Weaver phenotype diversity, calculated from phenotype frequencies, and H', the unshared alleles measure of diversity (Figure 4.2C and Figure 4.2E respectively).

The genotype-based differentiation statistic (θ, calculated across isoloci) did not vary appreciably with polyploid level or differentiation between isoloci (Figure 4.2B). Although there was some variation in the differentiation statistics calculated from phenotypes (based on Shannon-Weaver phenotype diversity, and H', the number of alleles by which individuals differ) the effect was very small, and neither of these

Figure 4.2: Effect of disomic polyploidy on diversity and differentiation statistics



Diversity (graphs A, C and E) and differentiation (graphs B, D, and F) statistics are shown for polyploids with disomic inheritance, under an island model of population structure. Samples of 250 individuals (25 each from 10 demes) were drawn from a structured population of 500 demes of 250 individuals; values are the average of 20,000 replicates. Statistics were calculated from genotypic data (A and B), Shannon-Weaver diversity of phenotypes (C and D), and the average number of allele-differences between pairs of individuals (E and F), and are plotted with respect to polyploid level (2x-6x) and differentiation between isoloci. Phenotype-based diversity increased with polyploid level and differentiation between isoloci, while differentiation was largely unaffected by polyploidy.

differentiation statistics was strongly affected by disomic polyploidy (Figure 4.2D and Figure 4.2F respectively).

4.5.6.2 The utility of different differentiation statistics

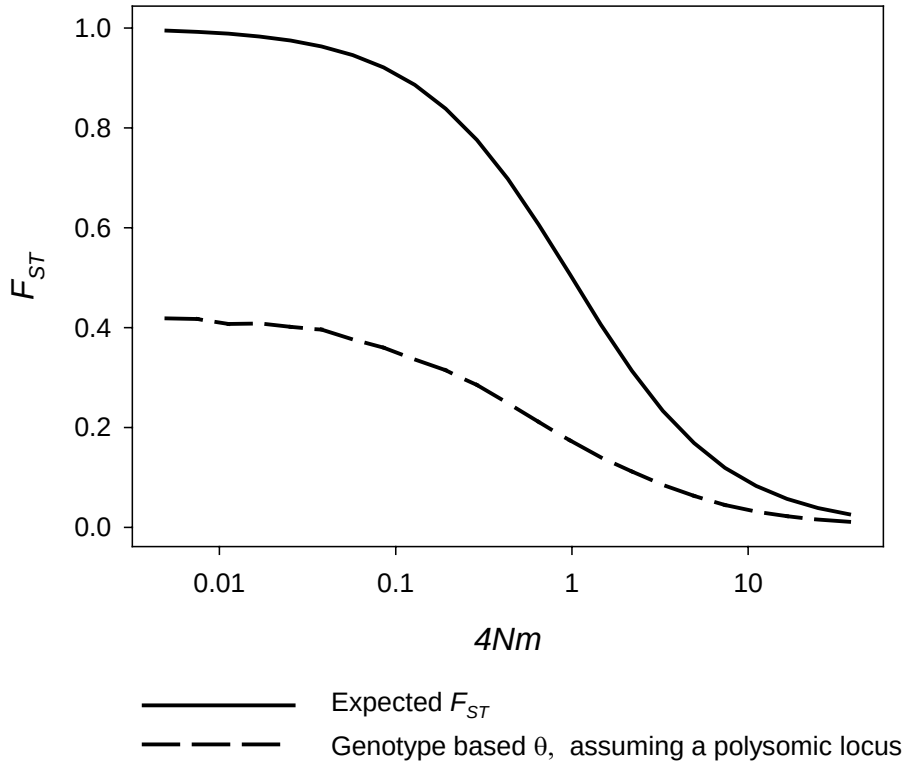
Three of the differentiation statistics deviated qualitatively from F_{ST} with respect to migration rate. When the polysomic genotype-based estimate θ (Ronfort *et al*. 1998) was calculated for a tetraploid genome with disomic inheritance, it did not tend toward one as the migration rate increased (Figure 4.3). Similarly, the differentiation statistic calculated from phenotype diversity H^{SW} did not approach zero as the migration rate increased (Figure 4.4). This was also true, but to a lesser extent, for the differentiation statistic calculated from H^{Phen}.

To examine the relative loss of information associated with the use of allelic phenotype data in place of genotype data, the phenotype-based differentiation statistic, F'_{ST}, was considered as an estimator of parametric F_{ST}. As expected, the tetraploid genotypic estimate (two isoloci, dashed line in Figure 4.5) was always better than the diploid genotypic estimate (one locus, dot-dash line in Figure 4.5). Under the parameters examined here, the tetraploid genotype-based statistic (calculated for both isoloci of the tetraploid) was a better estimator than F'_{ST} (the unshared-alleles statistic) when differentiation was low ($F_{ST} < 0.6$). However, when differentiation was high ($F_{ST} > 0.6$), F'_{ST} was a marginally better estimator (Figure 4.5). When $F_{ST} < 0.4$, the phenotype-based estimate was worse than the diploid genotype-based estimate (Figure 4.5).

4.6 DISCUSSION

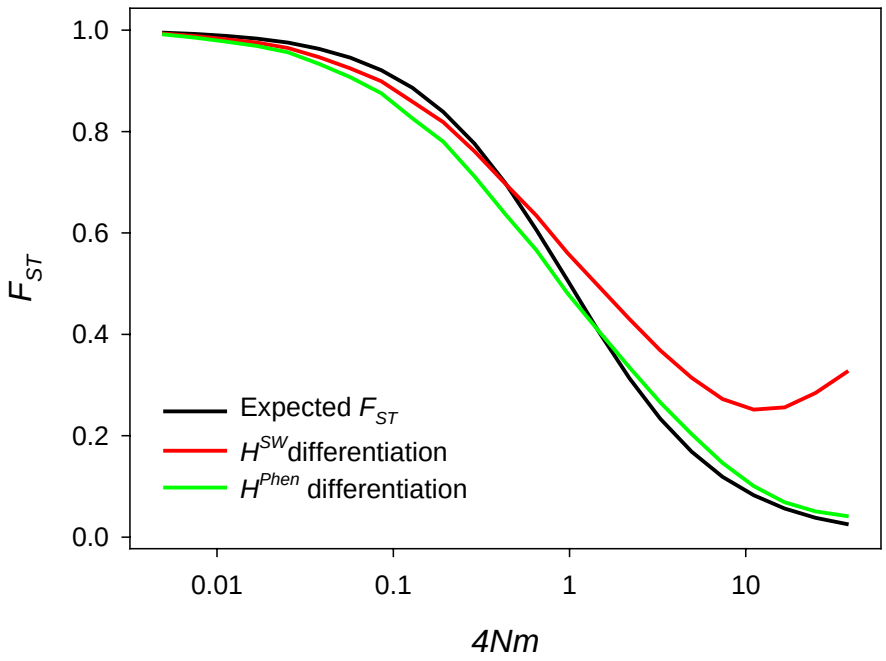
Polyploids with disomic inheritance are genetically diploid, so the framework for inference of population processes from genetic summary statistics is essentially the same as it is in diploids. Unfortunately, because duplicate loci co-migrate during electrophoresis, genotypic data is often unavailable, since genotypes cannot be inferred from banding patterns. When this is the case, diversity and differentiation statistics can be calculated using allelic phenotypes. However, it is not clear that these statistics are

Figure 4.3: The effect of assuming polysomic inheritance



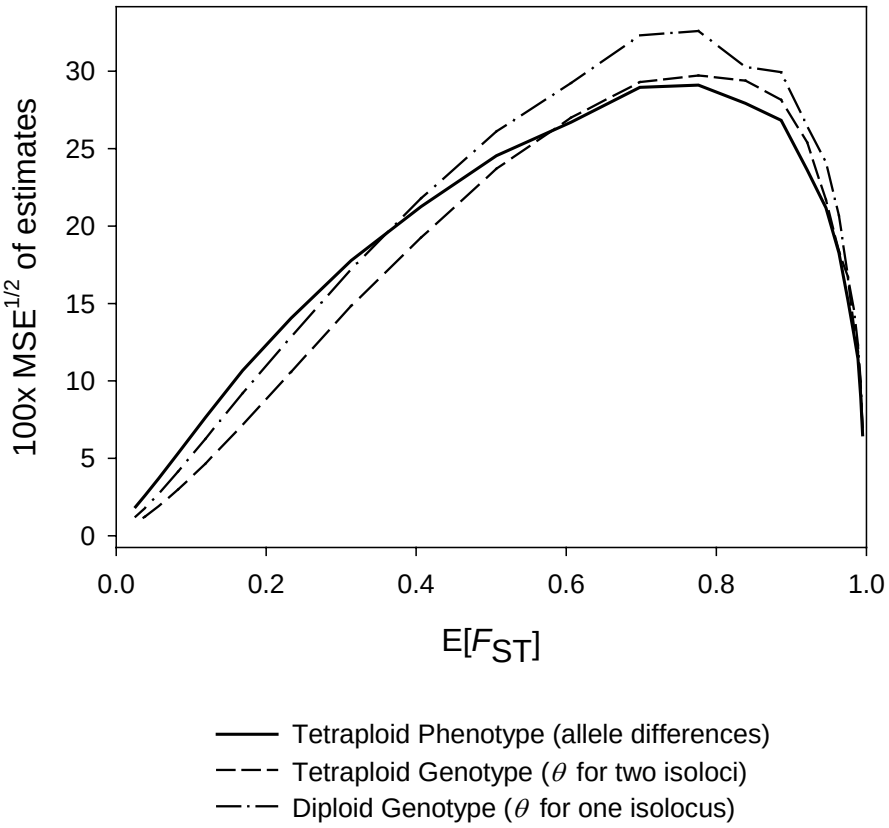
Samples of 250 individuals (25 each from 10 demes) were drawn from an island-model structured population of 500 demes, each of 250 tetraploid individuals; values are the average of 20,000 replicates. If the polyploid equivalent (Ronfort *et al*. 1998) of the differentiation statistic θ is inappropriately calculated as if inheritance were polysomic (i.e. each locus having four alleles), rather than disomic, (i.e. two isoloci with two alleles each) , θ does not behave as expected (dashed line versus solid line). This is because apparent heterozygosity, actually due to differences between isoloci, is treated as if it were genuine diversity, thereby inflating subpopulation diversity, and leading to low estimates of differentiation (see main text).

Figure 4.4: Differentiation calculated from phenotype diversity



Samples of 250 individuals (25 each from 10 demes) were drawn from an island-model structured population of 500 demes, each of 250 tetraploid individuals; values are the average of 20,000 replicates. Differentiation statistics, calculated from H^{Phen} and H^{SW} are plotted along with the expected value of F_{ST}. Neither is asymptotic to zero as migration rates increase (see main text for details)

Figure 4.5: Differentiation statistics as estimators of F_{ST}



The MSE (with respect to expected F_{ST}) of genotype-based (θ) and phenotype based (F'_{ST}) differentiation statistics, are plotted for a range of expected F_{ST} values (high to low migration rates). Statistics are calculated from 20,000 replicates, Samples were of 250 individuals (25 from each of 10 demes), drawn from an island-model structured population of 500 demes, each of 250 tetraploid individuals. Under the parameters used here (see main text), the information-loss associated with using allelic phenotypes in place of genotypic data outweighs the gain associated with the presence of more isoloci (solid versus dashed lines), when migration is high. When migration rates are low, phenotype and genotype-based differentiation statistics are approximately equal in their ability to estimate F_{ST}.

as informative regarding processes such as migrations as genotype-based statistics.

In the simulations presented here, phenotype-based diversity statistics for polyploids with disomic inheritance were strongly dependent on details such as the polyploid level (number of isoloci) and the differentiation between isoloci (Figure 4.2). However, differentiation statistics were not strongly affected by polyploid level (Figure 4.2). Differentiation statistics, when calculated as if inheritance were polysomic, and when calculated from allelic phenotype diversity, differed qualitatively from the expectation of F_{ST} with respect to migration rate (Figure 4.3 and Figure 4.4). Below, I discuss the likely reason for these effects, and the implications for quantifying diversity and differentiation in polyploids with disomic inheritance.

4.6.1 The need for a correct model of inheritance

When there is fixed heterozygosity, it is clear that inheritance must be disomic. However, when isoloci share a large proportion of their alleles (i.e. in the hexaploid *aa*|*ab*|*cc*, allele *a* is shared by two isoloci), the great inter-individual variation in the number of distinct alleles can make gel banding-patterns look superficially as if inheritance is polysomic. There is a danger that, if allele copy number can be identified, these data may be analysed using computer packages intended for autopolyploids (e.g. SPAGEDi: Ronfort *et al.* 1998; Hardy and Vekemans 2002). This procedure is inappropriate, because the apparent excess of heterozygotes (due to disomic inheritance) will inflate within-population diversity (H_S), so that it is non-zero even when there are no differences between individuals within populations. If polysomic inheritance is assumed, the analogue of θ (Ronfort *et al*. 1998) may be small in a polyploid with disomic inheritance, even when migration rates are almost zero (Figure 4.3).

4.6.2 Phenotype-based genetic diversity in disomic polyploids

Phenotype-based diversity statistics are strongly dependent on the number of isoloci (i.e. the polyploid level) and the degree of differentiation between isoloci (Figure 4.2). This is expected, because phenotype-based diversity statistics are simultaneously recording the diversity at several duplicate isoloci. If isoloci were to share no alleles, the overall phenotype diversity would be an additive function of diversity at each of the (diploid) isoloci, and thus increase with the polyploid level. By contrast, genetic

differentiation statistics do not vary much with polyploid level, given that other population parameters are the same (Figure 4.2). This is because differentiation statistics, such as F_{ST}, are a ratio of within-population diversity to total diversity (e.g. Hartl and Clark 1997), and they will be largely unaffected by factors that simultaneously increase both. This means that, while direct comparisons of diversity statistics such as H' and H^{SW} cannot be made between polyploid levels, comparisons of differentiation statistics derived from them are probably valid.

Some of the phenotype-based differentiation statistics behave unexpectedly in response to migration, i.e. they are qualitatively different to F_{ST} or genotype-based statistics (Figure 4.4). In particular, differentiation statistics calculated from diversity based on phenotype frequencies, such as H^{SW} and H^{Phen}, are not asymptotic to zero as migration rates increase towards panmixis. This is probably an effect of sample size. F_{ST} may be considered as a standardised variance in allele frequencies between populations (e.g. Weir 1996), and, for a given sample size, the variance in phenotype frequencies is larger than the variance in allele-frequencies. This is because alleles will be distributed differently between individuals in different samples, and unless the sample is very large, many rare phenotypes will not included. The effect is particularly strong when differentiation is based on Shannon-Weaver diversity, because this diversity index weights rare phenotypes disproportionately highly. These results suggest that inference regarding relative migration rates, when based on differentiation statistics calculated from phenotype frequencies, should be treated with caution, as even panmictic gene flow is likely to give differentiation statistics much greater than zero. F'_{ST}, the differentiation statistic based on allele-differences, does not suffer from this limitation.

It is interesting to ask whether the information lost when allelic phenotype data are used instead of genotype data outweighs the information-gain available from the presence of more (iso)loci. Under the parameter ranges examined above, there was an overall loss in information when migration rates were high (i.e. low differentiation), but a slight gain in information when migration rates were low. Since both θ and F'_{ST} are worse estimators of F_{ST} when migration rates are low, F'_{ST} appears to be a relatively good statistic for differentiation across a range of migration rates (Figure 4.5).

4.6.3 Polyploid population genetics

Here I have addressed the behaviour of phenotype-based statistics in a simple island model, for outcrossing polyploids with disomic inheritance. Under more complicated models of population-structure and genetic-system, important questions remain. For example, it is likely that phenotype-based statistics will behave differently under polysomic inheritance, as well as when other population processes, such as selfing or local extinction, are included. The coalescent approach to simulation used here is ideally suited to making these extensions, as both processes can easily be incorporated (Nordborg and Donnelly 1997; Wakeley and Aliacar 2001).

Although the quantification of genetic diversity and differentiation in polyploid organisms is more problematic than in diploids, there are now approaches that can be used for several different classes of polyploid. If genotypes can be scored, standard statistics such as H_e and θ can be applied to polyploids with polysomic inheritance (Ronfort *et al.* 1998; Hardy and Vekemans 2002), and if genotypes cannot be scored, allelic phenotypes can provide suitable alternative statistics, at least under disomic inheritance where inference is the same as for diploids. I have shown that, for many purposes, the diversity statistic H' is an informative way of summarising genetic diversity, and that the differentiation statistic derived from it (F'_{ST}) behaves in a way very similar to other, more widely used, differentiation statistics. Furthermore, F'_{ST} is affected very little by polyploid level in polyploids with disomic inheritance, making values comparable between polyploid levels.

5 EVIDENCE FOR METAPOPULATION PROCESS AND POST-GLACIAL MIGRATION

5.1 INTRODUCTION

Sexual-system evolution is of particular interest because the sexual-system controls the way genes are passed from one generation to the next. The flowering plants exhibit an enormous diversity of sexual systems (Richards 1997; Barrett 2002), making them a valuable study group for sexual-system evolution. In particular, questions regarding selection for combined versus separate sexes can be asked at a range of taxonomic levels, because the incidence of dioecy is widely scattered across the angiosperm phylogeny (Renner and Ricklefs 1995). The frequency, polarity and ecological correlates of sexual-system transitions have been identified by phylogenetic approaches (e.g. Weller and Sakai 1999; Weiblen *et al.* 2000; Vamosi *et al.* 2003); theoretical and experimental studies have identified the selective forces likely to bring such changes about (reviewed in Charlesworth 1999; Webb 1999). As dioecious individuals are necessarily outcrossing, and as functionally hermaphroditic (e.g. monoecious) individuals have at least the potential to self fertilise, selection for combined versus separate sexes can depend partly on selection for the mating-system. For example, selection for reproductive assurance can favour self-fertile hermaphrodites (e.g., Pannell and Barrett 1998), whilst the avoidance of inbreeding depression has probably been an important force in the evolution of dioecy (e.g., Charlesworth and Charlesworth 1979; Miller and Venable 2000).

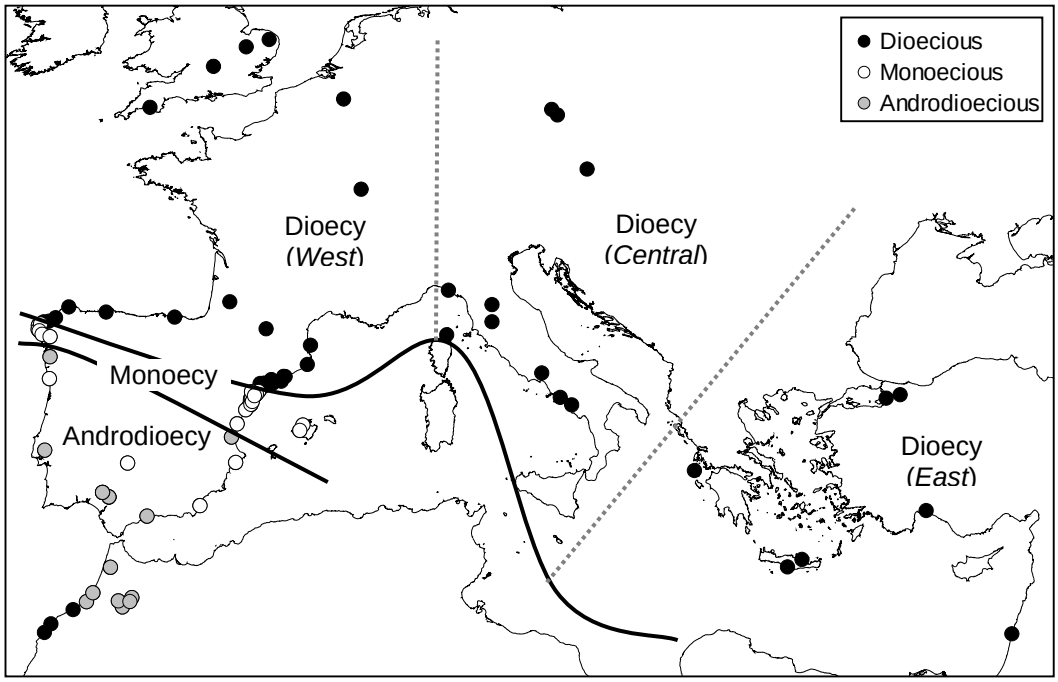
Amongst the most valuable study systems for sexual-system evolution are groups of closely related lineages that share similar life history and ecological traits but differ in sexual system (e.g. Dorken *et al*. 2002). Close relatedness reduces the number and extent of confounding traits, and allows predictions regarding evolutionary and ecological differences between sexual systems to be tested. Such systems are scarce, but those that have been studied have proved

valuable in understanding gender variation, life-history variation, patterns of genetic diversity and phylogenetic transitions in the sexual system; e.g. *Ecballium elaterium* (Costich and Meagher 1992, 2001) and *Sagittaria latifolia* (Dorken *et al.* 2002; Dorken and Barrett 2003, 2004).

The *Mercurialis annua* L. (Euphorbiaceae) polyploid complex, which presents both monoecious and dioecious populations, is another rare example of closely related lineages with different sexual systems (Thomas 1958; Durand 1963; Durand and Durand 1985; Krahenbuhl *et al.* 2002). *Mercurialis annua* is a wind-pollinated annual ruderal, native to central and Western Europe and the Mediterranean basin. It is dioecious in the north and east, and strictly monoecious in large areas of Iberia (Figure 5.1). From an experimental point of view, it is a particularly valuable model because sexual-system clines are replicated on the east and northwest coasts of Iberia (Figure 5.2). Although dioecious populations are diploid and monoecious populations are hexaploid, other differences in life history and vegetative morphology are very slight (Durand and Durand 1985). What makes *M. annua* an exceptional model for the study of sexual-system ecology and evolution, is the widespread occurrence of males in the otherwise monoecious populations of southern Iberia and north Africa (Durand 1963; Pannell 1997c)(Figure 5.2).

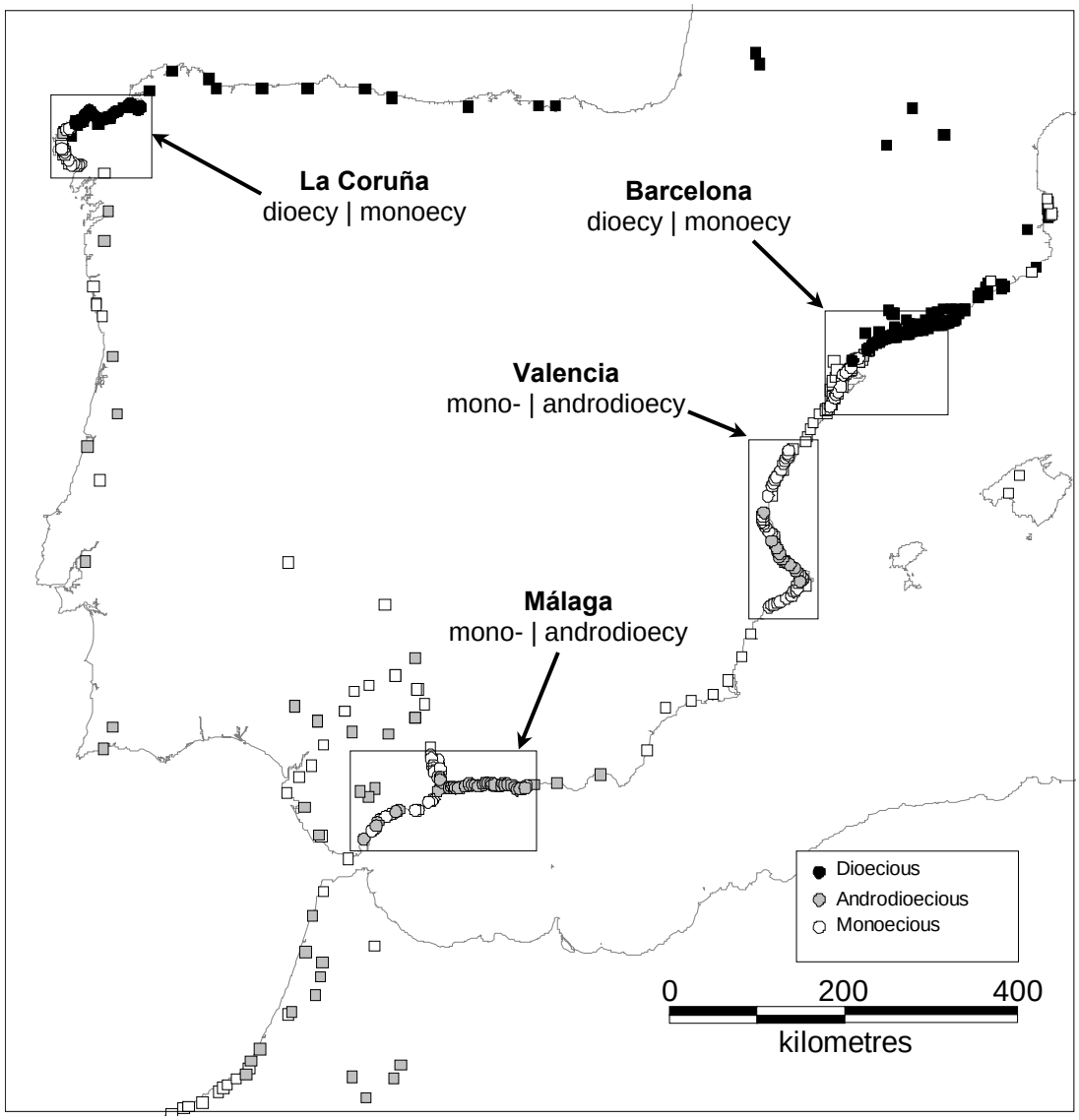
The co-occurrence of males and hermaphrodites (androdioecy) is a particularly rare sexual-system, previously thought to be a potential (but unlikely) intermediate step in the evolution of dioecy from hermaphroditism (Charlesworth 1984). Its rarity is thought to derive from the need for males to have more than twice the siring success of hermaphrodites if they are to invade a hermaphroditic population (Charlesworth and Charlesworth 1978; Charlesworth 1984). Although simple models imply highly restrictive conditions for the evolution of androdioecy from hermaphroditism, it has been shown more recently that pollen limitation can facilitate the evolution of androdioecy from dioecy (Wolf and Takebayashi 2004). In *M. annua*, the wide range of sexual-systems requires a model than can explain the maintenance and distribution of monoecy, dioecy and androdioecy. It has been hypothesised that in colonising species (such as *M. annua*), androdioecy may be favoured by a balance between selection for reproductive assurance during colonisation, and selection for gender

Figure 5.1: Distribution of sexual systems in *Mercurialis annua*



Mercurialis annua occurs all over northern Europe and around the Mediterranean Basin. In the north and east of this range it is dioecious and diploid; while in Iberia and North Africa it is monoecious (and androdioecious) and polyploid. Circles indicate the location of seed collections used in this study (see Appendix 8.7). Regions marked "Dioecy", "Monoecy", and "Androdioecy" denote zones of different sexual system (see text). The "west", "central" and "east" subdivisions of the dioecious zone are arbitrary regions used to illustrate the effect of range expansion on genetic diversity.

Figure 5.2: Fine scale distribution of sexual systems in Iberia and Morocco



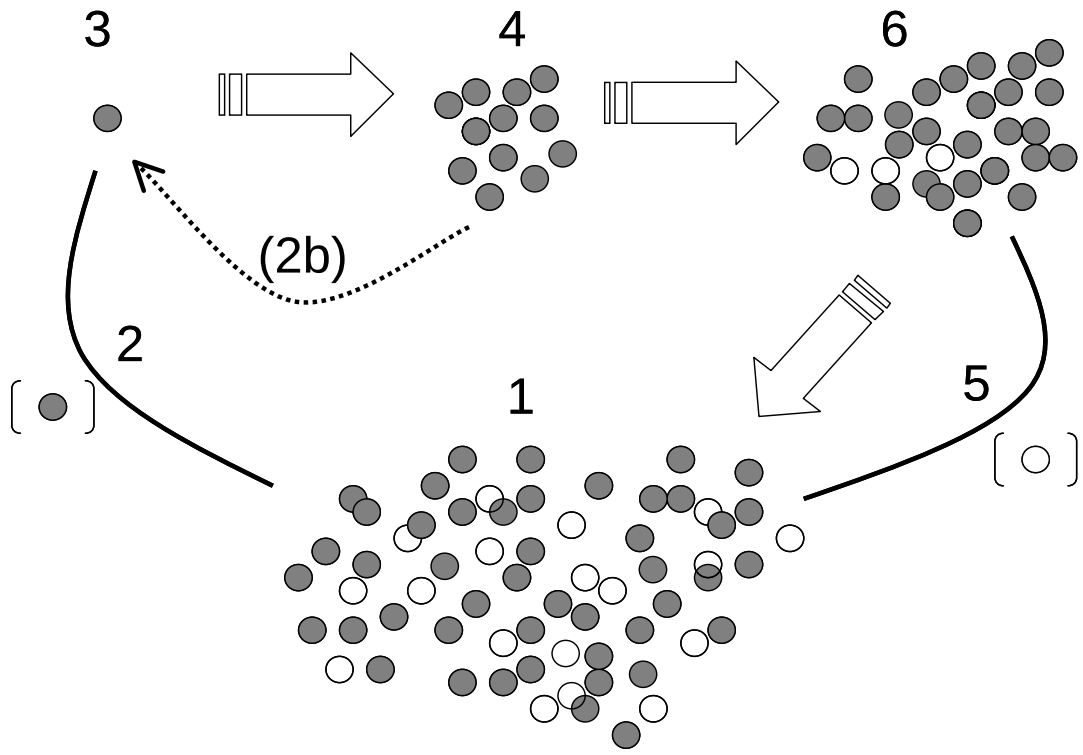
Squares indicate populations visited in the course of seed-collections, circles represent populations visited for demographic surveys. In the north of Iberia, *M. annua* populations are dioecious and diploid, in the south they are polyploid and either monoecious or androdioecious. Demographic surveys were performed across the sexual-system transitions marked by solid rectangles: (1) La Coruña on the west coast, (2) Barcelona on the east coast, (3) Valencia in the southeast and (4) Malaga on the south coast. The transitions between monoecy and dioecy are very abrupt, overlapping for less than 10 km at the La Coruña transition and (with the exception of scattered monoecious populations in the far north) less than 20 km on the Barcelona transition. [The demographic survey presented here is unpublished data provided by SM Eppley].

specialisation in large populations (Pannell 2001; Pannell 2002) (Figure 5.3).

Recurrent local extinction and recolonisation is a defining feature of metapopulations. In a metapopulation context, hermaphrodites capable of self-fertilisation have an advantage over males or females, which are unable to found new populations alone ('Baker's Law', Baker 1955; Pannell and Barrett 1998). This can lead to selection for self-compatible hermaphroditism at the level of the metapopulation (Pannell and Barrett 1998; Barrett and Pannell 1999). However, because the selfing rate is likely to be context-dependent (e.g. Routley *et al*. 1999; Vogler and Stephenson 2001), older and larger populations may have lower selfing rates, allowing specialist males with sufficient siring success (e.g. through greater pollen production and dispersal), to invade. Over ecological timescales, the metapopulation is stably androdioecious (Pannell 2001); in this hypothetical scenario, new populations are frequently founded by selfing hermaphrodites, but as each new population grows, the selfing rate decreases and immigrant males are able to invade from large long-established populations (Figure 5.3). In areas with higher levels of population-turnover the model is short-circuited (dotted line 2b in Figure 5.3), populations never persist long enough for males to gain a foothold (i.e. step 5 of Figure 5.3 never has an opportunity to occur), colonists come from monoecious populations (population marked ‘4’ in Figure 5.3), creating regions of strict monoecy.

This verbal model makes testable predictions about population structure, sex ratios, and patterns of genetic diversity in androdioecious species (Pannell 2001). Comprehensive surveys of population structure and species-wide genetic diversity can provide indirect evidence to test the model. High population turnover is expected to be associated with lower site occupancy (e.g. Levins 1970; Hanski 1997) and smaller (Hanski and Gilpin 1997), younger (Wade and McCauley 1988), populations. According to the model outlined above, this suggests that androdioecious populations will occur in more densely populated regions that have larger populations. Metapopulation processes will also leave a mark in patterns of genetic diversity. If extinction rates are sufficiently larger than migration rates, it is expected that regular local extinction will reduce both total genetic diversity and within-population diversity (reviewed in Pannell and Charlesworth 2000). Although genetic differentiation between demes may, in

Figure 5.3: A graphical depiction of a model for the maintenance of monoecy and androdioecy in a metapopulation



Males and hermaphrodites are symbolised by open and closed circles, respectively; block arrows represent population growth; simple arrows represent dispersal. According to the model, established populations (1) disperse male and hermaphrodite propagules across the metapopulation (2 and 5, respectively). Only hermaphrodites can self, and thus establish populations (2). New populations (3) are small and sparse, and are initially highly selfing (4). As populations grow, outcrossing with neighbours becomes easier and density-dependent selfing rates decline, allowing males to invade (5). A balance between extinction and recolonisation maintains the age structure of the metapopulation. If population turnover is very high, populations do not persist long enough for males to immigrate; (5) does not occur. The largest populations are monoecious (4) rather than androdioecious (1 or 6), so colonists come from monoecious populations (2b) rather than androdioecious populations (2); the cycle is 'short-circuited', giving rise to monoecy rather than androdioecy. Adapted from Pannell (2003).

principle, increase or decrease, depending on whether founders come from one deme or many (Wade and McCauley 1988; Whitlock and McCauley 1990), empirical evidence suggests that differentiation is often increased by local extinction (e.g. Whitlock 1992; Antrobus and Lack 1994; Giles and Goudet 1997). Thus, patterns of lower regional diversity, and especially lower within-population diversity, plus higher genetic differentiation, are expected in monoecious regions compared to those in androdioecious or dioecious regions.

Of course, patterns of genetic diversity result not only from ongoing processes, but also from historic ones (Hewitt 1999; Charlesworth *et al.* 2003). In particular, differences between monoecious and dioecious lineages may only be interpretable in the broader context of post-glacial range-expansion. For example, in *Ecballium elaterium* genetic differentiation between monoecious and dioecious populations can be explained by the existence of separate glacial refugia (Costich and Meagher 1992). In Europe, many species survived the last glaciation in the extreme south. Iberia, southern Italy and the Balkans frequently providing the source populations for recolonisation of northern Europe (e.g. Hewitt 1999; Petit *et al*. 2003). Therefore, the possibility that monoecious and dioecious lineages have undergone recent range expansion from one or more of these refugia also needs to be examined.

Here I use two sources of evidence to test the metapopulation hypothesis for sexual-system distribution in hexaploid *M. annua*. Firstly, a survey of populations in Iberia, where sexual-system is polymorphic, is used to identify differences between sexual systems in terms of population size and abundance (unpublished data provided by SM Eppley and JR Pannell). Secondly, a species-wide survey of genetic diversity is used to identify differences in genetic diversity and differentiation. Finally, I relate species-wide genetic diversity to historic range expansion. In doing so, I identify the likely glacial refugia for monoecious and dioecious *M. annua*.

5.2 METHODS

5.2.1 Metapopulation survey

To identify differences in population size and abundance between sexual systems, a survey of natural populations was conducted in Spain in January and February of 2003 A total of four transects were made, two across transitions between dioecy and monoecy in northern Spain, and two across transitions between monoecy and androdioecy in south-eastern Spain (Figure 5.2). The first dioecy-monoecy transect was on the eastern Spanish coast south of Barcelona; 47 sites were surveyed along 187 km of the N340 from Sitges to Peníscola ("Barcelona" cline in Figure 5.2). The second transect was on the north western Spanish coast near La Coruña; 58 sites were surveyed along 217 km of the N651 and the C552 from north of La Coruña to Carballo; small coastal highways through Malpica and Muxia; and the C550 from Cée to Noia ("La Coruña" Figure 5.2). The first monoecy-androdioecy transect along the southeastern Spanish coast comprised 64 survey sites along 216 km of the N340 and N332 highways from Castellón to Vila Joiosa ("Valencia" Figure 5.2). The second transect in the province of Andalucia in southern Spain comprised 90 survey sites along 317 km on the N331 and N340 outward from Málaga (Figure 5.2).

To assess population abundance in the landscape, the proportion of survey sites at which *M. annua* was present were recorded. Survey sites were placed at regular intervals along roads in each transect, every 3.2 km for the dioecy-monoecy transects and every 3.0 km for the monoecy-androdioecy transects. Survey sites were chosen as the first available open space (i.e. no building, concrete, or manicured garden or park) that was large enough for a five-minute survey and was more than 3 km from the last site. At each site, an active search for *M. annua* plants was made for five minutes to establish presence/absence, then the following information was recorded: (1) GPS coordinates of the site; (2) whether *M. annua* was found within the five-minute survey; (3) an estimate of the total number of *M. annua* plants encountered while surveying the site; (4) the breeding system of the population (dioecious, monoecious, or androdioecious), androdioecious populations defined as those with at least 1% males; (5) the proportion of males in androdioecious populations.

For comparisons of population abundance in the dioecy-monoecy and androdioecy-monoecy transects, a logistic regression analysis was used to determine the effect of transect and regional breeding system on the likelihood that a population was found at each survey site. For comparisons between populations from dioecious and androdioecious areas, a logistic regression analysis was used to determine the effect of sexual system and transect (nested in sexual system) on the likelihood that a population was found at each survey site. In this case, transect was nested in sexual system because dioecious and androdioecious populations never occur along the same transect, as is the case for the other comparisons. The results are reported as likelihood-ratio Chi-squared statistics, with *p* values included. All statistical analyses were done using the JMP statistical package (SAS 2003).

For comparisons of population size in the dioecy-monoecy transects and the androdioecy-monoecy transects, mixed-model ANOVAs were used to determine whether the number of plants recorded for a population was affected by transect (random) or sexual system (fixed). For comparisons of dioecious and androdioecious populations, a mixed-model ANOVA was used to determine whether the number of plants in a population was affected by sexual system (fixed) or transect (random, nested in sexual system). In this case, transect was nested in sexual system rather than treated as a block, as in the other ANOVAs, because dioecious and androdioecious populations never occur along the same transect. In all of these ANOVAs, the data were log transformed in order to ensure that the residuals were normally distributed and variances were homogeneous (data not shown). All statistical analyses were done using the JMP statistical package (SAS 2003)

5.2.2 Sampling for isozyme survey

Bulk seed collections were made from populations across the species' natural range (Figure 5.1, Appendix 8.7). For sexual system comparisons, the range was divided into geographic zones according to the predominant sexual system (marked 'Dioecy', 'Androdioecy' and 'Monoecy' on Figure 5.1). One androdioecious population (0616a) was included within the "monoecious" zone, and within the androdioecious zone, there were a mix of androdioecious and

monoecious populations. Comparisons were made between these zones, and between monoecious and androdioecious populations regardless of zone. Androdioecious populations were conservatively defined as those in which at least 1% of individuals were male. To analyse the phylogeographic component of diversity (i.e. that part of the pattern that may result from post-glacial migration) in the dioecious populations, the dioecious zone was further divided into 'west', 'central' and 'east' (Figure 5.1).

Bulk seed collections of approximately 20-40 (minimum 10, maximum *ca.* 100) seed-bearing individuals were made, individuals being chosen haphazardly from each population. Plants for isozyme analysis were grown from seed, under glass in Oxford (UK), at various times between November 2000 and August 2003. Approximately 45 plants grown from each bulk seed collection were used in the isozyme analysis, averaging 48 individuals per dioecious population and 42 individuals per monoecious or androdioecious population. Isozyme data from 89 populations are reported here: 45 dioecious and 44 monoecious or androdioecious.

5.2.3 Isozyme extraction and visualisation

Six enzyme systems could be consistently resolved in diploid *M. annua*: AAT (aspartate aminotransferase E.C. 2.6.1.1), PGI (glucose-6-phosphate isomerase E.C. 5.3.1.9), PGM (two loci, phosphoglucomutase, E.C. 5.4.2.2), ME (malic enzyme $NADP^+$ E.C. 1.1.1.40), 6-PGD (two loci, phosphogluconate dehydrogenase E.C. 1.1.1.44), and IDH (isocitrate dehydrogenase E.C.1.1.1.42). Only five of these enzymes could be reliably interpreted in hexaploid *M. annua*: AAT, PGI, ME, 6-PGD and IDH. Of these, ME was monomorphic in both diploids and polyploids, resulting in eight potentially informative loci for dioecious populations and six for the monoecious and androdioecious populations.

Protocols and solutions are adapted from Wendel and Weeden (1990). For each individual, approximately 2 cm^2 of fresh young leaf tissue was ground with extraction buffer (1.21 g Tris-HCl, 0.04 g EDTA, 0.076 g KCl, 0.2g $MgCl_2.6H_20$, 4 g PVP, 0.5 g PVPP in 100 ml stock, made up with 90:10:0.5 stock:DMSO:ß-mercaptoethanol) on a pre-chilled ceramic block. The resulting

paste was soaked into 3mm by 8 mm Whatmann No. 3 paper wicks for loading into starch gels, made from 14% w/v hydrolysed potato starch (StarchArt Corporation, Texas) and 0.25% w/v sucrose. Three buffer systems were used. Lithium-borate gels (electrode buffer: 0.039 M LiOH, 0.263 M boric acid; gel buffer: 33 mM Tris-base, 5 mM citric acid, 4 mM LiOH, 30 mM boric acid, pH 7.6, lithium borate), stained for AAT, PGI, PGM and ME. Morpholine-citrate gels (electrode buffer: 0.04 M citric acid, 0.068 M N-(3-aminopropyl)-morpholine pH 6.4; gel buffer: 1:14 dilution of the electrode buffer), were stained for 6-PGD. Tris-citrate gels (electrode buffer: 0.135 M Tris-base, 0.03 M citric acid, pH 8.3; gel buffer: 1:19 dilution of the electrode buffer), stained for IDH. All gels were run at 4 ºC for approximately 6 hours. Lithium-borate gels were run at 300 V, morpholine-citrate gels at 250 V, and Tris-citrate at 150 V. Gels were cut into 1 mm thick slices and stained at 40 ºC. Staining solutions are adapted from Wendel and Weeden (1990). Reactions were stopped after staining, and gels were stabilised by removing stain solution and adding 30 ml 25% v/v glycerol. Gels were scored either from photographs or at the time of staining.

5.2.4 Isozyme data analysis

5.2.4.1 Diploid populations

I used published data on enzyme structure and compartmentalisation in plants (Weeden and Wendel 1990; Wendel and Weeden 1990), in conjunction with artificial crosses where necessary (data not shown), to make full genetic interpretations for diploid isozyme banding patterns. Nei's gene diversity H_e (expected heterozygosity, Nei 1987), allelic richness corrected for sample size (El Mousadik and Petit 1996), and estimates of F_{IS} and F_{ST} (Weir and Cockerham 1984; Weir 1996) were calculated using FSTAT version 2.9.3.2 (Goudet 1995). Genetic diversity and differentiation were calculated: (1) across all populations; and (2) for each geographic zone in Figure 5.1. Significant differences between geographic regions were assessed using randomisation tests. For comparison with hexaploids, F'_{ST} was also calculated (see below).

5.2.4.2 Hexaploid populations

In polyploids, gene duplication, multiple alleles and the mode of inheritance can lead to practical and statistical complications in scoring genetic data (e.g. Barrett and Shore 1989; Weeden and Wendel 1990; Meerts *et al.* 1998; Rogers 2000) and in interpreting summary statistics (e.g. Bever and Felber 1992; Ronfort *et al*. 1998). For polyploids with disomic inheritance, the expectations for underlying genetic diversity and differentiation are the same as for a diploid, given a model of population structure. However, duplicate isozyme loci ('isoloci', Waples 1988) often co-migrate on the gel, and have to be treated together as a single 'locus', preventing the calculation of standard diversity statistics such as Nei's (1987) gene diversity. Nevertheless, if the mode of inheritance is known, difficulties in scoring genetic data can largely be overcome using a approaches based on 'allelic phenotypes'(e.g. Murdy and Carter 1985; Bayer and Crawford 1986; Rogers 2000; Berglund and Westerbergh 2001). Allelic phenotypes take account of which alleles are present in each individual, but not the number of copies of each allele and to which duplicate locus each allele belongs. Allelic phenotype-based summary statistics can be calculated using the program FDASH (Chapter 4).

In summary, I defined a measure of genetic diversity for use with allelic phenotypes, H', as the average number of alleles by which pairs of individuals differ, i.e., a pair in which one individual carries alleles *abc* and the other carries alleles *bd* differ by three alleles, *a*, *c* and *d*. This diversity measure can be calculated both within local populations (H'_S, here the *average* within-population diversity) and across many populations (H'_T) allowing a measure of genetic differentiation F'_{ST} to be calculated as $(H'_T - H'_S) / H'_T$, (Chapter 4).

Hexaploid *M. annua* has disomic inheritance (Chapter 3) and allele-dosage could not be consistently estimated from isozyme gels. Genetic diversity indices (H'_S, H'_T, number of alleles per population, number of phenotypes per population) and genetic differentiation (F'_{ST}) were calculated for hexaploids using FDASH. For hexaploids, genetic diversity and differentiation were calculated: (1) for all populations; (2) for each of the two sexual-system zones (monoecy and androdioecy) in Figure 5.1; (3) for monoecious and

androdioecious populations regardless of sexual-system zone; (4) separately for the different sexual systems present within the androdioecious zone.

5.2.4.3 Post-glacial range expansion and isolation-by-distance

To look for potential effects of range expansion, allelic richness (for diploids), H'_S and observed allele numbers (for hexaploids), were regressed on latitude or longitude, as appropriate. For hexaploid populations, residuals from this regression were examined for differences between monoecious and androdioecious populations. Regression analysis of F'_{ST} and geographic distance was used to identify isolation-by-distance effects (Rousset 1997). The significance of isolation-by-distance effects were assessed with Mantel tests (Mantel 1967) as implemented in FSTAT (Goudet 1995). Having examined the possibility of isolation-by-distance effects, pairwise population differentiation was compared between sexual systems. Significance was assessed using ANOVA with the degrees of freedom conservatively calculated from the numbers of population analysed, not the number of pairwise measures. Regression analyses and other statistical tests were performed using MINITAB (Release 12.1 © Minitab Inc.).

5.2.4.4 Comparisons between different polyploid levels

Allelic phenotype diversity statistics (such as H') are not comparable between different allopolyploid levels, because they are a function of diversity at multiple independent isoloci; coalescent simulation of a subdivided population with disomic inheritance suggests that the difference in H'_T between polyploid levels depends on the degree of differentiation between isoloci (Chapter 4). Since no information about allele frequencies for the different isoloci is available in *M. annua*, comparison of genetic diversity between dioecious (diploid) and monoecious or androdioecious (polyploid) populations could not be made.

The comparison of statistics that quantify patterns of genetic diversity, such as population differentiation (F_{ST}) and within-population deviation from panmixis (F_{IS}), is also complicated by use of allelic phenotypes. In particular, no statistic corresponding to F_{IS} can be calculated using allelic phenotypes, because F_{IS} essentially relies on estimating the relative numbers of heterozygotes and

homozygotes, which cannot be done when isoloci are indistinguishable and allele dosage is unknown. However, simulations indicate that if inheritance is disomic, F'_{ST} is almost unaffected by polyploid level (Chapter 4, Figure 4.2). This is expected because F'_{ST} (like F_{ST}) is a ratio of the diversity due to differences between populations, relative to the total diversity.

5.3 RESULTS

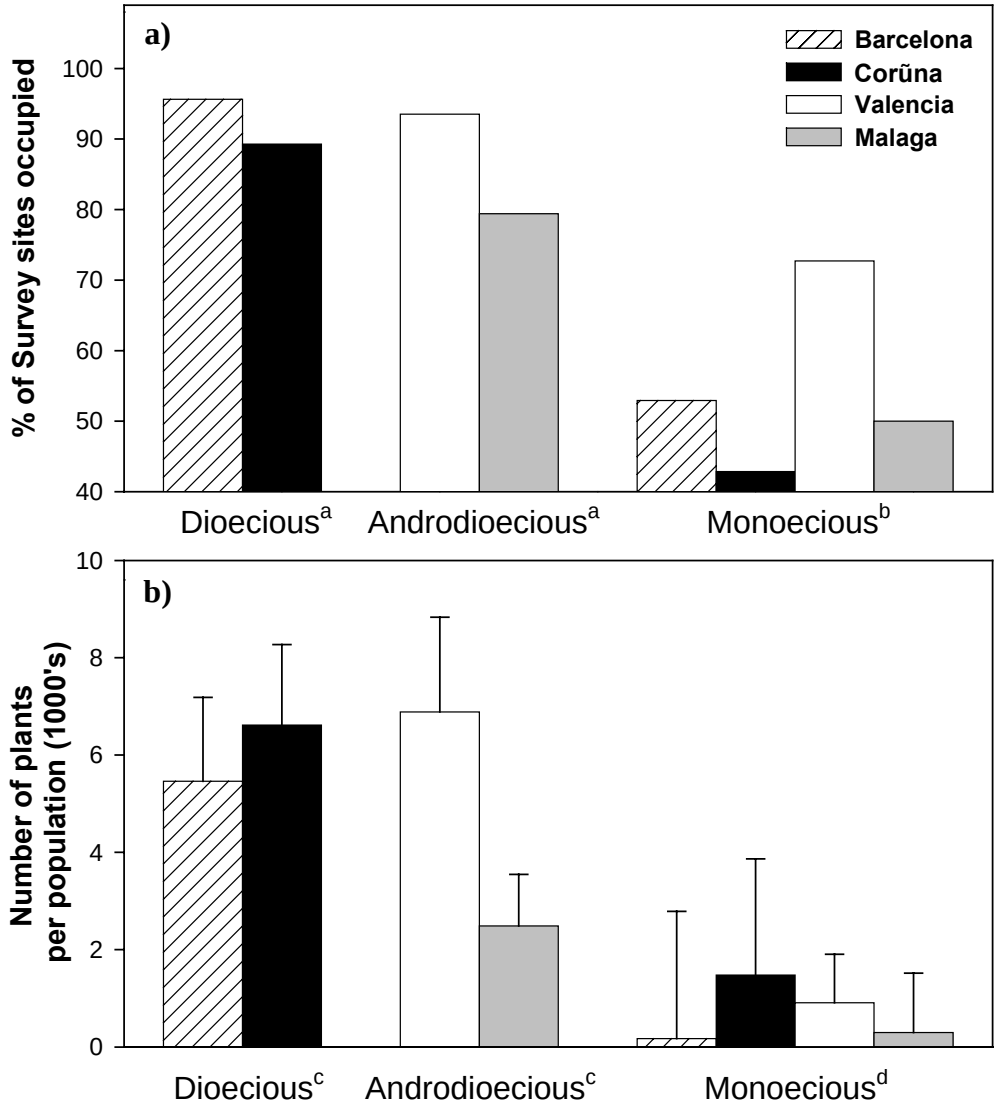
5.3.1 Population surveys

Populations were more abundant and larger in dioecious and androdioecious zones than in monoecious zones (Figure 5.4a and Figure 5.4b). In the monoecy-dioecy transects (La Coruña and Barcelona, Figure 5.2), populations were significantly more abundant ($p < 0.0001$) and larger ($p < 0.0001$) in the dioecious zone than the monoecious zone. Neither population abundance nor population size (number of individuals) differed significantly between transects ($p = 0.3189$ and $p = 0.1885$, respectively).

In the androdioecy-monoecy transects (Valencia and Málaga, Figure 5.2), populations were significantly more abundant ($p < 0.0006$) and larger ($p < 0.0002$) in the androdioecious zone than the monoecious zone (Figure 5.4a and Figure 5.4b). The transects differed significantly from one another in the abundance of populations and in the number of plants per population ($p < 0.0114$, $p < 0.0364$, respectively). Interestingly, not only did the size of a population reflect whether males were absent (monoecious) or present (androdioecious) in a population, but population size was also significantly positively correlated with the percentage of males in androdioecious populations (data not shown, linear regression: $r^2 = 0.13$; $p < 0.0108$).

While dioecious and androdioecious populations in the survey both differed from monoecious populations, they did not differ from each another (Figure 5.4a and Figure 5.4b). Dioecious and monoecious populations were equally abundant ($X^2 = 0.73$, $p = 0.3922$) and had similar numbers of plants ($p = 0.6843$).

Figure 5.4: Population abundance and size according to sexual system



(a) Population abundance measured as the proportion of occupied survey sites. **(b)** Population size measured by the number of individuals (with one standard error). See Figure 5.2 for the location of the survey transects. Different superscripts indicate significant differences between sexual systems (see main text for details). [This data and the associated analysis is unpublished material provided by SM Eppley].

Transects did not differ in the frequency of populations ($X^2 = 4.35$, $p = 0.1136$), but they did differ in plant number ($p < 0.0121$).

5.3.2 Isozyme survey

5.3.2.1 Dioecy

Total gene diversity (H_T) for all dioecious populations was 0.11; average within-deme gene diversity (H_S) within dioecious populations was 0.09 (Table 5.1, see Appendix 8.9 page 197 for illustrative isozyme gels). Allelic richness and H_T differed significantly between geographic regions (Table 5.1, $p = 0.001$), decreasing toward the west. Allelic richness and genetic diversity were highly correlated with longitude ($p < 0.0001$ for both, Figure 5.5, for the linear regression of allelic richness), but the effect was slightly stronger for allelic richness than diversity ($r^2 = 0.55$, and $r^2 = 0.50$ respectively). A comparison of allele frequencies between geographic regions shows this was due to a loss of rare alleles (Figure 5.6). Genetic differentiation (F_{ST}) over the whole range was 0.21, and differences in F_{ST} between geographic regions were not significant (Table 5.1, $p = 0.13$).

5.3.2.2 Monoecy and androdioecy

Differences in genetic diversity and differentiation were analysed in two ways: (1) for the zones defined in Figure 5.1, regardless of the sexual system of each population (e.g. the androdioecious zone contains many monoecious populations); and (2) for the androdioecious populations and the monoecious populations as separate groups, regardless of the zone in which they occur. See Appendix 8.9 page 197 for illustrative isozyme gels

Differences in diversity between zones. — In comparisons between the monoecious and androdioecious zones (Figure 5.1), total diversity (H'_T) did not differ between zones, whilst average within-population diversity (H'_S), the number of alleles per population, and the number of allelic phenotypes per population were all higher in the androdioecious zone than the monoecious zone (Table 5.2). There were no significant differences between monoecious regions on the east and west coasts of Iberia (data not shown).

Table 5.1: Genetic diversity and differentiation in dioecious populations

Location (populations)	Allelic richness	H_S	H_T	F_{IS}	F_{ST}	(95% bounds of F_{ST})
West (27)	*1.218*	*0.063*	*0.086*	*0.061*	*0.258*	*0.172, 0.285*
Central (10)	*1.422*	*0.116*	*0.126*	*-0.025*	*0.090*	*0.043, 0.149*
East (7)	*1.594*	*0.149*	*0.167*	*0.073*	*0.167*	*0.094, 0.225*
All (44)		**0.090**	**0.111**	**0.046**	**0.206**	**0.140, 0.235**

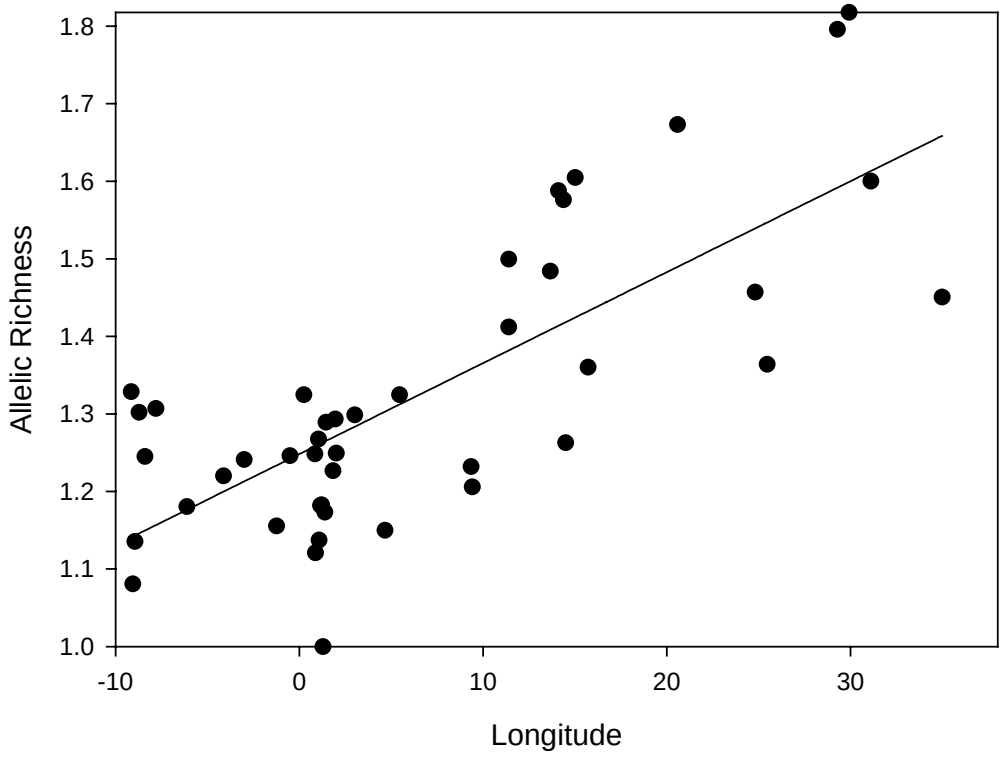
Values are averages across all 8 loci. Allelic richness is corrected for sample size, and quoted for a sample of 8 individuals. H_S is average Nei's gene diversity calculated within local populations. H_T is total gene diversity. F_{ST} and F_{IS} are Weir and Cockerham's (1984) estimators θ and f respectively. The 95% bounds of F_{ST} are estimated by bootstrapping across loci. Significant differences between west, central and eastern Europe were inferred by randomising populations between zones.

Table 5.2: Genetic diversity and differentiation in monoecious and androdioecious populations

Populations (sample size)	Alleles / population	Allelic phenotypes / population	H'_S	H'_T	F'_{ST}
Monoecious zone (24)	2.337	1.590	0.136	0.636	0.629
Androdioecious zone (18)	3.253	3.718	0.453	0.649	0.236
- of which monoecious (4)	2.327	1.516	0.112	0.695	0.562
- of which androdioecious (14)	3.488	4.279	0.528	0.612	0.106
All monoecious populations (29)	*2.098*	*1.475*	*0.111*	*0.537*	*0.519*
All androdioecious populations (16)	*3.007*	*3.576*	*0.426*	*0.507*	*0.109*
All (45)	**2.424**	**2.227**	**0.217**	**0.554**	**0.436**

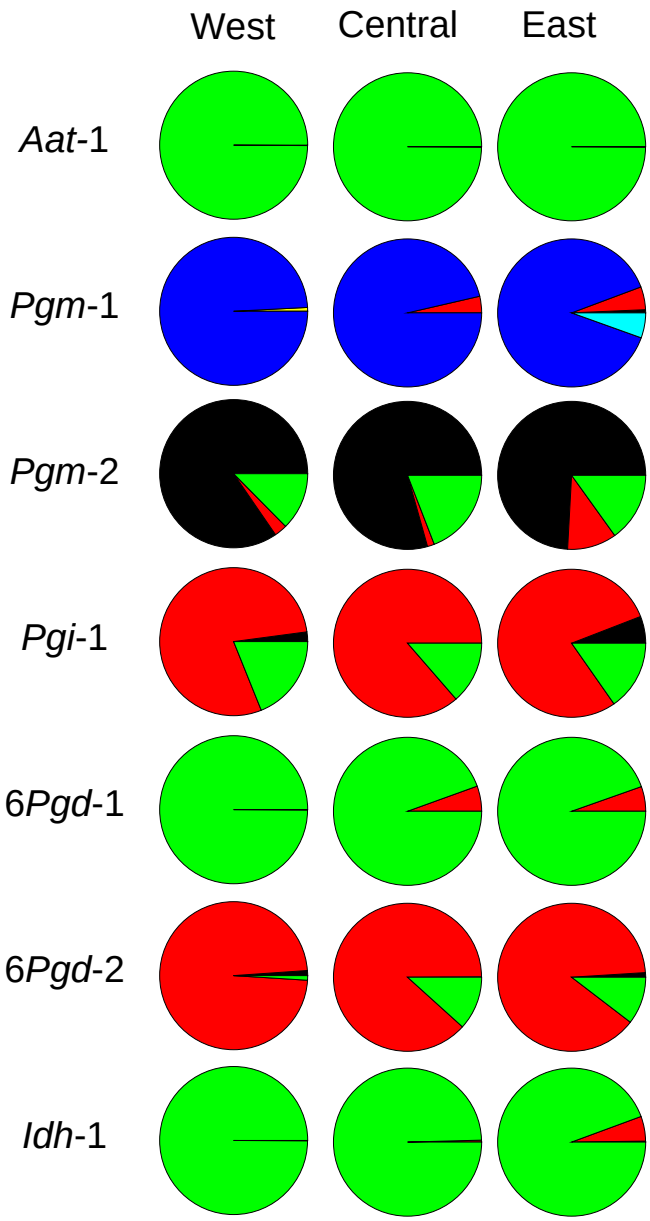
Values are averages across all 6 loci, weighted for sample size. H'_S and H'_T are the average number of alleles by which pairs of individuals differ, calculated within populations and over all populations respectively. F'_{ST} is calculated as $(H'_T - H'_S) / H'_T$. Significant differenced were inferred by randomising populations between analysis groups.

Figure 5.5: Correlation between allelic richness and longitude in dioecious populations



The average allelic richness per locus (corrected for sample size, El Mousadik and Petit 1996) of diploid populations is highly correlated with longitude (r^2 = 0.55, $p < 0.0001$). An almost identical, but weaker effect is seen for genetic diversity, H_S.

Figure 5.6: Allele frequencies in different dioecious regions



The pie charts show allele frequencies for seven isozyme loci that are variable in at least one population. The dioecious zone is arbitrarily divided into three geographic regions (East, Central and West, see Figure 5.1). Allele frequencies are very similar between the three regions. While this could be due to ongoing long-range gene flow, the progressive loss of rare alleles from east to west suggests a recent range expansion.

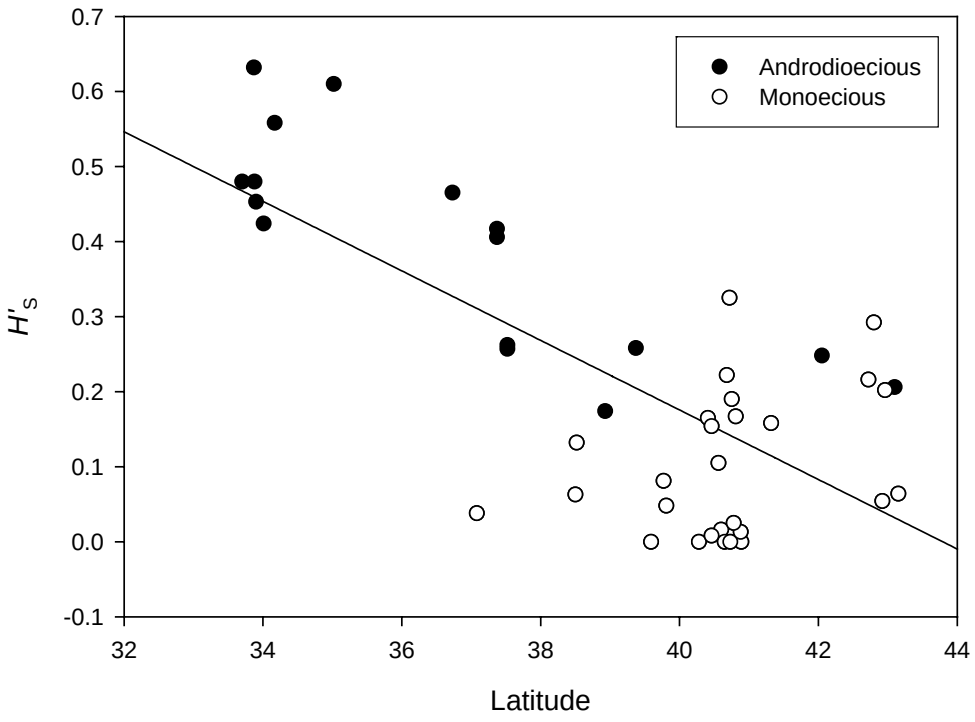
Differences in diversity regardless of zone. — Total diversity (H'_T) was not significantly different between monoecious and androdioecious populations (Table 5.2, H'_T = 0.54 vs. H'_T = 0.51, p = 0.66). However, within-population diversity (H'_S) was much higher for androdioecious populations than for monoecious populations (Table 5.2, $p < 0.001$). The number of alleles and the number of allelic phenotypes seen in each population were also significantly higher for androdioecious populations (Table 5.2, $p < 0.001$ for each).

Diversity within the androdioecious zone. — The largest differences in within-population diversity were for comparisons between different sexual-systems within the androdioecious zone: H'_T did not differ significantly (p = 0.493) while the within-population diversity measures were higher for androdioecious populations than for monoecious populations within the androdioecious zone ($p < 0.006$ for alleles per population, $p < 0.001$ for H'_S and the number of phenotypes per population).

The effect of latitude on diversity. — Genetic diversity within monoecious and androdioecious populations is significantly negatively correlated with latitude (linear regression, $r^2 = 0.50$, $p < 0.0001$, Figure 5.7). The correlation also applies for the number of alleles per population (linear regression, $r^2 = 0.17$, $p = 0.003$, data not shown). The residuals from these regressions differ significantly between sexual systems; diversity is greater (after accounting for latitude) in androdioecious populations than monoecious populations ($p < 0.0001$ for each measure of diversity). When sexual system was accounted for first, i.e. the regression analysis was done using residuals from an ANOVA to explain genetic diversity in terms of sexual system, there was still a significant correlation with latitude ($r^2 = 0.09$, $p = 0.02$).

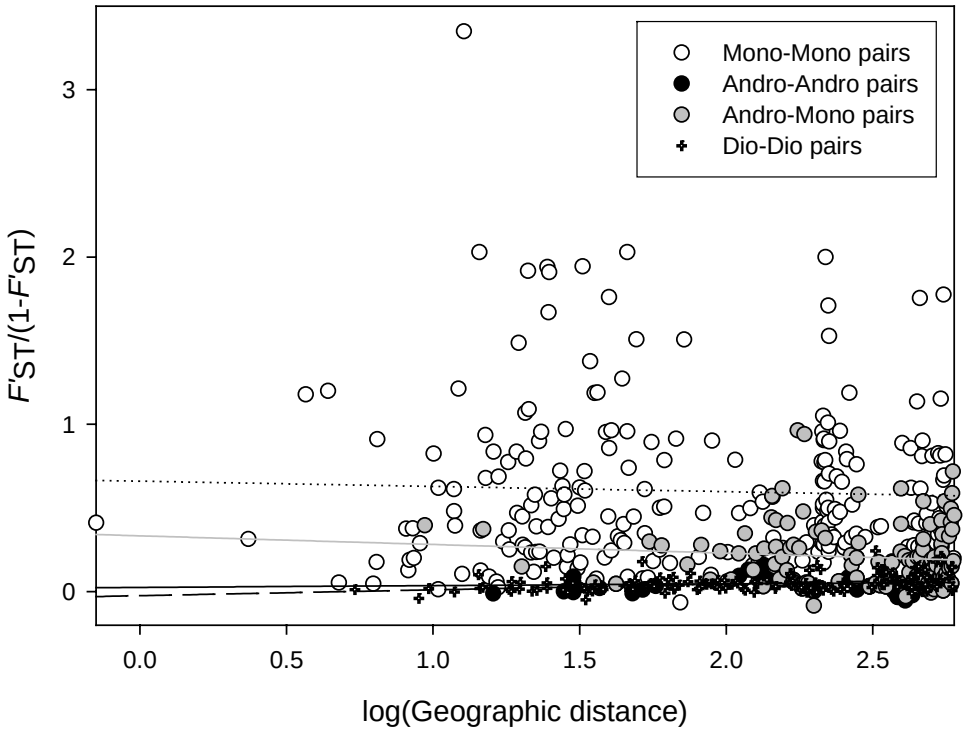
Genetic differentiation and isolation-by-distance. — Genetic differentiation (F'_{ST}) was higher in the monoecious zone than in the androdioecious zone (Table 5.2, F'_{ST} = 0.63 versus 0.24, $p < 0.001$). It was also higher for monoecious populations than androdioecious populations, when geographic zone was disregarded (Table 5.2, F'_{ST} = 0.52 versus 0.11, $p < 0.001$). F'_{ST} for all dioecious (diploid) populations in Iberia was similar to that of androdioecious populations (F'_{ST} = 0.15). There was no evidence for isolation by distance between pairs of monoecious populations (p = 0.24, dotted line in Figure 5.8). However, there was

Figure 5.7: Correlation between genetic diversity and latitude in monoecious- and androdioecious populations



Genetic diversity (H'_S) of hexaploid populations is highly correlated with latitude $r^2 = 0.50$, $p < 0.0001$). Two monoecious populations from the Canary Islands (0093a and 0101a) are not shown, as latitude is unlikely to be a good proxy for their migration history, if they are included $r^2 = 0.27$, $p < 0.001$.

Figure 5.8: Pairwise genetic differentiation between populations (transformed data)



Over distances of less than 200km there is no significant isolation by distance for monoecious populations: $p = 0.24$, but there is a significant distance effect for androdioecious populations $r^2 = 0.23$ $p = 0.008$, and a marginal effect for dioecious populations $r^2 = 0.03$ $p = 0.062$. All p-values are for $F'_{ST}/(1-F'_{ST})$ on *log*(distance), using 10,000 randomisations in the mantel procedure implemented in FSTAT. All isolation-by-distance effects are insignificant if distances up to 600km are considered.

The y-intercept of the linear regressions indicates that the genetic differentiation between androdioecious populations and their androdioecious neighbours (black circles, solid black line) is lower than that between androdioecious populations and their monoecious neighbours (grey circles, grey line). Conversely, the pairwise differentiation between monoecious populations and their androdioecious neighbours (grey circles, grey line) is lower than that between monoecious populations and their monoecious neighbours (white circles, dotted line). Pairwise F'_{ST} values are very similar for dioecious populations (black crosses, dashed line) and androdioecious populations (black circles, solid black line).

a small effect of geographic distance for pairs of androdioecious populations (r^2 = 0.23, p = 0.008, black line in Figure 5.8), and there was a small but marginally insignificant effect for pairs of dioecious populations (r^2 = 0.03 p = 0.062, dashed line Figure 5.8). An ANOVA for the effect of sexual system on average pairwise differentiation between populations showed differentiation to be significantly higher between monoecious populations than between androdioecious populations (F = 7.03, df = 40, p = 0.012), while pairwise F'_{ST} values did not differ between dioecious and androdioecious populations (F = 1.99, df = 41, p = 0.166). These differences in pairwise genetic differentiation between sexual systems are illustrated by the differing y-intercepts in Figure 5.8.

5.4 DISCUSSION

The distribution of sexual-systems within the *M. annua* species complex has been hypothesised to result from an interaction between selection for gender-specialisation and selection for reproductive assurance in a metapopulation (Pannell 2001; Pannell 2002). Here I have shown that *M. annua* populations are larger and more frequent in dioecious and androdioecious zones than they are in monoecious zones (Figure 5.4), and that genetic differentiation is lower between dioecious populations and between androdioecious populations than it is between monoecious populations (Figure 5.8). In addition, although there are strong geographic trends in genetic diversity within *M. annua* (Figure 5.5 and Figure 5.7), suggesting post-glacial range expansion, within-population diversity is also affected by sexual system in a predictable way (Figure 5.7). These data are consistent with the metapopulation model for sexual-system evolution in *M. annua*, and identify separate glacial refugia for the different sexual-system lineages.

5.4.1 Geographic patterns in species-wide genetic diversity

Gradients in genetic diversity suggest that diploid *M. annua* has recently spread from the eastern end of Europe to occupy its current pan-European distribution, and that hexaploid *M. annua* has spread northward through Iberia, from a North African or southern Iberian glacial refugium. Because rare alleles

are likely to be lost through sampling effects when a species expands rapidly into new territory, islands of high genetic diversity may to be a good indicator of refugia. This allows genetic data to be used to infer range-expansion and the location of glacial refugia (Hewitt 1999; Petit *et al*. 2003). However, when there are many refugia, the mixing of alleles where lineages meet can lead to a lack of correlation between refugia and high diversity (Petit *et al*. 2003), and more generally inference of such historical processes may only be possible when population history is simple (Charlesworth *et al*. 2003). The clearest cases are those in which there are distinct gradients in genetic diversity (Charlesworth *et al*. 2003), such as that displayed by *Lophocereus schottii* in Baja California (Nason *et al*. 2002), by oak gall wasps in Europe (Rokas *et al*. 2003), and in the data presented here (Figure 5.5 and Figure 5.7).

In Europe, post-glacial expansion patterns have been loosely classified into three broad categories, exemplified by 'hedgehogs', 'grasshoppers', and 'bears' (Hewitt 1999). Respectively, these are: (1) equal expansion into northern Europe from Iberia, Italy, and Greece/Turkey; (2) recolonisation predominantly from Greece/Turkey with Iberian and Italian lineages failing to leave their respective peninsulas; and (3) recolonisation predominantly from Greece/Turkey and Iberia, with lineages failing to leave the Italian peninsula. Several plant species conform to these paradigms; *Quercus* spp. and *Abies alba* to 'hedgehogs', and *Alnus glutinosa* and *Fagus sylvatica* to 'grasshoppers' (reviewed in Hewitt 1999). However, there have been few studies on weedy annuals such as *M. annua* (for a rare example see Koch and Bernhardt 2004).

5.4.1.1 Dioecious populations

In dioecious *M. annua* populations, allelic richness (and to a lesser extent gene diversity) is lower at the western end of the species range than at the eastern end (Table 5.1) and in general, richness and diversity are strongly correlated with longitude (Figure 5.5). This gradient in diversity appears to be due to a loss of rare alleles in central and western populations (Figure 5.6). This strongly suggests a recent range expansion from the eastern end of Europe with no centres of diversity in the common European refugia of Iberia and Italy. However, at the eastern end of the range, sampling was not dense enough to distinguish between

Greece, Turkey, or more easterly refugia (Figure 5.1). This corresponds most strongly to the 'grasshopper' paradigm seen in alder (King and Ferris 1998) and beech (Demesure *et al*. 1996), though in the case of diploid *M. annua* there is no evidence for refugial populations in the Iberian and Italian peninsulas.

Although there are other possible explanations for this longitudinal correlation in *M. annua* (e.g., low-diversity refugia, sampling effects and differences in on-going population processes) I believe that evidence favours post-glacial range expansion. The similarity in allele frequencies across the range (Figure 5.6) argues against the existence of low-diversity refugia in the west and high-diversity refugia in the east. The effect cannot be attributed to differences in sampling, as genetic diversity is highest where sampling intensity was lowest. Nor can it easily be attributed to a gradient in population census size, as *M. annua* is very common across this entire range: even where diversity is low (e.g. northern Iberia), populations occur at 92% of roadside sites (Figure 5.4). It is possible that differences in effective population size, brought about through differences in migration rate or population turnover, could account for the latitudinal gradient. However, there is no significant difference in population differentiation between the geographic regions (Table 5.1), which would be expected if a gradient in population structure or turnover were present.

5.4.1.2 Monoecious and androdioecious populations

In monoecious and androdioecious populations, genetic diversity and allele-numbers correlate strongly with latitude (Figure 5.7). In principle, this effect could result from differences in sexual system and demography rather than range expansion, as they are (partially) confounded with latitude. However, residuals from an ANOVA accounting for variation due to sexual system are still correlated with latitude. Thus, this gradient supports a recent range expansion from the south, i.e. a southern Iberian or North African refugium. Although not surveyed in this study, the widespread occurrence of monoecious polyploid (tetraploid, hexaploid and octoploid) *M. annua* along the north coast of Africa and on the Mediterranean islands of Corsica and Sardinia (Durand 1963), supports a North African refugium.

This pattern is not represented amongst the paradigms outlined by Hewitt (1999), possibly because few studies have considered species that are widespread in both North Africa and Europe. Where sampling does include North Africa, it is more commonly cited as a target for migration than as a source (e.g., Wall Lizards, Harris et al. 2002; Wood mice, Michaux et al. 2003). A striking exception is that of *Ecballium elaterium*, where separate refugia for the monoecious and dioecious races have been proposed (Costich and Meagher 1992). Although, the locations of refugia were not addressed explicitly, the geographic distribution of *E. elaterium* suggests a north African refugium for the dioecious race, and a European refugium for the monoecious race (Figure 1 of Costich and Meagher 1992); an inversion of the distribution seen in *M. annua*. Also in agreement with the results for *M. annua*, separate glacial refugia for diploid and polyploid races have been hypothesised in other European species (e.g. Trewick *et al.* 2002; Koch and Bernhardt 2004). It is tempting to speculate that such differences in polyploid level and sexual system were only able to establish under the geographic isolation imposed by glaciation.

5.4.2 Evidence for metapopulation processes

Mercurialis annua only occurs in recently disturbed ruderal and anthropogenic habitats, and exists as a patchwork of spatially discrete patches. This could be consistent with different ongoing processes; populations may (1) be stable in size, (2) fluctuate significantly in size over time, or (3) be the result of frequent recolonisation after local extinction. Gene flow between patches may be high, preventing differentiation in neutral allele frequencies, or low, promoting it. Similarly, colonists involved in recolonisation may be few, or many.

In order to explain the evolution of androdioecy and the distribution of male frequencies in hexaploid *M. annua*, the metapopulation model proposed by Pannell (2001) requires that populations are frequently recolonised by small numbers of individuals and that there is not panmictic gene-flow (see Figure 5.3). Otherwise, if populations were long-lived, colonist numbers large, or gene flow high, all populations would quickly contain the same (high) frequency of males. Furthermore, the extension of the model used to explain the absence of males in

some regions of Iberia (Figure 5.3, 'short-circuit' 2b), requires higher population turnover, and/or lower migration rates in monoecious regions than in androdioecious regions. Under this model, dioecy can be maintained where population turnover is very low or pollen movement very high, removing selection for reproductive assurance. Thus, dioecious populations are expected to display population structure more similar to androdioecious populations than monoecious populations. Below, I discuss (1) the survey of population size and abundance, and (2) the survey of genetic diversity in Iberia, in terms of the population structure predicted by the metapopulation model.

5.4.2.1 Population size and abundance

Mercurialis annua populations in the dioecious and androdioecious zones of Iberia were larger and more common than populations in monoecious zones (Figure 5.2, Figure 5.4), but dioecious and androdioecious regions did not differ significantly from each other. Under simple metapopulation models, high local-extinction rates are expected to reduce the occupied proportion of available sites (e.g. Levins 1970; Hanski 1997). Thus, the scarcity of populations in the monoecious zone compared to the dioecious and androdioecious zones is consistent with the metapopulation model, although it is difficult to exclusively attribute low population abundance in the landscape to high population turnover.

Many processes causing local extinction will be affected by population size (e.g. stochastic fluctuation in population size), with large populations being more likely to persist that small ones; indeed, many metapopulation models assume that the probability of local extinction is directly related to population size, (e.g. Hanski 1997). Alternatively, if population size generally increases with population age, a preponderance of small populations may be indicative of a younger age structure. In either case, small population size will be correlated with high population turnover. Thus, the smaller size of *M. annua* populations in monoecious regions is also consistent with a higher rate of population turnover there than in the androdioecious regions.

The most reliable way to assess the relative frequencies of population turnover in zones of different sexual system is by direct observation. Anecdotal evidence of population turnover in *M. annua* does exist (JR Pannell pers. obs.), and a

comprehensive survey is ongoing. However, at present, there remains the possibility that population turnover does not differ between sexual systems; i.e., in the monoecious zone populations are smaller and rarer, but do not experience more frequent extinction.

If hexaploid *M. annua* populations in Iberia are not at equilibrium, strictly monoecious regions could be explained by invoking a rapid range expansion in which males, being unable to found populations, lagged behind hermaphrodites. However, this is not consistent with the 'patchy' distribution of monoecy (Figure 5.2). I am not aware of any hypothesis that better explains the large regions of monoecy than the one of higher population turnover and/or lower gene flow in those areas.

5.4.2.2 Patterns of genetic diversity

Total diversity. — Total genetic diversity (H'_T) in *M. annua* was the same in monoecious and androdioecious zones, for monoecious and androdioecious populations regardless of zone, and for monoecious and androdioecious population in the androdioecious zone (Table 5.2). Since it is expected that population turnover will reduce regional genetic diversity (e.g. Hedrick and Gilpin 1997; Pannell and Charlesworth 2000; Wakeley and Aliacar 2001; Rousset 2003), the observation of equal genetic diversity in monoecious and androdioecious zones is superficially at odds with prediction. The apparent contradiction in *M. annua* may stem from the equilibrium conditions and "closed system" (meta)populations of the models. If the high-turnover regions exist within a larger matrix of low-turnover populations, as is probable in *M. annua*, then an influx of alleles from the wider metapopulation would balance regional diversity losses due to local extinction. That is, instead of considering Iberian *M. annua* in terms of distinct monoecious and androdioecious metapopulations, they should be considered as one extended metapopulation, with the potential for geographic variation in local extinction rates.

Within-population diversity. — Within-population diversity (H'_S) was significantly lower, compared to androdioecy, for populations in the monoecious zone, monoecious populations regardless of zone, and monoecious populations in the androdioecious zone (Table 5.2). This was true even when the possibly

confounding effect of latitude (i.e. northward post-glacial migration) was accounted for (Figure 5.7). As outcrossing rates in established hexaploid *M. annua* populations are probably greater than 50% (SM Eppley, unpublished data), such a decrease in within-population diversity cannot be entirely explained by a difference in selfing rate, which alone could not reduce within-population diversity by a factor of four (H'_S for all androdioecious populations = 0.43, monoecious populations = 0.11, for all, Table 5.2).

On the other hand, genetic bottlenecks at colonisation are expected to strongly reduce within-population diversity, (unless colonists are many, or come from many different populations e.g. Wade and McCauley 1988; Pannell and Charlesworth 2000). If there is any post-colonisation migration, diversity in the newly formed population will increase over time, as immigrants bring new alleles. Such a relationship between local diversity and population age has been reported in several species; *Silene dioica* (Giles and Goudet 1997), *S. alba* (McCauley *et al*. 1995), and the beetle *Bolitotherus cornutus* (Whitlock 1992). Thus, the very low within-population diversity observed for monoecious *M. annua* populations, particularly in the androdioecious zone (Table 5.2), is most easily explained by monoecious populations being younger than androdioecious populations, as predicted by the metapopulation model. Whilst it could also be attributed to extreme fluctuations in local population size (i.e. bottlenecks without colonisation), fluctuations would have to be exceptionally large in order to remove males (a selected trait) from the population.

Genetic differentiation. — Relative changes in total diversity (H_T) and within-population diversity (H_S) are quantified by F_{ST}, which measures genetic differentiation between populations (e.g. Hartl and Clark 1997). Population turnover may either increase or decrease differentiation, depending on the number and source of colonists (Wade and McCauley 1988). However, empirical studies suggest that differentiation is usually increased by population turnover because colonist numbers are small (e.g. Whitlock 1992; McCauley *et al.* 1995; Giles and Goudet 1997). Selfing also tends to increase genetic differentiation (Hamrick and Godt 1990; Charlesworth 2003). Reasons for this include reduced migration rate, increased genetic hitchhiking effects, and increased background selection (reviewed in Charlesworth and Pannell 2001; Charlesworth 2003).

Selfing and population turnover can also interact to increase differentiation (Ingvarsson 2002).

In *M. annua*, low within-population diversity of monoecious populations resulted in much greater genetic differentiation amongst monoecious populations than androdioecious populations (Table 5.2). This effect was not an artefact of the sampling regime (for example, the geographic area of sampling might differ between sexual systems), as it was also apparent in the magnitude of pairwise differentiation at all spatial scales (Figure 5.8). Instead, it is likely to result from different rates of selfing, population turnover, or migration.

According to the metapopulation model (Figure 5.3), monoecious populations within androdioecious zones are those that have not yet been invaded by males. Having been formed by very small numbers of colonists (Step 2 in Figure 5.3), they will have lower within-population diversity (Table 5.2), and they will be highly differentiated (through the sub-sampling of alleles) from surrounding populations (Table 5.2, Figure 5.8). Since such differentiation would be eroded by later migration, indicating that monoecious populations are younger than, or have experienced lower migration rates than, androdioecious populations. Dioecious populations display the same low level of genetic differentiation as androdioecious populations; consistent with the hypothesis that dioecy is maintained by selection for gender specialisation in the absence of regular population turnover, which would select for reproductive assurance.

Clearly, the situation in *M. annua* is complicated, and it is impossible to rule out all other possible models that could account for the observed patterns of sex allocation, male distribution, population size, population abundance, genetic diversity, and genetic differentiation. However, none of the simple alternative population structures outlined above (e.g. stable or fluctuating population size without colonisation) give rise to the patterns observed, which seem to be consistent in all important respects with the metapopulation model invoked to explain the evolution and maintenance of androdioecy in *M. annua*.

Perhaps equally importantly, other androdioecious species provide similar evidence, albeit anecdotal, consistent with metapopulation dynamics (reviewed in Pannell 2002). For example, populations often appear discrete and partially isolated, both of which are pre-requisites for metapopulation dynamics to impact upon biology, e.g. *Datisca glomerata* (Liston *et al*. 1990). Males are not present

in all populations and androdioecious populations vary widely in male frequency, e.g., *Datisca glomerata* (Liston *et al*. 1990), *Schizopepon bryoniaefolius* (Akimoto *et al*. 1999), and populations containing males have higher genetic diversity than hermaphroditic ones, *Schizopepon bryoniaefolius* (Akimoto *et al*. 1999). In addition, many of these observations are also true for the androdioecious invertebrates *Eulimnadia texana* and *Triops longicaudatus* (e.g., Sassaman 1989, 1995), suggesting that Pannell's (2001) model for the maintenance of androdioecy may well operate in many androdioecious species.

6 GENERAL DISCUSSION

The primary aim of this study was to elucidate the evolutionary origin and maintenance of sexual-system variation within the *Mercurialis annua* polyploid complex, using information available from patterns of genetic diversity. To do this, I addressed specific questions regarding (1) the phylogenetic origin of androdioecy in *M. annua*, and (2) the ecological maintenance of dioecy, monoecy, and androdioecy in *M. annua*. To address question (1), I conducted a phylogenetic analysis of the annual mercuries using chloroplast and ITS sequence data, supported by morphometric analysis, chromosome number, and genome-size data (see Chapter 2). To address question (2) I first had to establish the mode of marker inheritance (Chapter 3) and to identify suitable summary statistics for use in hexaploid populations (Chapter 4). I then used a species-wide survey of isozyme diversity to infer post-glacial range expansion, and to test a metapopulation model of sexual-system dynamics in *M. annua* (Chapter 5).

Although dioecious (diploid) *M. annua* has been widely used as a model species for sex determination and expression (Yampolsky 1919; Gabe 1939; Durand and Durand 1991; Khadka *et al*. 2002), and the association between polyploidy and sexual system in the complex is well known (Thomas 1958; Durand 1963), the value of *M. annua* as a model for sexual-system evolution has only recently been highlighted (Pannell 1997c; Pannell 2002; Pannell *et al*. 2004). Below, I briefly outline the findings of earlier work on sexual-system and polyploid variation in *M. annua*, and discuss the main results of this study in terms of (1) the apparent hybrid origin of androdioecy, (2) the significance of separate glacial refugia for the sexual systems, and (3) evidence for the metapopulation model of sexual-system dynamics. In doing so I note some of the limitations to this study, and suggest additional work that could be performed.

6.1 A HYBRID ORIGIN FOR ANDRODIOECY

Krahenbuhl *et al*. (2002) recently used ITS sequences to reconstruct the phylogeny of *Mercurialis*. They both confirmed the base chromosome number in the genus (n = 8), and showed that dioecy is ancestral to monoecy. The phylogenetic analysis

presented in Chapter 2 extends these results in two ways. First, through better lineage sampling it was possible to identify the phylogenetic relationship between androdioecious lineages and monoecious and dioecious lineages. Second, by actively searching for intra-individual variation in ITS sequences, and additionally sampling chloroplast sequences, it was possible to identify past hybridisation. This joint chloroplast and ITS analysis, in conjunction with morphological and genome-size data, shows that hybridisation has played an important role in the evolution of the annual mercuries; the Tenerife mercury is an allopolyploid, *M. huetii* has experienced chloroplast capture from a woody-perennial species, and most significantly in the context of sexual-system evolution, hexaploid *M. annua* is probably allopolyploid in origin.

An allopolyploid origin for hexaploid *M. annua* has important implications for the use of *M. annua* as a model for the evolution of androdioecy. Theoretical models differ in whether androdioecy originates from dioecy or from monoecy (compare Charlesworth 1984; Wolf and Takebayashi 2004), but it appears that an origin within dioecy may be evolutionarily easier (Pannell 2002). In support of this, androdioecy appears to have evolved from dioecy in several other androdioecious groups, e.g. *Datisca* and *Schizopepon* (reviewed in Pannell 2002). If androdioecious hexaploid *M. annua* is a hybrid between monoecious tetraploid *M. annua* and dioecious diploid *M. huetii*, as the evidence presented here suggests, it is possible that the evolution of androdioecy was facilitated by the inheritance of male traits (e.g. the pedunculate inflorescence) from *M. huetii* (Chapter 2). Such a hybrid origin would be unique amongst known androdioecious species, and is worthy of further study; it is increasingly being recognised that hybridisation can be an important source of evolutionary novelty (e.g. Rieseberg *et al.* 2003).

At present, this scenario is highly speculative. A hybrid origin for androdioecious *M. annua* is supported by ITS and isozyme data, and to some extent by morphological data (Chapter 2). However, the conclusion of an autopolyploid origin for monoecious tetraploid *M. annua* is based on the absence of *M. huetii*-like ITS repeats. It is possible these were once present and have since been lost through concerted evolution, or failed to amplify under the chosen the PCR conditions (e.g. Alvarez and Wendel 2003). In addition, although Durand and Durand (1992) asserted that males are absent from tetraploid *M. annua*, a more detailed survey is needed.

A more intensive attempt to uncover 'hidden' ITS-types in both tetraploid and hexaploid *M. annua*, using a cloning-based approach, might prove informative, as it would be able to identify further sequences that are present but failed to amplify (for whatever reason). However, an approach using multiple single-copy nuclear genes would be the most robust, as it would be less subject to the problems of concerted evolution. Identification of the genes for sex-determination and male-associated sex-expression from the lineages in question would provide the best possible test (see Chapter 2), but such an approach would be technically challenging. In addition, the inclusion of the strictly monoecious polyploid lineages from Tunisia, Sardinia and Corsica would be useful, to identify whether they are derived from hexaploid *M. annua* and have lost pure males, or are derived directly from monoecious *M. annua*.

6.2 SEPARATE GLACIAL REFUGIA FOR MONOECY AND DIOECY

Durand (1963) established the geographic distribution of polyploid and sexual-system races in *M. annua* during an extensive survey of Western Europe and North Africa. Based on this survey, he suggested that although the distribution was associated with climate, it was conditional upon post-glacial dispersal (Durand 1963). He proposed (1963, pages 616-620) that *M. annua* had a north African refugium, and recolonised northern Europe from Morocco into Iberia, and from Tunisia to Italy.

The data presented in Chapter 5 summarise genetic diversity in *M. annua* across Europe. The strong continent-wide gradients in diversity (allelic richness) with longitude and latitude are indicative of post-glacial range expansion (e.g. Hewitt 1999), and are consistent with an eastern European refugium for the dioecious diploid lineage, and a North African refugium for the polyploid races of *M. annua*. Durand (1963) does not seem to have considered the possibility of such an Eastern European refugium, possibly because his survey did not extend east of Italy. Unfortunately, the geographic sampling in Chapter 5 is insufficient to identify refugia with any precision. Although *M. annua* in Egypt is dioecious, and in Tunisia is predominantly monoecious, nothing is known about the cline that presumably occurs in between. Additionally, although it seems likely that dioecious diploid *M. annua* present in northern Tunisia has recently invaded from northern Europe, it is possible that this in fact represents a relictual population. Further samples from the North African coast

between Morocco and Egypt, and from the Middle East (e.g. Syria and Iraq) would probably resolve the remaining issues regarding glacial refugia in *M. annua*. In this context, and given the previously unrecognised presence of an allopolyploid mercury endemic to the Canary Islands (Chapter 2 and Appendix 8.4, page 172), samples from the Macaronesian islands would be particularly interesting.

To some extent, the additional genetic differentiation between sexual-system races that is implied by separate glacial refugia may limit the potential for comparative studies, as sexual system differentiation is associated with considerable genetic differentiation. However, separate glacial refugia for the dioecious and monoecious races of *M. annua* are in line with what is known for other species with similar variation in polyploid level or sexual-system. For example, monoecious and dioecious races of *Ecballium elaterium*, are thought to have different refugia (Costich and Meagher 1992), and different refugia for polyploid races are a common feature of several polyploid complexes (e.g. Trewick *et al.* 2002; Koch and Bernhardt 2004). This observation may be informative in itself, for example, novel traits such as polyploidy or a new sexual-system may be more likely to establish in an isolated refugium than when the species occupies a continent-wide range.

6.3 METAPOPULATION STRUCTURE AND SEXUAL SYSTEM VARIATION

Despite the widespread use of molecular markers in studies of plant population structure (Hamrick and Godt 1990, 1996), and the suggestion that sexual-system variation makes *M. annua* a prime candidate for such a study (Costich and Meagher 1992, page 598), to date, molecular genetic investigations of *M. annua* have been limited to a search for sex-linked sequences (e.g. Yang *et al.* 1998; Khadka *et al.* 2002). Pannell's (1995; 2001) demographic model of sexual system dynamics provides an evolutionary explanation for sexual system variation in *M. annua*, and the genetic diversity data presented in Chapter 5 provide an indirect test of this model. The low genetic diversity of monoecious populations, in association with high genetic differentiation, is consistent with their having been formed by very few colonists, and having experienced little subsequent immigration. This is a key prediction of the metapopulation model, as only frequent local extinction, associated with recolonisation largely by hermaphroditic plants, can explain the presence at equilibrium of

monoecious populations, and the variation in male frequencies between androdioecious populations.

The need for reproductive assurance clearly has the potential to favour self-fertilising hermaphrodites over obligate outcrossing individuals (Pannell and Barrett 1998). There is considerable evidence that individuals capable of self-fertilization are favoured in long-distance dispersal events, such as the colonisation of islands (e.g. Baker 1967; Bawa 1982). From a theoretical point of view it has been established that the need for regular colonisation in the context of a metapopulation can in principle select for the ability to self-fertilize, and the sexual-system variation within the *M. annua* complex provides an opportunity to test this prediction.

Although the genetic data, the survey of population size and abundance (Chapter 5), and observed local sex-ratio variation is consistent with the metapopulation model in *M. annua*, they are not conclusive evidence that population turnover is an ongoing process. Direct observation of many populations over a number of seasons is required to confirm this. Ideally, such a study would also include a study of potential habitat sites, to confirm that colonisation is also occurring at a rate consistent with equilibrium metapopulation conditions, and of potential seed-bank dynamics. A survey of population presence and absence is ongoing (M.E. Dorken and J.R. Pannell, unpublished data) but it would be strengthened by an accompanying genetic survey, which could allow the colonist source-population (including the seed-bank) to be distinguished.

6.4 THE GENUS *MERCURIALIS* AS A MODEL

Mercurialis annua benefits from many characteristics that make it a good model organism; not only is there variation in traits of interest, such as gender and polyploid level, but it is small, short-lived, and very easy to culture. Additionally, at a time when whole-scale genome sequencing is an increasingly viable approach, a relatively small genome (in the lower 15% of diploid angiosperm species, Bennett and Leitch 2003), counts in its favour. However, unlike well-established model genera, such as *Drosophila* or *Arabidopsis*, which were selected in part because intensive research in one species (*D. melanogaster* and *A. thaliana*) increased the value of comparative biology in sister taxa, the genus *Mercurialis* as a whole presents interesting questions

regarding polyploid variation and gender and sexual dimorphism that are worthy of pursuit at a higher taxonomic level.

The increased availability of informative molecular techniques, and new discoveries regarding the significance and mechanism of epigenetic changes, have sparked a renewed interest in the ecological and evolutionary significance of polyploidy (e.g., Comai 2000; Soltis and Soltis 2000; Wendel 2000; Hodkinson *et al.* 2002; Rauscher *et al.* 2002; Osborn *et al.* 2003). The extended polyploid series (2x to 12x) available in the *M. annua* complex clearly gives it potential as a model for polyploid evolution, particularly because of the presence of extensive polyploidisation across the entire genus. The rhizomatous perennials *M. perennis* and *M. ovata* form a second polyploid series, also 2x to 12x, and *M. leiocarpa* has been recorded in both 2x and 6x forms (Krahenbuhl and Kupfer 1995). The woody perennial species (*M. elliptica*, *M. corsica*, *M. tomentosa* and *M. reverchonii*) all have chromosome numbers that suggest polyploidisation (Table 2.1, page 21), and variation in chromosome number is known in *M. elliptica* (Krahenbuhl *et al.* 2002). Thus, from a base chromosome number of 2*n* = 16, at least three clades, and possibly more within the woody perennials, have independently evolved into polyploid complexes. It is unknown whether these are autopolyploid series, as was previously thought in *M. annua*, or whether they represent allopolyploids and infra-specific hybrids. However, hybridisation occurs easily between the woody perennial taxa (Krahenbuhl *et al.* 2002), and many hybrids have been reported (listed by Güemas 1997). This wealth of closely related polyploid lineages, including two intact series from 2x to 12x, provides ample opportunity for comparative studies regarding (1) the evolutionary fate of duplicated genes, (2) inter-ploidy hybridisation and cytotype exclusion, and (3) the factors that contribute to the formation of polyploid complexes.

The genus *Mercurialis* is almost universally dioecious, making it a potential model of sexual dimorphism and sex determination. Sex-linked markers have been found in *M. annua* (Yang *et al.* 1998; Khadka *et al.* 2002), but the occurrence of highly skewed progeny sex ratios indicates that sex determination is not straightforward (JR Pannell, pers obs). Genetic studies indicate that sex determination in dioecious *M. annua* is achieved through three independent biallelic loci, labelled A, B_1 and B_2 (case indicates dominance); such that (*a*|*a*, -|-, -|-) and (-|-, $b_1|b_1$, $b_2|b_2$) are female, and all other genotypes, such as (*A*|*a*, $B_1|\ b_1$, $b_1|\ b_1$) or (*a*|*a*, $b_1|\ b_1$, $B_1|\ B_1$), are male (Louis 1989). This model of sex determination in *M. annua* has entered the secondary literature

(Richards 1997; Ainsworth 2000), but its evolutionary maintenance is not well understood. A few moments thought reveals that, since all three loci are equivalent, with any one locus being able to determine sex on its own, allele frequencies are free to drift until they reach fixation at two of the three loci. This raises two interesting questions. First, what maintains this polymorphism over ecological timescales? And second, is this mode of sex determination conserved throughout the genus? To maintain such diversity in *M. annua*, even after rapid post-glacial range expansion, would require strong selection. It seems likely that negative frequency-dependent selection, such as that which maintains self-incompatibility alleles, would play a role, and thus it is possible that the sex-determination alleles in *M. annua* may be very old. Certainly, *Mercurialis* could provide a model of non-chromosomal sex determination in plants that would provide an insight into important aspects of evolutionary biology.

In this thesis I have used patterns of genetic diversity to address key questions regarding the origin and maintenance of sexual-system variation in the *M. annua* polyploid complex. The data I present suggest that androdioecy in *M. annua* may have had an allopolyploid origin, and support the idea that androdioecy is selectively maintained by regular population turnover. However, many questions regarding sexual-system variation in *M. annua* remain, and more extensive fieldwork and sophisticated molecular approaches will clearly be required. The genus *Mercurialis* also poses wider questions regarding the origin and maintenance of sex-determination mechanisms, and the evolutionary significance of polyploidy. It seems this genus has great potential for the future as a model for addressing fundamental questions in plant evolution.

7 REFERENCES

Ainouche, ML, RJ Bayer, JP Gourret, A Defontaine, and MT Misset. 1999. The allotetraploid invasive weed *Bromus hordeaceus* L. (Poaceae): Genetic diversity, origin and molecular evolution. Folia Geobotanica 34:405-419.

Ainsworth, C. 2000. Boys and girls come out to play: The molecular biology dioecious plants. Annals of Botany 86:211-221.

Akimoto, J, T Fukuhara, and K Kikuzawa. 1999. Sex ratios and genetic variation in a functionally androdioecious species, *Schizopepon bryoniaefolius* (Cucurbitaceae). American Journal of Botany 86:880-886.

Alvarez, I, and JF Wendel. 2003. Ribosomal ITS sequences and plant phylogenetic inference. Molecular Phylogenetics and Evolution 29:417-434.

Antrobus, S, and AJ Lack. 1994. Genetics of colonizing and established populations of *Primula veris*. Heredity 71:252-258.

Arft, AM, and TA Ranker. 1998. Allopolyploid origin and population genetics of the rare orchid *Spiranthes diluvialis*. American Journal of Botany 85:110-122.

Bailey, CD, TG Carr, SA Harris, and CE Hughes. 2003. Characterization of angiosperm nrDNA polymorphism, paralogy, and pseudogenes. Molecular Phylogenetics and Evolution 29:435-455.

Baker, HG. 1955. Self-compatibility and establishment after "long-distance" dispersal. Evolution 9:347-348.

Baker, HG. 1967. Support for Baker's Law - as a rule. Evolution 21:853-856.

Baldwin, BG, MJ Sanderson, JM Porter, MF Wojciechowski, CS Campbell, and MJ Donoghue. 1995. The ITS region of nuclear ribosomal DNA - a valuable source of evidence on angiosperm phylogeny. Annals of the Missouri Botanical Garden 82:247-277.

Balloux, F, and J Goudet. 2002. Statistical properties of population differentiation estimators under stepwise mutation in a finite island model. Molecular Ecology 11:771-783.

Barrett, SCH. 2002. The evolution of plant sexual diversity. Nature Reviews Genetics 3:274-284.

Barrett, SCH, and JR Pannell. 1999. Metapopulation dynamics and mating-system evolution in plants. Pp. 74-100 *in* P. Hollingsworth, R. Bateman and R. Gornall, eds. Molecular Systematics and Plant Evolution. Chapman and Hall, London.

Barrett, SCH, and JS Shore. 1987. Variation and evolution of breeding systems in the *Turnera ulmifolia* complex (Turneraceae). Evolution 41:340-354.

Barrett, SCH, and JS Shore. 1989. Isozyme variation in colonising plants. Pp. 106-126 *in* D. E. Soltis and P. S. Soltis, eds. Isozymes in plant biology. Dioscorides, Portland, Oregon.

Bawa, KS. 1980. Evolution of dioecy in flowering plants. Annual Review of Ecology and Systematics 11:15-39.

Bawa, KS. 1982. Outcrossing and the incidence of dioecism in island floras. American Naturalist 119:866-871.

Bayer, RJ, and DJ Crawford. 1986. Allozyme divergence among five diploid species of *Antennaria* (Asteraceae: Inuleae) and their allopolyploid derivatives. American Journal of Botany 73:287-296.

Bennett, M, and I Leitch. 2003. Plant DNA C-values Database, Kew Gardens (Release 4.0, Jan. 2003), http://www.rbgkew.org.uk/cval/homepage.html.

Berglund, ABN, A Saura, and A Westerbergh. 2001. Genetic differentiation of a polyploid plant on ultramafic soils in Fennoscandia. South African Journal of Science 97:533-535.

Berglund, ABN, and A Westerbergh. 2001. Two postglacial immigration lineages of the polyploid *Cerastium alpinum* (Caryophyllaceae). Hereditas 134:171-183.

Bever, JD, and F Felber. 1992. The theoretical population genetics of autopolyploidy. Oxford Surveys of Evolutionary Biology 8:185-217.

Birky, CW. 1995. Uniparental inheritance of mitochondrial and chloroplast genes: Mechanisms and evolution. Proceedings of the National Academy of Sciences of the United States of America 92:11331-11338.

Birky, CW. 2001. The inheritance of genes in mitochondria and chloroplasts: Laws, mechanisms, and models. Annual Review of Genetics 35:125-148.

Bouza, C, J Castro, L Sanchez, and P Martinez. 2001. Allozymic evidence of parapatric differentiation of brown trout (*Salmo trutta* L.) within an Atlantic river basin of the Iberian Peninsula. Molecular Ecology 10:1455-1469.

Bretagnolle, F, and JD Thompson. 1995. Tansley Review No-78 - Gametes with the somatic chromosome number - mechanisms of their formation and role in the evolution of autopolyploid plants. New Phytologist 129:1-22.

Brochmann, C, and R Elven. 1992. Ecological and genetic consequences of polyploidy in Arctic *Draba* (Brassicaceae). Evolutionary Trends in Plants 6:111-124.

Brochmann, C, DE Soltis, and PS Soltis. 1992. Electrophoretic relationships and phylogeny of Nordic polyploids in *Draba* (Brassicaceae). Plant Systematics and Evolution 182:35-70.

Brochmann, C, QY Xiang, SJ Brunsfeld, DE Soltis, and PS Soltis. 1998. Molecular evidence for polyploid origins in *Saxifraga* (Saxifragaceae): The narrow arctic endemic S-svalbardensis and its widespread allies. American Journal of Botany 85:135-143.

Brunet, J. 1992. Sex allocation in hermaphroditic plants. Trends in Ecology and Evolution 7:79-84.

Bruvo, R, NK Michiels, TG D'Souza, and H Schulenburg. 2004. A simple method for the calculation of microsatellite genotype distances irrespective of ploidy level. Mol Ecol 13:2101-2106.

Buckler, ES, A Ippolito, and TP Holtsford. 1997. The evolution of ribosomal DNA: Divergent paralogues and phylogenetic implications. Genetics 145:821-832.

Camerarius, RJ. 1694. *De sexu plantarum epistola*, Tübingen.

Case, AL, and SCH Barrett. 2001. Ecological differentiation of combined and separate sexes of *Wurmbea dioica* (Colchicaceae) in sympatry. Ecology 82:2601-2616.

Case, AL, and SCH Barrett. 2004. Environmental stress and the evolution of dioecy: *Wurmbea dioica* (Colchicaceae) in Western Australia. Evolutionary Ecology 18:145-164.

Cha, RS, H Zarble, P Keohavong, and WG Thilly. 1992. Mismatch amplification mutation assay (MAMA): application to the C-H-*ras* gene. PCR methods and applications 2:14-20.

Charlesworth, B, D Charlesworth, and NH Barton. 2003. The effects of genetic and geographic structure on neutral variation. Annual Review of Ecology Evolution and Systematics 34:99-125.

Charlesworth, D. 1984. Androdioecy and the evolution of dioecy. Biological Journal of the Linnean Society 23:333-348.

Charlesworth, D. 1999. Theories of the evolution of dioecy. Pp. 33-60 *in* M. A. Geber, T. E. Dawson and L. F. Delph, eds. Gender and Sexual Dimorphism in Flowering Plants. Springer, Heidelberg.

Charlesworth, D. 2003. Effects of inbreeding on the genetic diversity of populations. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 358:1051-1070.

Charlesworth, D, and B Charlesworth. 1978. A model for the evolution of dioecy and gynodioecy. American Naturalist 112:975-997.

Charlesworth, D, and B Charlesworth. 1979. The evolution of genetics of sexual systems in flowering plants. Proceedings of the Royal Society of London, B 205:513-530.

Charlesworth, D, and B Charlesworth. 1987. Inbreeding Depression and Its Evolutionary Consequences. Annual Review of Ecology and Systematics 18:237-268.

Charlesworth, D, and JR Pannell. 2001. Mating systems and population genetic structure in the light of coalescent theory. Pp. 73-95 *in* J. Silvertown and J. Antonovics, eds. Integrating Ecological and Evolutionary Processes in a Spatial Context. Blackwells, Oxford.

Charnov, EL, J Maynard Smith, and JJ Bull. 1976. Why be an hermaphrodite? Nature 263:125-126.

Cheptou, PO, and A Mathias. 2001. Can varying inbreeding depression select for intermediary selfing rates? American Naturalist 157:361-373.

Chung, MG, JL Hamrick, SB Jones, and GS Derda. 1991. Isozyme variation within and amoung populations of *Hosta* (Liliaceae) in Korea. Systematic Botany 16:667-684.

Cockerham, CC, and BS Weir. 1993. Estimation of Gene Flow From F-Statistics. Evolution 47:855-863.

Comai, L. 2000. Genetic and epigenetic interactions in allopolyploid plants. Plant Molecular Biology 43:387-399.

Connor, HE. 1996. Breeding systems in Indomalesian *Spinifex* (Paniceae: Gramineae). Blumea 41:445-454.

Costich, DE, and TR Meagher. 1992. Genetic Variation in *Ecballium elaterium* (Cucurbitaceae): Breeding System and Geographic-Distribution. Journal of Evolutionary Biology 5:589-601.

Costich, DE, and TR Meagher. 2001. Impacts of floral gender and whole-plant gender on floral evolution in *Ecballium elaterium* (Cucurbitaceae). Biological Journal of the Linnean Society 74:475-487.

Cottrell, JE, RC Munro, HE Tabbener, ACM Gillies, GI Forrest, JD Deans, and AJ Lowe. 2002. Distribution of chloroplast DNA variation in British oaks (*Quercus robur* and *Q. petraea*): The influence of postglacial colonisation and human management. Forest Ecology and Management 156:181-195.

Crawford, DJ. 1989. Enzyme electrophoresis and plant systematics. Pp. 146-164 *in* D. E. Soltis and P. S. Soltis, eds. Isozymes in plant biology. Dioscorides, Portland, Oregon.

Cruzan, MB, and ML Arnold. 1993. Ecological and genetic associations in an Iris hybrid zone. Evolution 47:1432-1445.

Darwin, C. 1837. Letter to the Revd. Prof. Henslow., Saturday 23rd September, 1837. Pp. 137-138 *in* N. Barlow, ed. Darwin and Henslow. The growth of an idea (1967). John Murray, London.

Darwin, C. 1877. The Different Forms of Flowers on Plants of the Same Species. Appleton, New York.

Delph, LF. 1990a. Sex-differential resource allocation patterns in the subdioecious shrub *Hebe subalpina*. Ecology 71:1342-1351.

Delph, LF. 1990b. Sex-ratio variation in the gynodioecious shrub *Hebe strictissima* (Scrophulariaceae). Evolution 44:134-142.

Demasure, B, N Sodzi, and RJ Petit. 1995. A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. Molecular Ecology 4:129-131.

Demesure, B, B Comps, and RJ Petit. 1996. Chloroplast DNA phylogeography of the common beech (*Fagus sylvatica* L) in Europe. Evolution 50:2515-2520.

Dolezel, J, P Binarova, and S Lucretti. 1989. Analysis of nuclear DNA content in plant cells by flow cytometry. Biologia Plantarum 31:113-120.

Dorken, ME, and SCH Barrett. 2003. Life-history differentiation and the maintenance of monoecy and dioecy in *Sagittaria latifolia* (Alismataceae). Evolution 57:1973-1988.

Dorken, ME, and SCH Barrett. 2004. Sex determination and the evolution of dioecy from monoecy in *Sagittaria latifolia* (Alismataceae). Proceedings of the Royal Society of London Series B-Biological Sciences 271:213-219.

Dorken, ME, J Friedman, and SCH Barrett. 2002. The evolution and maintenance of monoecy and dioecy in *Sagittaria latifolia* (Alismataceae). Evolution 56:31-41.

Doyle, JJ, and JL Doyle. 1987. A rapid DNA isoloation procedure for small amounts of fresh leaf tissue. Phytochemistry Bulletin 19:11-15.

Durand, B. 1963. Le complexe *Mercurialis annua* L. *s.l.*: une étude biosystématique. Annales des Sciences Naturelles, Botanique, Paris 12:579-736.

Durand, B, and R Durand. 1991. Sex Determination and Reproductive Organ Differentiation in *Mercurialis*. Plant Science 80:49-65.

Durand, B, JP Louis, S Hamdi, E Cabre, LX Yu, B Guerin, and G Teller. 1987. Major regulator genes, phytohormone levels and specific gene expression for reproductive organogenesis in *Mercurialis annua* L. (2n=16). Journal of Cellular Biochemistry:18-20.

Durand, R, and B Durand. 1985. *Mercurialis*. Pp. 376-387 *in* A. H. Halevey, ed. CRC Handbook of Flowering. CRC Press, Boca Raton, Florida.

Durand, R, and B Durand. 1992. Dioecy, monoecy, polyploidy and speciation in annual Mercuries. Bulletin de la Societé Botanique de France, Lettres Botaniques 139:377-399.

El Mousadik, A, and RJ Petit. 1996. High level of genetic differentiation for allelic richness among populations of the argan tree *Argania spinosa* (L) Skeels endemic to Morocco. Theoretical and Applied Genetics 92:832-839.

Ennos, RA. 1994. Estimating the relative rates of pollen and seed migration among plant populations. Heredity 72:250-259.

Ennos, RA, WT Sinclair, X-S Hu, and A Langdon. 1999. Using organelle markers to elucidate the history, ecology and evolution of plant populations. Pp. 1-19 *in* P. M. Hollingsworth, R. M. Bateman and R. J. Gornall, eds. Molecular systematics and plant evolution. Taylor & Francis, London.

Gabe, DR. 1939. Inheritance of sex in *Mercurialis annua*. Comptes Rendus, Academie des Sciences URSS:478-481.

Garcia, P, MI Morris, LE Saenzdemiera, RW Allard, MP Delavega, and G Ladizinsky. 1991. Genetic diversity and adaptedness in tetraploid *Avena*

barbata and its diploid ancestors *Avena hirtula* and *Avena wiestii*. Proceedings of the National Academy of Sciences of the United States of America 88:1207-1211.

Gaur, PK, RW Lichtwardt, and JL Hamrick. 1980. Isozyme variation among soil isolates of *Histoplasma capsulatum*. Experimental Mycology 5:69-77.

Gielly, L, and P Taberlet. 1994. The use of chloroplast DNA to resolve plant phylogenies - noncoding versus Rbcl sequences. Molecular Biology and Evolution 11:769-777.

Giles, BE, and J Goudet. 1997. A case study of genetic structure in a plant metapopulation. Pp. 429-454 *in* I. Hanski and M. E. Gilpin, eds. Metapopulation Biology: Ecology, Genetics, and Evolution. Academic Press, San Diego.

Glover, BJ, and RJ Abbott. 1995. Low genetic diversity in the Scottish endemic *Primula scotica* Hook. New Phytologist 129:147-153.

Goloboff, P. 1999. NONA ver. 2. Published by the author, Tucumán, Argentina.

Goudet, J. 1995. Fstat v-1.2: a computer program to calculate F-statistics. Journal of Heredity 86:485-486.

Graham, SW, and RG Olmstead. 2000. Utility of 17 chloroplast genes for inferring the phylogeny of the basal angiosperms. American Journal of Botany 87:1712-1730.

Güemas, J. 1997. *Mercurialis* L. Pp. 201-210 *in* S. Castroviejo, C. Aedo, C. Benedí, M. Laínz, F. Muñoz Garmendia, G. Nieto Feliner and J. Paiva, eds. Flora Iberica. Vol. III. Haloragaceae - Euphorbiaceae. Real Jardín Botánico, C.S.I.C, Madrid.

Hall, TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41:95-98.

Hamdi, S, G Teller, and JP Louis. 1987. Master regulatory genes: auxin levels and sexual organogeneses in the dioecious plant *Mercurialis annua*. Plant Physiology 85:393-399.

Hamrick, JL, and MJW Godt. 1990. Allozyme diversity in plant species. Pp. 43-63 *in* A. H. D. Brown, M. T. Clegg, A. L. Kahler and B. S. Weir, eds. Plant Population Genetics, Breeding, and Genetic Resources. Sinauer, Sunderland.

Hamrick, JL, and MJW Godt. 1996. Effects of life history traits on genetic diversity in plant species. Philosophical Transactions of the Royal Society of London, B 351:1291-1298.

Hanski, I. 1997. Metapopulation dynamics. Pp. 69-92 *in* I. Hanski and M. E. Gilpin, eds. Metapopulation biology: Ecology, Genetics, Evolution. Academic Press, San Diego.

Hanski, IA, and ME Gilpin. 1997. Metapopulation Biology: Ecology, Genetics and Evolution. Academic Press, San Diego.

Hardy, OJ, M De Loose, X Vekemans, and P Meerts. 2001. Allozyme segregation and inter-cytotype reproductive barriers in the polyploid complex *Centaurea jacea*. Heredity 87:136-145.

Hardy, OJ, and X Vekemans. 2001. Patterns of allozyme variation in diploid and tetraploid *Centaurea jacea* at different spatial scales. Evolution 55:943-954.

Hardy, OJ, and X Vekemans. 2002. SPAGEDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Molecular Ecology Notes 2:618-620.

Harris, DJ, S Carranza, EN Arnold, C Pinho, and N Ferrand. 2002. Complex biogeographical distribution of genetic variation within *Podarcis* wall lizards across the Strait of Gibraltar. Journal of Biogeography 29:1257-1262.

Harris, SA, and R Ingram. 1991. Chloroplast DNA and biosystematics: the effects of intraspecific diversity and plastid transmission. Taxon 40:393.

Hartl, DL, and AG Clark. 1997. Principles of Population Genetics. Sinauer Associates, Sunderland, Massachusetts.

Hedrick, PW. 1970. A new approach to measuring genetic similarity. Evolution 25:276.

Hedrick, PW, and ME Gilpin. 1997. Genetic effective size of a metapopulation. Pp. 165-181 *in* I. Hanski and M. E. Gilpin, eds. Metapopulation biology: Ecology, Genetics, Evolution. Academic Press, San Diego.

Hedrick, PW, ES Hutchinson, and M Mesler. 1991. Estimation of self-fertilization rate and allelic frequencies in diploidized tetraploids. Heredity 67:259-264.

Heslop-Harrison, J, and Y Heslop-Harrison. 1957. The Effect of Carbon Monoxide on Sexuality in Mercurialis ambigua L. fils. New Phytologist 56:352-355.

Hewitt, GM. 1999. Post-glacial re-colonization of European biota. Biological Journal of the Linnean Society 68:87-112.

Heyer, F. 1884. Untersuchungen über das Verhältnis des Geschlechts bei einhäusigen und zweihäusigen Pflanzen, unter Berücksichtigung des Geschlechtsverhältniss bei Thieren und Menschen. Ber. physiolog. Lab. u. Versuchsan. landw. Inst. Univ. Halle. 1:43.

Hodkinson, TR, MW Chase, C Takahashi, IJ Leitch, MD Bennett, and SA Renvoize. 2002. The use of DNA sequencing (ITS and trnL-F), AFLP, and fluorescent *in situ* hybridization to study allopolyploid *Miscanthus* (Poaceae). American Journal of Botany 89:279-286.

Holsinger, KE. 2000. Reproductive systems and evolution in vascular plants. Proceedings of the National Academy of Sciences of the United States of America 97:7037-7042.

Horandl, E, and J Greilhuber. 2002. Diploid and autotetraploid sexuals and their relationships to apomicts in the *Ranunculus cassubicus* group: insights from DNA content and isozyme variation. Plant Systematics and Evolution 234:85-100.

Hsiao, C, NJ Chatterton, KH Asay, and KB Jensen. 1994. Phylogenetic relationships of 10 grass species - an assessment of phylogenetic utility of the Internal transcribed spacer region in nuclear ribosomal DNA in monocots. Genome 37:112-120.

Hu, XS, and RA Ennos. 1997. On estimation of the ratio of pollen to seed flow among plant populations. Heredity 79:541-552.

Huang, H, F Dane, Z Wang, Z Jiang, R Huang, and S Wang. 1997. Isozyme inheritance and variation in *Actinidia*. Heredity 78:328-336.

Hudson, RR. 1990. Gene genealogies and the coalescent process. Pp. 1-44 *in* D. J. Futuyma and J. Antonovics, eds. Oxford Surveys in Evolutionary Biology. Oxford University Press, Oxford.

Hughes, CE, CD Bailey, and SA Harris. 2002. Divergent and reticulate species relationships in *Leucaena* (Fabaceae) inferred from multiple data sources: Insights into polyploid origins and nrDNA polymorphism. American Journal of Botany 89:1057-1073.

Hutchinson. 1959. Cornaceae to Rubiaceae *in* C. C. Townsend and E. Guest, eds. Flora of Iraq. Ministry of Agriculture and Agrarian Reform, Iraq, Baghdad.

Ingvarsson, PK. 2002. A metapopulation perspective on genetic diversity and differentiation in partially self-fertilizing plants. Evolution 56:2368-2373.

Iwata, H, H Nesumi, S Ninomiya, Y Takano, and Y Ukai. 2002. The evaluation of genotype x environment interactions of citrus leaf morphology using image analysis and elliptic Fourier descriptors. Breeding Science 52:243-251.

Iwata, H, S Niikura, S Matsuura, Y Takano, and Y Ukai. 1998. Evaluation of variation of root shape of Japanese radish (Raphanus sativus L.) based on image analysis using elliptic Fourier descriptors. Euphytica 102:143-149.

Iwata, H, and Y Ukai. 2002. SHAPE: A computer program package for quantitative evaluation of biological shapes based on elliptic Fourier descriptors. Journal of Heredity 93:384-385.

Jain, SK, and RS Singh. 1979. Population biology of *Avena*. VII. Allozyme variation in relation to the genome analysis. Botanical Gazette 140:356-363.

Jenczewski, E, K Alix, and E Jenczewski. 2004. From diploids to allopolyploids: The emergence of efficient pairing control genes in plants. Critical Reviews in Plant Sciences 23:21-45.

Kahler, AL, RW Allard, M Krzakowa, CF Wehrhahn, and E Nevo. 1980. Associations between isozyme phenotypes and Environment in the slender Wild Oat (*Avena barbata*) in Israel. Theoretical and Applied Genetics 56:31-47.

Khadka, DK, A Nejidat, M Tal, and GA Golan. 2002. DNA markers for sex: molecular evidence for gender dimorphism in dioecious *Mercurialis annua* L. Molecular Breeding 9:251-257.

King, RA, and C Ferris. 1998. Chloroplast DNA phylogeography of *Alnus glutinosa* (L.) Gaertn. Molecular Ecology 7:1151-1161.

Koch, M, and KG Bernhardt. 2004. Comparative biogeography of the cytotypes of annual *Microthlaspi perfoliatum* (Brassicaceae) in europe using isozymes and cpDNA data: Refugia, diversity centers, and postglacial colonization. American Journal of Botany 91:115-124.

Kohout, V, and P Hamouz. 2000. Annual mercury (*Mercurialis annua* L.): Reasons for expansion on arable land. Zeitschrift fuer Pflanzenkrankheiten und Pflanzenschutz:143-144.

Krahenbuhl, M, and P Kupfer. 1995. Le genre *Mercurialis* (Euphorbiaceae): cytogeographie et evolution du complexe polyploide des *M. perennis* L., *M. ovata* Sternb. & Hoppe et *M. leiocarpa* Sieb. & Zucc. Candollea 50:411.

Krahenbuhl, M, YM Yuan, and P Kupfer. 2002. Chromosome and breeding system evolution of the genus *Mercurialis* (Euphorbiaceae): implications of ITS molecular phylogeny. Plant Systematics and Evolution 234:155-170.

Krebs, SL, and JF Hancock. 1989. Tetrasomic inheritance of isoenzyme markers in the highbush blueberry, *Vaccinium corymbosum* L. Heredity 63:11-18.

Kuhl, FP, and CR Giardina. 1982. Elliptic Fourier Features of a Closed Contour. Computer Graphics and Image Processing 18:236-258.

Lack, AJ, and QON Kay. 1988. Allele frequencies, genetic relationships and heterozygosity in *Polygala vulgaris* populations from contrasting habitats in southern Britain. Biological Journal of the Linnean Society 34:119-147.

Lande, R, and DW Schemske. 1985. The evolution of self-fertilization and inbreeding depression in plants: I. Genetic models. Evolution 39:24-40.

Laushman, RH, A Schnabel, and JL Hamrick. 1996. Electrophoretic evidence for tetrasomic inheritance in the dioecious tree *Maclura pomifera* (raf.) Schneid. Journal of Heredity 87:469-473.

Lavin, M, M Thulin, JN Labat, and RT Pennington. 2000. Africa, the odd man out: Molecular biogeography of dalbergioid legumes (Fabaceae) suggests otherwise. Systematic Botany 25:449-467.

Legatt, RA, and GK Iwama. 2003. Occurance of polyploidy in the fishes. Reviews in fish biology and fisheries 13:237-246.

Leitch, IJ, and MD Bennett. 1997. Polyploidy in angiosperms. Trends in Plant Science 2:470-476.

Levin, DA. 2002. The role of chromosomal change in plant evolution. Oxford University Press, New York.

Levins, R. 1970. Extinction. Pp. 77-107 *in* M. Gerstenhaber, ed. Some mathematical questions in biology. American Mathematical Society, Prividence.

Lewontin, RC. 1974. The Genetic Basis of Evolutionary Change. Cambridge University Press, Cambridge.

Lexer, C, ME Welch, O Raymond, and LH Rieseberg. 2003. The origin of ecological divergence in *Helianthus paradoxus* (Asteraceae): Selection on transgressive characters in a novel hybrid habitat. Evolution 57:1989-2000.

Liston, A, LH Rieseberg, and TS Elias. 1990. Functional androdioecy in the flowering plant *Datisca glomerata*. Nature 343:641-642.

Lloyd, DG. 1975a. Breeding systems in *Cotula* III. Dioecious populations. New Phytologist 74:109-123.

Lloyd, DG. 1975b. Breeding systems in *Cotula* IV. Reversion from dioecy to monoecy. New Phytologist 74:125-145.

Louis, JP. 1989. Genes for the regulation of sex differentiation and male fertility in *Mercurialis annua* L. Journal of Heredity 80:104-111.

Louis, J-P, and B Durand. 1978. Studies with the dioecious angiosperm Mercurialis annua L. (2n=16). Correlation between genic and cytoplasmic male sterility, sex segregation and feminising hormones (cytokinins). Molecular and General Genetics 165:309-322.

Luo, ZW, RM Zhang, and MJ Kearsey. 2004. Theoretical basis for genetic linkage analysis in autotetraploid species. PNAS 101:7040-7045.

Machon, N, M Lefranc, I Bilger, and JP Henry. 1995. Isoenzymes as an Aid to Clarify the Taxonomy of French Elms. Heredity 74:39-47.

Mahy, G, LP Bruederle, B Connors, M Van Hofwegen, and N Vorsa. 2000. Allozyme evidence for genetic autopolyploidy and high genetic diversity in tetraploid cranberry, *Vaccinium oxycoccos* (Ericaceae). American Journal of Botany 87:1882-1889.

Maki, M, M Masuda, and K Inoue. 1996. Tetrasomic segregation of allozyme markers in an endangered plant, *Aster kantoensis*. Journal of Heredity 87:378-380.

Manly, BFJ. 1986. Multivarient statistical methods. A primer. Chapman and Hall, London.

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

Marsden, JE, SJ Schwager, and B May. 1987. Single locus Inheritance in the tetraploid treefrog *Hyla versicolor* with an analysis of expected progeny ratios in tetraploid organisms. Genetics 116:299-311.

Mason, RJ, KE Holsinger, and RK Jansen. 1994. Biparental Inheritance of the chloroplast genome in *Coreopsis* (Asteraceae). Journal of Heredity 85:171-173.

Mazet, F, and SM Shimeld. 2002. Gene duplication and divergence in the early evolution of vertebrates. Current Opinion in Genetics & Development 12:393-396.

McCauley, DE. 1995. The use of chloroplast DNA polymorphism in studies of gene flow in plants. Trends in Ecology & Evolution 10:198-202.

McCauley, DE, J Raveill, and J Antonovics. 1995. Local founding events as determinants of genetic structure in a plant metapopulation. Heredity 75:630-636.

Meerts, P, T Baya, and C Lefebvre. 1998. Allozyme variation in the annual weed species complex *Polygonum aviculare* (Polygonaceae) in relation to ploidy level and colonizing ability. Plant Systematics and Evolution 211:239-256.

Meirmans, PG. 2004. GenoType and GenoDive: Users Manual. Published by the Author, http://staff.science.uva.nl/~meirmans/softindex.html.

Meirmans, PG, and PH van Tienderen. 2004. GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. Molecular Ecology Notes (In Press)

Michaux, JR, E Magnanou, E Paradis, C Nieberding, and R Libois. 2003. Mitochondrial phylogeography of the Woodmouse (*Apodemus sylvaticus*) in the Western Palearctic region. Molecular Ecology 12:685-697.

Miller, JS, and DL Venable. 2000. Polyploidy and the evolution of gender dimorphism in plants. Science 289:2335-2338.

Milligan, BG. 1992. Is Organelle DNA Strictly Maternally Inherited - Power Analysis of a Binomial-Distribution. American Journal of Botany 79:1325-1328.

Murawski, DA, TH Fleming, K Ritland, and JL Hamrick. 1994. Mating system of *Pachycereus pringlei* - an autotetraploid cactus. Heredity 72:86-94.

Murdy, WH, and MEB Carter. 1985. Electrophoretic study of the allopolyploidal origin of *Talinum teretifolium* and the specific status of *T. appalachianum* (Portulacaceae). American Journal of Botany 72:1590-1597.

Nason, JD, JL Hamrick, and TH Fleming. 2002. Historical vicariance and postglacial colonization effects on the evolution of genetic structure in Lophocereus, a Sonoran Desert columnar cactus. Evolution 56:2214-2226.

Nassar, JM, JL Hamrick, and TH Fleming. 2003. Population genetic structure of Venezuelan chiropterophilous columnar cacti (Cactaceae). American Journal of Botany 90:1628-1637.

Nei, M. 1987. Molecular Evolutionary Genetics. Columbia University Press, New York.

Neigel, JE. 1997. A comparison of alternative strategies for estimating gene flow from genetic markers. Annual Review of Ecology and Systematics 28:105-128.

Nixon, KC. 1999. Winclada (BETA) ver. 0.9.9. PUBLISHED BY THE AUTHOR, ITHACA, NY.

Nordborg, M. 2001. Coalescent Theory. Pp. 179-212 *in* D. J. Balding, ed. Handbook of Statiscical Genetics.

Nordborg, M, and P Donnelly. 1997. The coalescent process with selfing. Genetics 146:1185-1195.

Nuismer, SL, and JN Thompson. 2001. Plant polyploidy and non-uniform effects on insect herbivores. Proceedings of the Royal Society of London Series B-Biological Sciences 268:1937-1940.

Ohri, D. 1998. Genome size variation and plant systematics. Annals of Botany 82 (Supplement A):75-83.

Olson, MS. 1997. Bayesian procedures for discriminating among hypotheses with discrete distributions: Inheritance in the tetraploid *Astilbe biternata*. Genetics 147:1933-1942.

Osborn, TC, JC Pires, JA Birchler, DL Auger, ZJ Chen, HS Lee, L Comai, A Madlung, RW Doerge, V Colot, and RA Martienssen. 2003. Understanding mechanisms of novel gene expression in polyploids. Trends in Genetics 19:141-147.

Otto, SP, and J Whitton. 2000. Polyploid incidence and evolution. Annual Review of Genetics 34:401-437.

Palmer, JD, CR Shields, DB Cohen, and TJ Orton. 1983. Chloroplast DNA evolution and the origin of amphidiploid *Brassica* species. Theoretical and Applied Genetics 65:181-189.

Pannell, J. 1997a. Mixed genetic and environmental sex determination in an androdioecious population of Mercurialis annua. Heredity 78:50-56.

Pannell, J. 1997b. Variation in sex ratios and sex allocation in androdioecious Mercurialis annua. Journal of Ecology 85:57-69.

Pannell, J. 1997c. Widespread functional androdioecy in Mercurialis annua L (Euphorbiaceae). Biological Journal of the Linnean Society 61:95-116.

Pannell, JR. 1995. Models of androdioecy and studies of *Mercurialis annua* L. Oxford.

Pannell, JR. 2001. A hypothesis for the evolution of androdioecy: The joint influence of reproductive assurance and local mate competition in a metapopulation. Evolutionary Ecology 14:195-211.

Pannell, JR. 2002. The evolution and maintenance of androdioecy. Annual Review of Ecology and Systematics 33:397-425.

Pannell, JR, and SCH Barrett. 1998. Baker's law revisited: Reproductive assurance in a metapopulation. Evolution 52:657-668.

Pannell, JR, and B Charlesworth. 2000. Effects of metapopulation processes on measures of genetic diversity. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 355:1851-1864.

Pannell, JR, DJ Obbard, and RJA Buggs. 2004. Polyploidy and the sexual system: what can we learn from *Mercurialis annua*? Biological Journal of the Linnean Society (In Press)

Pax, F. 1914. Euphorbiaceae - Acalypheae - Mercurialinae. Engelmann, Leipzig.

Petit, RJ, I Aguinagalde, JL de Beaulieu, C Bittkau, S Brewer, R Cheddadi, R Ennos, S Fineschi, D Grivet, M Lascoux, A Mohanty, GM Muller-Starck, B Demesure-Musch, A Palme, JP Martin, S Rendell, and GG Vendramin. 2003. Glacial refugia: Hotspots but not melting pots of genetic diversity. Science 300:1563-1565.

Popp, M, and B Oxelman. 2001. Inferring the history of the polyploid Silene aegaea (Caryophyllaceae) using plastid and homoeologous nuclear DNA sequences. Molecular Phylogenetics and Evolution 20:474-481.

Prince, VE, and FB Pickett. 2002. Splitting pairs: The diverging fates of duplicated genes. Nature Reviews Genetics 3:827-837.

Prober, SM, LH Spindler, and AHD Brown. 1998. Conservation of the grassy white box woodlands: Effects of remnant population size on genetic diversity in the allotetraploid herb Microseris lanceolata. Conservation Biology 12:1279-1290.

Ramsey, J, and DW Schemske. 1998. Pathways, mechanisms, and rates of polyploid formation in flowering plants. Annual Review of Ecology and Systematics 29:467-501.

Ramsey, J, and DW Schemske. 2002. Neopolyploidy in flowering plants. Annual Review of Ecology and Systematics 33:589-639.

Rauscher, JT, JJ Doyle, and AHD Brown. 2002. Internal transcribed spacer repeat-specific primers and the analysis of hybridization in the Glycine tomentella (Leguminosae) polyploid complex. Molecular Ecology 11:2691-2702.

Raybould, AF, AJ Gray, MJ Lawrence, and DF Marshall. 1991. The Evolution of Spartina-Anglica Hubbard,C.E. (Gramineae) - Genetic-Variation and Status of the Parental Species in Britain. Biological Journal of the Linnean Society 44:369-380.

Raymond, O, F Piola, and C Sanlaville-Boisson. 2002. Inference of reticulation in outcrossing allopolyploid taxa: caveats, likelihood and perspectives. Trends in Ecology & Evolution 17:3-6.

Renner, SS, and RE Ricklefs. 1995. Dioecy and Its Correlates in the Flowering Plants. American Journal of Botany 82:596-606.

Renner, SS, and HS Won. 2001. Repeated evolution of dioecy from monoecy in Siparunaceae (Laurales). Systematic Biology 50:700-712.

Richards, AJ. 1997. Plant Breeding Systems. Chapman and Hall, London.

Ridout, MS, JA Bell, and DW Simpson. 2001. Analysis of segregation data from selfed progeny of allopolyploids. Heredity 87:537-543.

Rieseberg, LH. 1997. Hybrid origins of plant species. Annual Review of Ecology and Systematics 28:359-389.

Rieseberg, LH, and MF Doyle. 1989. Tetrasomic Segregation in the Naturally-Occurring Autotetraploid Allium-Nevii (Alliaceae). Hereditas 111:31-36.

Rieseberg, LH, and NC Ellstrand. 1993. What Can Molecular and Morphological Markers Tell Us About Plant Hybridization. Critical Reviews in Plant Sciences 12:213-241.

Rieseberg, LH, MA Hanson, and CT Philbrick. 1992. Androdioecy is derived from dioecy in Datiscaceae: evidence from restriction site mapping of PCR-amplified chloroplast DNA fragments. Systematic Botany 17:324-336.

Rieseberg, LH, O Raymond, DM Rosenthal, Z Lai, K Livingstone, T Nakazato, JL Durphy, AE Schwarzbach, LA Donovan, and C Lexer. 2003. Major ecological transitions in wild sunflowers facilitated by hybridization. Science 301:1211-1216.

Rogers, DL. 2000. Genotypic diversity and clone size in old-growth populations of coast redwood (Sequoia sempervirens). Canadian Journal of Botany-Revue Canadienne De Botanique 78:1408-1419.

Rohlf, FJ, and JW Archie. 1984. A Comparison of Fourier Methods for the Description of Wing Shape in Mosquitos (Diptera, Culicidae). Systematic Zoology 33:302-317.

Rokas, A, RJ Atkinson, L Webster, G Csoka, and GN Stone. 2003. Out of Anatolia: longitudinal gradients in genetic diversity support an eastern origin for a circum-Mediterranean oak gallwasp Andricus quercustozae. Molecular Ecology 12:2153-2174.

Ronfort, J. 1999. The mutation load under tetrasomic inheritance and its consequences for the evolution of the selfing rate in autotetraploid species. Genetical Research 74:31-42.

Ronfort, J, E Jenczewski, T Bataillon, and F Rousset. 1998. Analysis of population structure in autotetraploid species. Genetics 150:921-930.

Ronquist, F, and JP Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-1574.

Rousset, F. 1997. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. Genetics 145:1219-1228.

Rousset, F. 2003. Effective size in simple metapopulation models. Heredity 91:107-111.

Routley, MB, K Mavraganis, and CG Eckert. 1999. Effect of population size on the mating system in a self- compatible, autogamous plant, *Aquilegia canadensis* (Ranunculaceae). Heredity 82:518-528.

Rumsey, FJ, JC Vogel, SJ Russell, JA Barrett, and M Gibby. 1999. Population structure and conservation biology of the endangered fern Trichomanes speciosum Willd. (Hymenophyllaceae) at its northern distributional limit. Biological Journal of the Linnean Society 66:333-344.

Sakai, AK, K Karoly, and SG Weller. 1989. Inbreeding depression in Schiedea globosa and S. salicaria (Caryophyllaceae), subdioecious and gynodioecious Hawaiian species. American Journal of Botany 76:437-444.

Sakai, AK, and SG Weller. 1991. Ecological aspects of sex expression in subdioecious Schiedea globosa (Caryophyllaceae). American Journal of Botany 78:1280-1288.

Sakai, S. 2001. Thrips pollination of androdioecious *Castilla elastica* (Moraceae) in a seasonal tropical forest. American Journal of Botany 88:1527-1534.

SAS. 2003. JMP for Windows. SAS Institute, Cary, North Carolina, USA.

Sassaman, C. 1989. Inbreeding and sex ratio variation in female-biased populations of a clam shrimp, *Eulimnada texana*. Bulletin of Marine Science 45:425-432.

Sassaman, C. 1995. Sex determination and the evolution of unisexuality in the Conchostraca. Hydrobiologia 298:45-65.

Schaal, BA, DA Hayworth, KM Olsen, JT Rauscher, and WA Smith. 1998. Phylogeographic studies in plants: problems and prospects. Molecular Ecology 7:465-474.

Schemske, DW, and R Lande. 1985. The evolution of self-fertilization and inbreeding depression in plants: II. Empirical observations. Evolution 39:41-52.

Schierenbeck, KA, JL Hamrick, and RN Mack. 1995. Comparison of Allozyme Variability in a Native and an Introduced Species of Lonicera. Heredity 75:1-9.

Seehausen, O. 2004. Hybridization and adaptive radiation. Trends in Ecology and Evolution 19:198-207.

Shoemaker, JS, IS Painter, and BS Weir. 1999. Bayesian statistics in genetics - a guide for the uninitiated. Trends in Genetics 15:354-358.

Shore, JS. 1991. Tetrasomic Inheritance and Isozyme Variation in Turnera- Ulmifolia Vars Elegans Urb and Intermedia Urb (Turneraceae). Heredity 66:305-312.

Shore, JS, and M Triassi. 1998. Paternally biased cpDNA inheritance in Turnera ulmifolia (Turneraceae). American Journal of Botany 85:328-332.

Simmons, MP, and H Ochoterena. 2000. Gaps as characters in sequence-based phylogenetic analyses. Systematic Biology 49:369-381.

Sokal, RR, and FJ Rohlf. 1995. Biometry. W.H. Freeman and Company, New York.

Soltis, DE, LA Johnson, and C Looney. 1996. Discordance between ITS and chloroplast topologies in the Boykinia group (Saxifragaceae). Systematic Botany 21:169-185.

Soltis, DE, and PS Soltis. 1988. Electrophoretic Evidence for Tetrasomic Segregation in Tolmiea- Menziesii (Saxifragaceae). Heredity 60:375-382.

Soltis, DE, and PS Soltis. 1989. Tetrasomic Inheritance in Heuchera-Micrantha (Saxifragaceae). Journal of Heredity 80:123-126.

Soltis, PS, GM Plunkett, SJ Novak, and DE Soltis. 1995. Genetic-Variation in Tragopogon Species - Additional Origins of the Allotetraploids T-Mirus and T-Miscellus (Compositae). American Journal of Botany 82:1329-1341.

Soltis, PS, and DE Soltis. 2000. The role of genetic and genomic attributes in the success of polyploids. Proceedings of the National Academy of Sciences of the United States of America 97:7051-7057.

Soltis, PS, DE Soltis, and PG Wolf. 1991. Allozymic and Chloroplast DNA Analyses of Polyploidy in Polystichum (Dryopteridaceae) .1. The Origins of P-Californicum and P-Scopulinum. Systematic Botany 16:245-256.

Swensen, SM, JN Luthi, and LH Rieseberg. 1998. Datiscaceae revisited: Monophyly and the sequence of breeding system evolution. Systematic Botany 23:157-169.

Swofford, DL. 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). Version 4. Sinaur Associates, Sunderland, Massachusetts

Sybenga, J. 1975. Meiotic Configurations. Springer Verlag, Berlin.

Sybenga, J. 1996. Chromosome pairing affinity and quadrivalent formation in polyploids: Do segmental allopolyploids exist? Genome 39:1176-1184.

Taberlet, P, L Gielly, G Patou, and J Bouvet. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Molecular Biology 17:1105-1109.

Taylor, DJ, and DO Foighil. 2000. Transglobal comparisons of nuclear and mitochondrial genetic structure in a marine polyploid clam (Lasaea, Lasaeidae). Heredity 84:321-330.

Thomas, RG. 1958. Sexuality in diploid and hexaloid races of *Mercurialis annua* L. Annals of Botany 22:55-72.

Thompson, JD, Gibson,T.J., Plewniak,F., Jeanmougin,F. and Higgins,D.G. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research 24:4876-4882.

Thrall, PH, and A Young. 2000. AUTOTET: A program for analysis of autotetraploid genotypic data. Journal of Heredity 91:348-349.

Trewick, SA, M Morgan-Richards, SJ Russell, S Henderson, FJ Rumsey, I Pintér, JA Barrett, M Gibby, and JC Vogel. 2002. Polyploidy, phylogeography and

Pleistocene refugia of the rockfern Asplenium ceterach: Evidence from chloroplast DNA. Molecular Ecology 11:2003-2012.

Tsitrone, A, M Kirkpatrick, DA Levin, and A Tsitrone. 2003. A model for chloroplast capture. Evolution 57:1776-1782.

Tutin, TG, VH Heywood, NA Burges, DM Moore, DH Valentine, SM Walters, and DA Webb. 1968. Flora Europeae. Cambridge University Press, Cambridge.

Vamosi, JC, SP Otto, and SCH Barrett. 2003. Phylogenetic analysis of the ecological correlates of dioecy in angiosperms. Journal of Evolutionary Biology 16:1006-1018.

Vassiliadis, C, M Valero, P Saumitou-Laprade, and B Godelle. 2000. A model for the evolution of high frequencies of males in an androdioecious plant based on a cross-compatibility advantage of males. Heredity 85:413-422.

Vogel, JC, FJ Rumsey, SJ Russell, CJ Cox, JS Holmes, W Bujnoch, C Starks, JA Barrett, and M Gibby. 1999. Genetic structure, reproductive biology and ecology of isolated populations of Asplenium csikii (Aspleniaceae, Pteridophyta). Heredity 83:604-612.

Vogler, DW, and AG Stephenson. 2001. The potential for mixed mating in a self-incompatible plant. International Journal of Plant Sciences 162:801-805.

Wade, MJ, and DE McCauley. 1988. Extinction and recolonization: their effects on the genetic differentiation of local populations. Evolution 42:995-1005.

Wagner, A. 2002. Asymmetric functional divergence of duplicate genes in yeast. Molecular Biology and Evolution 19:1760-1768.

Wakeley, J, and N Aliacar. 2001. Gene genealogies in a metapopulation. Genetics 159:893-905.

Waples, RS. 1988. Estimation of Allele Frequencies at Isoloci. Genetics 118:371-384.

Watson, LE, WJ Elisens, and JR Estes. 1991. Electrophoretic and Cytogenetic Evidence for Allopolyploid Origin of Marshallia mohrii (Asteraceae). American Journal of Botany 78:408-416.

Webb, CJ. 1999. Empirical studies: evolution and maintenance of dimorphic breeding systems. Pp. 61-95 *in* M. A. Geber, T. E. Dawson and L. F. Delph, eds. Gender and Sexual Dimorphism in Flowering Plants. Springer, Berlin Heidelberg.

Webster, GL, and EA Rupert. 1973. Phylogenetic Significance of Pollen Nuclear Number in the Euphorbiaceae. Evolution 27:524-531.

Weeden, NF, and JF Wendel. 1990. Genetics of plant isozymes. Pp. 46-72 *in* D. E. Soltis and P. S. Soltis, eds. Isozymes in plant biology. Chapman and Hall, London.

Weeks, SC, BR Crosser, R Bennett, M Gray, and N Zucker. 2000. Maintenance of androdioecy in the freshwater shrimp, Eulimnadia texana: Estimates of inbreeding depression in two populations. Evolution 54:878-887.

Weiblen, GD, RK Oyama, and MJ Donoghue. 2000. Phylogenetic analysis of dioecy in monocotyledons. American Naturalist 155:46-58.

Weir, BS. 1996. Genetic Analysis II. Sinauer, Sunderland MA.

Weir, BS, and CC Cockerham. 1984. Estimating *F*-statistics for the analysis of population structure. Evolution 38:1358-1370.

Weller, SG, and AK Sakai. 1999. Using phylogenetic approaches for the analysis of plant breeding system evolution. Annual Review of Ecology and Systematics 30:167-199.

Weller, SG, AK Sakai, AE Rankin, A Golonka, B Kutcher, and KE Ashby. 1998. Dioecy and the evolution of pollination systems in Schiedea and Alsinidendron (Caryophyllaceae : Alsinoideae) in the Hawaiian Islands. American Journal of Botany 85:1377-1388.

Weller, SG, WL Wagner, and AK Sakai. 1995. A Phylogenetic Analysis of Schiedea and Alsinidendron (Caryophyllaceae, Alsinoideae) - Implications For the Evolution of Breeding Systems. Systematic Botany 20:315-337.

Wendel, J, A Schnabel, and T Seelanan. 1995. Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). Proceedings of the National Acadamy of Science, USA 92:280-284.

Wendel, JF. 2000. Genome evolution in polyploids. Plant Molecular Biology 42:225-249.

Wendel, JF, CL Brubaker, and AE Percival. 1992. Genetic Diversity in Gossypium-Hirsutum and the Origin of Upland Cotton. American Journal of Botany 79:1291-1310.

Wendel, JF, and NF Weeden. 1990. Visualisation and interpretation of plant isozymes. Pp. 5-45 *in* D. E. Soltis and P. S. Soltis, eds. Isozymes in plant biology. Chapman and Hall, London.

Whitlock, MC. 1992. Nonequilibrium population structure in forked fungus beetles: extinction, colonization, and the genetic variance among populations. American Naturalist 139:952-970.

Whitlock, MC, and DE McCauley. 1990. Some Population Genetic Consequences of Colony Formation and Extinction - Genetic Correlations within Founding Groups. Evolution 44:1717-1724.

Whitlock, MC, and DE McCauley. 1999. Indirect measures of gene flow and migration: F-ST not equal 1/(4Nm+1). Heredity 82:117-125.

Widen, B, and M Widen. 2000. Enzyme variation and inheritance in Glechoma hederacea (Lamiaceae), a diploidized tetraploid. Hereditas 132:229-241.

Wolf, DE, and N Takebayashi. 2004. Pollen limitation and the evolution of androdioecy from dioecy. American Naturalist 163:122-137.

Wolf, PG, PS Soltis, and DE Soltis. 1989. Tetrasomic Inheritance and Chromosome-Pairing Behavior in the Naturally-Occurring Autotetraploid Heuchera-Grossulariifolia (Saxifragaceae). Genome 32:655-659.

Wolfe, KH. 2001. Yesterday's polyploids and the mystery of diploidization. Nature Reviews Genetics 2:333-341.

Wu, R, C-X Ma, G Casella, R Wu, and R Wu. 2004. A Bivalent Polyploid Model for Mapping Quantitative Trait Loci in Outcrossing Tetraploids. Genetics 166:581-595.

Wu, RL, M Gallo-Meagher, RC Littell, and ZB Zeng. 2001. A general polyploid model for analyzing gene segregation in outcrossing tetraploid species. Genetics 159:869-882.

Yampolsky, C. 1919. Inheritance of Sex in Mercurialis annua. American Journal of Botany 6:410-442.

Yampolsky, C. 1930. Induced Alteration of Sex in the Male Plant of Mercurialis annua. Bulletin of the Torrey Botanical Club 57:51-58.

Yampolsky, C. 1957. Further Experiments with Male and Female Grafts of Mercurialis annua. Bulletin of the Torrey Botanical Club 84:1-8.

Yang, TW, YA Yang, and ZG Xiong. 2000. Paternal inheritance of chloroplast DNA in interspecific hybrids in the genus Larrea (Zygophyllaceae). American Journal of Botany 87:1452-1458.

Yang, Z, J El Aidi, T Ait-Ali, C Augur, G Teller, F Schoentgen, R Durand, and B Durand. 1998. Sex-specific marker and trans-zeatin ribosidase in female annual Mercury. Plant Science 139:93-103.

Young, AG, AHD Brown, and FA Zich. 1999. Genetic structure of fragmented populations of the endangered daisy Rutidosis leptorrhynchoides. Conservation Biology 13:256-265.

Young, ND, and J Healy. Gap-coder. http://www-home.cr.duq.edu/~youngnd/GapCoder/.

Yunus, AG, MT Jackson, and JP Catty. 1991. Phenotypic Polymorphism of 6 Enzymes in the Grasspea (Lathyrus- Sativus L). Euphytica 55:33-42.

8 APPENDICES

8.1 CHLOROPLAST HAPLOTYPE VARIATION IN *MERCURIALIS ANNUA*

The plastid genome is widely used for both phylogenetic reconstruction (e.g. Gielly and Taberlet 1994; Graham and Olmstead 2000) and population-level surveys of genetic diversity (Ennos 1994; McCauley 1995; Hu and Ennos 1997). Its widespread use stems from three factors: (1) there are highly conserved genes that can be used as annealing sites for 'universal' PCR primers, which successfully amplify variable DNA sequences across a wide range of species (e.g. Taberlet *et al.* 1991; Demasure *et al.* 1995); (2) the plastid genome exists in very high copy number, making PCR amplification easier than it is for single-copy sequences; and (3), it is usually uniparentally inherited, reducing the chance of more than one copy-type in any one individual, and providing a measure of gender-specific gene flow (Ennos 1994; Birky 1995; Ennos *et al.* 1999). This Appendix presents evidence regarding the inheritance of chloroplasts in diploid *Mercurialis annua*, and the results of a survey of chloroplast variation across the species' European range.

8.1.1 Chloroplast inheritance in *M. annua*

In the majority of angiosperms chloroplasts are inherited exclusively from the maternal parent (Harris and Ingram 1991; Birky 1995, 2001). However, there are a number of species with paternal chloroplast inheritance, and a very small minority in which chloroplasts are inherited from both parents (e.g. Harris and Ingram 1991; Mason et al. 1994; Shore and Triassi 1998; Yang et al. 2000). If chloroplasts are uniparentally inherited, they can be used to infer the maternal and paternal parents of hybrid lineages (Chapter 2), and if there is also sufficient variation in chloroplast types, they can be used to assess the relative rate of gene flow by pollen and seeds (Ennos 1994; Hu and Ennos 1997; Ennos *et al.* 1999). Thus, it is essential to identify whether there is an appreciable rate of bi-parental chloroplast transmission.

To infer strict uniparental inheritance by examining the progeny of artificial crosses requires very large sample sizes, because there is little power to exclude the possibility of a low rate of transmission from the alternative parent (Milligan 1992). However, an

upper limit to the rate of transmission by a parental-type that is not represented amongst the progeny can be found. When using a molecular marker to survey cpDNA inheritance, the sensitivity of the marker screen may also affect the power of the test (Milligan 1992). For example, the paternal parent may contribute 5% of the plastids to some offspring (the maternal parent contributing the other 95%), so that they have a mixed cpDNA type, but the test fails to identify the minority chloroplast type. Thus, some measure of the power of the screen to identify mixed chloroplast types is also desirable.

8.1.1.1 Materials and methods

To identify the mode of chloroplast inheritance in *M. annua*, I analysed the progeny of four artificial crosses between plants known to differ in the sequence of the chloroplast *trn*L-*trn*F spacer region (populations 0059 and 0068, see Appendix 8.1 for population locations). Pairs of male and female plants were placed together in pollen-proof boxes for 4-6 weeks, followed by harvesting and drying the female plant to collect seeds (as described for the crosses in Chapter 3). The *trn*L-*trn*F region was amplified using the universal primers of Taberlet *et al.* (1991). PCR materials and methods are as described in Chapter 2, except that no DNA extraction or purification was performed. Instead, small (2-4mm^2) pieces of fresh very young leaf were macerated directly into each PCR reaction. PCR products were then digested with *Hin*fI (10 U μl^{-1}, New England Biolabs) for 2-4 hours at 37°C, according to the manufacturers instructions (1 µl enzyme; 2.5 µl manufacturers 10x buffer; 12 µl PCR product; 9.5 µl deionised H_2O). *Hin*fI cuts at sequence `GAnT|C`, resulting in four fragments (largest fragment ~145 bp) for individuals from population 0068a and only three fragments for population 0059a (largest fragment ~250bp). The digest products were resolved on a 2.5% (w/v) agarose gel, allowing the two parental chloroplast types to be distinguished. To test the power of the screen to detect mixed chloroplast types, PCR and restriction digests were performed on mixed templates, composed of complete genomic DNA extractions mixed at ratios between 1:1 and 50:1.

8.1.1.2 Results

All offspring were of the same chloroplast type as their maternal parent (total n = 144 across four sets of progeny, see Table 8.1). This was true for crosses in both directions. From this, a maximum rate of paternal transmission (measured as the

proportion of offspring with detectable paternal chloroplasts) for a given statistical power can be calculated (Milligan 1992). For the largest progeny (0068a08 x 0059a66, $n = 45$, Table 8.1) this gives a maximum paternal transmission of 9.7% (power = 0.99) or 3.5% (power = 0.80). If the inheritance mechanism can be assumed to be identical in all four progenies (i.e. the four datasets are informative about the same process), and the results are bulked ($n = 144$), the maximum rate of paternal transmission consistent with the results is 3.2% (power = 0.99) or 1.1% (power = 0.80).

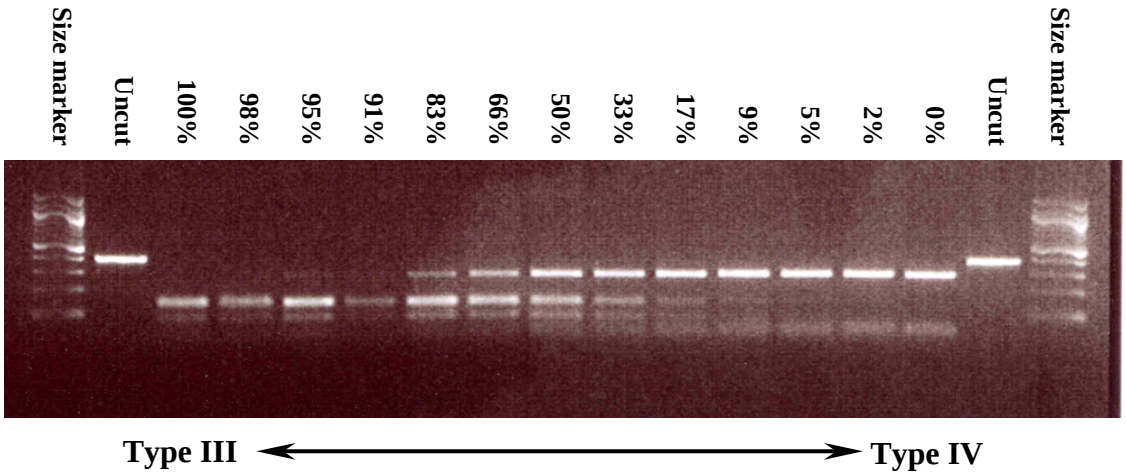
The restriction digest of PCR products amplified from a mixed template indicates that this approach had some power to detect mixed chloroplast DNA within offspring; both DNA sequences were detected when the minority template was 20% of the total (Figure 8.1). Thus, it is possible that the paternal chloroplast type was present at low concentrations in some samples, despite the PCR and restriction digest identifying only the maternal chloroplast type.

The complete absence of any evidence to suggest paternal or biparental chloroplast inheritance in *M. annua* means that it is safe to conclude that most chloroplast inheritance in this group is maternal.

Table 8.1: Chloroplast Inheritance offspring frequencies

Parent		Offspring cpDNA	
Maternal	Paternal	Type IV	Type III
0068a08	0059a66	0	45
0059a57	0068a04	40	0
0068a06	0059a53	0	44
0068a19	0059a55	0	15

Figure 8.1: Restriction digest of mixed *trn*L-*trn*F PCR, using *Hin*fI



This gel shows the products of a *Hin*fI restriction digest of the *trn*L-*trn*F PCR product from mixed a DNA template (percentages refer to the proportion of the template mixtured that was Type III, see main text for details). This is a test of the power of the PCR-digest screen to detect a low rate of biparental chloroplast inheritance. The range of band intensity indicates that although both products are detectable with careful examination at 5%, the lowest reliable rate at which the minority DNA can be detected is approximately 15-20%.

8.1.2 The spatial distribution of chloroplast haplotypes

At a species-wide scale, the spatial distribution of chloroplast haplotypes can be informative for phylogeographic studies (Demesure *et al.* 1996; Schaal *et al.* 1998; Hewitt 1999; Cottrell *et al.* 2002; Petit *et al.* 2003), whilst at smaller scales the variation within and between populations can be used to make inferences about population processes, such as migration (Ennos 1994; Ennos *et al.* 1999). It was hoped that the chloroplast variation identified by sequencing (Chapter 2) would be informative to both aspects of the genetic-diversity study in *M. annua* (Chapter 4). To this end, PCR and restriction digests were used to survey four to six (occasionally 10) individuals from each of the *ca.* 95 populations represented in the isozyme study (Chapter 4, population details listed in Appendix 8.1).

8.1.2.1 Materials and methods

Using the DNA sequence variation identified in the phylogenetic analysis (Chapter 2), restriction enzymes were selected to distinguish between chloroplast Types I – V; (I) polyploid *M. annua* 0020a, (II) polyploid *M. annua* 0060a, (III) diploid *M. annua* 0002a, (IV) diploid *M. annua* 0059a, (IV) *M. huetii* 0719a (see Appendix 8.1 for population identifier locations, and see Figure 2.4, page 35). Some sequence variants could not be distinguished using restriction digests, e.g. tetraploids versus hexaploid type II, and diploid population 0080 (Turkey) versus diploid type III, thus this variation was ignored.

DNA extractions and purification were as described in Chapter 2 (page 23). PCR ingredients and conditions were as follows: 2.5 μM each of dATP, dTTP, dGTP and dTCP, 0.8 μM of each primer, 1 unit of DNA polymerase and 10-100 ng DNA. Reaction volumes were 10 μl (see Table 2.2, page 24 for primer sequences). (1) *trn*L-*trn*F: 2 minutes at 94 °C, then 21 cycles of 25 s at 94 °C, 25 s at 53 °C, 25 s at 72 °C, with a final 2 minutes at 72 °C. (2) *mat*K-*trn*K: 2 minutes at 95 °C, then 25 cycles of 30 s at 95 °C, 30 s at 50 °C, 30 s at 72 °C, with a final 3 minutes at 72 °C. The *trn*L-*trn*F PCR products were digested with the *Hin*fI DNA restriction enzyme (10 U μl^{-1}, New England Biolabs) and the *mat*K-*trn*K PCR products were digested with the *Hha*I DNA restriction enzyme (10 U μl^{-1}, New England Biolabs). Digests were incubated for 2-4 hours at 37°C, according to the manufacturer's instructions (0.15 μl enzyme; 1.5 μl

manufacturers 10x buffer; 10 µl PCR product; 3.4 µl deionised H_2O). The digest products were resolved on a 2.5% (w/v) agarose gel. The *Hin*fI digest (cut site GAnT|C) of the *trn*L-*trn*F fragment distinguished Types II, IV and V from each other, and from Types I or III. The *mat*K-*trn*K fragment was then amplified from samples that proved to be of Type I or III, and digested with *Hha*I (cut site GCG|C), which distinguishes between them (see Table 8.2 for fragment sizes).

8.1.2.2 Results

Little chloroplast diversity was revealed in the *M. annua* complex. Of the two chloroplast types distinguishable by restriction digest in diploid *M. annua*, one was limited to a single population in Israel (0059a), whilst the other occurred in all other diploid populations from Turkey to Galicia (Figure 8.3A). There was more variation present in hexaploid *M. annua*; Type II was found all over Iberia and in Morocco, whilst type I was limited to Iberia, but occurred in Galicia, Catalonia, and Seville (Figure 8.3B). Three mixed populations of Type I and Type II were found, one in each of these three regions. Chloroplast haplotypes corresponded strictly to polyploid level, Types I and II found only in polyploids, Types III and IV found only in diploids, and Type V found only in *M. huetii*. For full results see Table 8.3.

8.1.2.3 Discussion

The low level of diversity in *M. annua* means there is little information available for quantifying population structure in relation to population processes such as gene flow or local extinction. However, the variation available is informative with respect to (1) inter-ploidy gene flow, and (2) the location of glacial refugia. The complete association between chloroplast type and polyploid level, particularly in Iberia where polyploid races of *M. annua* grow close proximity to each other and to *M. huetii* (see Figure 8.3B), suggests that the introgression of chloroplasts between these races is relatively rare.

The European-wide distribution of diversity in diploids broadly supports the suggestion that diploids *M. annua* has recently spread from an Eastern European refugium (Chapter 4), as what little variation there is occurs at the eastern end of the range. The two Types, IV (Israel) and III (others), differ across *mat*K-*trn*K and *trn*L-*trn*F by five substitutions and a 1 bp indel in a poly-T region. The extent of this differentiation perhaps suggests that intermediate variants may be found at the eastern

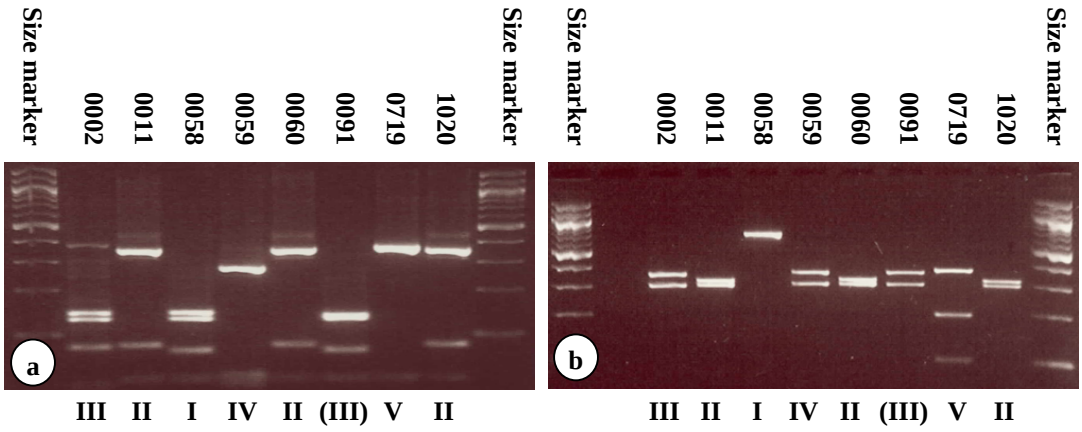
end of the species range. Certainly, the distribution of chloroplast types, and the differentiation between them suggests further fieldwork and more extensive sequencing would be informative to the post-glacial history of *M. annua*.

Table 8.2: Restriction fragment sizes used to distinguish chloroplast haplotypes

Hinf I					*Hha* I	
I	II	III	IV	V	III	I
0020a	*0060a*	*0002a*	*0059a*	*0719a*	*0002a*	*0020a*
145	360	145	281	(not cut)	372	(not cut)
135	120	135	42		344	
74		80	38			
38		32	32			

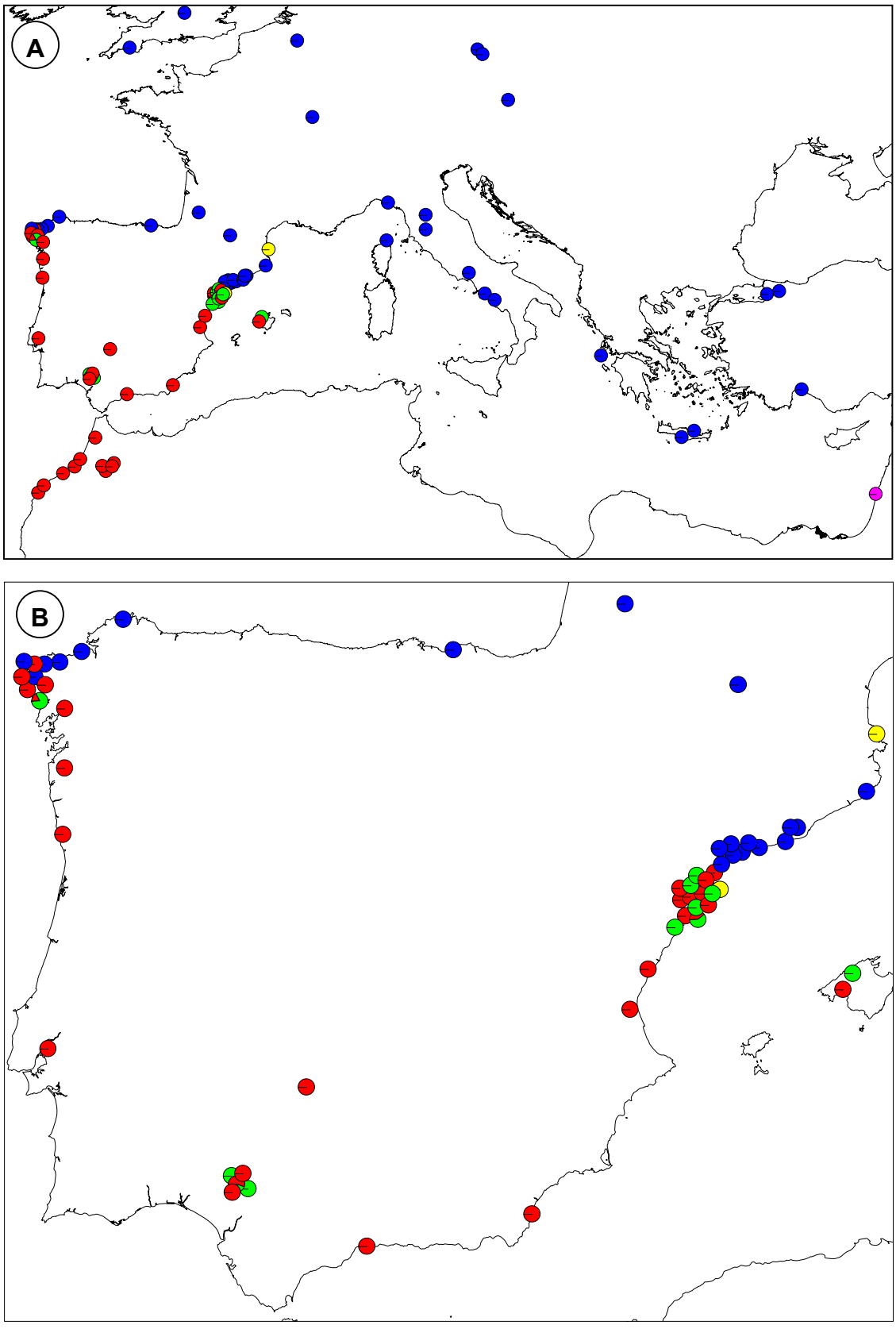
This table show the approximate restriction fragment sizes (in bases) for the five chloroplast Types (I-V). Sizes are approximate because the fragment lengths are taken from sequences, and thus exclude 15-25 bp at the extreme 3' and 5' ends. Below each type identifier is the population number from which the corresponding sequenced individuals (Chapter 2) were taken.

Figure 8.2: *trn*L-*trn*F and *mat*K-*trn*K restriction digests



(a) Restriction fragments from the *Hin*fI digest of *trn*L-*trn*F; (b) Restriction fragments from the *Hha*I digest of *mat*K-*trn*K. Population identifiers are given above the gel, and corresponding chloroplast types below the gel. Note that Types I and III are indistinguishable in the *Hin*fI digest of *trn*L-*trn*F, and that the Tenerife mercury type (population 0091), whilst subtly different, was not consistently distinguishable using this assay.

Figure 8.3: *Mercurialis annua* chloroplast haplotype distribution in Europe



[For captions see over]

Figure 8.3: *Mercurialis annua* chloroplast haplotype distribution in Europe

[See previous page for figure]

Samples were between 4 and 6 individuals per population (rarely 8-10). Colours correspond to haplotypes: I Green, II Red, III Blue, IV Purple, V yellow. I and II are restricted to polyploid *M. annua*, III and IV to diploid *M. annua*. V is restricted to *M. huetii*. Map (A) shows the Europe-wide distribution; map (B) the distribution in Iberia, with pie chart positions altered by the minimum amount necessary to uncover populations obscured by their neighbours.

Table 8.3: Chloroplast haplotype frequencies

For population locations see Appendices 8.1 and 8.8. Chloroplast types are arbitrary labels, Type I and Type II are as used in the phylogenetic analysis (Chapter 2), Type III is the widespread chloroplast haplotype found in most diploid *Mercurialis annua*, (and not distinguished from *M. canariensis* in this survey), Type IV was limited to one Israeli population (0059a), and Type V is the *M. huetii* chloroplast type.

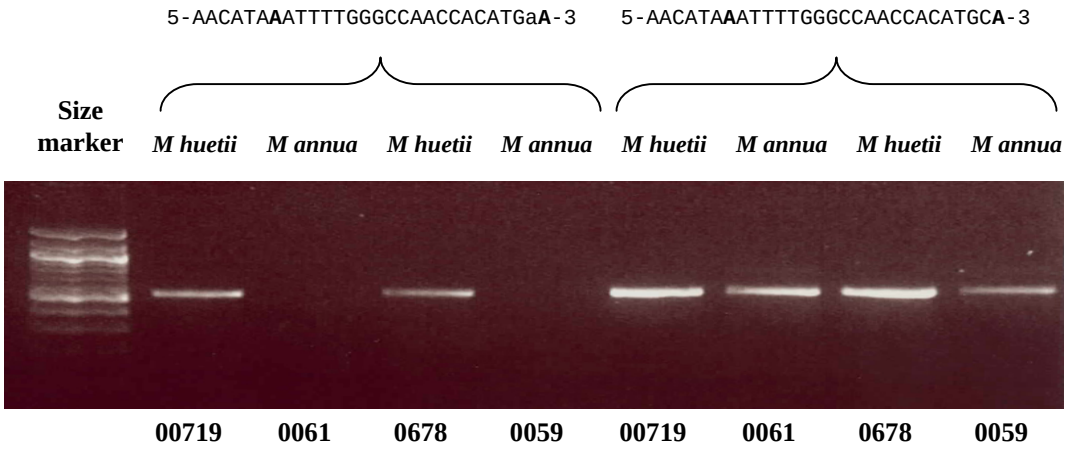
Ploidy	Population	Country	I	II	III	IV	V
2	0002a	Spain	0	0	9	0	0
2	0003a	Spain	0	0	5	0	0
2	0015a	Spain	0	0	5	0	0
2	0018a	Spain	0	0	6	0	0
2	0019a	Spain	0	0	5	0	0
2	0028a	Spain	0	0	5	0	0
2	0039a	Spain	0	0	5	0	0
2	0042a	Spain	0	0	6	0	0
2	0048a	Spain	0	0	5	0	0
2	0049a	Spain	0	0	5	0	0
2	0059a	Israel	0	0	0	5	0
2	0061a	Italy	0	0	5	0	0
2	0062a	UK	0	0	5	0	0
2	0063a	Czech Republic	0	0	6	0	0
2	0064a	Czech Republic	0	0	6	0	0
2	0065a	Austria	0	0	6	0	0
2	0067a	Italy	0	0	6	0	0
2	0068a	Greece	0	0	5	0	0
2	0070a	Belgium	0	0	6	0	0
2	0072a	Crete	0	0	6	0	0
2	0074a	Crete	0	0	6	0	0
2	0075a	UK	0	0	5	0	0
2	0077a	UK	0	0	6	0	0
2	0078a	UK	0	0	6	0	0
2	0079a	Turkey	0	0	6	0	0
2	0080a	Turkey	0	0	5	0	0
2	0081a	Turkey	0	0	6	0	0
2	0225a	Italy	0	0	6	0	0
2	0228a	Italy	0	0	6	0	0
2	0231a	Italy	0	0	6	0	0
2	0232a	Italy	0	0	6	0	0
2	0240a	France	0	0	6	0	0
2	0581a	Spain	0	0	6	0	0

2	0591a	Spain	0	0	6	0	0
2	0593a	Spain	0	0	6	0	0
2	0596a	Spain	0	0	6	0	0
2	0599a	Spain	0	0	6	0	0
2	0601a	Spain	0	0	6	0	0
2	0607a	Spain	0	0	6	0	0
2	0678b	Spain	0	0	0	0	6
2	0695a	Spain	0	0	5	0	0
2	0719a	Spain	0	0	0	0	6
2	1562a	France	0	0	6	0	0
2	1564a	France	0	0	6	0	0
2	1569a	France	0	0	5	0	0
4	0090a	Canary Islands	0	0	3	0	0
4	0091a	Canary Islands	0	0	7	0	0
4	0092a	Canary Islands	0	1	4	0	0
4	0200a	Canary Islands	0	0	6	0	0
4	0206a	Canary Islands	0	1	5	0	0
4	0209a	Canary Islands	0	0	6	0	0
4	0213a	Canary Islands	0	0	6	0	0
4	1018a	Morocco	0	6	0	0	0
4	1020a	Morocco	0	5	0	0	0
4	1031a	Morocco	0	6	0	0	0
6	0004f	Spain	0	5	0	0	0
6	0008a	Spain	0	15	0	0	0
6	0009a	Spain	0	5	0	0	0
6	0010a	Spain	0	5	0	0	0
6	0011a	Spain	0	5	0	0	0
6	0012a	Spain	14	0	0	0	0
6	0020a	Spain	5	0	0	0	0
6	0021a	Spain	0	15	0	0	0
6	0022a	Spain	13	8	0	0	0
6	0055a	Spain	0	5	0	0	0
6	0056a	Spain	5	0	0	0	0
6	0057a	Spain	5	0	0	0	0
6	0058a	Spain	5	6	0	0	0
6	0060a	Morocco	0	5	0	0	0
6	0076a	Spain	0	5	0	0	0
6	0083a	Spain	0	5	0	0	0
6	0084a	Morocco	0	6	0	0	0
6	0085a	Morocco	0	6	0	0	0
6	0086a	Mallorca	6	0	0	0	0
6	0087a	Mallorca	0	6	0	0	0
6	0088a	Morocco	0	6	0	0	0
6	0093a	Canary Islands	0	5	0	0	0
6	0101a	Canary Islands	0	4	0	0	0
6	0506a	Spain	0	4	0	0	0

6	0598a	Spain	0	6	0	0	0
6	0605a	Spain	0	7	0	0	0
6	0608a	Spain	0	6	0	0	0
6	0609a	Spain	0	6	0	0	0
6	0612a	Spain	5	1	0	0	0
6	0614a	Spain	0	6	0	0	0
6	0616a	Spain	0	6	0	0	0
6	0620a	Portugal	0	6	0	0	0
6	0625a	Spain	0	6	0	0	0
6	0630a	Spain	0	6	0	0	0
6	0636a	Spain	0	6	0	0	0
6	0648a	Spain	0	6	0	0	0
6	0650a	Spain	0	6	0	0	0
6	0655a	Spain	6	0	0	0	0
6	0658a	Spain	5	0	0	0	0
6	0660a	Spain	0	5	0	0	0
6	0682a	Spain	6	0	0	0	0
6	1036a	Morocco	0	6	0	0	0
6	1039a	Morocco	0	5	0	0	0
6	1044a	Morocco	0	6	0	0	0

8.2 PCR FOR INTRA-INDIVIDUAL ITS VARIATION

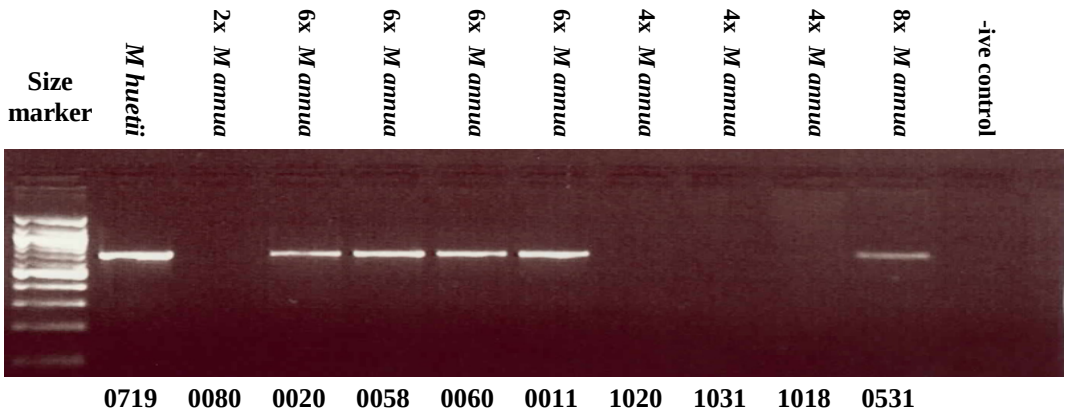
Figure 8.4: The effect of including additional primer mismatches



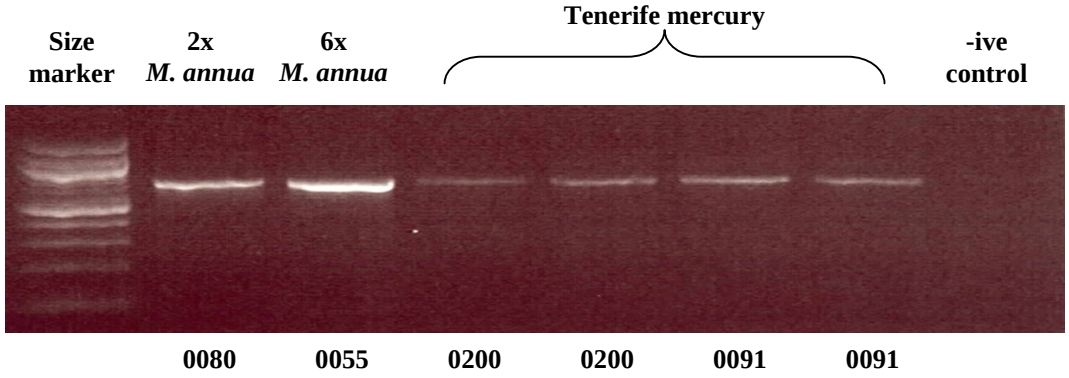
To test the hypothesis that hexaploid *M. annua* is an allopolyploid, with *M. huetii* as a parent, primers were designed to amplify the *M. huetii* ITS sequence but exclude the *M. annua* ITS sequence. Following Cha *et al.* (1992) an additional mismatch to both potential templates was included at the penultimate 3' base. This figure shows the effect of not including this mismatch in the primer Huet468MMF. On the left is the primer including the mismatch (C replaced by A, indicated in lower case), on the right, the same PCR using a primer that does not include the mismatch (PCR conditions and other primers are as shown in Table 2.2, page 24).

Figure 8.5: Intra-individual variation in hexaploid *M. annua* and Tenerife mercury

a) *M. huetii*-like sequence is present in hexaploid *M. annua*



b) *M. annua*-like sequence is present in tetraploid Tenerife mercury



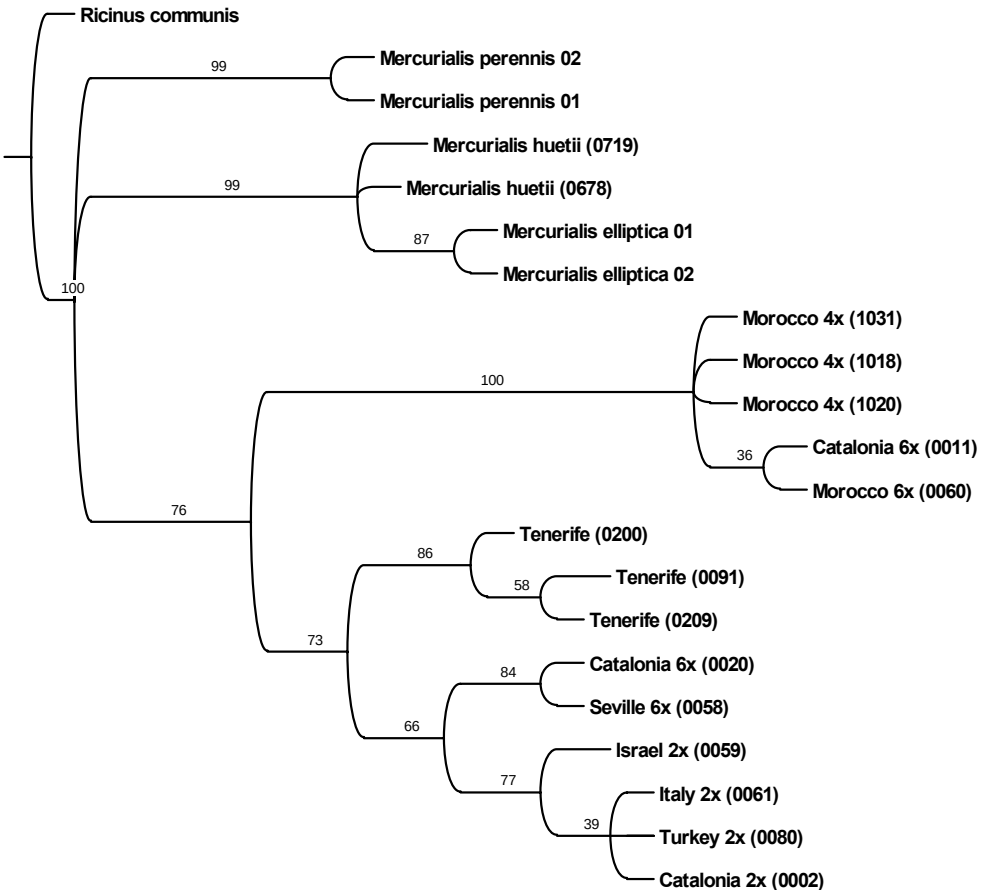
Specific primers were designed to identify mixed ITS types within polyploid individuals (see Chapter 2, page 25). Descriptions are given above the gel, population numbers below. PCR primers and conditions are given in Table 2.2, page 24. (a) A *Mercurialis huetii*-like ITS type was amplified from hexaploid *M. annua*, but not from diploid or tetraploid *M. annua*. The octoploid individual was from an aberrant population in Catalonia. (b) An *M. annua*-like ITS sequence was amplified from the Tenerife mercury.

8.3 ITS AND CPDNA GENE TREES NOT PRESENTED IN THE MAIN TEXT

These analyses were done using NONA (Goloboff 1999) within Winclada (Nixon 1999), using a heuristic search based on tree bisection and reconnection (TBR). 1000 replicates were run, each with 100 starting trees.

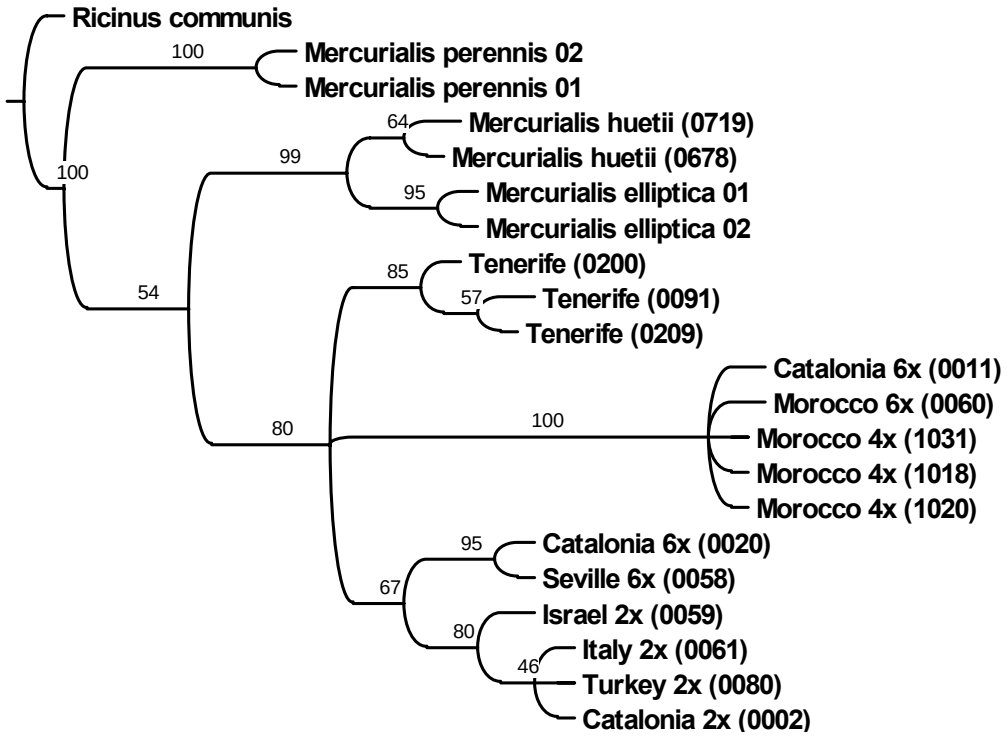
8.3.1 Chloroplast DNA gene trees

Figure 8.6: Parsimony analysis of *mat*K-*trn*K, gaps treated as missing data



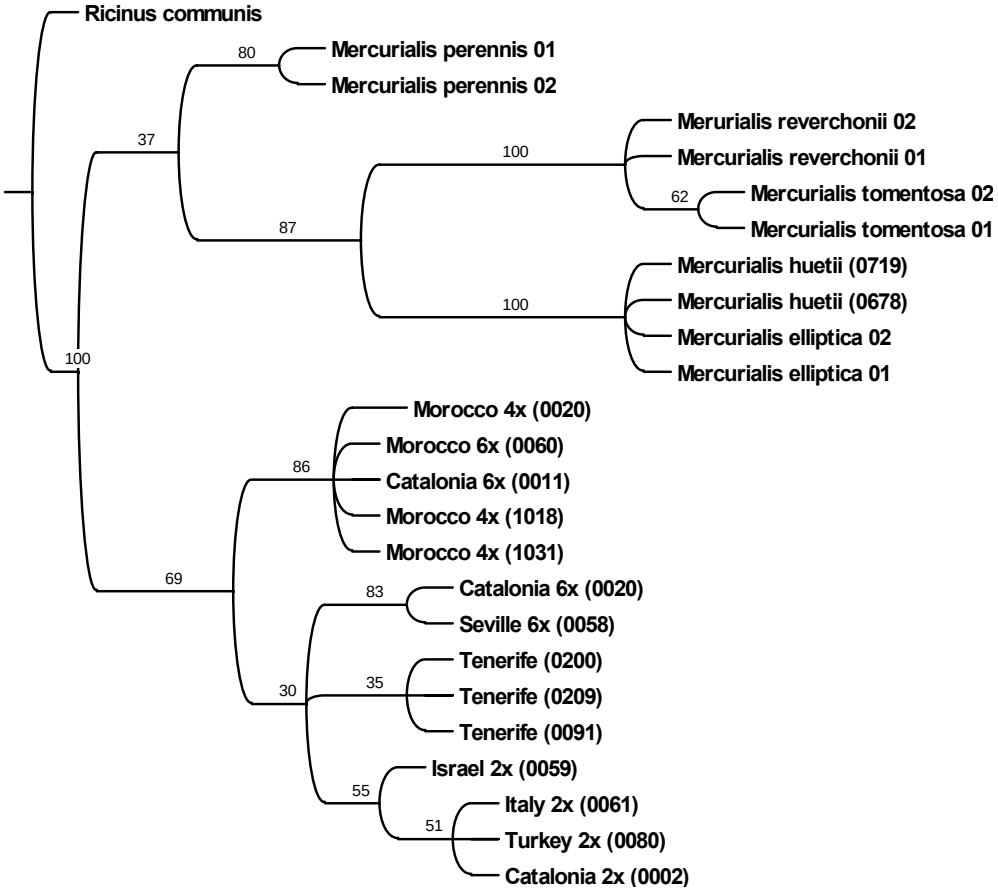
For materials and methods, and an explanation of the terminal labels, see Chapter 2. This is a strict consensus of the two equally most parsimonious trees based on the *mat*K-*trn*K chloroplast sequence alone. Gaps were treated as missing data; numbers above the nodes show percentage bootstrap support (1000 replicates). There were 55 informative characters (L=76, Ci=80, Ri=92).

Figure 8.7: Parsimony analysis of *mat*K-*trn*K, gaps treated as characters



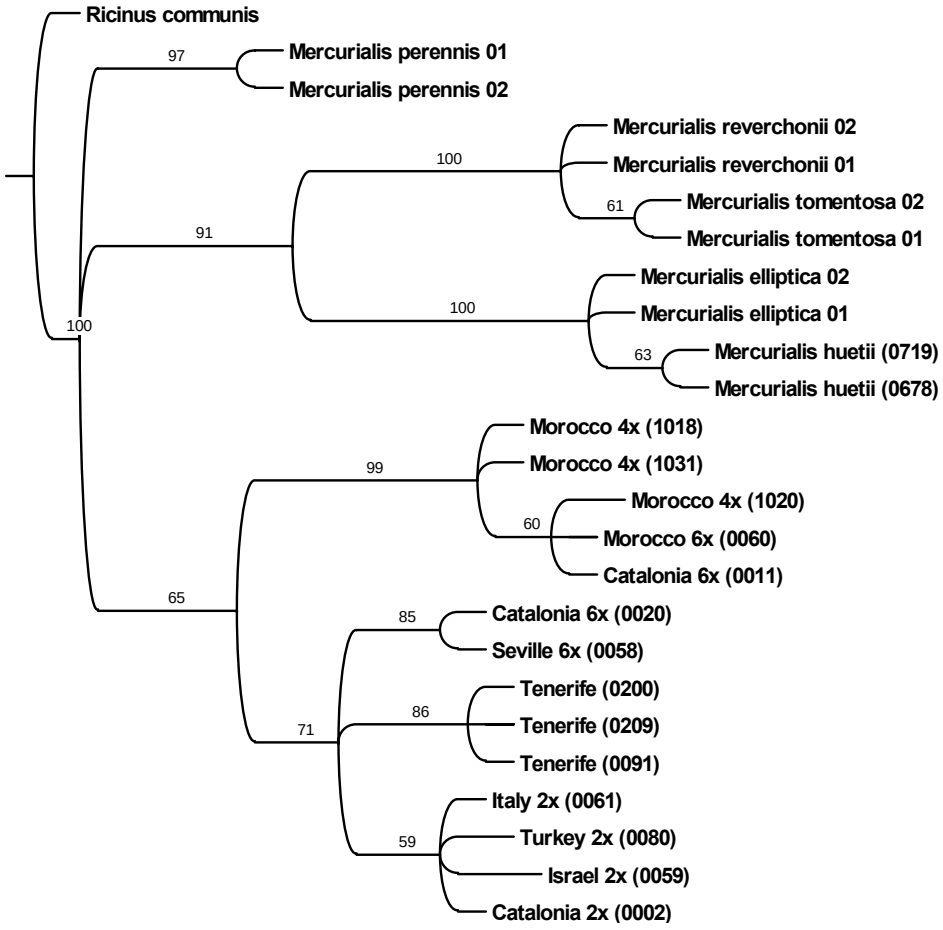
For materials and methods, and an explanation of the terminal labels, see Chapter 2. This is a strict consensus of the four equally most parsimonious trees based on the *mat*K-*trn*K chloroplast sequence alone. Gaps were coded as characters (see Chapter 2 page 26); numbers above the nodes show percentage bootstrap support (1000 replicates). There were 68 informative characters, of which 13 were indel characters (L=91, Ci=81, Ri=92).

Figure 8.8: Parsimony analysis of *trn*L-*trn*F, gaps treated as missing data



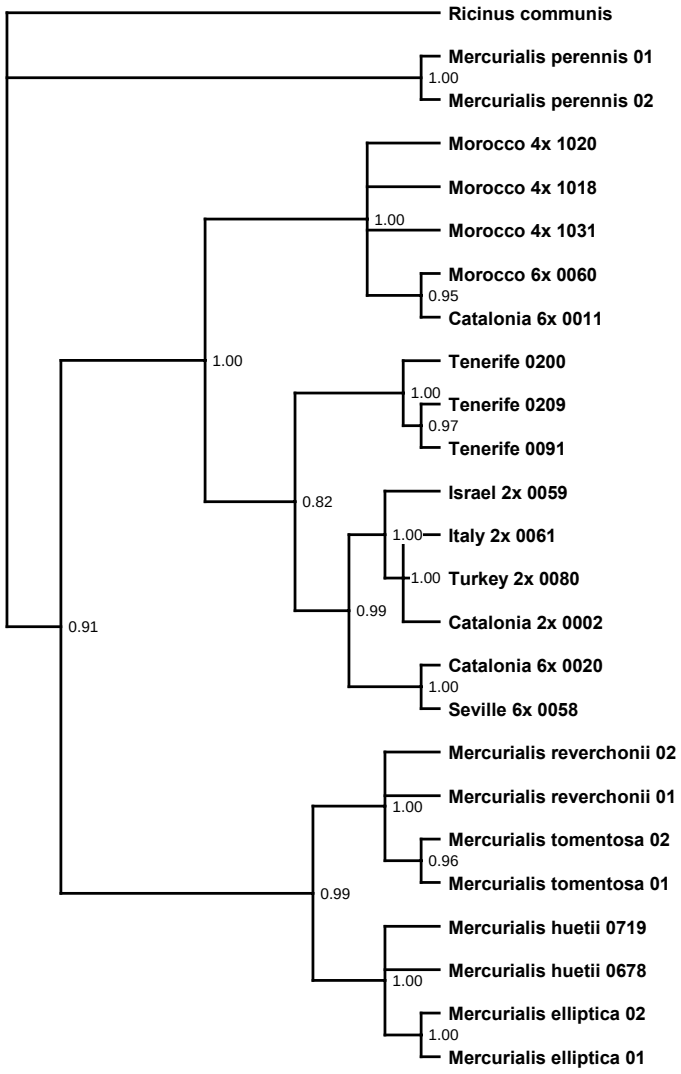
For materials and methods, and an explanation of the terminal labels, see Chapter 2. This is the single most parsimonious tree based on the *trn*L-*trn*F chloroplast sequence alone. Gaps were treated as missing data; numbers above the nodes show percentage bootstrap support (1000 replicates). There were 38 informative characters (L=52, Ci=86, Ri=95).

Figure 8.9: Parsimony analysis of *trn*L-*trn*F, gaps coded as characters.



For materials and methods, and an explanation of the terminal labels, see Chapter 2. This is a strict consensus of the six equally most parsimonious trees based on the *trn*L-*trn*F chloroplast sequence alone. Gaps were coded as characters (see Chapter 2 page 26); numbers above the nodes show percentage bootstrap support (1000 replicates). There were 57 informative characters, of which 19 were indel characters (L=78, Ci=82, Ri=94).

Figure 8.10: Bayesian analysis of joint cpDNA data

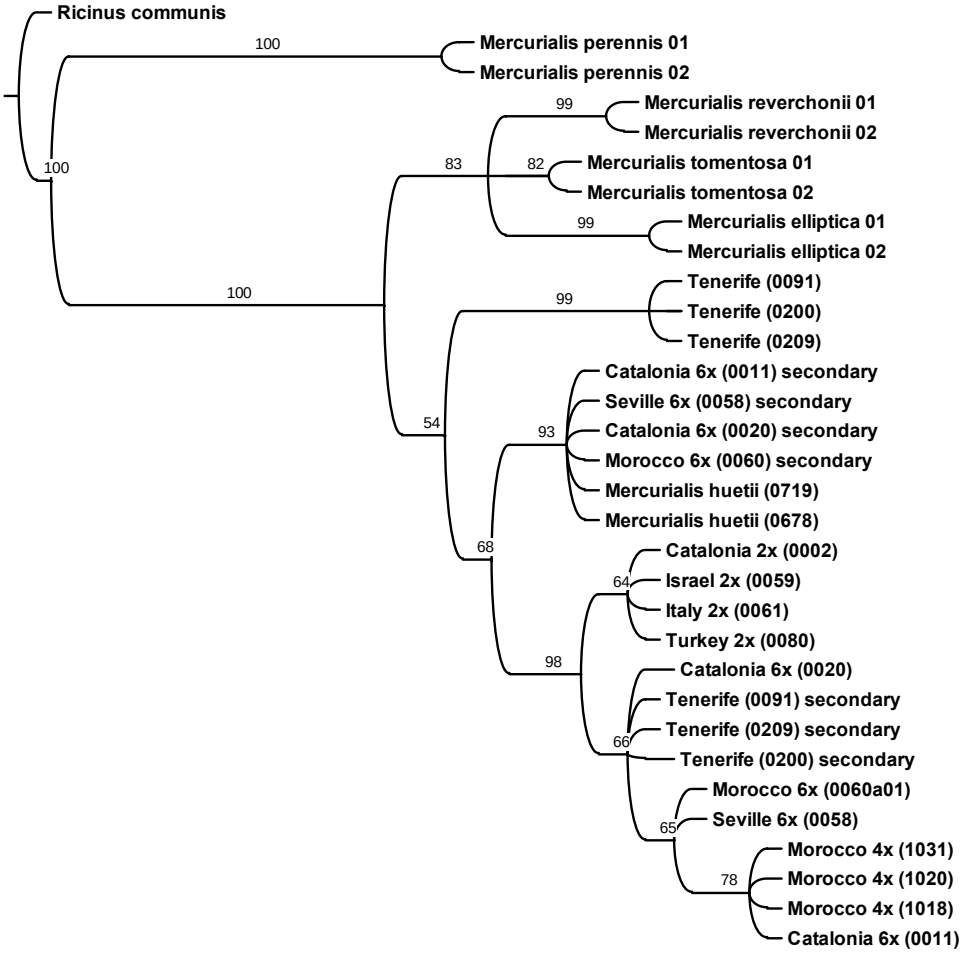


For full materials and methods, and an explanation of the terminal labels, see Chapter 2. For this analysis a GTR model with Gamma-distributed mutation rates (without invariant sites) was specified. All other parameters were as default in MrBayes 3.0.

The data were divided into 5 regions with unlinked model parameters: (1) *trn*L-*trn*F; (2) *mat*K-*trn*K non-coding intron; (3) *mat*K coding sequence 1st codon position; (4) *mat*K coding sequence 2nd codon position; (5) *mat*K coding sequence 3rd codon position. The markov chain ran for 500,000 steps, converging after approximately 10,000 steps. Sampling was done every 100th step starting after the first 50,000. *Ricinus* was specified as the outgroup. The figure shows a majority-rule consensus of sampled trees with approximate posterior probability values for the clades shown by the numbers above the nodes.

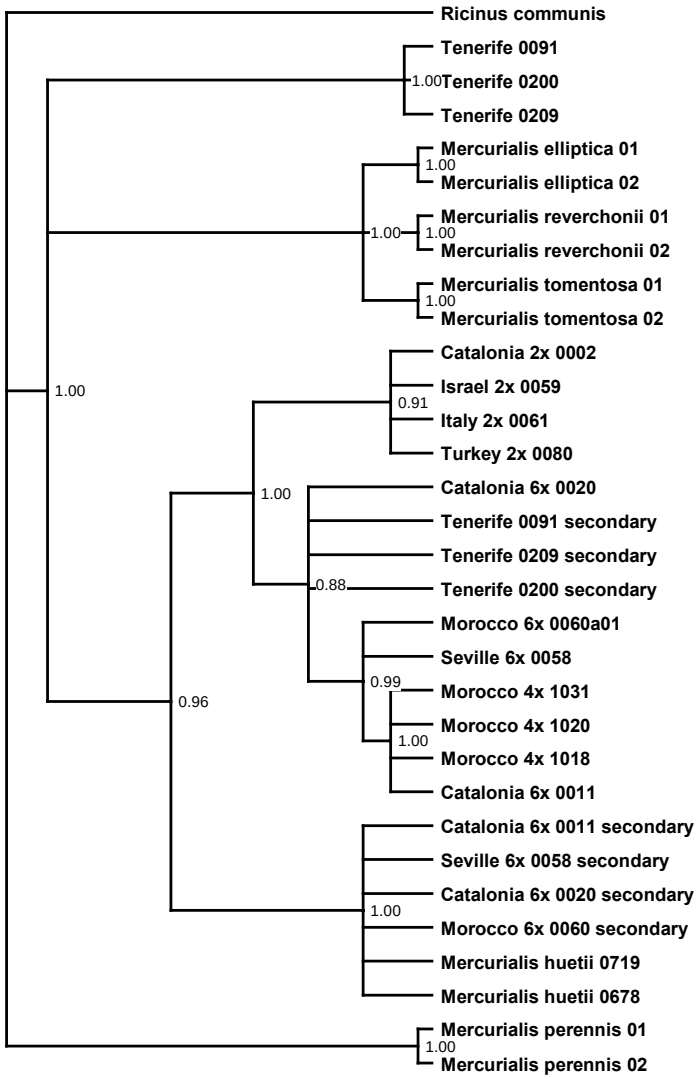
8.3.2 ITS DNA gene trees

Figure 8.11: Parsimony analysis of ITS data, gaps treated as missing data



For materials and methods, and an explanation of the terminal labels, see Chapter 2. This is the single most parsimonious tree based on the joint ITS1:5.8S:ITS2 sequence. Gaps were treated as missing data; numbers above the nodes show percentage bootstrap support (5000 replicates). There were 102 informative characters (L=139, Ci=89, Ri=94).

Figure 8.12: Bayesian analysis of ITS sequence data



For full materials and methods, and an explanation of the terminal labels, see Chapter 2. For this analysis a GTR model with Gamma-distributed mutation rates (with invariant sites) was specified. All other parameters were as default in MrBayes 3.0.

The Markov chain ran for 600,000 steps, converging after approximately 50,000 steps. Sampling was done every 100th step starting after the first 100,000. *Ricinus* was specified as the outgroup. The figure shows a majority-rule consensus of sampled trees with approximate posterior probability values for the clades shown by the numbers above the nodes.

8.4 MANUSCRIPT: "A NEW SPECIES OF *MERCURIALIS* ENDEMIC TO THE CANARY ISLANDS"

[This appendix is a reproduction of jointly authored manuscript currently in preparation. Headings and figures have been re-numbered to conform to the rest of the thesis. References are included after the text]

A NEW SPECIES OF *MERCURIALIS* (EUPHORBIACEAE) ENDEMIC TO THE CANARY ISLANDS

Darren J. Obbard, John R. Pannell and Stephen A. Harris

Department of Plant Sciences, University of Oxford, South Parks Road, Oxford, OX1 3RB, UK.

8.4.1 Abstract

A new species in the genus *Mercurialis* is described in the present paper. This species is known only from the Canary Islands and can be distinguished from the morphologically similar *M. annua* by long stipules (more than 4 mm), clusters of male flowers subtended by a bract and fruits with 3-4 locules. A key to the species in the genus *Mercurialis* is provided.

8.4.2 Introduction

The last taxonomic monograph of the genus *Mercurialis* (Euphorbiaceae) was completed by Pax in 1914, and eight species were recognised (Table 8.4) distributed in the Mediterranean and temperate Eurasia to Japan. However, the number of species, and especially infraspecific taxa, has varied in the intervening 90 years depending upon the treatment of variation in annual *M. annua* L. (Govaerts *et al*., 2000; Durand 1963; Durand and Durand 1984; Durand and Durand 1985; Durand and Durand 1992; Güemas, 1997; Tutin, 1968).

Mercurialis annua comprises a polyploid complex (2x-12x) of anthropogenic ruderals (Durrand, 1963). This complex is occasionally, wholly or partially, spilt into *M. annua sensu stricto*, *M. monoica* (Moris) Durand and *M. ambigua* L. *sensu* Durand (1963), according to polyploid level and sexual system. Dioecy in the *M. annua* complex is limited to the diploids, and monoecy is limited to the polyploids but morphological differentiation between the types is slight (Durand and Durand 1985, and this paper). *Mercurialis annua* was instrumental in the recognition of dioecy amongst plants (Camerarius, 1694), and was used as an early experimental model for sex expression (Heyer 1884; Yampolsky 1919; Gabe 1939; Heslop-Harrison and Heslop-Harrison 1957) and the molecular basis of sex determination (Louis and Durand 1978; Hamdi *et al*. 1987; Louis 1989; Durand and Durand 1991). Furthermore, since the recognition that polyploid *M. annua* is monoecious or androdioecious (Thomas 1958; Durand 1963; Pannell 1997c), it has become an important model for the evolution and ecology of androdioecy (Pannell 2001; Pannell 2002).

In addition to *M. annua*, two dioecious annual taxa can readily be distinguished in the genus, *M. huetii* Hanry (Güemas, 1997; Obbard, 2004) and a taxon found only in the Canary Islands (referred to as Tenerife mercury in Obbard, 2004). It is this latter taxon that is the subject of the present paper.

8.4.3 Materials and methods

All specimens of annual *Mercurialis* species were examined from the following herbaria OXF, BM, RDN, BC, LY, MACB, BCN, TFMC, ORT, TFC. To confirm

Table 8.4: Taxa in the genus *Mercurialis*

Taxon	Ploidy (x=8)	Sexual system	Habit	Pax (1914)
M. annua L.	2x *(x=8)*	Dioecious	Annual	*
	4x—12x *(x=8)*	Monoecious / Androdioecious	Annual	
M. canariensis D.J.Obbard, J.Pannell & S.A.Harris	4x *(x=8)*	Dioecious	Annual	
M. huetii Hanry.	2x *(x=8)*	Dioecious	Annual	
M. perennis L.	6x—12x *(x=8)*	Dioecious	Rhizomatous perennial	*
M. ovata Sternb. & Hoppe.	2x—4x *(x=8)*	Dioecious	Rhizomatous perennial	*
M. leiocarpa Sieber & Zucc.	2x *(x=8)*	Monoecious	Rhizomatous perennial	*
	6x *(x=8)*	*(Not reported)*	Rhizomatous perennial	
M. elliptica Lam.	*2n=42, 2n=220*	Dioecious	Woody Perennial	*
M. corsica Cosson.	*2n=66*	Dioecious	Woody Perennial	*
M. tomentosa L.	*2n=26*	Dioecious	Woody Perennial	*
M. reverchonii Rouy.	*2n=26*	Dioecious	Woody Perennial	*

morphological differentiation under common conditions, plants from a range of populations were grown in Oxford using a randomised block design (sample size 179; details in Obbard, 2004). Thirteen morphological characters were measured and analysed and detailed in Obbard (2004).

8.4.4 Results

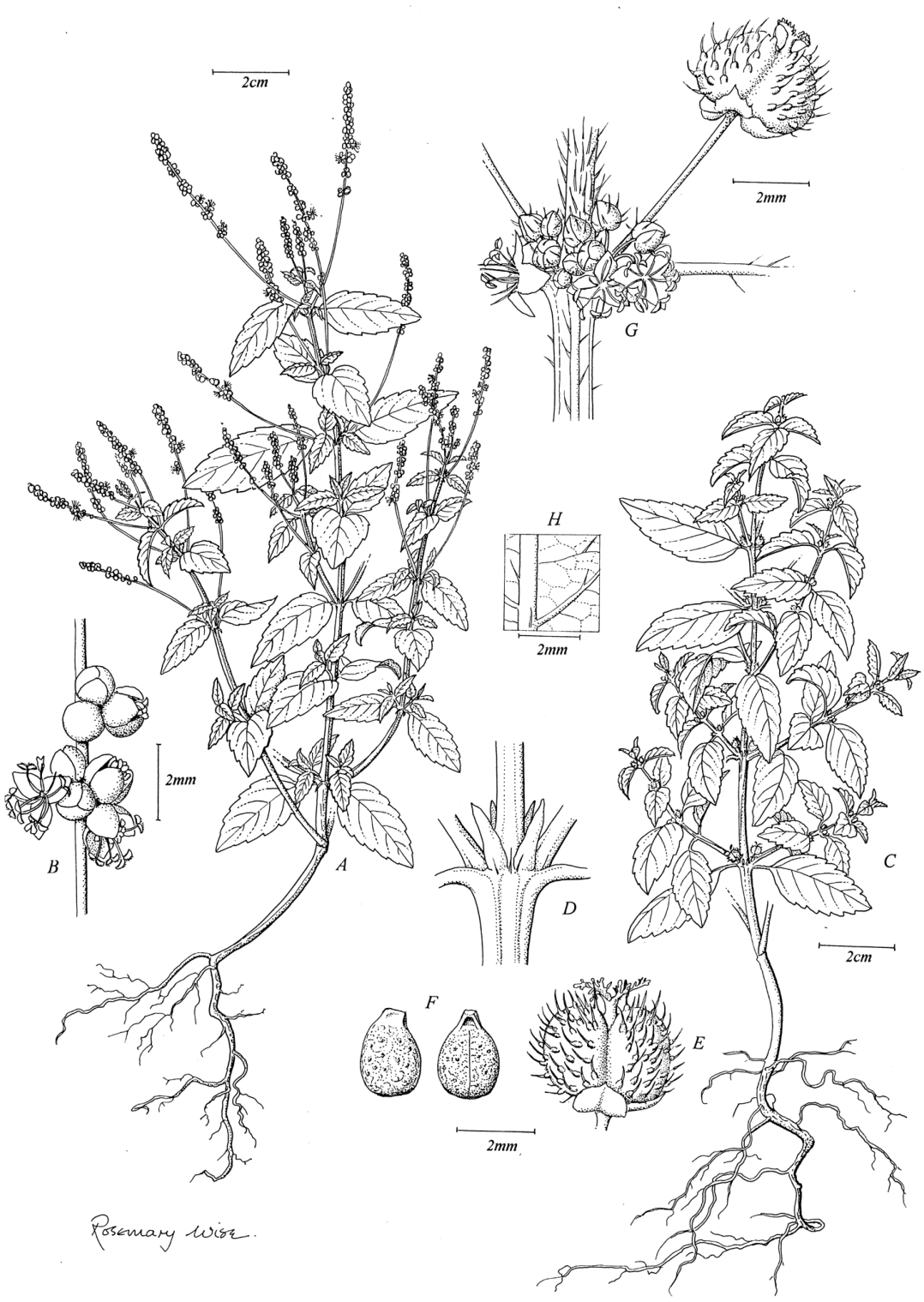
Cluster analysis of glasshouse-grown material grouped the two *M. huetii* populations together and the discriminant functions correctly assigned 100% of *M. huetii* individuals, supporting Durand and Durand's (1985) view that, despite the absence of discrete diagnostic characters, *M. huetii* is morphologically distinct from *M. annua*. Furthermore, Durand and Durand's observations that, in the field, *M. huetii* is generally smaller than *M. annua*, with relatively longer branches and smaller leaves is also true for glasshouse-grown material ($p < 0.001$ for each).

The Tenerife mercury is morphologically distinct (Obbard, 2004; Figure 8.13, Figure 8.14). Morphological characters associated with Canarian *Mercurialis* included (Table 8.5) large size, occasionally branched male inflorescences, large stipules (Figure 8.16a), male flower bracts, and a high frequency of tricapsulate female flowers (Figure 8.17b). Of these characters, the presence of male flower bracts and stipules > 4 mm long are diagnostic.

Table 8.5: Comparison of *Mercurialis annua*, *M. huetii*, and *M. canariensis*.

Species	***M. annua***		***M. huetii***	***M. canariensis***
Chromosome number (x = 8)	2x	4x – 12x	2x	4x
Distribution	Northern Spain, Central and Northern Europe, Middle East. Naturalised in N. America, S. Africa and Japan.	Iberia, North Africa, Sardinia, Tunisia, Canary Islands (including southern Tenerife)	South eastern France, north eastern Spain. Rare.	Canary Islands (La Palma, Lanzarote northern Tenerife,)
Habitat	Common in untended urban gardens, on road verges, on waste ground and as a horticultural weed. Recently disturbed, well-drained, anthropogenic sites, rarely found growing in shade or amongst competition.		Rocky, disturbed ground. Not usually found in anthropogenic environments.	As *M. annua*, but additionally found growing in competition with other herbs, and in shady conditions as an understorey herb.
Sexual system	Dioecious (rarely inconstant females)	Monoecious, males often present in 6x populations.	dioecious (rarely inconstant females)	dioecious (rarely inconstant females)
Plant height (cm)	25-40(-60)	25-40(-60)	10-20	(15-)20-50(-100)
Leaf size				
Male inflorescences	Ebracteate, unbranched	Ebracteate, unbranched	Ebracteate, unbranched	Bracteate [1(-4) mm], often branched
Fruits	Bi-locular (rarely tri-) capsules	Bi- or tri-locular capsules	Bi-locular capsules	Tri- or tetra-locular capsules (rarely bi-)
Stipules (mm)	1.5-3.5		1-2	4-8

Figure 8.13: *Mercurialis annua*



[See next page for legend]

Figure 8.13: *Mercurialis annua*

[see previous page for figure]

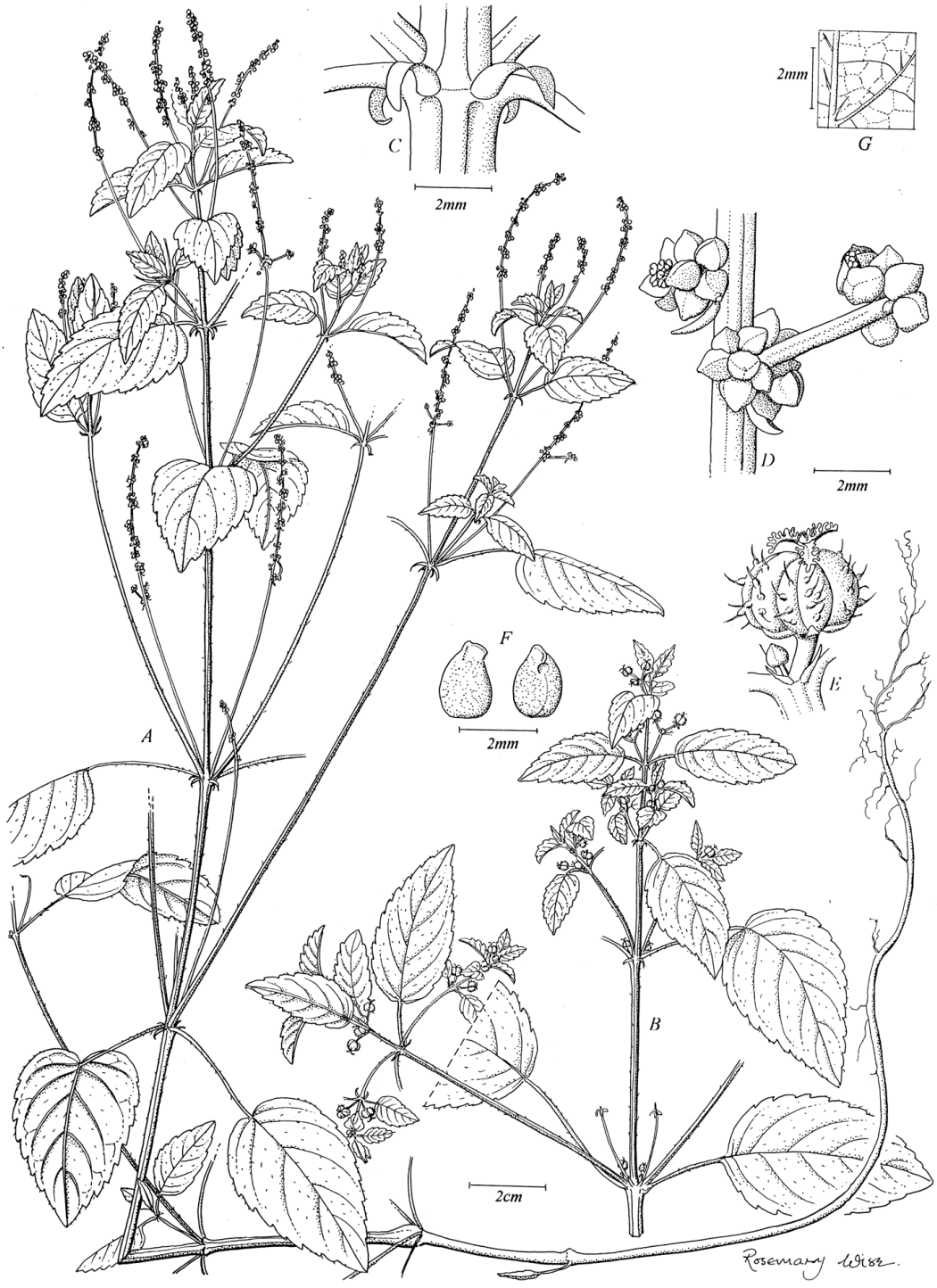
Mercurialis annua. A) Male diploid *Mercurialis annua* with pedunculate inflorescences. B) Detail of male inflorescence. Buds are usually glabrous (as shown) but polyploid males (i.e. from androdioecious populations) often have trichomes. Inflorescences are almost never branched. C) Female diploid *Mercurialis annua*. D) Detail of leaf node. Stipules are always upright and 1.5 to 3.5 mm long. E) Seed capsule. Trichome density is very variable, with some individuals producing almost glabrous capsules. In diploid *M. annua* flowers are usually bicapsulate, with some plants producing a low rate of tricapsulate flowers. F) Seed. G) Male and female flowers on a cosexual polyploid. The occurrence of trichomes on male flowers is highly variable, being common Iberia and North Africa. In cosexual hexaploid and tetraploid individuals tricapsulate female flowers are common. H) Leaf surface, trichome density on leaves and stems is highly variable, but considerably higher in polyploids than diploids. Material illustrated: male, *J. Carbonell s.n.* (5th May 1981; BCN) and female, *Josep Vicens Fandos s.n*. (13th Match 1995; BCN).

Figure 8.14: *Mercurialis canariensis*

[see next page for figure]

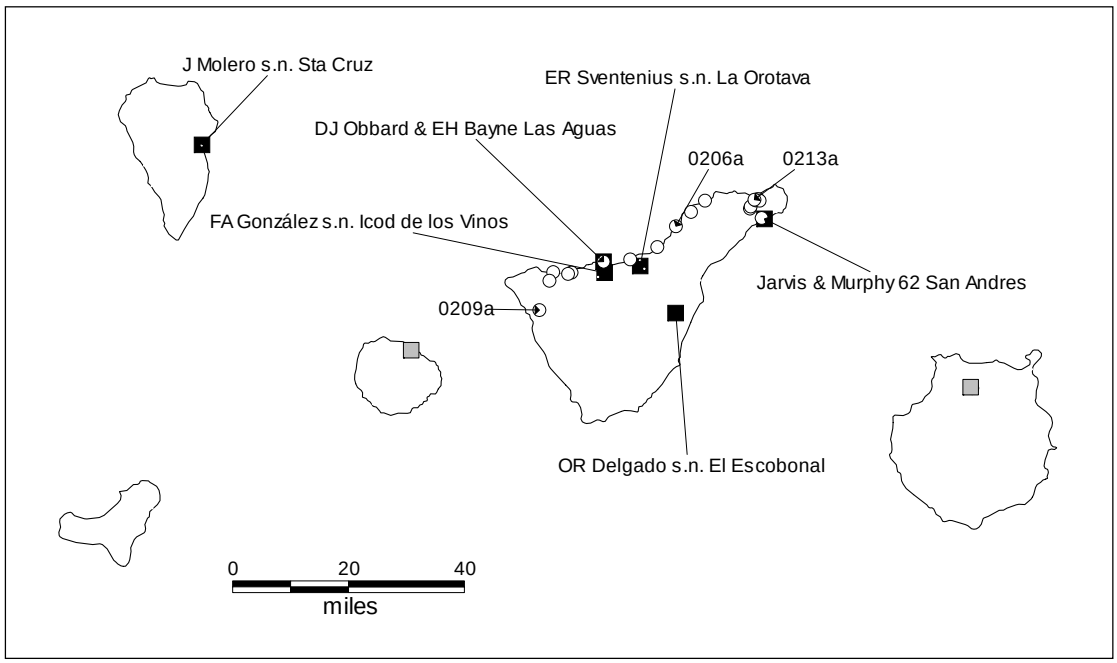
Mercurialis canariensis. While height is highly variable, plants are generally taller than diploid *M. annua*, frequently up to 100cm. They are more robust in appearance, with longer internodes and larger leaves, and unlike diploid or other polyploid *M. annua*, they are often found growing in shade and amongst dense horticultural and roadside weeds. The plants are very common on horticultural land across the northern half of the island: all the dioecious plants examined seem to belong to this group. A) Male Tenerife accession with pedunculate inflorescences. B) Female Tenerife accession. Pedicels are longer than in the (usually) subsessile diploid *M. annua*, and frequently branched. C) Stipules are 4 to 7mm long and recurved. D) Detail of male inflorescence: Flower clusters are subtended by a bract, and peduncles are frequently branched. E) Seed capsule. Trichomes always present, but restricted to the edge. Female flowers are most commonly tricapsulate. F) Seed. G) Leaf surface, trichome density is variable, but generally higher than in diploid *M. annua*. Material illustrated: *DJ Obbard & EH Bayne 2* (15th March 2003).

Figure 8.14: *Mercurialis canariensis*



[See previous page for legend]

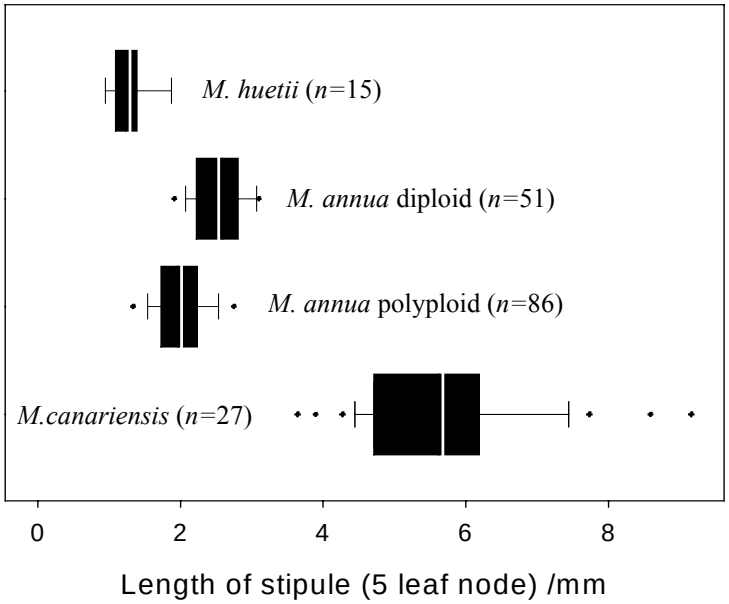
Figure 8.15: Map of the distibution of *Mercurialis* collections in the Canary Islands



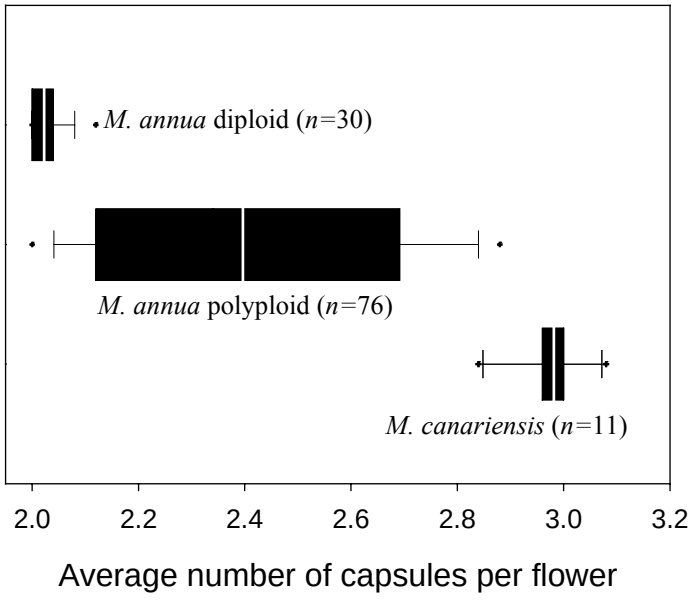
Map of the distribution of *Mercurialis canariensis*. Black squares represent herbarium specimens; white circles represent *M. canariensis* seed collections; grey squares represent hexaploid *M. annua* seed collections. Herbarium specimens are refered to in the main text. Numbers are population identifiers for the *M. canariensis* populations used in the morphometric analysis.

Figure 8.16: Diagnostic character differences between *Mercurialis annua*, *M. huetii*, and *M. canariensis*

(a)



(b)



(a) Stipule length (in greenhouse-grown material) at the fifth leaf node from the base and (b) the proportion of trilocular fruits in annual *Mercurialis* species, where *n* is the sample size. For (b) the sample size for *M. huetii* (fewer than five plants with fruits) was too small to be meaningful. All comparisons are statistically different from each other (Tukey test; $p < 0.0001$)

8.4.5 Discussion

The Canary Island *Mercurialis* considered in this investigation is morphologically distinct in the field having long stipules (more than 4 mm), clusters of male flowers subtended by a bract and fruits with 3-4 locules compared to other annual *Mercurialis* species (Table 8.5). Furthermore, these characters are maintained in cultivation. In addition, the taxon is common in the north of Tenerife and probably on the other islands of the Macaronesian Archipelago. These factors justify recognising the taxon as a distinct entity and giving it formal species name.

Mercurialis canariensis D.J.Obbard, J.Pannell & S.A.Harris

Diagnosis: Affinis *Mercurialis annua* L. affinis sed stipulae plus quam 4 mm, flores mas spicas interrupto-glomeruliflores cum bracteae subtentus et capsulae 3-4-locularis.

Holotype: *DJ Obbard & EH Bayne* 1 Canary Islands: Tenerife, Las Aguas, 28°23.64' N 16°38.17' W, 15 - 30m above sea level. 15th March 2003 [FHO (male)].

Other specimens: *DJ Obbard & EH Bayne* 2 Canary Islands: Tenerife, Las Aguas, 28°23.64' N 16°38.17' W, 15 - 30m above sea level. 15th March 2003 [FHO (female)]. *DJ Obbard & EH Bayne* 3 Canary Islands: Tenerife, Las Aguas, 28° 23.64' N 16° 38.17' W, 15 - 30m above sea level. 15th March 2003 [FHO (male)]. *DJ Obbard & EH Bayne* 4 Canary Islands: Tenerife, Las Aguas, 28°23.64' N 16°38.17' W, 15 - 30m above sea level. 15th March 2003 [FHO (female)]. *Borgeau* 320 Canary Islands: Lanzerote. 1845 [OXF (male), BM (male, female). *J Molero s.n.* Canary Islands: La Palma, Sta Cruz. 7th June 1996 [BCN (male, female)]. *Jarvis & Murphy* 62 Canary Islands: Tenerife, San Andres. 4th April 1977 [RDN (male), BM (male)]. *K Lems* 2219 Canary Islands: Tenerife, Taganana 17th August 1954 [RDN (male)]. *FA González s.n.* Canary Islands: Tenerife, Icod de los Vinos. 15th May 1988 [TFC (male)]. *OR Delgado s.n.* Canary Islands: Tenerife, El Escobonal. 22nd January 1977 [TFC (male)]. *PL Perez s.n.* Canary Islands: Tenerife, Bailedero (Anarga). 14th January 1972 [TFC (male)]. *ER Sventenius s.n.* Canary Islands: Tenerife, Valle de la Orotava. 21st November 1943 [TOR (female)]. *ER Sventenius s.n.* Canary Islands: Tenerife, La Orotava. 12th March 1949 [TOR (female)].

Annual, dioecious, glabrous occasionally with sparse adpressed hairs on abaxial leaf surface (especially veins) and along leaf margins, (15-)20-50(-100) cm tall, erect, branching, herbaceous with erect-patent branches. Leaves (15-)20-70(-150) mm x (15-)20-35(-80) mm, planar, petiolate, lanceolate-elliptic, shallowly cordate base, acuminate-acute tip, shallowly crenate-serrate margin. Petiole (5-)15-50(-70) mm, half to one times length of leaf blade; persistent stipules linear or triangular, reflexed, greenish, 4 – 7 mm long. Male flowers in clusters subtended by a persistent, brownish (greenish on fresh material) bract, 1(-4) mm, 3 – 7 flowers per cluster and 3 – 5 mm diameter, in an axillary, spicate inflorescence (often branched) longer than subtending leaf (15 – 30 mm long, excluding peduncle), peduncle 10 – 80 mm long, 4 – 8 separate floral peduncle longer than gap between the clusters; female flowers axillary, solitary or in verticles of 2-3, shortly pedunculate. Sepals ovate, glabrous, whitish. Ovary in general with (2-)3(-4) carpels. Fruit 2 – 3 mm x 1 – 2 mm, hirsute along valve margins. Seeds c. 2 mm x 1.5 mm, ovoid, smooth-rugulose, grey to brownish-grey. 2n = 32. Flowering and fruiting: November to August. Figure 8.14.

Mercurialis canariensis is known only from the Canary Islands, the islands of Tenerife, Lanzerota and La Palma (Figure 8.15). On Tenerife, the distribution of *M. canariensis* is limited to the moister northern half of the island. However, the distribution of *M. canariensis* on the islands of the Canarian Archipelago is unclear due to undercollection and confusion with the superficially similar *M. annua. Mercurialis canariensis* has been collected from cultivated land, open scrub and rough grassland (*Jarvis & Murphy* 62) and is widespread but scattered in the climax forest of the cloud belt (*Lems* 2219). However, *M. canariensis* is most common in horticultural and ruderal environments all along the north coast of Tenerife. *Mercurialis canariensis* has been recorded from sea-level to 1000 m above sea level in Tenerife. However, as altitude increases *M. canariensis* becomes rarer. The earliest known collection (*Borgeau* 320; OXF, BM; collected in 1845) of *M. canariensis* is from Lanzerote, although the exact location is not given on the specimen.

The specific epithet reflects the known occurrence of the species on the Canary Islands.

Mercurialis canariensis can be readily separated from other members of the genus *Mercurialis* using the key below.

1. Rhizomatous; aerial stem simple	**2**
1. Not rhizomatous; stems branched	**4**
2. Ovary hairy. Fruit pubescent. European	**3**
2. Ovary glabrous. Fruit glabrous or sparsely setose, muricate. Central and East Asian	***leiocarpa***
3. Lower leaves usually scale-like; upper leaves ovate-elliptic to elliptic-lanceolate; petiole (3-)5-10(-18)mm	***perennis***
3. Lower leaves like the upper, but smaller; upper leaves suborbicular to ovate; petiole 1-2 mm	***ovata***
4. Plant densely tomentose	***tomentosa***
4. Plant glabrous or sparsely hairy	**5**
5. Annual, without thick woody stock	**6**
5. Perennial, with thick woody stock	**8**
6. Stipules < 4 mm long; male flowers without subtending bract	**7**
6. Stipules ≥ 4 mm long; male flowers with subtending bract	***canariensis***
7. Plant < 20 cm tall, fruit < 2 mm long	***huetii***
7. Plant ≥ 25 cm tall, fruit 2-4 mm long	***annua***
8. Sparsely hairy; leaves incise-dentate	***reverchonii***
8. Glabrous; leaves crenate-dentate or shallowly sinuate-dentate	**6**
9. Leaves crenate-dentate; fruit 3-4 x 5-6 mm; seed smooth	***elliptica***
9. Leaves shallowly sinuate-dentate; fruit c. 2 x 3 mm; seed rugulose	***corsica***

In addition to *M. canariensis*, monoecious *M. annua* also grows in the Canary Islands, although it is rare in north of Tenerife (Obbard, pers. obs.). Some seeds grown from bulked seed lots of *M. canariensis* produced plants with disturbed floral morphology, which may be F_1 hybrids between *M canariensis* and monoecious *M. annua* (Obbard, pers. obs.).

The Canarian *Mercurialis* is an allotetraploid with a 4C genome size of 6.43 pg to 6.46 pg and has distinct distributions of alleles at five allozyme loci (*Aat*-1, *Pgi*-2, 6*Pgd*-1, 6*Pgd*-2, *Idh*-1) compared to other annual *Mercurialis* species (Obbard, 2004). Furthermore, analysis of both nrDNA internal transcribed spacer sequences and cpDNA-encoded *trn*L-*trn*F and *mat*K-*trn*K sequences shows that *M. canariensis* is monophyletic with respect to other annual *Mercurialis* (Obbard, 2004). Obbard (2004) has proposed that *M. canariensis* is a tetraploid derivative of dioecious diploid or monoecious tetraploid mother and an unidentified diploid *Mercurialis* species father.

Most of the species in the genus *Mercurialis* are dioecious, the exceptions are polyploid *M. annua* and possibly some *M. leiocarpa* (Table 8.4). Furthermore, within the annual species, dioecy and diploidy always occur together except in the case of *M. canariensis*. The uncoupling of dioecy and diploidy in *M. canariensis* means that *M. canariensis* will be a useful model for future studies of sexual-system/ ploidy interaction.

8.4.6 Acknowledgements

We thank the curators of OXF, BM, RDN, BC, LY, MACB, BCN, TFMC, ORT and TFC for the loan of specimens in their care, and Rosemary Wise for the two excellent plates that accompany this article. We particularly thank Cecilia Duraes (University of Edinburgh) for bringing the existence of *M. canariensis* to our attention, and providing the first seed sample. We also thank K. & S. Obbard, whose holiday was spent making seed collections. Fieldwork in Tenerife was partly funded by The Queens College (Oxford).

8.4.7 References

Camerarius, R. J. (1684) *Rudolphi Jacobi Camerarii ad Thessalum de sexu plantarum epistola*. Tubingæ.

Durand, B. 1963. Le complexe *Mercurialis annua* L. *s.l.*: une étude biosystématique. Annales des Sciences Naturelles, Botanique, Paris 12:579-736.

Durand, R, and B Durand. 1984. Sexual differentiation in higher plants. Physiologia Plantarum 60:267-274.

Durand, R, and B Durand. 1985. *Mercurialis*. Pp. 376-387 *in* A. H. Halevey, ed. CRC Handbook of Flowering. CRC Press, Boca Raton, Florida.

Durand, B, and R Durand. 1991. Sex Determination and Reproductive Organ Differentiation in *Mercurialis*. Plant Science 80:49-65.

Durand, R, and B Durand. 1992. Dioecy, monoecy, polyploidy and speciation in annual Mercuries. Bulletin de la Societé Botanique de France, Lettres Botaniques 139:377-399.

Govaerts, R. Frodin, D., G. And Radcliffe-Smith, A. (2000) *World Checklist and bibliography if Euphorbiaceae. Vol. 3*. The Royal Botanic Gardens, Kew.

Güemas, J. (1997) *Mercurialis* L. in Castroviejo, S., Aedo, C., Benedí, C., Laínz, M., Muñoz Garmendia, F., Nieto Feliner, G. and Paiva, J. (eds) *Flora Iberica. Vol. III. Haloragaceae – Euphorbiacea*e, pp. 201-210. Real Jardín Botánico, CSIC, Madrid.

Hamdi, S, G Teller, and JP Louis. 1987. Master regulatory genes: auxin levels and sexual organogeneses in the dioecious plant *Mercurialis annua*. Plant Physiology 85:393-399.

Louis, JP. 1989. Genes for the Regulation of Sex Differentiation and Male Fertility in Mercurialis-Annua L. Journal of Heredity 80:104-111.

Louis, J-P, and B Durand. 1978. Studies with the dioecious angiosperm Mercurialis annua L. (2n=16). Correlation between genic and cytoplasmic male sterility, sex segregation and feminising hormones (cytokinins). Molecular and General Genetics 165:309-322.

Obbard, 2004. D.Phil. Thesis,. University of Oxford. "Genetic variation and sexual-system evolution in the annual mercuries" (Chapter 2 and Appendices)

Pannell, JR. 2001. A hypothesis for the evolution of androdioecy: The joint influence of reproductive assurance and local mate competition in a metapopulation. Evolutionary Ecology 14:195-211.

Pannell, JR. 2002. The evolution and maintenance of androdioecy. Annual Review of Ecology and Systematics 33:397-425.
Pax, F. (1914) *Das Pflanenreich. IV.147.VII. Euphorbiaceae-Acalypheae-Mercurialinae*. Verlag von Wilhelm Engelmann, Leipzig & Berlin.
Tutin, T. G. (1968) *Mercurialis* L. in Tutin, T. G., Heywood, V. H., Burges, N. A., Moore, D. M., Valentine, D. H., Walters, S. M. & Webb, D. A. (eds.) *Flora Europaea. Vol. 2. Rosaceae to Umbellifereae*, p. 212. Cambridge University Press, Cambridge.
Thomas, RG. 1958. Sexuality in diploid and hexaloid races of *Mercurialis annua* L. Annals of Botany 22:55-72.
Gabe, DR. 1939. Inheritance of sex in *Mercurialis annua*. Comptes Rendus, Academie des Sciences URSS:478-481.
Heslop-Harrison, J, and Y Heslop-Harrison. 1957. The Effect of Carbon Monoxide on Sexuality in Mercurialis ambigua L. fils. New Phytologist 56:352-355.
Heyer, F. 1884. Untersuchungen über das Verhältnis des Geschlechts bei einhäusigen und zweihäusigen Pflanzen, unter Berücksichtigung des Geschlechtsverhältniss bei Thieren und Menschen. Ber. physiolog. Lab. u. Versuchsan. landw. Inst. Univ. Halle. 1:43.
Yampolsky, C. 1919. Inheritance of Sex in Mercurialis annua. American Journal of Botany 6:410-442.

8.5 PHOTOGRAPHS OF PLANT MORPHOLOGY

Figure 8.17: Photograph of morphological variation within diploid *M. annua*

Morphological variation within diploid *Mercurialis. annua.* Plants were grown from seed samples collected from (left to right) France (1564), Turkey (0079) and Italy (0232). Photographs are from plants used in the morphometric analysis presented in Chapter 2.

Figure 8.18: Photograph of hexaploid *M. annua* and its putative parents

Superficial morphological series between hexaploid *Mercurialis annua* (centre), and its putative parents, tetraploid *M. annua* (left) and *M. huetii* (right) Plants were grown from seed samples collected from (left to right) Morocco (1020), and Spain (0648) and (0678). Plants were used in the morphometric analysis presented in Chapter 2.

Figure 8.19: Photograph of *M. annua* and *M. canariensis*

Gross morphological differentiation between diploid *M. annua* (left, population 1564) and the allotetraploid *M. canariensis*. (right, population 0206). Both plants were used in the morphometric analysis presented in Chapter 2.See Appendix 8.4 for further information regarding *M. canariensis*.

Figure 8.20: Photographs of the diagnostic characters of *M. canariensis*

Diagnostic morphological characters to distinguish between *Mercurialis. annua* and *M.* canariensis. Top panel, male flower bracts are absent in *M. annua* (left) but present in *M. canariensis* (right). Note that those depicted here are larger than usual, with the bracts toward the top of the inflorescence being more representative. Bottom panel, stipules are usually 1.5-2.5 mm and upright in *M. annua*, but 4-8 mm and recurved in *M. canariensis*. Photographs are from plants used in the morphometric analysis presented in Chapter 2. See also Appendix 8.4 for details on *M. canariensis*.

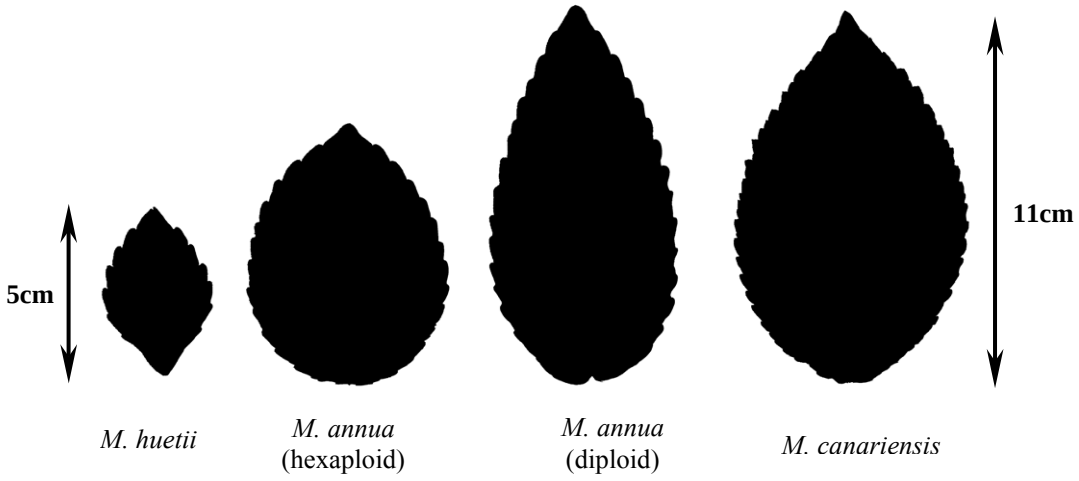
8.6 LEAF SHAPE VARIATION IN THE ANNUAL MERCURIES

In his taxonomic survey of the *Mercurialis annua* polyploid complex, and in describing the progeny of artificial crosses, Durand (1963, pages 637-659) made extensive use of leaf outlines to distinguish between *M. huetii* and polyploid races of *M. annua* (for illustrative examples from our glasshouse dataset see Figure 8.21). However, his numerical analysis was limited to petiole length, leaf length, leaf width, leaf dentition, and the angle formed by the base of the leaf, and there was no statistical analysis of this data.

Since Durand performed this analysis in *ca.* 1960, more sophisticated techniques have become available for the analysis of shape. In particular, the elliptic Fourier analysis of outlines can be used to capture subtle differences in shape, and submit them to statistical analysis (Kuhl and Giardina 1982; Rohlf and Archie 1984; Iwata *et al.* 2002; but see particularly Iwata and Ukai 2002).

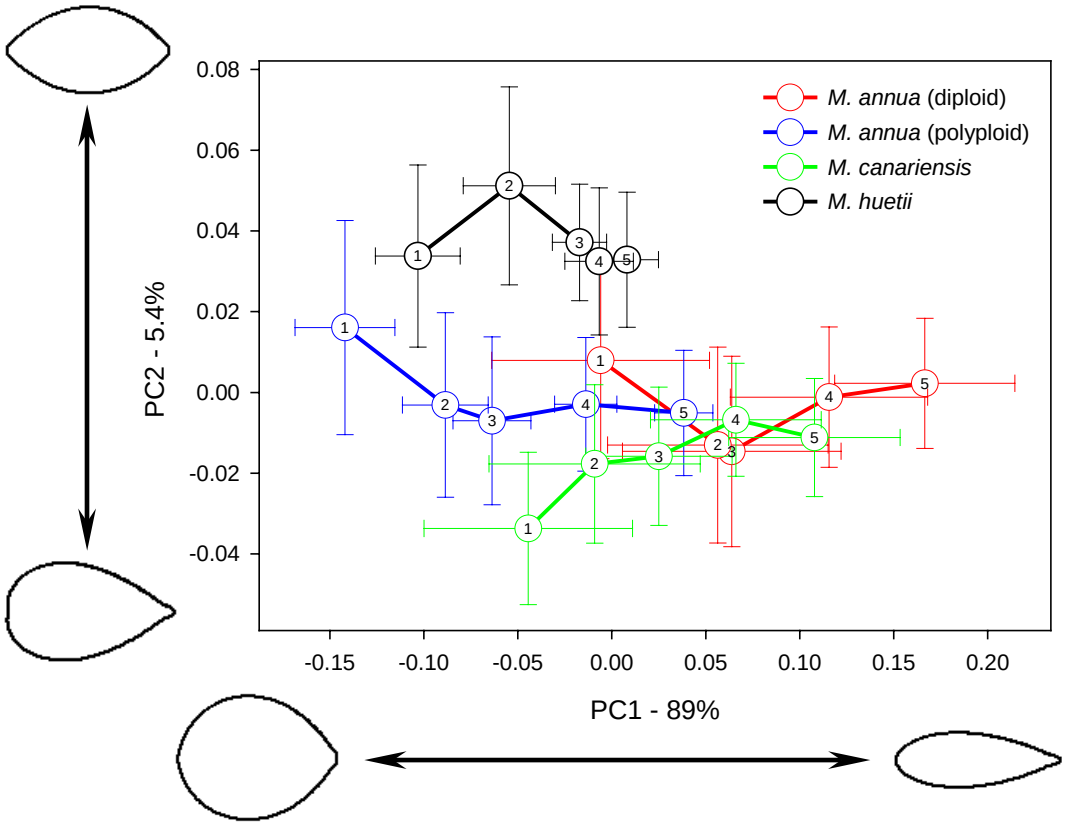
Such an analysis of leaf shape in *M. annua* reveals that differentiation within the *M. annua* polyploid complex is slight, although statistically significant differences exist between polyploid races (analysis not shown). There is more variation within plants than between polyploid levels for some shape-components, making it essential that leaves are compared from corresponding developmental stages (Figure 8.22). With the possible exception of the lower leaf length-width ratio in polyploid *M. annua* versus diploid *M. annua*, (i.e. more 'round' leaves, Principal Component 1, see Figure 8.22) the immense phenotypic plasticity seen in the field makes it doubtful whether leaf shape is of much use in distinguishing *M. annua* races.

Figure 8.21: Representative leaf outlines of the annual mercuries



These are taken from plants grown as part of the morphometric analysis in Chapter 2 (details of method on page 27); see the main text for details of materials and methods. These leaves have been selected (by eye) to be typical of each group as grown in the glasshouse. However, having been grown in lower light conditions than in their natural environment, all are much larger than is usually seen in the field. These leaves were taken from the 4th node on the main stem (cotyledons denoted as node zero) Populations (left to right) are Spain 0678, 0636, Crete 0074 and Tenerife 0206.

Figure 8.22: The first two principle components of leaf shape, by taxonomic group and by leaf postion



The first two principal components of leaf shape (accounting for 94% of the variance), as analysed and calculated using 'SHAPE' (Iwata and Ukai 2002). All data is taken from the morphometric dataset presented in Chapter 2. Data points are averages for each leaf node (numbered 1 to 5, 0 being the cotyledons), with error bars +/- 1 standard deviation. The components of leaf shape variation are illustrated with outlines that depict +/- 2 standard deviations around the grand mean. It is clear that the greatest variation is due to a developmental series within plants; leaves broaden as they develop.

8.7 THE ORIGIN OF SAMPLES USED IN THE ISOZYME STUDY

Table 8.6: Origin of the seed samples used in the isozyme analyis

'ID' refers to the population identifier used throughout this thesis; longitude and latitude are in decimal degrees, negative longitude values being west of the Greenwich meridian; *n* is the number of progeny grown from the bulk seed collection that were used in the isozyme analysis.

Ploidy	ID	Locality	Country	Longitude	Latitude	*n*
2	0002a	5km south of Tarragona on N340	Spain	1.20962	41.11530	77
2	0003a	3km north of Cambrils on n340, Vilafortay turnoff	Spain	1.07540	41.08300	80
2	0015a	L'Almadrava	Spain	0.85422	40.94028	80
2	0018a	N340 outside Reus, sign "Riera de Mas Pujols"	Spain	1.04740	41.16740	73
2	0019a	Entrance to Pradell on A420	Spain	0.87758	41.15747	80
2	0028a	N340, turning to Roda de Bara	Spain	1.45912	41.16793	80
2	0039a	C246 west of Rodonya	Spain	1.39160	41.28758	80
2	0042a	C246 east of Sitges, above rail tunnel	Spain	1.84010	41.23568	74
2	0048a	South of Sant Vincenc	Spain	2.01348	41.39168	80
2	0049a	West of Cervello on N340	Spain	1.94878	41.39320	60
2	0059a	Fivon, near Haifa (Israel)	Israel	34.98333	32.80000	55
2	0061a	Montestigliano Rosia (Siena, Italy)	Italy	11.40000	43.21667	44
2	0062a	Norwich	UK	1.29222	52.64222	40
2	0063a	Slapanice (Czech Republic)	Czech Republic	14.10900	50.30800	41
2	0064a	Suchdol (Prague, Czech Republic)	Czech Republic	14.36900	50.12200	41
2	0065a	Soigewerk Getzerdorf (Austria)	Austria	15.71667	48.31667	40
2	0067a	Sieci (20km east of Florence, Italy)	Italy	11.39700	43.79000	39
2	0068a	Cephalonia (Greece)	Greece	20.58333	38.25000	27
2	0070a	Corbais (Belgium)	Belgium	4.66528	50.65444	40
2	0072a	Malia (Crete)	Crete	25.46222	35.28417	39
2	0074a	Festos (Crete)	Crete	24.80500	35.03389	53
2	0075a	East Oxford	UK	-1.23750	51.74639	30
2	0077a	Ely	UK	0.25139	52.39889	40
2	0078a	Plymouth	UK	-4.11611	50.37639	40
2	0079a	Aspendos (Near Serik, Turkey)	Turkey	31.10000	36.91667	45
2	0080a	Izmit (Turkey)	Turkey	29.91667	40.78333	40
2	0081a	Yalova (Turkey)	Turkey	29.28333	40.66667	40

2	0225a	Near Castellenare	Italy	14.49736	40.69331	40
2	0228a	Pastrun	Italy	15.01204	40.44338	40
2	0231a	Near Pontecorvo	Italy	13.65782	41.50782	40
2	0232a	Sestri Levante	Italy	9.42143	44.27127	40
2	0240a	Nonza (Corsica)	France	9.34528	42.78472	40
2	0581a	Sopelana	Spain	-2.99630	43.37175	50
2	0586a	Muros de Nalon	Spain	-6.10919	43.54185	40
2	0591a	Espasante	Spain	-7.80015	43.71207	40
2	0593a	La Coruna	Spain	-8.40293	43.35226	40
2	0596a	Cances	Spain	-8.71939	43.23508	50
2	0599a	Borneiro	Spain	-8.94477	43.21199	40
2	0601a	Vilaseco	Spain	-9.06108	43.07217	20
2	0607a	Leis	Spain	-9.15203	43.11745	40
2	0678b	L'aldea	Spain	0.64945	40.77016	40
2	0695a	Sant Feliu de Guixols	Spain	3.01816	41.79326	40
2	0719a	Portbou	Spain	3.16716	42.43448	44
2	1562a	Mont de Marsan	France	-0.49516	43.88421	40
2	1564a	St Ligier	France	1.15148	42.98389	40
2	1569a	e of Dijon	France	5.46351	47.64496	40
6	0004f	Camping Nautique in Ametlla de Mar	Spain	0.80633	40.88863	28
6	0008a	Nr. T331 south East of Ulldecona	Spain	0.44272	40.59662	60
6	0009a	South West of Godall, olive grove	Spain	0.46693	40.65167	30
6	0010a	North East of La Galera	Spain	0.46892	40.68477	10
6	0011a	North East of Sta. Barbara	Spain	0.50248	40.71933	80
6	0012a	SW of L'Aldea, on N340	Spain	0.59775	40.73457	40
6	0020a	Xerta on C230	Spain	0.48903	40.87507	79
6	0021a	N340, after junction43 of A7	Spain	0.38495	40.40713	49
6	0022a	SE of bridge on N340 south of Vinnaros	Spain	0.45775	40.45717	69
6	0055a	Hinjos del Duque	Spain	-5.13333	38.50000	40
6	0057a	Generalisimo Bridge (over Guadalquirir? Seville)	Spain	-5.98333	37.36667	40
6	0058a	Pabellon de Cuba (Seville)	Spain	-5.98333	37.36667	39
6	0060a	El Hajeb (Morocco)	Morocco	-5.36667	33.70000	52
6	0076a	Anzalcollar School (Seville)	Spain	-6.26667	37.51667	22
6	0083a	Anzalcollar	Spain	-6.26667	37.51667	20
6	0084a	5 km S of Fes (Morocco)	Morocco	-4.94200	34.00600	70
6	0085a	Meknes (Morocco)	Morocco	-5.55000	33.90000	68
6	0086a	Soller (Mallorca)	Mallorca	2.81667	39.76667	36
6	0087a	Palma (Mallorca)	Mallorca	2.67500	39.58700	40
6	0088a	Ain Chigag (Morocco)	Morocco	-5.03300	33.88300	40
6	0093a	Los Tilos (Gran Canaria)	Canary Islands	-15.60144	28.07683	30
6	0101a	La Gomera A	Canary Islands	-17.18350	28.18313	40
6	0506a	Camarles	Spain	0.68018	40.78002	40
6	0530a	L'Ampolla	Spain	0.70680	40.81393	10
6	0598a	Cee	Spain	-9.19217	42.96318	40
6	0605a	Pasarela	Spain	-9.08387	43.14527	40

6	0608a	Molinos	Spain	-9.17756	43.09629	40
6	0609a	Ezaro	Spain	-9.14459	42.91723	40
6	0612a	Abelleira	Spain	-9.01221	42.80323	40
6	0614a	Enfesta	Spain	-8.65000	42.71830	40
6	0616a	Tui	Spain	-8.65275	42.05480	40
6	0620a	Vairao	Portugal	-8.67787	41.31515	40
6	0625a	Porto Alto	Spain	-8.89337	38.92688	40
6	0630a	A340 junc. For Benagalbon	Spain	-4.25359	36.72596	30
6	0636a	Torre del Penon	Spain	-1.85030	37.08492	39
6	0643a	San Antonio	Spain	-0.23493	38.51916	30
6	0648a	Silla	Spain	-0.42259	39.36520	40
6	0650a	Junc 49 of A7	Spain	-0.15964	39.81399	40
6	0655a	Alcala de Xivert	Spain	0.23041	40.27919	40
6	0658a	Vinaros (sud)	Spain	0.46339	40.46188	40
6	0660a	Les Cases d'Alcanar	Spain	0.52869	40.55652	40
6	0682a	L'Aldea	Spain	0.62023	40.74696	30
6	1036a	Nrth. Skirat	Morocco	-6.99530	33.87450	45
6	1039a	nr. Bouknadel	Morocco	-6.70500	34.16730	40
6	1044a	El Ksar el Kbir	Morocco	-5.91480	35.01630	50

8.8 THE ORIGIN OF SAMPLES USED IN THE PHYLOGENETIC ANALYSIS

Table 8.7: Origin of the samples used in the phylogenetic analysis

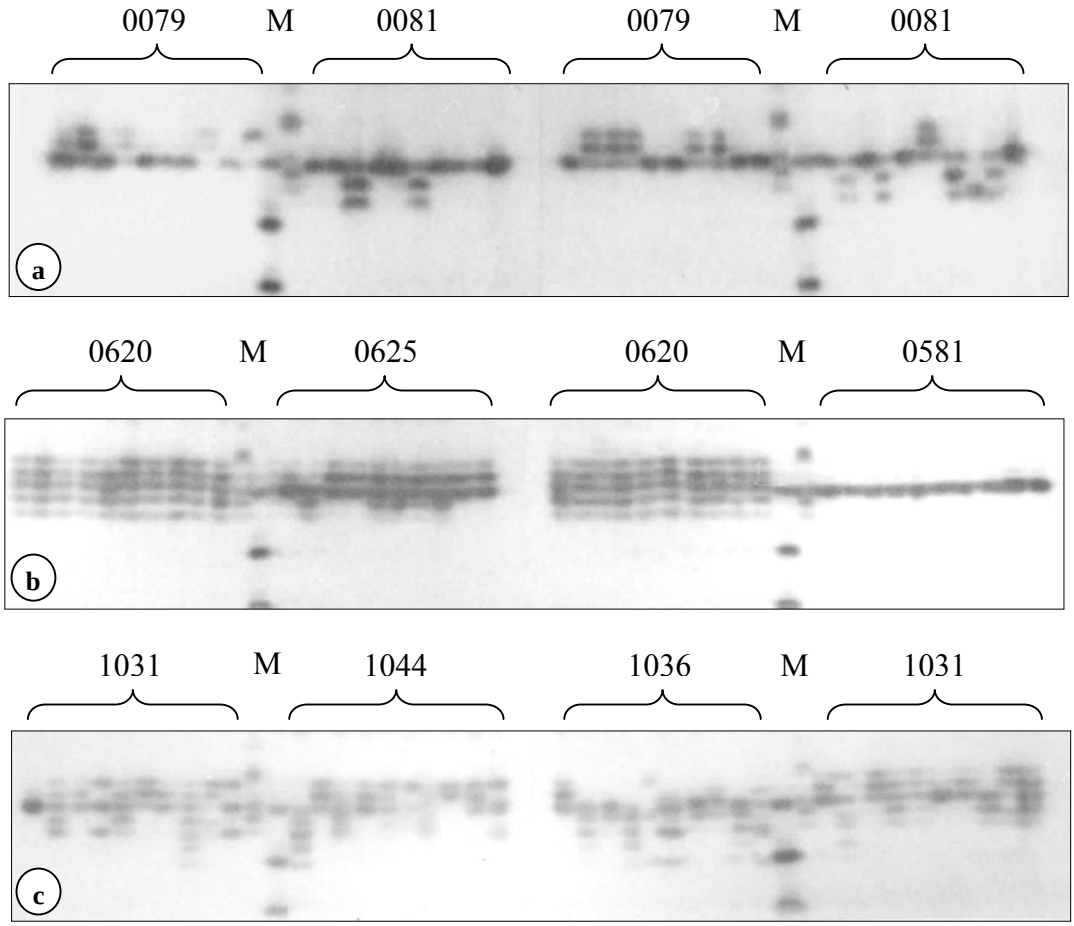
For the seed samples 'ID' refers to the population identifier used throughout this thesis, for the leaf and herbarium samples the numeric simply distinguishes between different samples of the same species.

ID	Species	Ploidy	Sample type	Source location
0002	*M. annua*	2x	Seed	(Spain) 01:12.577e, 41:06.918n
0059	*M. annua*	2x	Seed	(Israel) 0034:59e, 32:48n
0061	*M. annua*	2x	Seed	(Italy) 011:24e, 43:13n
0080	*M. annua*	2x	Seed	(Turkey) 029:55e, 40:47n
0062	*M. annua*	2x	Seed	(UK) 001:17:32e, 52:38:32n
1564	*M. annua*	2x	Seed	(France) 1.15148e, 42.98389n
0596	*M. annua*	2x	Seed	(Spain) 008.71939w, 43.23508n
0072	*M. annua*	2x	Seed	(Crete) 25:27:44e, 35:17:03n
0079	*M. annua*	2x	Seed	(Turkey) 31:06:0e, 36:55:0n
0228	*M. annua*	2x	Seed	(Italy) 15.01204e, 40.44338n
0232	*M. annua*	2x	Seed	(Italy) 009.42143e, 44.27127n
1018	*M. annua*	4x	Seed	(Morocco) 8.9051w, 32.8439n
1020	*M. annua*	4x	Seed	(Morocco) 8.6131w, 33.1324n
1031	*M. annua*	4x	Seed	(Morocco) 7.6025w, 33.6095n
0011	*M. annua*	6x	Seed	(Spain) 000:30.149e, 40:43.160n
0012	*M. annua*	6x	Seed	(Spain) 000:35.865e, 40:44.074n
0020	*M. annua*	6x	Seed	(Spain) 000:29.342e, 40:52.504n
0058	*M. annua*	6x	Seed	(Spain) 005:59w, 37:22n
0060	*M. annua*	6x	Seed	(Morocco) 005:22w, 33:42n
0085	*M. annua*	6x	Seed	(Morocco) 005:33w, 33:54n
0620	*M. annua*	6x	Seed	(Portugal) 008.67787w, 41.31515n
0631	*M. annua*	6x	Seed	(Spain) 004.09813w, 36.73395n
0636	*M. annua*	6x	Seed	(Spain) 001.85030w, 37.08492n
0648	*M. annua*	6x	Seed	(Spain) 000.42259w, 39.36520n
0660	*M. annua*	6x	Seed	(Spain) 000.52869e, 40.55652n
1044	*M. annua*	6x	Seed	(Morocco) 5.9148w, 35.0163n
1036	*M. annua*	6x	Seed	(Morocco) 6.9953w, 33.8745n

0091	Tenerife mercury	4x	Seed	(Tenerife) 16:12:07w, 28:32:57n
0200	Tenerife mercury	4x	Seed	(Tenerife) 16:38.17w, 28:23.64n
0209	Tenerife mercury	4x	Seed	(Tenerife) 16:49.01w, 28:16.44n
0213	Tenerife mercury	4x	Seed	(Tenerife) 16:12.70w, 28:32.85n
0678	*M. huetii*	2x	Seed	(Spain) 0.64945e, 40.77016n
0719	*M. huetii*	2x	Seed	(Spain) 3.16716e, 42.43448n
01	*M. elliptica*		Silica-dried Leaf	(Spain, Huelva prov.) Between Hinojos and Aliconte Collected by S. Talavera 4/10/2003
02	*M. elliptica*		Silica-dried Leaf	(Spain, Huelva prov.) Between Hinojos and Aliconte Collected by S. Talavera 4/10/2003
01	*M. tomentosa*		Silica-dried Leaf	Provided by Y-M Yuan, Collection M87-057-01 used in (Krahenbuhl *et al.* 2002)
02	*M. tomentosa*		Silica-dried Leaf	(Spain, Seville) Between Dos Hernanas and Alcala de Guadaiva Collected by S. Talavera, 1/12/2003
01	*M. reverchonii*		Herbarium	SEV 858683 M.A. Mareos 6985/95 Morocco 35:06n, 005:09w collected 23/07/1995
02	*M. reverchonii*		Herbarium	SEV S. Talavera 482/03M Morocco 35: 6:13.9n, 005:08w collected 14/06/2003
01	*M. perennis*		Silica-dried Leaf	East Sussex (UK) collected by DJ Obbard and EH Bayne, 2002. 51.093n, 0.165e
02	*M. perennis*		Silica-dried Leaf	Howford bank, Newcastle, UK collected by DJ Obbard Spring 2002.
	Ricinus communis		Cultivated material	University of Oxford Botanic Garden

8.9 EXAMPLE PGI ISOZYME GELS

Figure 8.23: Example high and low diversity *Pgi*-2 isozyme gels from *M. annua*



These isozyme gels illustrate the levels of genetic variation present in different *M. annua* populations. Figure (a) shows polymorphic diploid populations from Turkey; Figure (b) shows almost fixed heterozygosity in Catalonian hexaploid populations, and complete monomorphism in a Catalonian diploid population (0581); Figure (c) shows the high level of variation in Morrocan tetraploid and hexaploid populations. See Chapter 5, page 97 for details of materials and methods, and Appendix 8.7 for population details. 'M' indicates the same pair of marker individuals twice on each of the three gels.

8.10 ISOZYME ALLELE FREQUENCIES IN *MERCURIALIS ANNUA*

8.10.1 Hexaploid *M. annua* allele-carrier frequencies

Allele-carrier frequencies for hexaploid populations used in Chapter 5. Population identifiers are those used throughout this thesis, and population details can be found in Appendix 8.1. In hexaploids, allele frequencies cannot be calculated from allelic phenotypes, because the allele copy-number is hidden (see Chapter 4). Below are the Allele-carrier frequencies (i.e. the proportion of individuals carrying each allele), calculated using FDASH (Chapter 4).

Table 8.8: Allele-carrier frequencies for *Aat*-1 and *Pgi*-1

Locus	Aat-1						Pgi-1								
Allele	**100**	**115**	**56**	**68**	**75**	**140**	**77**	**100**	**122**	**50**	**89**	**136**	**151**	**112**	**10**
Average	0.94	0.99	0.15	0.02	0.02	0	0.53	0.95	0.69	0.1	0.02	0.05	0	0	0
0004f	1	1					1	1	1						
0008a	1	1					0.05	1	1						
0009a	1	1						1							
0010a	0.9	1					0.1	1	1						
0011a	0.31	1					0.04	1	1						
0012a	1	1	1					1	1						
0020a	1	1	0.99				0.01	1	0.99						
0021a	1	1					0.37	1	0.76						
0022a	1	1					0.81	1	0.35						
0055a	1	1						0.03	1	1					
0086a	1	1	1					1	1	0.64					
0087a	1	1					1	1			1				
0093a	0.63	1					0.97	0.8	0.4			0.03			
0101a	1	1					0.75	1	0.55						
0506a	1	1						1	1						
0530a	1	1					0.9	1	0.6						
0598a	1	0.96		0.42			0.98	1	0.8						
0605a	1	1					1	0.75	1						
0609a	1	1					1	1	0.88						
0612a	1	1					0.93	0.85	0.8	0.13					
0614a	1	0.74		0.72			0.97	0.97	1						
0620a	1	1					1	1	0.98						
0636a	1	1						1		0.13					
0643a	1	0.9	1				0.77	1	1						

0650a	1	0.98					0.03	0.9	1						
0655a	1	1	1					1	1						
0658a		1	1				1	1	0.03						
0660a	1	1					1	1							
0682a	1	1						1	1						
0608a	1	1					0.93	0.75	0.05						
0616a	1	1					0.92	0.77	0.54		0.87				
0625a	1	0.95					0.65	1	0.83						
0630a	1	1	0.2				0.93	1	0.33	0.13	0.03				
0648a	1	1		0.05			0.83	0.98	0.68	0.03					
1036a	1	0.98		0.02			0.58	0.96	0.36	0.64		0.02	0.09		
0057a	1	1					0.68	0.97	0.82	0.11	0.11				
0058a	1	1					0.54	1	0.84	0.03	0.03				
0060a	0.94	1					0.5	1	0.48	0.14	0.4	0.04		0.04	
0076a	0.95	0.91					0.92	1	0.75	0.33					
0083a	1	1	0.1				0.9	1	0.4	0.2					
0084a	1	0.98		0.08	0.02		0.36	1	0.84	0.2	0.04				
0085a	0.99	1		0.07			0.68	1	0.51	0.14	0.54				
0088a	1	1					0.73	1	0.6	0.38	0.05	0.03		0.13	
1039a	1	0.98		0.25			0.35	0.95	0.68	0.5					
1044a	1	0.98		0.46			0.76	0.9	0.9	0.07	0.07				

Table 8.9: Allele-carrier frequencies for 6*Pgd*-1, 6*Pgd*-2, *Idh*-1

	6Pgd-1					**6Pgd-2**					**Idh-1**				
Allele	**93**	**100**	**73**	**80**	**109**	**100**	**61**	**78**	**114**	**93**	**85**	**100**	**64**	**115**	**128**
Average	0.99	0.99	0.01	0.02	0.04	1	0.38	0.33	0.05	0	0.97	1	0.12	0.04	0
0004f	1	1				1					1	1			
0008a	1	1				1	1	1			1	1			
0009a	1	1				1	1					1	1		
0010a	1	1				1	0.9	0.9			0.4	1	1		
0011a	1	0.99				1	0.27	0.83			0.78	1	0.26		
0012a	1	1				1	1	1			1	1	1		
0020a	1	1				1					1	1			
0021a	1	1				1		0.94			1	1			
0022a	1	1				0.98	0.95	1			1	1			
0055a	1	1				1			0.2		1	1			
0086a	1	1				1	1	1			1	1			
0087a	1	1				1					1	1			
0093a	1	1				1					1	1			
0101a	1	1				1					1	1			
0506a	1	1				1	1	1			0.98	0.98	0.98		
0530a	1	1	0.1			1	1				1	1			
0598a	1	1				1	0.75				1	1			
0605a	1	1				1	1				1	1			

0609a	1	1				1	0.05				1	1			
0612a	1	1				1	0.08				1	1		0.25	
0614a	1	1				1	0.18				0.98	1	0.03		
0620a	1	1				1		0.85	0.9		1	1	0.33		
0636a	1	1				1	1				1	1			
0643a	1	1				1	0.9	0.05			1	1			
0650a	1	1				1	1	1			1	1			
0655a	1	1				1					1	1			
0658a	1	1				1		1			1	1			
0660a	1	1				1	1	0.64			0.95	1	0.05		
0682a	1	1				1	0.87	0.87			1	1	0.87		
0608a	1	1				1	0.05				1	1		0.5	
0616a	1	1				1	0.15				1	1			
0625a	1	1				1					1	1		0.1	
0630a	1	1	0.17			1	0.43	0.2	0.07		1	1			
0648a	1	1		0.03		1	0.13				1	1	0.2		
1036a	0.91	1	0.09	0.11	0.34	1	0.09	0.05	0.14	0.02	0.93	1	0.18		
0057a	1	0.98			0.03	1	0.13	0.43	0.2		1	0.98		0.08	
0058a	1	0.95		0.05	0.05	1	0.23	0.31	0.23		1	1		0.05	
0060a	0.98	0.98		0.02	0.11	1	0.48				1	1		0.02	
0076a	1	1				1	0.14	0.14	0.05		1	1			
0083a	1	1			0.2	1	0.15	0.05			1	1			
0084a	0.96	0.97	0.04	0.03	0.06	1	0.13	0.24	0.04		0.99	1	0.07	0.07	
0085a	1	0.98	0.02	0.2	0.15	1	0.19	0.03	0.06		1	1	0.15		0.01
0088a	1	1		0.07		1	0.38		0.03		1	1		0.23	
1039a	0.98	1	0.03	0.03	0.45	1			0.18		1	1		0.3	
1044a	0.96	0.96		0.38	0.23	1	0.63	0.5	0.17		1	1			

8.10.2 Diploid *M. annua* allele frequencies

Allele frequencies for diploid populations used in Chapter 5. Population identifiers are those used throughout this thesis, and population details can be found in Appendix 8.1.

Table 8.10: Allele frequencies for *Aat*-1, *Pgi*-2, *Pgm*-1 and *Pgm*-2

	Aat-1		**Locus: Pgi-2**					**Pgm-1**			**Pgm-2**		
Allele	**100**	**115**	**5**	**68**	**77**	**100**	**122**	**100**	**108**	**110**	**71**	**100**	**112**
0002a	1					1		0.99		0.01	0.03	0.53	0.45
0003a	1					1		1				0.72	0.28
0015a	1					1		0.78		0.23		0.33	0.67
0018a	1					1		0.98		0.02	0.01	0.69	0.31
0019a	1					1		1				1	
0028a	1					1		0.96		0.04	0.1	0.55	0.35

0039a	1					1		0.99	0.01	0.01	0.01	0.87	0.13
0042a	1					1		0.87		0.13		0.87	0.13
0048a	1					1		0.87		0.13	0.08	0.9	0.02
0049a	1					1		0.96		0.04	0.18	0.71	0.11
0059a	1			0.02		0.96	0.02	0.72	0.04	0.24	0.01	0.91	0.08
0061a	1			0.03		0.97		0.5		0.5		0.75	0.25
0062a	1					1		1				1	
0063a	1			0.15		0.85		0.9		0.1		0.74	0.26
0064a	1			0.04		0.96		0.87	0.04	0.09		0.8	0.21
0065a	1					1		0.86		0.14		0.85	0.15
0067a	1			0.08		0.92		0.9	0.05	0.05		0.96	0.04
0068a	1			0.06		0.94		0.85	0.06	0.09	0.02	0.93	0.06
0070a	1					1		0.65		0.35		1	
0072a	1					1		0.95	0.04	0.01		0.68	0.32
0074a	1		0.03	0.04		0.83	0.11	0.95		0.05		0.68	0.32
0075a	1					1		0.8	0.02	0.18		1	
0077a	1					1		0.63	0.05	0.33		0.39	0.61
0078a	1					1		0.32		0.68		0.92	0.08
0079a	1			0.01		0.79	0.2	0.49		0.51	0.29	0.7	0.01
0080a	1		0.01	0.09		0.9		0.48	0.5	0.03	0.05	0.78	0.16
0081a	1			0.16		0.81	0.03	0.73	0.1	0.17	0.04	0.9	0.06
0225a	1			0.03		0.98		0.8		0.2		1	
0228a	0.99	0.01		0.03		0.98		0.69		0.31		0.8	0.2
0231a	1			0.01		0.99		0.76		0.24		0.9	0.1
0232a	1					1		0.95	0.05			0.65	0.35
0240a	1					1		0.56		0.44		1	
0581a	1					1		0.65	0.19	0.16		1	
0586a	1					1		0.89	0.04	0.08		1	
0591a	1				0.23	0.78		0.83	0.17			0.96	0.04
0593a	1					1		0.83	0.17			0.77	0.23
0596a	1				0.01	0.99		0.87	0.09	0.04		0.85	0.15
0599a	1				0.01	0.99		0.93	0.06	0.01		1	
0601a	1					1		1				0.95	0.05
0607a	1				0.01	0.99		0.77	0.04	0.2		0.86	0.14
0695a	1					1		0.84	0.11	0.05		0.88	0.13
1562a	1					1		0.18		0.83		0.41	0.59
1564a	1					1		0.97		0.03		0.79	0.21
1569a	1					1		0.55		0.45		0.95	0.05
Average	1		0	0.02	0.01	0.97	0.01	0.8	0.04	0.16	0.02	0.82	0.17

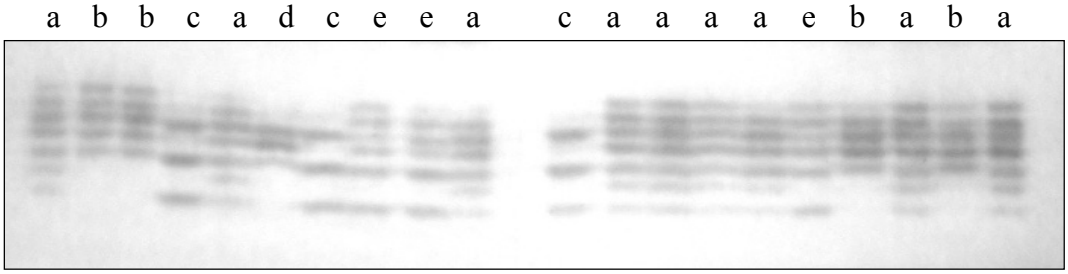
Table 8.11: Allele frequencies for 6*Pgd*-1, 6*Pgd*-2 and *Idh*-1

	6Pgd-1			**6Pgd-2**			**6Pgd-3**		
Allele	**93**	**100**	**109**	**78**	**100**	**112**	**73**	**85**	**100**
0002a		1			1				1
0003a		1			0.99	0.01			1
0015a		1			1				1
0018a		0.94	0.06		1				1

0019a		1		0.18	0.82				1
0028a		1			1				1
0039a		1		0.01	0.99				1
0042a		1			1				1
0048a		1			1				1
0049a		1			1				1
0059a	0.02	0.98		0.01	0.99		0.01		0.99
0061a	0.07	0.93			0.91	0.09			1
0062a		1			1				1
0063a	0.18	0.82			0.84	0.16			1
0064a	0.12	0.88			0.87	0.13			1
0065a		1			0.73	0.28			1
0067a		1			0.91	0.09			1
0068a	0.22	0.78			0.96	0.04		0.07	0.93
0070a		1			0.99	0.01			1
0072a	0.01	0.99			0.51	0.49			1
0074a		1		0.01	0.98	0.01			1
0075a		1			1				1
0077a		1			1				1
0078a		1			1				1
0079a		1			0.97	0.03		0.26	0.74
0080a	0.16	0.84			0.85	0.15		0.06	0.94
0081a	0.05	0.95		0.05	0.86	0.09			1
0225a		1			0.93	0.08			1
0228a	0.19	0.81		0.03	0.88	0.1			1
0231a	0.04	0.96			0.9	0.1		0.04	0.96
0232a		1			1				1
0240a		1			0.9	0.1			1
0581a		1			1				1
0586a		1			0.99	0.01			1
0591a		1			1				1
0593a		1			1				1
0596a		1			1				1
0599a		1			1				1
0601a		1			1				1
0607a		1			1				1
0695a		1			1				1
1562a		1			1				1
1564a		1			1				1
1569a		1			0.68	0.32			1
Average	0.02	0.97	0	0.01	0.94	0.05		0.01	0.99

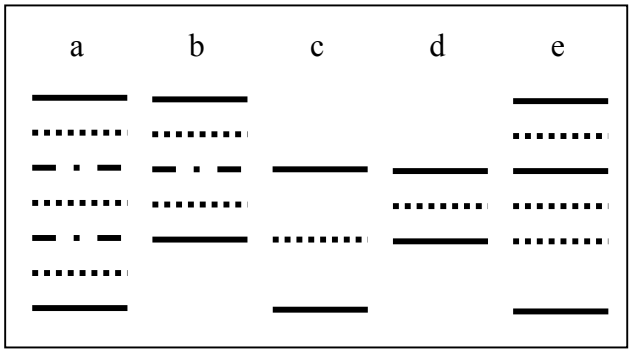
8.11 THE PROGENY FROM A SELF-FERTILISED HEXAPLOID

Figure 8.24: Self-fertilised progeny from a hexploid



Pgi-2 Isozyme gel of selfed progeny from a hexaploid individual, with five different offspring phenotypes visible (labelled a-e). For a genetic interpretation, see below. For materials and methods see Chapter 3 page 54.

Figure 8.25: Genetic interpretation



There are four alleles segregating, and including inter-allele heterodimers, up to seven bands are visible. Solid lines indicate homodimers, dotted lines indicate heterodimers, and dot-dash lines indicate superimposed homodimers and heterodmes, caused by the products co-migrating during electrophoresis. Phenotype ‘a’ has all four alleles, and thus four homodimers and six heterodimers (the central band represents two different, super-imposed heterodimers). Phenotypes ‘b’ and ‘e’ each have three alleles, and thus three homodimers and three heterodimes. Phenotypes ‘c’ and ‘d’ each have two alleles, and thus two homodimers and one heterodimers.

8.12 ALLELIC PHENOTYPE FREQUENCIES OF SELFED PROGENY

8.12.1 Individual I

Locus Pgi-2 (alleles abc)	Allelic phenotypes							Chi^2	df.	Posterior	Chi^2 Probability
Phenotypes:	c	b	bc	a	ac	ab	abc				
Observed:			***33.000***			***28.000***	***53.000***				
(ac bb bb)			28.500			28.500	57.000	1.000	3.000	0.999	0.801
(abbbbc)			4.560	23.940		23.940	61.560	9.868	4.000	0.001	0.043
(ab bb bc)			7.125	21.375		21.375	64.125	17.431	4.000	0.000	0.002
(ac bb bc)				28.500		7.125	78.375	70.086	3.000	0.000	0.000
(ab ac bb)				7.125		28.500	78.375	102.191	3.000	0.000	0.000
(ab bc bc)			1.781	26.719		1.781	5.344	78.375	109.312	5.000	0.000	0.000
(abbbcc)			0.285	28.215		0.285	4.275	80.940	142.693	5.000	0.000	0.000
(ab ab bc)			1.781	5.344		1.781	26.719	78.375	154.973	5.000	0.000	0.000
(aabbbc)			0.285	4.275		0.285	28.215	80.940	203.228	5.000	0.000	0.000
(ac ac bb)				7.125			7.125	99.750	177.038	3.000	0.000	0.000
(ab ac bc)				7.125		7.125	7.125	92.625	179.204	4.000	0.000	0.000
(ac bc bc)		1.781		26.719		5.344	1.781	78.375	402.739	5.000	0.000	0.000
(ab ab ac)				1.781	1.781	5.344	26.719	78.375	562.552	5.000	0.000	0.000
(aabbcc)				4.560		4.560	4.560	100.320	324.746	4.000	0.000	0.000
(abbccc)		0.285		28.215		4.275	0.285	80.940	2710.178	5.000	0.000	0.000
(ac ac bc)		1.781		5.344		26.719	1.781	78.375	565.770	5.000	0.000	0.000
(ab ac ac)				1.781	1.781	26.719	5.344	78.375	679.922	5.000	0.000	0.000
(aaabbc)				0.285	0.285	4.275	28.215	80.940	3769.544	5.000	0.000	0.000
(aabccc)		0.285		4.275		28.215	0.285	80.940	2926.319	5.000	0.000	0.000
(aaabcc)				0.285	0.285	28.215	4.275	80.940	3925.149	5.000	0.000	0.000

Locus 6Pgd-2 (alleles bd)	Allelic phenotypes			Chi^2	df.	Posterior	Chi^2 Probability
Phenotypes:	d	b	bd				
Observed:		***29.000***	***85.000***				
(bbbbbd)		28.500	85.500	0.012	2.000	0.500	0.994
(bb bb bd)		28.500	85.500	0.012	2.000	0.500	0.994
(bb bd bd)		7.125	106.875	71.637	2.000	0.000	0.000
(bbbbdd)		4.560	109.440	136.448	2.000	0.000	0.000
(bd bd bd)	1.781	1.781	110.438	423.562	3.000	0.000	0.000
(bbbddd)	0.285	0.285	113.430	2900.573	3.000	0.000	0.000

Locus Idh-1 (alleles bef)	Allelic phenotypes							Chi^2	df.	Posterior	Chi^2 Probability
Phenotypes:	f	e	ef	b	bf	be	bef				
Observed:						***49.000***	***50.000***				
(bb bf ee)						24.750	74.250	31.680	2.000	0.484	0.000
(bb ee ef)						24.750	74.250	31.680	2.000	0.484	0.000
(bbbeef)				0.247	0.247	3.712	24.503	70.290	34.557	5.000	0.019	0.000
(bb be ef)						6.188	24.750	68.062	34.741	3.000	0.006	0.000
(be bf ee)				6.188			24.750	68.062	34.741	3.000	0.006	0.000
(be be bf)				1.547	1.547	4.641	23.203	68.062	41.208	5.000	0.000	0.000
(be be ef)			1.547	4.641		1.547	23.203	68.062	41.208	5.000	0.000	0.000
(bb bb ef)						24.750	24.750	49.500	48.515	3.000	0.000	0.000
(bf ee ee)				24.750			24.750	49.500	48.515	3.000	0.000	0.000
(bbbbef)					3.960	20.790	20.790	53.460	63.252	4.000	0.000	0.000
(beeeef)			3.960	20.790			20.790	53.460	63.252	4.000	0.000	0.000
(bb be bf)					6.188	18.562	18.562	55.688	75.240	4.000	0.000	0.000
(be ee ef)			6.188	18.562			18.562	55.688	75.240	4.000	0.000	0.000
(bb ef ef)						6.188	6.188	86.625	317.900	3.000	0.000	0.000
(bf bf ee)				6.188			6.188	86.625	317.900	3.000	0.000	0.000
(be bf ef)				6.188		6.188	6.188	80.438	320.120	4.000	0.000	0.000
(bb bf ef)						24.750	6.188	68.062	325.771	3.000	0.000	0.000
(bf ee ef)				24.750			6.188	68.062	325.771	3.000	0.000	0.000
(bbeeff)				3.960		3.960	3.960	87.120	536.009	4.000	0.000	0.000
(be bf bf)				1.547	1.547	23.203	4.641	68.062	455.118	5.000	0.000	0.000

(be ef ef)		1.547	23.203	1.547	4.641	68.062	455.118	5.000	0.000	0.000
(bf bf ef)	1.547		4.641	23.203	1.547	68.062	1489.893	5.000	0.000	0.000
(bf ef ef)	1.547		23.203	4.641	1.547	68.062	1489.893	5.000	0.000	0.000
(bbefff)	0.247		3.712	24.503	0.247	70.290	9637.577	5.000	0.000	0.000
(beefff)	0.247		24.503	3.712	0.247	70.290	9637.577	5.000	0.000	0.000

8.12.2 Individual II

Locus Pgi-2 (alleles abc)	**Allelic Phenotypes**					**Chi^2**	**df.**	**Posterior Chi^2 Probability**	
Phenotypes:	bc	a	ac	ab	abc				
Observed:		***4.000***	***7.000***	***11.000***	***27.000***				
(aa ab ac)		3.062	9.188	9.188	27.562	1.177	4.000	0.700	0.882
(aaaabc)		1.960	10.290	10.290	26.460	3.235	4.000	0.300	0.519
(ab ab ac)	0.766	0.766	2.297	11.484	33.688	25.407	5.000	0.000	0.000
(ab ac ac)	0.766	0.766	11.484	2.297	33.688	50.485	5.000	0.000	0.000
(aaabbc)	0.122	0.122	1.837	12.127	34.790	139.211	5.000	0.000	0.000
(aaabcc)	0.122	0.122	12.127	1.837	34.790	172.457	5.000	0.000	0.000

Locus Idh-1 (alleles aef)	**Allelic Phenotypes**							**Chi^2**	**df.**	**Posterior Chi^2 Probability**	
Phenotypes:	f	e	ef	a	af	ae	aef				
Observed:						***11.000***	***36.000***				
(aa af ee)						11.750	35.250	0.064	2.000	0.469	0.969
(aa ee ef)						11.750	35.250	0.064	2.000	0.469	0.969
(aa ae ef)					2.938	11.750	32.312	3.406	3.000	0.020	0.333
(ae af ee)			2.938			11.750	32.312	3.406	3.000	0.020	0.333
(ae ae af)			0.734	0.734	2.203	11.016	32.312	4.093	5.000	0.010	0.536
(ae ae ef)		0.734	2.203		0.734	11.016	32.312	4.093	5.000	0.010	0.536
(aa ef ef)					2.938	2.938	41.125	25.705	3.000	0.000	0.000
(af af ee)			2.938			2.938	41.125	25.705	3.000	0.000	0.000
(ae af ef)			2.938		2.938	2.938	38.188	28.129	4.000	0.000	0.000
(aa ae af)				2.938	8.812	8.812	26.438	15.752	4.000	0.000	0.003
(ae ee ef)		2.938	8.812			8.812	26.438	15.752	4.000	0.000	0.003
(aaaaef)				1.880	9.870	9.870	25.380	16.323	4.000	0.000	0.003

(aeeeef)		1.880	9.870			9.870	25.380	16.323	4.000	0.000	0.003
(aaeeff)			1.880		1.880	1.880	41.360	48.696	4.000	0.000	0.000
(aa aa ef)					11.750	11.750	23.500	18.447	3.000	0.000	0.000
(af ee ee)			11.750			11.750	23.500	18.447	3.000	0.000	0.000
(aa af ef)					11.750	2.938	32.312	34.300	3.000	0.000	0.000
(af ee ef)			11.750			2.938	32.312	34.300	3.000	0.000	0.000
(ae af af)			0.734	0.734	11.016	2.203	32.312	48.030	5.000	0.000	0.000
(ae ef ef)		0.734	11.016		0.734	2.203	32.312	48.030	5.000	0.000	0.000
(af af ef)	0.734		2.203		11.016	0.734	32.312	157.874	5.000	0.000	0.000
(af ef ef)	0.734		11.016		2.203	0.734	32.312	157.874	5.000	0.000	0.000
(aaefff)	0.118		1.762		11.633	0.118	33.370	1021.625	5.000	0.000	0.000
(aeefff)	0.118		11.633		1.762	0.118	33.370	1021.625	5.000	0.000	0.000

8.12.3 Individual III

Locus Pgi-2 (alleles abc)				**Allelic phenotypes**				**Chi²**	**df.**	**Posterior**	**Chi² Probability**
Phenotypes:	c	b	bc	a	ac	ab	abc				
Observed:				***43.000***			***153.000***				
(ab bb cc)				49.000			147.000	0.980	2.000	0.500	0.613
(ac bb cc)				49.000			147.000	0.980	2.000	0.500	0.613
(abbbcc)			0.490	48.510		0.490	7.350	139.160	10.332	5.000	0.000	0.066
(abbccc)		0.490		48.510		7.350	0.490	139.160	10.332	5.000	0.000	0.066
(ac bb bc)				49.000			12.250	134.750	15.456	3.000	0.000	0.001
(ab bc cc)				49.000		12.250		134.750	15.456	3.000	0.000	0.001
(ab bc bc)			3.062	45.938		3.062	9.188	134.750	17.972	5.000	0.000	0.003
(ac bc bc)		3.062		45.938		9.188	3.062	134.750	17.972	5.000	0.000	0.003
(ab ab cc)				12.250		12.250		171.500	91.434	3.000	0.000	0.000
(ac ac bb)				12.250			12.250	171.500	91.434	3.000	0.000	0.000
(ab ac bc)				12.250		12.250	12.250	159.250	101.934	4.000	0.000	0.000
(aabbcc)				7.840		7.840	7.840	172.480	175.562	4.000	0.000	0.000
(ab bb bc)			12.250	36.750			36.750	110.250	66.639	4.000	0.000	0.000
(ac bc cc)		12.250		36.750		36.750		110.250	66.639	4.000	0.000	0.000
(abbbbc)			7.840	41.160			41.160	105.840	70.096	4.000	0.000	0.000

(abcccc)	7.840		41.160		41.160		105.840	70.096	4.000	0.000	0.000
(ac bb bb)			49.000			49.000	98.000	80.602	3.000	0.000	0.000
(ab cc cc)			49.000		49.000		98.000	80.602	3.000	0.000	0.000
(ab ac bb)			12.250			49.000	134.750	128.660	3.000	0.000	0.000
(ab ac cc)			12.250		49.000		134.750	128.660	3.000	0.000	0.000
(ab ab bc)		3.062	9.188		3.062	45.938	134.750	178.973	5.000	0.000	0.000
(ac ac bc)	3.062		9.188		45.938	3.062	134.750	178.973	5.000	0.000	0.000
(aabbbc)		0.490	7.350		0.490	48.510	139.160	223.781	5.000	0.000	0.000
(aabccc)	0.490		7.350		48.510	0.490	139.160	223.781	5.000	0.000	0.000
(ab ab ac)			3.062	3.062	9.188	45.938	134.750	581.477	5.000	0.000	0.000
(ab ac ac)			3.062	3.062	45.938	9.188	134.750	581.477	5.000	0.000	0.000
(aaabbc)			0.490	0.490	7.350	48.510	139.160	3745.686	5.000	0.000	0.000
(aaabcc)			0.490	0.490	48.510	7.350	139.160	3745.686	5.000	0.000	0.000

Locus 6Pgd-2 (alleles bef)	**Allelic phenotypes**					**Chi^2**	**df.**	**Posterior**	**Chi^2 Probability**
Phenotypes:	ef	b	bf	be	bef				
Observed:			***19.000***	***33.000***	***48.000***	***73.000***			
(bb be bf)			10.812	32.438	32.438	97.312	19.750	4.000	0.907	0.001
(bbbbef)			6.920	36.330	36.330	93.420	29.605	4.000	0.093	0.000
(be be bf)		2.703	2.703	8.109	40.547	118.938	196.466	5.000	0.000	0.000
(be bf bf)		2.703	2.703	40.547	8.109	118.938	316.328	5.000	0.000	0.000
(bbbeef)		0.432	0.432	6.487	42.818	122.830	926.738	5.000	0.000	0.000
(bbbeff)		0.432	0.432	42.818	6.487	122.830	1085.645	5.000	0.000	0.000

Locus Aat-2 (alleles ab)	**Allelic phenotypes**			**Chi^2**	**df.**	**Posterior**	**Chi^2 Probability**
Phenotypes:	b	a	ab				
Observed:	***22.000***		***172.000***				
(ab ab bb)	12.125		181.875	8.579	2.000	0.996	0.014
(aabbbb)	7.760		186.240	27.220	2.000	0.003	0.000
(abbbbb)	48.500		145.500	19.306	2.000	0.000	0.000
(ab bb bb)	48.500		145.500	19.306	2.000	0.000	0.000
(ab ab ab)	3.031	3.031	187.938	123.084	3.000	0.000	0.000
(aaabbb)	0.485	0.485	193.030	957.199	3.000	0.000	0.000

8.13 FDASH DOCUMENTATION

[This appendix is a reproduction of the documentation intended for distribution with 'FDASH', a text-based computer program for the analysis of allelic phenotype data (see Chapter 4). Headings have been re-numbered to conform to the rest of the thesis]

8.13.1 What is FDASH?

FDASH is a text-interface program that calculates a number of the *ad hoc* statistics that have been used to summarise genetic diversity and differentiation in polyploids, such as the Shannon-Weaver diversity of phenotypes.

FDASH gets its name from F'_{ST}, a genetic-differentiation statistic based on allelic phenotype* diversity within and between populations (introduced in Chapter 4 of this thesis)

> ***NOTE** An allelic phenotype arises if you cannot distinguish between polyploid individuals (a,b,b,b) and (a,a,b,b) and are forced to score them both as (a,b) regardless of the number of copies of each allele.*

While the program was written for use with isozyme data, there is no reason why it could not also be used for microsatellites. However, it does not take account of any information that could be gained by the use of a stepwise mutation model. FDASH is intended for use in disomic polyploids (allopolyploids and diploidised autopolyploids), as inference is straightforward (see Chapter 4). Although the summary statistics may also be used to quantify genetic variation in autopolyploids, it seems likely that they will be more strongly affected by processes such as selfing, which increase homozygosity.

To make sense of what FDASH is intended for, it helps to have a reasonable understanding of the issues involved with polyploid population genetics. At present this documentation should be read in conjunction with Chapter 4.

There are many papers that have dealt with polyploid population genetics, some of which are very helpful in terms of suggested methodology, in particular:

For polysomic polyploids I recommend

Bever JD, Felber F (1992) The theoretical population genetics of autopolyploidy. *Oxford Surveys of Evolutionary Biology* **8**, 185-217.

Ronfort J, Jenczewski E, Bataillon T, Rousset F (1998) Analysis of population structure in autotetraploid species. *Genetics* **150**, 921-930.

Hardy OJ, Vekemans X (2001) Patterns of allozyme variation in diploid and tetraploid Centaurea jacea at different spatial scales. *Evolution* **55**, 943-954.

Hardy OJ, Vekemans X (2002) SPAGEDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes* **2**, 618-620

For *ad hoc* statistics with allelic phenotypes I recommend

Rogers DL (2000) Genotypic diversity and clone size in old-growth populations of coast redwood (Sequoia sempervirens). *Canadian Journal of Botany-Revue Canadienne De Botanique* **78**, 1408-1419.

Meerts P, Baya T, Lefebvre C (1998) Allozyme variation in the annual weed species complex Polygonum aviculare (Polygonaceae) in relation to ploidy level and colonizing ability. *Plant Systematics and Evolution* **211**, 239-256.

Chung MG, Hamrick JL, Jones SB, Derda GS (1991) Isozyme variation within and amoung populations of *Hosta* (Liliaceae) in Korea. *Systematic Botany* **16**, 667-684.

Gaur PK, Lichtwardt RW, Hamrick JL (1980) Isozyme variation among soil isolates of *Histoplasma capsulatum*. *Experimental Mycology* **5**, 69-77.

If you are aware of literature that is relevant, either bad or good, please tell me about it.

8.13.2 What can you do with FDASH?

FDASH is very limited in its application. It is simply intended to make the calculation of phenotypic statistics faster than it is by hand. Included are a series of summary statistics for allelic phenotype data, including (but not limited to):

- The total number of alleles
- Total number of phenotypes
- The proportion of polymorphic loci
- Within-sample diversity measures
- Overall diversity measures
- F_{ST}–like differentiation statistics based on the above diversity measures
- The average number of distinct alleles carried by each individual
- The most alleles carried by an individual
- The fewest alleles carried by an individual
- The average number of alleles in each phenotype

Additionally, given a list of points (e.g. population locations) with longitude and latitude co-ordinates in decimal degrees, FDASH can calculate the pairwise geographic distances between them.

Genetic summary statistics are calculated per locus, and as a cross-locus average.

- They can be calculated:
    - As a single average for a set of population samples
    - Where appropriate, for each population sample individually
    - As an average for each of a number of user-defined groups of population samples (e.g. regional groups, groups that differ in an interesting biological trait)
- All averages are weighted by sample-size
- F'_{ST} can be calculated pairwise between all pairs of populations.

- The significance of differences between groups for some statistics (especially F'_{ST} and related diversity measures) can be assessed using randomisation tests. However, **no correction is made for multiple tests**.

8.13.3 The summary statistics

All cross-locus averages are weighted by the sample size (there may be missing data for some loci).

The following is a list of all the statistics that FDASH calculates, some are unlikely to have any application; they are included only because they were used in an earlier simulation study.

- **Total number of alleles seen**
  - The total number of alleles seen in each locus for the unit in question (e.g. over all population samples, per population sample, per group)

- **Av. number of alleles per sample**
  - The average number of alleles seen in each locus in each sample (e.g. over all population samples, per population sample, per group). The average is weighted by sample size, but does not account for the fact that more alleles will be seen in larger samples.

- **Total number of phenotypes seen**
  - The total number of phenotypes seen in each locus for the unit in question (e.g. over all population samples, per population sample, per group)

- **Av. number of phenotypes per sample**
  - The average number of phenotypes seen in each locus in each sample (e.g. over all population samples, per population sample, per group). The average is weighted by sample size, but does not account for the fact that more phenotypes will be seen in larger samples.

- ***H'*S**
  - The average number of alleles by which pairs of individuals differ, with pairs on individuals being taken within samples, e.g. individuals [100,122,136] and [077,122,136] differ by 2 alleles, 100 and 077.
  - This is a measure of within-sample diversity H_S': If all individuals in each sample are the same it will be zero, if they are all different it will

$$H' = \frac{1}{n(n-1)} \sum_{i}^{n} \sum_{j>i}^{n} \sum_{k \in alleles} x_{ijk}$$

n is the number of individuals
$x_{ijk} = 1$ if only one of the individuals i and j carry allele k; otherwise $x_{ijk} = 0$

    be large. The more differences there are between individuals, the larger the number.
  - The average across samples is weighted by sample-size.

- ***H'*T**
  - The average number of alleles by which pairs of individuals differ, with pairs on individuals being taken from all samples, e.g. individuals [100,122,136] and [077,122,136] differ by 2 alleles, 100 and 077.
  - This is a measure of over-all diversity H_T': If all individuals are the same it will be zero, if they are all different it will be large. The more

$$H' = \frac{1}{n(n-1)} \sum_{i}^{n} \sum_{j>i}^{n} \sum_{k \in alleles} x_{ijk}$$

n is the number of individuals
$x_{ijk} = 1$ if only one of the individuals i and j carry allele k; otherwise $x_{ijk} = 0$

    differences there are between individuals, the larger the number.

- ***F'ST***
  - A differentiation statistic: the proportion of total diversity that is found between samples, calculated using the two "unshared alleles" diversity statistics from above.

$$F'_{ST} = \frac{H'_T - \bar{H}'_S}{H'_T}$$

$H'_T \equiv$ total diversity, $\overline{H'_S} \equiv$ mean subpopulation diversity

  - No corrections for sample size (other than average H'_S being weighted) are made.

- $H^{SW}{}_S$
  - Shannon-Weaver diversity of phenotypes, calculated within samples, and averaged (weighted by sample-size) across samples.

$$H' = \sum_{i=1}^{m} p_i \log(p_i)$$ m is the number of alleles; p_i is the frequency of the ith phenotype

  - This is a measure of within-sample diversity H_S'

- $H^{SW}{}_T$
  - Shannon-Weaver diversity of phenotypes, calculated across samples
  - This is a measure of over-all diversity H_T'

$$H' = \sum_{i=1}^{m} p_i \log(p_i)$$ m is the number of alleles; p_i is the frequency of the ith phenotype

- H^{SW}**-based** F_{ST}
  - A differentiation statistic: the proportion of total diversity that is found between samples, calculated using the two "unshared alleles" diversity statistics from above.

$$F'_{ST} = \frac{H'_T - \bar{H}'_S}{H'_T}$$ $H'_T \equiv$ total diversity, $\overline{H'_S} \equiv$ mean subpopulation diversity

  - No corrections for sample size (other than average H'_S being weighted) are made.

- $H^{Phen}{}_S$
  - Shannon-Weaver diversity of phenotypes, calculated within samples, and averaged (weighted by sample-size) across samples.
  - This is a measure of within-sample diversity H_S'

$$H' = \sum_{i=1}^{m} p_i(1-p_i)$$ m is the number of alleles, p_i is the frequency of the ith phenotype

- H^{Phen}_{T}
  - Shannon-Weaver diversity of phenotypes, calculated across samples

$$H' = \sum_{i=1}^{m} p_i(1-p_i)$$ m is the number of alleles, p_i is the frequency of the ith phenotype

  - This is a measure of over-all diversity H_T'

- H^{Phen} **–based** F_{ST}
  - A differentiation statistic: the proportion of total diversity that is found between samples, calculated using the two "unshared alleles" diversity statistics from above.

$$F'_{ST} = \frac{H'_T - \bar{H}'_S}{H'_T}$$ $H'_T \equiv$ total diversity, $\overline{H'_S} \equiv$ mean subpopulation diversity

  - No corrections for sample size (other than average H'_S being weighted) are made.

- **Average number of different alleles carried by each individual**
  - The number of different alleles carried by each individual is counted, and the total divided by the number of individuals.

- **Variance in the number of different alleles carried by each individual**
  - The number of different alleles carried by each individual is counted, and the variance calculated.

- **Most alleles carried by an individual**

- **Fewest alleles carried by an individual**

- **Av. number of alleles in each phenotype**

- The number of alleles in each phenotype is counted, and the total divided by the number of different phenotypes.

- **Variance in the number of alleles in each phenotype**
  - The number of alleles in each phenotype is counted, and the variance calculated. Note this is a within-locus allelic phenotype, not a multilocus-phenotype.

- **Av. number of SHARED alleles between pairs of individuals within samples**
  - For each pair of individuals (drawn within samples) the number of alleles they share is counted (e.g. individuals [100,122,136] and [077,122,136] share 2 alleles, 122 and 136), and this is averaged over pairs.

- **Av. number of SHARED alleles between pairs of individuals from different samples**
  - For each pair of individuals (drawn from different samples) the number of alleles they share is counted (e.g. individuals [100,122,136] and [077,122,136] share 2 alleles, 122 and 136), and this is averaged over pairs.

- **Av. number of UNSHARED alleles between pairs of individuals from different samples**
  - For each pair of individuals (from different samples) the number of alleles they do not share is counted (e.g. individuals [100,122,136] and [077,122,136] do not share 2 alleles, 077 and 122), and this is averaged over pairs.

- **Av. number of SHARED alleles between pairs of individuals across all samples**
  - For each pair of individuals (drawn from all samples) the number of alleles they share is counted (e.g. individuals [100,122,136] and

[077,122,136] share 2 alleles, 122 and 136), and this is averaged over pairs.

- **Is the locus polymorphic? (1/0)**
  - A one or zero (true or false) record of whether the locus was polymorphic, i.e. whether there was more then one allelic phenotype.
  - Note that if **all** individuals have the same alleles at a locus, there is no polymorphism.
  - The cross-locus average of this value is a measure of the proportion of polymorphic loci.

- **Could PsLD be calculated? (1/0)**
  - A one or zero (true or false) record of whether the PseudoLD statistic (described below) could be calculated – i.e. were there more than two variable alleles.

- **PsLD**
  - PseudoLD; this is a very odd statistic that has very little power, and even less use.
  - If the organism is a disomic 2*n*-ploid, a 'locus' is in fact *n* different 'sub-loci' that have the same alleles. If there is some form of admixture and/or some selfing, Linkage Disequilibrium (LD) can be generated and maintained between the sub-loci. This statistic was an attempt to get a handle on that LD. However, since it needs many variable loci, and will always have some non-zero value by chance, it is not much help. It would be much better to have a measure of inter-locus LD.

- **Av. Number of carriers of each allele across all samples**
  - The average proportion of individuals that carry each allele.
  - i.e. if there are 6 alleles, a,b,c,d,e & f, and 97% of individuals carry a, 50% carry b, 30% carry c, 1% carry d and 0.5% carry e and f, then this statistic will be 0.2983

- **Variance in the number of carriers of each allele across all samples**
    - The variance in the proportion of individuals that carry each allele.
    - i.e. if there are 6 alleles, a,b,c,d,e & f, and 97% of individuals carry a, 50% carry b, 30% carry c, 1% carry d and 0.5% carry e and f, then this statistic will be 0.1245

- **Frequency of carriers of the commonest allele across all samples**
    - The proportion of individuals carrying the allele that has the most carriers
    - i.e. if there are 6 alleles, a,b,c,d,e & f, and 97% of individuals carry a, 50% carry b, 30% carry c, 1% carry d and 0.5% carry e and f, then this statistic will be 0.970

- **Frequency of carriers of the rarest allele across all samples**
    - The proportion of individuals carrying the allele that has the fewest carriers
    - i.e. if there are 6 alleles, a,b,c,d,e & f, and 97% of individuals carry a, 50% carry b, 30% carry c, 1% carry d and 0.5% carry e and f, then this statistic will be 0.005

- **Av. Number of carriers of each allele within each sample**
    - The average proportion of individuals that carry each allele calculated within samples, and averaged (weighted by sample-size) across samples

- **Variance in the number of carriers of each allele within each sample**
    - The variance in the proportion of individuals that carry each allele, calculated within samples, and averaged (weighted by sample-size) across samples

- **Frequency of carriers of the commonest allele within each sample**
    - The proportion of individuals carrying the allele that has the most carriers, calculated within samples, and averaged (weighted by sample-size) across sample

- **Frequency of carriers of the rarest allele within each sample**
    - The proportion of individuals carrying the allele that has the fewest carriers, calculated within samples, and averaged (weighted by sample-size) across sample

- **Av. Number of carriers of each phenotype across all samples**
    - This and the following phenotype statistics are analogous to the equivalent eight allele-statistics given above.
    - Note that while individuals can carry many alleles, they can each only carry one phenotype (per locus), and thus the frequencies must sum to one.

8.13.4 Limitations of the program

For ease of programming there are a small number of limitations to the data:

- A single locus can have no more than 40 alleles
- A single locus can have no more than 150 phenotypes
- A single phenotype can have no more than 10 alleles

These limits could all be increased very easily – email me if you have a problem.

- The significance of differences between groups is only reported for a few statistics. While a small amount of re-programming would be needed to report this for other statistics, it would not be too difficult to do.
- There is no limit (other than the size and speed of your computer) to the number of loci, the number of groups, the number of samples, or the number of individuals.
- However, be aware that when doing randomisations for significance testing, the program can be slow. It was written quickly without much though for speed optimisation, and with >2000 individuals doing randomisation tests takes a significant amount of time.

8.13.5 What the input file looks like

I have attempted to make the input as user-friendly as possible. FDASH reads tab-delimited text files as output by the "export" function of MS Access and the "save as tab-delimited txt" option in MS Excel. The first row contains column-titles (remember to choose this option when exporting from access!) These can be as wide as you want, and include any characters (except tabs) you want, but a tab must separate columns, and there must be no tab after the last column e.g.:

```
GroupNames<tab>SampleNames<tab>Other_data<tab>Locus1<tab>Locus2
```

- Each subsequent row corresponds to an individual.
- The data in column 1 are used as a group identifier, and may consist of any alphanumeric characters.
- The data in column 2 are used as the sample (e.g. population) identifiers, and again may consist of any alphanumeric characters.
- The data in column 3 is unused, and can contain any other notes you wish.
- Column 3 must be present, but can be empty (e.g. `<tab><tab>`).
- All other columns are assumed to contain genetic data.

Genetic data is in the form of allelic phenotypes, alleles identified with a 3-digit code (e.g. 110, 456, or 001) and separated by commas. Thus the phenotype of an individual that carries alleles 010, 100, and 136 is recorded as: 010,100,136. The order does not matter, and duplication will be ignored, thus 010,100,136,136 will be treated as: 010,100,136. A missing locus for an individual can be coded with a single '?' or left blank (i.e `<tab><tab>`)

A suitable access query (two groups, two samples, 16 individuals for two loci: Aat and Pgi) ready for export could look like this (note the missing data):

Country	Population	Location	Aat-2	Pgi-2
Canary Islands	0206a	El Sauzal	100,115	077,100
Canary Islands	0206a	El Sauzal	115	100
Canary Islands	0206a	El Sauzal	115	100

Country	Population	Location	Aat-2	Pgi-2
Canary Islands	0206a	El Sauzal	115	100
Canary Islands	0206a	El Sauzal	?	100
Canary Islands	0206a	El Sauzal	115	100
Canary Islands	0206a	El Sauzal	115	100,122
Canary Islands	0206a	El Sauzal	115	100
Morocco	1020a	nr. Dar-ed-Dou	100,115	077,100,122
Morocco	1020a	nr. Dar-ed-Dou		077,100,122
Morocco	1020a	nr. Dar-ed-Dou	100,115	077,100,122
Morocco	1020a	nr. Dar-ed-Dou	100,115	077,100,122
Morocco	1020a	nr. Dar-ed-Dou		077,100,122,136
Morocco	1020a	nr. Dar-ed-Dou	100,115	?
Morocco	1020a	nr. Dar-ed-Dou	100,115	077,100,122,136
Morocco	1020a	nr. Dar-ed-Dou	100,115	077,100,122,136

8.13.6 What the output files look like

The output files should be completely self-explanatory.

Both files are in tab-delimited text format, and can be easily imported into MS excel (Choose “Open”, then “Text Files” then “Tab delimited”, or simply drag-and-drop the files into an open Excel window).

<name> is the stem the user chooses at the start of program execution.

File 1: **`name_summary.txt`**

- This file contains a summary of allele-carrier counts and frequencies, and phenotype counts and frequencies, overall, and per sample, arranged by locus.
- This data could be used for further analysis, or simply as an easy way of checking for mistakes in the input data.

File 2: **`name_stats.txt`**

- This file contains all the summary statistics generated by FDASH.

- There are up to 4 sections to the file, depending on which selections you make at the beginning:
    1. **Statistics calculated across all samples**: All the summary statistics, calculated across all samples
    2. **Statistics calculated for each sample individually:** All the summary statistics are calculated for each sample individually. Many statistics (e.g. F'_{ST}) cannot be calculated for a single sample. In these cases either zero is returned, or the multi-sample statistic takes the same value as the single-sample equivalent.
    3. **Pairwise F'_{ST} (unshared alleles):** F'_{ST} calculated using the unshared-alleles measure of diversity is calculated pairwise between all samples. Since there is no well-defined expectation as to how F'_{ST} *should* behave under an isolation-by-distance model, it is not entirely clear why you would wish to do this.
    4. **Group Statistics:** All the summary statistics, calculated across all samples for each user-defined group. These are followed by p-values for the hypothesis of "no difference between groups", for a few key summary statistics. Randomly shuffling the samples between groups; then recording the proportion of occasions on which the difference between shuffled groups is greater than between the original groups is used to calculate these p-values.

8.13.7 Conditions

This program and source code are distributed free of charge for any academic use, but if you publish statistics calculated using it, please cite this program.

Please contact me (Darren.Obbard@plants.ox.ac.uk) if you wish to employ either the program or any part of the code for a non-academic purpose.

If you should have some strange desire to re-write it completely and make a nice GUI for it, then that would be very welcome, and I only ask that you cite the original source.

8.13.8 Disclaimer

This is a largely untested beta-version of FDASH, and there is no guarantee that any part of it works correctly (or indeed at all). It is provided as-is, to be used entirely at your own risk, and I can accept no responsibility for any loss or damage that arises as a consequence of its use.

If you find bugs, errors, or omissions then please email me (Darren.Obbard@plants.ox.ac.uk) and I shall have a look at how hard they are to correct.

8.14 PUBLISHED PAPER: "PROBING THE PRIMACY OF THE PATCH: WHAT MAKES A METAPOPULATION?"

[The following co-authored paper was published in 2003, in the Journal of Ecology vol. 91 pages 485-488. Here it is presented in manuscript form, with the headings re-numbered to conform to the style of this thesis. References are listed after the text.]

PROBING THE PRIMACY OF THE PATCH:
WHAT MAKES A METAPOPULATION?

J. R. Pannell1 and D. J. Obbard

Department of Plant Sciences
University of Oxford,
Oxford OX1 3RB
U.K.

1: Corresponding author: john.pannell@plants.ox.ac.uk

The use of the term 'metapopulation' has broadened substantially since its inception to include, for example, subdivided populations that are not necessarily prone to local extinction, and populations with a locally patchy distribution that are not clearly subdivided into discrete demes. This broadened scope has coincided with an exponential increase in the number of articles applying the term (Hanski & Simberloff, 1997; Pannell & Charlesworth, 2000), but it has arguably come at a cost of precision. In their useful review of the literature on large-scale spatial dynamics in plants, Freckleton and Watkinson (2002) have blown the whistle on this trend. They argue that metapopulations ought only to describe an array of populations (1) that are prone to local extinction, and (2) that also inhabit discrete and recognisable habitat patches. The first of these criteria is uncontroversial, although the rate of local population extinction that is biologically significant will depend on the nature of the questions being addressed. We believe that the second criterion, however, fails to recognise the utility of the metapopulation approach in studies that are not focused specifically on patch occupancy rates. In this sense, our view thus differs from that of Freckleton and Watkinson (2002), as well as from views expressed recently by Bullock *et al*. (2002), who also emphasise the importance of fixed habitat patches in defining a metapopulation.

It is of course true that many applications of the metapopulation concept in ecology and conservation need to address patch occupancy rates explicitly, and thus require the a priori identification of habitat patches. In their reply to Ehrlén and Eriksson's (2003) critique of their review, Freckleton and Watkinson (2003) emphasise this point. If we are interested in ensuring the regional conservation of a metapopulation, then efforts must be directed towards the conservation of habitat, whether currently occupied or not. But the dynamics of population turnover in a metapopulation affect not only the regional persistence or survival of a species, but also its population genetics (reviewed in Pannell & Charlesworth, 2000) and evolution (reviewed in Ronce & Olivieri, 2003)- whether or not we can identify its habitat. Bullock *et al*. (2002) state that "a basic premise of metapopulation theory and models [is] that extinctions make habitat patches available for colonisation" (p. 291). We agree with the implication that an empirical definition of a metapopulation ought to be consistent with its use in the theoretical literature. However, not all theoretical metapopulation models assume the existence of fixed habitat patches.

It seems that our differing views stem, at least in part, from a failure to integrate genetic or evolutionary aspects of metapopulation biology with ecological or demographic ones. For example, Freckleton and Watkinson (2002) adopted an entirely demographic perspective in their review and did not consider population genetic structure, because "metapopulation theory is not concerned with the movement of genes per se" (p. 421). However, the metapopulation perspective has in fact been used fruitfully in both ecology and population genetics, and indeed the concept was first considered by Wright (1940) in the context of population genetics long before Levins (1969; 1970) explicitly introduced the term. Ives and Whitlock (2002) have recently noted that 'population genetic metapopulations' may not necessarily equate with 'ecological metapopulations', because extinction-colonisation dynamics need not affect both the demography and the genetic structure of a species to the same extent. However, because the same underlying processes make the metapopulation concept valuable in population genetics and ecology, consistency in applying the term is desirable.

8.14.1 Subdivided populations versus habitat patches

The main point we wish to make is that, whilst it is clearly true that species that occupy discrete habitat patches will occur in discrete groups, the discrete nature of the groups themselves will affect important aspects of a species' biology, irrespective of the underlying causes. Indeed, Levins (1970) stressed the importance of a patchy habitat principally in its creation of an insular distribution of organisms. Reduced insistence on the identification of discrete habitat patches recognises that what really matters in a metapopulation is that the extinction of groups is balanced by founding of new groups within the range of the metapopulation, i.e., not necessarily within a set of fixed and recognisable patches, which may by nature be temporary and difficult to observe. From an ecological point of view, an inability to identify unoccupied habitat will make the metapopulation approach difficult to apply, because the proportion of occupied patches will often be an important state variable. From a population-genetics perspective, by contrast, the most useful information for the application of metapopulation theory will be the colony age distribution (Wade & McCauley, 1988). In situations where colony age correlates with the successional stage of the patch (e.g.,

Giles & Goudet, 1997a), patch characterisation may be useful, but this will not always be possible or necessary.

Ultimately, group discreteness depends on limited mixing between groups through dispersal (migration). The point at which dispersal erodes group identity enough to make the metapopulation concept unhelpful is to some extent arbitrary, but essentially it will depend upon the biological question being addressed. For example, neutral gene frequencies will begin to differ appreciably between demes in a subdivided population without local extinction, when m, the proportion of individuals in extant demes that were migrants in the previous generation, exceeds 1/N, the reciprocal of the local population size. Thus, groups can be viewed as being relatively discrete when $Nm < 1$ (Wright, 1951). In a metapopulation with extinction, patterns of neutral diversity are only affected appreciably by population turnover when e, the population extinction rate, exceeds m (Slatkin, 1977; Pannell & Charlesworth, 1999, 2000). Subdivided populations in which local processes are dominant, described by Freckleton and Watkinson (2002) as 'regional ensembles', will meet the criterion $e < m < 1/N$. In subdivided populations with migration dominant over both extinction ($m > e$) and the local effects of drift ($m > 1/N$), the effects of genetic bottlenecks that follow colonisation events are quickly eroded, so that the average effects of extinction can be ignored. Similarly, the effects of selection, e.g., in fuelling local adaptation or manifest in inbreeding depression, depend on the genetic identity within and differentiation between populations, which in turn depend on relative rates of extinction, migration and drift (Whitlock & McCauley, 1990; Whitlock, 2002). As a final example, the relative degree of group identity, quantified in terms of the relatedness of interacting individuals, directly determines the efficacy of kin selection (Hamilton, 1964).

Note that none of these theoretical scenarios is spatially explicit, and none assumes fixed habitat patches. Nor have several empirical tests of metapopulation genetic theory needed to heed the existence of discrete habitat patches (e.g., Antonovics *et al*., 1994; Richards *et al*., 1999; McCauley *et al*., 2000; Richards, 2000), although they may occur (see Gaggiotti *et al*., 2002, for a population-genetic study where the explicit recognition of habitat patches was useful). An interesting example is that of *Silene dioica* on islands in the Baltic Sea (Giles & Goudet, 1997b), which Bullock *et al*. (2002) cite as a rare instance of a true plant metapopulation-even though this system violates their stated premise that "extinctions make habitat available for recolonisation". The habitat of *S. dioica* on these Baltic islands is indeed patchy in the

extreme, but it is the patchiness of the groups that has been the focus of analysis, and the same sort of analysis could equally have been employed if the species had occurred as a 'shifting cloud' of populations in a continuous habitat (sensu Freckleton & Watkinson, 2002) or a system of populations involved in 'habitat tracking' (sensu Harrison & Taylor, 1997). Comparisons of metapopulations with and without a patchy habitat are badly needed from both a demographic and a population genetic perspective, but in the absence of empirical data it is not clear that they must be different.

8.14.2 The question of scale

Freckleton and Watkinson (2002) and Bullock *et al*. (2002) rightly emphasised the need to distinguish between local and regional scales: a metapopulation approach might be appropriate for analysis of processes occurring at a regional scale, but not for processes operating below the local scale. Ehrlén and Eriksson (2003) have replied that any subdivided population might be regarded as a metapopulation if the appropriate spatial and temporal scale of analysis is adopted. We agree with Freckleton and Watkinson (2003) that this perspective seems unhelpful. What matters is whether the biological attribute under study is affected by population turnover at a regional scale. It therefore seems clear that the local dynamics in Freckleton and Watkinson's (2002) 'spatially extended population', where the spatial dimensions of groups of organisms are extended and moved by the diffusive effects of local dispersal, do not constitute an important component of the (regional) dynamics of a metapopulation.

Nevertheless, the potentially extendable and mobile groups of such spatially extended populations may usefully be regarded as the sub-populations or demes of a metapopulation if they are sufficiently isolated from other such patches and their origin is the result of a colonisation event through non-local dispersal. This may be abiotically (e.g., by wind or water) or biotically assisted and may thus differ qualitatively from local dispersal. Even where colonisation and gene flow amongst extant populations are due to the same process of dispersal, however, Ibrahim *et al*. (1996) have shown that discretely subdivided populations can result in a continuous habitat as a result of the stochasticity of dispersal and colony establishment when the dispersal curve has a long tail. The important point is not the possible significance of the mode or absolute distance of dispersal, but whether dispersal leads to a

colonisation event, i.e., to the establishment of a new discrete group. In population-genetic and evolutionary models, it is typically the repeated bottlenecks associated with colonisation that give rise to the metapopulation effects (Olivieri *et al*., 1997) that are not seen in (spatially extended) local populations, such as particular patterns of genetic (reviewed in Pannell & Charlesworth, 2000), mating-system (reviewed in Barrett & Pannell, 1999) and life-history variation (reviewed in Ronce & Olivieri, 2003). Migration amongst established groups tends to erode these effects, which cannot be produced by local dispersal.

Population turnover that gives rise to metapopulation effects may of course occur in species with discrete habitat patches (e.g., Ebert *et al*., 2002), but discrete and identifiable habitat patches are not essential. It is difficult, for example, to make sense of variation in sex ratios and sex allocation in several gynodioecious (e.g., van Damme, 1986; Manicacci *et al*., 1997; Taylor *et al*., 1999; and see Frank, 1989) and androdioecious plant species (reviewed in Pannell, 2002) without invoking population turnover in a metapopulation-even though these species may occupy continuous habitat as a 'shifting cloud'. The temporal scale over which population turnover occurs may make the metapopulation dynamics difficult to study directly, but their effects may be starkly apparent just the same. Indeed, under certain circumstances metapopulation processes may be inferred indirectly on the basis of these effects (Pannell, 2001, and unpublished ms).

Freckleton and Watkinson (2002) argued that plants such as *Vulpia ciliata* and *Silene alba* should not be regarded as metapopulations, because it is difficult to determine what constitutes a suitable habitat patch for these arable weeds (see their paper, pp. 430-431, for the relevant references). However, both of these species may occur as spatially discrete groups of individuals. In the case of *V. ciliata*, they note that "the only regional-scale phenomenon that cannot be predicted [on the basis of local-scale processes] is the origin of new populations" (p. 430). The low rate at which this occurs may be uninteresting from an ecological or demographic point of view, and in this sense investigators would be right in rejecting a metapopulation approach to analysis or management. However, from a genetic perspective these rare metapopulation events may be important in defining the structure of a species (Pannell & Charlesworth, 1999). Similarly, although S. alba appears to occupy a continuous habitat in roadside vegetation, the spatially discrete nature of groups of individuals has been shown to have important genetic consequences concerning the genetic rescue of

inbred populations by immigrants (e.g., Richards, 2000), in much the same way that it occurs in species with discrete habitat (e.g., Ebert *et al*., 2002).

8.14.3 Conclusions

Both Freckleton and Watkinson (2002) and Bullock *et al*. (2002) recommended a more precise terminology for describing the structure and regional dynamics of subdivided plant populations. Their taxonomies are useful in that they focus much needed attention on the important differences that occur between structures that have otherwise been referred to broadly as metapopulations. These differences are not just semantic, and it therefore seems useful to recognise them with an appropriate nomenclature. Nevertheless, we believe that consistency in the use of the term metapopulation is needed. This would be served by affording priority to the discrete and ephemeral nature of groups of individuals in defining a metapopulation, rather than to the presence of discrete habitat patches that may or may not underlie a given population structure. The identification of habitat patches will doubtless make metapopulation analysis easier from certain points of view, and patches may be of direct relevance to the ecology and conservation of species. However, the discrete nature of the groups of organisms involved is more fundamental. The successful adoption of the metapopulation terminology and approach in evolutionary and population-genetic analysis of species that do not occupy readily identifiable habitat patches underscores this point.

8.14.4 Acknowledgements

We thank R. P. Freckleton for helpful discussion, and Jonathan Silvertown and three anonymous referees for constructive comments on the manuscript. We gratefully acknowledge a Long Studentship (The Queen's College, Oxford) awarded to DJO and funding to JRP through NERC grant NER/B/S/2002/00225.

8.14.5 References

Antonovics, J., Thrall, P.H., Jarosz, A., & Stratton, D. (1994). Ecological genetics of metapopulations: the Silene-Ustilago plant-pathogen system. Ecological Genetics (ed L. Real). Princeton University Press, Princeton, NJ.

Barrett, S.C.H. & Pannell, J.R. (1999). Metapopulation dynamics and mating-system evolution in plants. Molecular Systematics and Plant Evolution (eds P. Hollingsworth, R. Bateman & R. Gornall), pp. 74-100. Chapman and Hall, London.

Bullock, J.M., Moy, I.L., Pywell, R., Coulson, S.J., Nolan, A.M., & Caswell, H. (2002). Plant dispersal and colonisation processes at local and landscape scales. Dispersal Ecology (eds J.M. Bullock, R.E. Kenward & R. Hails), pp.279-302. Blackwell Science, Oxford.

Ebert, D., Haag, C., Kirkpatrick, M., Riek, M., Hottinger, J.W., & Pajunen, V.I. (2002) A selective advantage to immigrant genes in a Daphnia metapopulation. Science, 295, 485-488.

Ehrlén, J. & Eriksson, O. (2003) Large-scale spatial dynamics of plants: a response to Freckleton & Watkinson. Journal of Ecology, 91, 000-000.

Frank, S.A. (1989) The evolutionary dynamics of cytoplasmic male sterility. American Naturalist, 133, 345-376.

Freckleton, R.P. & Watkinson, A.R. (2002) Large-scale spatial dynamics of plants: metapopulations, regional ensembles and patchy populations. Journal of Ecology, 90, 419-434.

Freckleton, R.P. & Watkinson, A.R. (2003) Are all plant populations metapopulations? Journal of Ecology, 91, 000-000.

Gaggiotti, O.E., Jones, F., Lee, W.M., Amos, W., Harwood, J., & Nichols, R.A. (2002) Patterns of colonization in a metapopulation of grey seals. Nature, 416, 424-427.

Giles, B.E. & Goudet, J. (1997a). A case study of genetic structure in a plant metapopulation. Metapopulation Biology: Ecology, Genetics, and Evolution (eds I. Hanski & M.E. Gilpin), pp. 429-454. Academic Press, San Diego.

Giles, B.E. & Goudet, J. (1997b) Genetic differentiation in Silene dioica metapopulations: Estimation of spatiotemporal effects in a successional plant species. American Naturalist, 149, 507-526.

Hamilton, W.D. (1964) The genetical evolution of social behaviour, I and II. Journal of Theoretical Biology, 7, 1-52.

Hanski, I. & Simberloff, D. (1997). The metapopulation approach, its history, conceptual domain, and application to conservation. Metapopulation Biology: Ecology, Genetics, Evolution (eds I. Hanski & M.E. Gilpin), pp. 5-26. Academic Press, San Diego.

Harrison, S. & Taylor, A.D. (1997). Empirical evidence for metapopulation dynamics. Metapopulation biology: Ecology, Genetics, Evolution (eds I. Hanski & M.E. Gilpin), pp. 27-42. Academic Press, San Diego.

Ibrahim, K.M., Nichols, R.A., & Hewitt, G.M. (1996) Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. Heredity, 77, 282-291.

Ives, A.R. & Whitlock, M.C. (2002) Inbreeding and metapopulations. Science, 295, 454-455.

Levins, R. (1969) Some demographic and genetic consequences of environmental heterogeneity for biological control. Bulletin of the Entomological Society of America, 15, 237-240.

Levins, R. (1970). Extinction. Some mathematical questions in biology (ed M. Gerstenhaber), pp. 77-107. American Mathematical Society, Prividence.

Manicacci, D., Atlan, A., & Couvet, D. (1997) Spatial structure of nuclear factors involved in sex determination in the gynodioecious Thymus vulgaris L. Journal of Evolutionary Biology, 10, 889-907.

McCauley, D.E., Olson, M.S., Emery, S.N., & Taylor, D.R. (2000) Population structure influences sex ratio evolution in a gynodioecious plant. American Naturalist, 155, 814-819.

Olivieri, I., Gouyon, P.H., & Gilpin, M.E. (1997). Evolution of migration rate and other traits: The metapopulation effect. Metapopulation Biology: Ecology, Genetics, and Evolution (ed I. Hanski), pp. 293-323. Academic Press, San Diego.

Pannell, J.R. (2001) A hypothesis for the evolution of androdioecy: the joint influence of reproductive assurance and local mate competition in a metapopulation. Evolutionary Ecology, 14, 195-211.

Pannell, J.R. (2002) The evolution and maintenance of androdioecy. Annual Review of Ecology and Systematics, 33, 397-425.

Pannell, J.R. & Charlesworth, B. (1999) Neutral genetic diversity in a metapopulation with recurrent local extinction and recolonization. Evolution, 53, 664-676.
Pannell, J.R. & Charlesworth, B. (2000) Effects of metapopulation processes on measures of genetic diversity. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences, 355, 1851-1864.
Richards, C.M. (2000) Inbreeding depression and genetic rescue in a plant metapopulation. American Naturalist, 155, 383-394.
Richards, C.M., Church, S., & McCauley, D.E. (1999) The influence of population size and isolation on gene flow by pollen in Silene alba. Evolution, 53, 63-73.
Ronce, O. & Olivieri, I. (2003). What makes life in a metapopulation so special? Metapopulation Biology (eds I. Hanski & O.E. Gaggiotti). Academic Press, San Diego.
Slatkin, M. (1977) Gene flow and genetic drift in a species subject to frequent local extinction. Theoretical Population Biology, 12, 253-262.
Taylor, D.R., Trimble, S., & McCauley, D.E. (1999) Ecological genetics of gynodioecy in Silene vulgaris: Relative fitness of females and hermaphrodites during the colonization process. Evolution, 53, 745-751.
van Damme, J.M.M. (1986) Gynodioecy in Plantago lanceolata L. IV. Frequencies and spatial distribution of nuclear and cytoplasmic genes. Heredity, 56, 355-364.
Wade, M.J. & McCauley, D.E. (1988) Extinction and recolonization: their effects on the genetic differentiation of local populations. Evolution, 42, 995-1005.
Whitlock, M.C. (2002) Selection, load, and inbreeding depression in a large metapopulation. Genetics, 160, 1191-1202.
Whitlock, M.C. & McCauley, D.E. (1990) Some population genetic consequences of colony formation and extinction: genetic correlations within founding groups. Evolution, 44, 1717-1724.
Wright, S. (1940) Breeding structure of populations in relation to speciation. American Naturalist, 74, 232-248.
Wright, S. (1951) The genetical structure of populations. Annals of Eugenics, 15, 323-354.

8.15 IN PRESS: "POLYPLOIDY AND THE SEXUAL SYSTEM: WHAT CAN WE LEARN FROM *MERCURIALIS ANNUA*?"

[The following co-authored paper appears in the *Biological Journal of the Linnean Society*, vol. 82, p547-560, August 2004. Here it is presented in manuscript form, with headings re-numbered to conform to the style of this thesis. References are listed after the text.]

POLYPLOIDY AND THE SEXUAL SYSTEM: WHAT CAN WE LEARN FROM MERCURIALIS ANNUA?

John R. Pannell, Darren J. Obbard & Richard J.A. Buggs
Department of Plant Sciences
University of Oxford
South Parks Road
Oxford, OX1 3RB, UK.

Corresponding author: john.pannell@plants.ox.ac.uk

8.15.1 Abstract

The evolutionary success of polyploidy most directly requires the ability of polyploid individuals to reproduce and transmit their genes to subsequent generations. As a result, the sexual system (i.e., the breeding system and the sex allocation of a species) will necessarily play a key role in determining the fate of a new polyploid lineage. The effects of the sexual system on the evolution of polyploidy are complex and interactive. They include both aspects of the genetic system, the genetic load maintained in a population, and the ecological context in which selection takes place. Here, we explore these complexities and review the empirical evidence for several potentially important genetic and ecological interactions between ploidy and the sexual system in plants. We place particular emphasis on work in our lab on the European

annual plant *Mercurialis annua*, which offers promising scope for detailed investigations on this topic. *M. annua* forms a polyploid complex that varies in its sexual system from dioecy (separate sexes) through androdioecy (males and hermaphrodites) to functional hermaphroditism.

8.15.2 Introduction

For many species, the most important factor affecting the initial spread, and evolutionary maintenance, of a polyploid lineage is its sexual system. This is because it determines both the transmission of genes from one generation to the next and the genetic architecture upon which natural selection acts. The sexual system encompasses those aspects of a species' biology that regulate (1) the allocation of resources to male, female and ancillary functions such as pollinator attraction and reward (the sex allocation), and (2) the rates of self-fertilisation and outcrossing in the population (the mating system) (Barrett, 2002).

Phenotypic models have shown that the invasion of a population by sex-allocation or mating-system modifiers depends on aspects of pollination biology and the relative fitness of selfed versus outcrossed progeny, often irrespective of how the phenotypic characteristics are determined genetically (Lloyd, 1975. 1983; Charnov, 1982; Zhang, 2000). The initial spread of such modifiers should therefore not depend on a species' ploidy per se. In contrast, the way the mating system evolves subsequent to the initial spread of a modifier may indeed depend rather strongly on the underlying genetics (e.g., Charlesworth & Charlesworth, 1978; Lande & Schemske, 1985; Lande *et al*., 1994; Schultz, 1999) and thus may differ under different ploidy backgrounds. It is thus important to distinguish between the evolution of a new or modified sexual system in a population, and its subsequent maintenance through time.

A principle aim of this paper is to explore the complex interactions we expect to find between the evolution of polyploid lineages and their sexual systems and to review relevant studies of that bear on this issue. We distinguish between the direct effects of polyploidisation on the reproductive system, and the subsequent evolution of the sexual and genetic systems once polyploid populations have become established. The former effects can generally only be studied through analysis of very recently polyploidised lineages in nature, or of artificially induced polyploids (Ramsey & Schemske, 2002), whilst the latter may be addressed through replicated comparisons of

established polyploid lineages with their diploid progenitors. For clarity of discussion, we also discriminate between the ecological and genetic interactions between polyploidy and the sexual system. Although these two aspects are interrelated, it is useful to distinguish the evolution of the genetic system (including the genetic load) under polyploidy from the often indirect effects of genome duplication on the sexual system through changes in a species' ecological context.

Our other major aim is to summarise past and ongoing work on the *Mercurialis annua* species complex, which displays unusually broad variation in both its sexual system and its ploidy (Durand & Durand, 1992). Whilst a taxonomically broad comparative analysis of ploidy and the mating system may one day prove revealing (but see Charlesworth, 2001), the sort of data required for such studies is still very limited. A great deal therefore remains to be learnt from detailed analysis of individual species or genera, such as *Mercurialis*, within which appropriate variation is displayed.

8.15.3 *Mercurialis annua* as a model system

Mercurialis annua L. (Euphorbiaceae) is a wind-pollinated annual plant that occupies ruderal and roadside habitat throughout central and western Europe and around the Mediterranean Basin (Tutin *et al*., 1964). It is naturalised in North America, the Caribbean, South Africa, and Japan (Durand, 1963). In mesic climates, plants flower throughout the year, but the species is a winter annual in the Mediterranean region. Primary seed dispersal is ballistic, with secondary dispersal by ants (Lisci & Pacini, 1997), although seeds are also doubtless moved in soil by humans and may be blown substantial distances by wind (pers. obs.).

There is a long history of research on *M. annua* that dates back to early investigations of sex determination and sex expression in dioecious populations by Heyer (1884), Yampolsky (1919; 1930) and Gabe (1939), and the species has continued to be a model for studies of sex determination in plants (Delaigue *et al*. 1984; Durand *et al*. 1987; Durand and Durand 1991; Pannell 1997b). A firm foundation for research on the evolution of ploidy and the sexual system was established by Durand (1963) in a biosystematic study of the species complex.

The genus *Mercurialis* comprises seven to ten European species and one Asian species; all except the *M. annua* group are rhizomaceous or woody perennials (Tutin *et al*., 1964; Krahenbuhl *et al*., 2002;). On the basis of morphology, ploidy and sexual

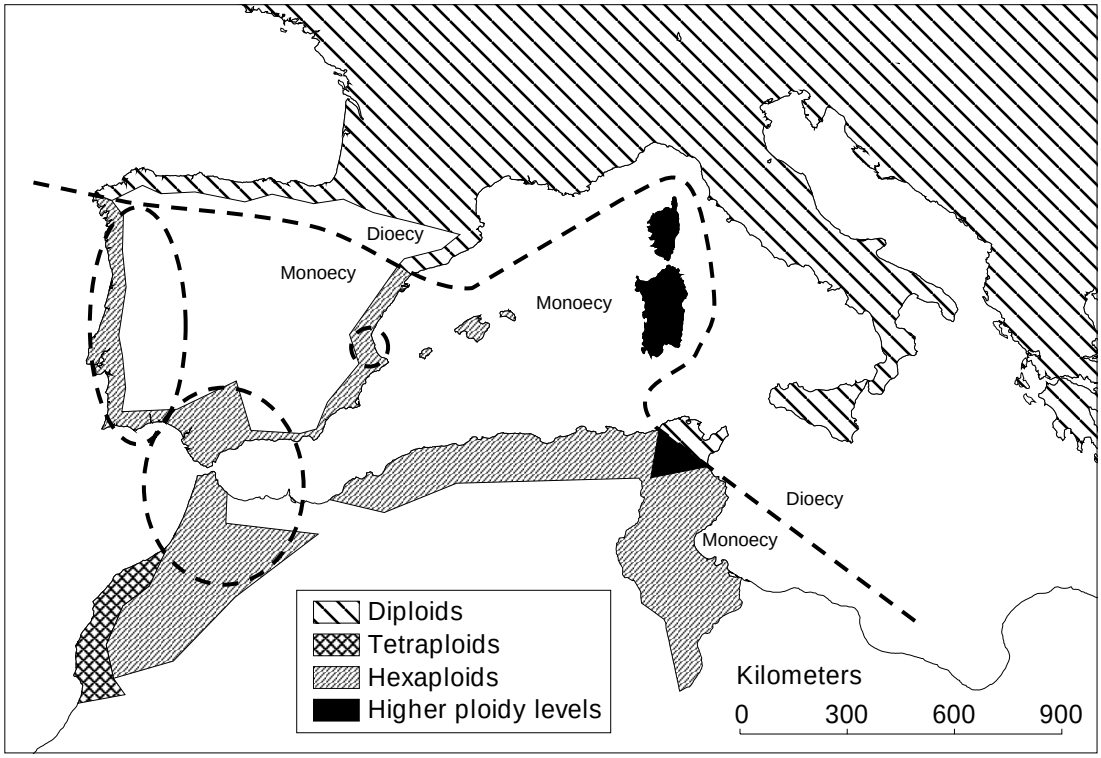
system, Durand (1963) recognised four annual species within this group: *M. huetii* Hanry, *M. annua* L., *M. ambigua* L. fil. and *M. monoica* (Moris) Durand. Of these, *M. huetii* is quite distinct whilst it is difficult to identify diagnostic morphological characters amongst the (Durand &Durand, 1985). Given the reported monophyly of the clade into which *M. annua*, *M. ambigua* and *M. monoica* fall (Krahenbuhl *et al*., 2002) and the difficulty in species delimitation between them, we refer to these three putative taxa together as *M. annua* in the broad sense.

Diploid populations of *M. annua* (2n = 16) range from Israel throughout central and western Europe into southern France and northern Spain; these populations are uniformly dioecious (Durand, 1963, and Figure 1). In northern Spain, diploid populations give way to hexaploid populations (2n = 48) across two abrupt transitions on the Mediterranean and Atlantic coasts. In Catalonia in the east, this transition coincides with a shift from dioecy to self-compatible monoecy, while in Galicia in the west, dioecy gives way to androdioecy (i.e., the co-occurrence of males with functional hermaphrodites). South of these transitions, hexaploid *M. annua* populations are variously monoecious or androdioecious (Durand, 1963, and Figure 1). In northern Morocco, around Fes and Meknes, hexaploid populations are sub-dioecious, with males and females co-occurring with female-biased monoecious plants. Tetraploid populations of *M. annua* (2n = 32) are found on the coast in central western Morocco, where individuals are generally monoecious, and higher ploidy levels (up to 12x) are found further east in north Africa and on the Mediterranean islands of Corsica and Sardinia; these populations are all monoecious (Durand, 1963, and Figure 1).

In the *M. annua* species complex, there is evidently a correspondence between polyploidy and monoecy. This raises the question of whether monoecy is a direct consequence of polyploidisation, whether polyploidy has only been able to arise in a selfing lineage, or whether polyploidy has allowed selection on the sexual system to favour monoecy subsequent to genome duplication. Although ploidy and the sexual system are broadly confounded in *M. annua*, variation in the sexual system among hexaploid populations in particular offers an opportunity to conduct replicated comparisons between sexual systems in isolation of ploidy effects.

In short, *M. annua* presents valuable material with which to address questions concerning the establishment and evolution of polyploid races with contrasting sexual systems. Its notable features can be summarised as: (1) broad correspondence between

Figure 1: The distribution of different ploidy levels in the *Mercurialis annua* complex around the Mediterranean Basin.



The dashed ellipses designate regions in which androdioecious populations are found. Map modified after Durand (1963).

polyploidy and monoecy, derived from dioecy; (2) within-ploidy variation in the sexual system amongst hexaploid populations; and (3) two effectively independent diploid-hexaploid contact zones in northern Spain. Being a fast-growing annual, the plant is easy to manipulate under controlled conditions, and natural populations are accessible and highly abundant in the field. Artificial neo-polyploids can be produced under laboratory conditions (Durand, 1963), monoecious plants can be readily self-fertilised by isolating them in pollen-proof growth boxes (Pannell, 1997c), and the self-fertilisation of unisexual plants is possible by altering their gender expression through the exogenous application of phytohormones (Durand & Durand, 1991).

8.15.4 Hypothesis for the evolution of the sexual system in *M. annua*

Given the universal occurrence of dioecy in the other species of *Mercurialis*, and particularly in diploid *M. annua*, there is little doubt that monoecy in the polyploid populations is a derived trait (Krahenbuhl *et al*., 2002). Whether androdioecy and subdioecy are derived from dioecy or secondarily from monoecy is not yet known, but it seems clear that the genetic basis for male and female floral and inflorescence development has been conserved throughout the species complex and genus. Males in diploid and hexaploid populations uniformly disperse their pollen from staminate flowers held on erect pedunculate inflorescences, while pistillate flowers are typically borne on subsessile axillary pedicles. In monoecious plants, staminate flowers usually cluster around a single pistillate flower in each leaf axil, although we have occasionally found monoecious hexaploid populations that additionally have pedunculate inflorescences similar to those of males (see also Durand, 1963). It therefore seems most likely that monoecious individuals of *M. annua* are effectively modified females. The same conclusion has been reached for the hermaphrodites of other plants and animals in which androdioecy is derived from dioecy (reviewed in Pannell, 2002).

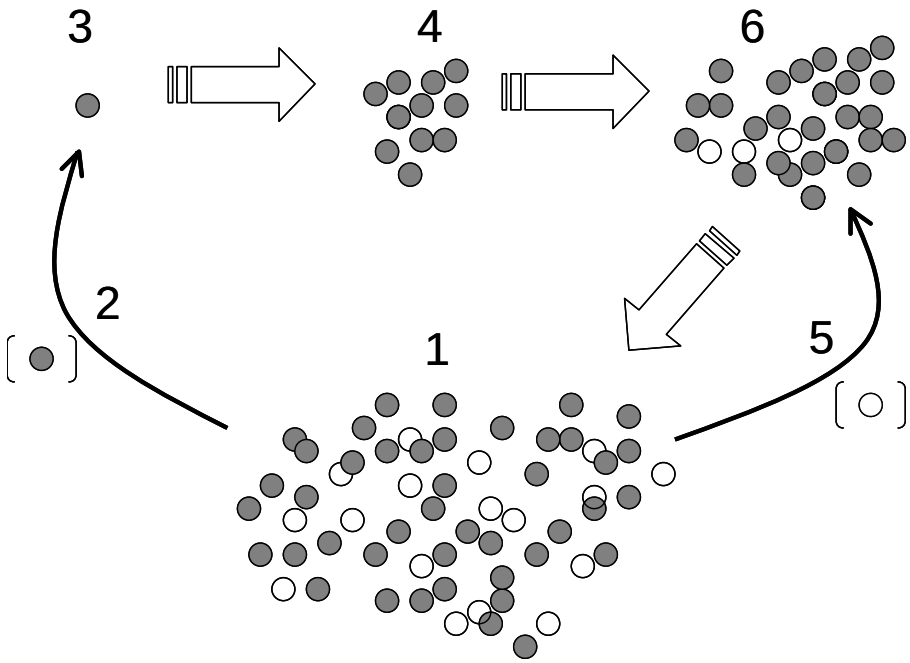
Androdioecy is a particularly rare sexual system in plants, and its evolution and maintenance was until recently difficult to explain (Charlesworth, 1984; Pannell, 2002). A key prediction made by theoretical models is that males must enjoy at least twice the siring success of hermaphrodites. This condition is most likely to be met if male pollen production is high relative to that of hermaphrodites, and if the hermaphrodite selfing rate is low (Lloyd, 1975; Charlesworth & Charlesworth, 1978;

Charlesworth, 1984). Pollen production in *M. annua* is between four and ten times higher in males than in monoecious individuals (Pannell, 1997b; Pannell, 1997c), but the species is self-compatible (Durand, 1963), and patterns of sex allocation in monoecious populations are consistent with a history of selection under repeated bouts of inbreeding (Pannell, unpubl. ms). Similarly, other androdioecious plants and animals are able to self and may experience selfing rates as high as unity in the absence of males (Pannell, 2002). This would seem to be inconsistent with the prediction of high outcrossing rates for the evolution and maintenance of males.

Studies of variation in sex allocation in *M. annua* conducted in the mid 1990's (Pannell, 2002) inspired a reappraisal of theoretical models for androdioecy. In particular, apart from the importance of recognising dioecy rather than hermaphroditism as the ancestral trait (see also Fritsch & Rieseberg, 1992), the population structure and demography of a ruderal weed such as *M. annua* suggest selection on the sexual system at both the population and the metapopulation levels (Pannell, 2001; Pannell, unpubl. ms). The most likely hypothesis for the maintenance of androdioecy in *M. annua*, and indeed in several other species (reviewed in Pannell, 2002), is that functional self-compatible hermaphroditism evolved from dioecy and is maintained by selection for reproductive assurance at the metapopulation level as a result of repeated bouts of mate limitation during colonisation (Pannell, 2001, and Figure 3). These functionally hermaphroditic populations will be female-biased in their sex allocation as a result of selection under self-fertilisation (Hamilton, 1967; Lloyd, 1987). With local demographic growth and a concomitant reduction in the selfing rate following colonisation, populations become susceptible to the invasion and spread of males, which can be maintained at the metapopulation level if gene flow amongst populations is sufficiently high and local extinction rates are low (Figure 2). The observed metapopulation structure, and the high among-population variation in male frequencies in several androdioecious species, appear to conform with this model.

The above scenario invokes differences in demography and population structure between dioecious, androdioecious and monoecious regions occupied by *M. annua*, with more ephemeral and genetically isolated populations found toward the monoecious end of this gradient. The evolution of monoecy (and androdioecy) from dioecy also requires a shift in the reproductive and/or demographic parameters that regulate the stability of the sexual system. Because the transition from dioecy to monoecy coincides with polyploidisation in *M. annua*, the effects of polyploidy need

Figure 2: A graphical depiction of a model for the maintenance of androdioecy in a metapopulation



A graphical depiction of a model for the maintenance of androdioecy in a metapopulation. Males and hermaphrodites are symbolised by open and closed circles, respectively; block arrows represent population growth; simple arrows represent dispersal. According to the model, established androdioecious populations (1) disperse male and hermaphrodite propagules across the metapopulation (2 and 5). Only hermaphrodites can establish populations as sole colonisers due to reproductive assurance (2). New populations (3) and small and sparse and are initially highly selfing (4). As populations grow, they become denser, outcrossing with neighbours becomes easier and density-dependent selfing rates decline; the increased levels of outcrossing allow males to invade, because they have greater access to ovules (5). The age structure of the metapopulation is maintained through a balance of extinction and recolonisation. Inbreeding during repeated colonisation events selects for female-biased sex allocation in the hermaphrodites as a result of local mate competition. From Pannell (2003).

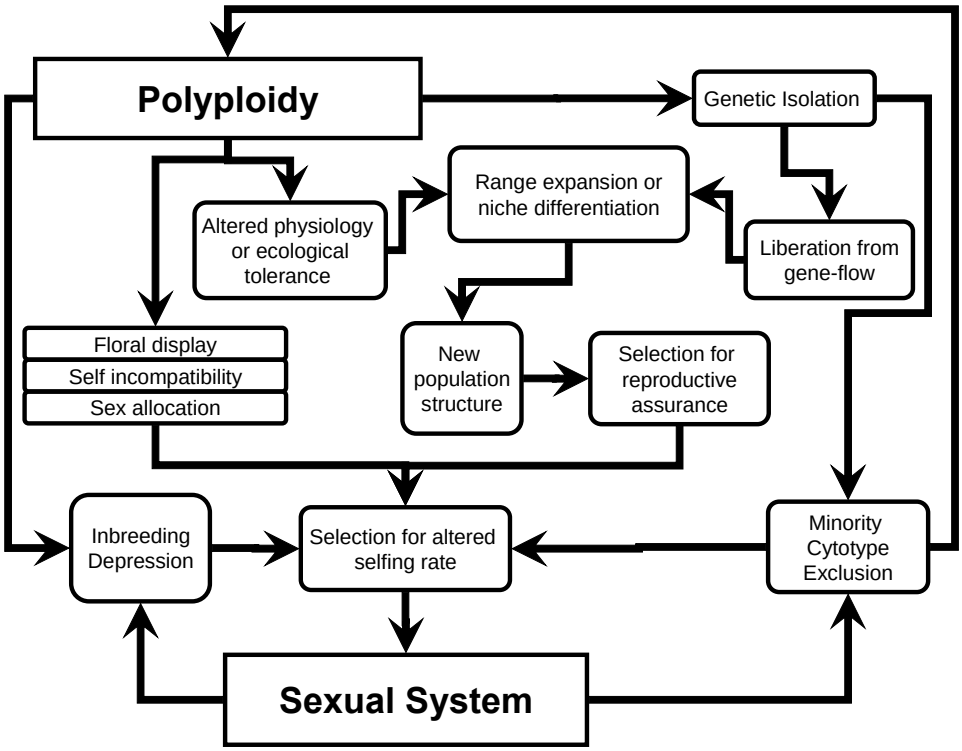
to be integrated into hypotheses regarding both the evolution of sexual systems in *M. annua* and their longer-term maintenance. Such an integrated framework may be built around quite general considerations of the interactions between polyploidy and the sexual system.

8.15.5 Polyploidy and the sexual system

Some of the expected genetic and ecological interactions brought about by polyploidisation, and their direct and indirect effects on the sexual system, are depicted in Figure 3. A useful point of departure in interpreting this figure is the recognition that transitions in the sexual system depend on changes in the context in which it is selected. From the genetic point of view, chromosome doubling may directly reduce inbreeding depression in neopolyploids, and, under potentially modified selfing rates, polyploidy may modify the total genetic load maintained in the population in the longer term (Lande & Schemske, 1985; Ronfort, 1999). The genetic upheaval associated with polyploidisation might also be expected to interfere with processes in which relative allele dosage is important, such as self-incompatibility mechanisms in hermaphrodites (Lewis, 1960; Stone, 2002) and sex determination in dioecious species (Westergaard, 1958). From an ecological perspective, the expansion of a species' range, or its invasion into new habitat following polyploidisation, may reduce the availability of prospective mating partners or pollinators, thus giving rise to selection for reproductive assurance and the evolution of self-fertilisation (Baker, 1955; Pannell & Barrett, 1998; and see Brochmann *et al*., 2004, for empirical examples). Selfers may be more likely to create polyploid lineages by increasing the chances of union between unreduced gametes, and selfing polyploid lineages are also likely to establish more easily than out-crossing ones as they avoid fertilisation by diploid progenitors and its associated fitness costs.

Along with phenotypic variation or abrupt shifts in traits associated with the sexual system, which may be caused by direct genetic effects of polyploidisation, inbreeding depression and reproductive assurance probably represent the two most fundamental factors affecting the evolutionary stability of the plant mating systems in general (Barrett & Harder, 1996; Holsinger, 1991; Morgan & Schoen, 1997). We address each of these three issues in the context of polyploidy below. We then consider the ways in

Figure 3. Hypothesised interactions between polyploidy and the sexual system.



Polyploidisation can have direct genetic consequences for the mating-system (e.g. floral display, sex expression) which may be selected for. It can also change the genetic and ecological context in which selection acts (e.g., altered levels of inbreeding depression, and possibilities for range expansion due to altered ecophysiology and genetic isolation). Conversely, since self-fertilisation increases the probability of fusion of unreduced gametes and helps to avoid the cost minority cytotype exclusion, the sexual system can affect the likelihood of polyploid origin and establishment (dotted line). The sexual system feeds back on intermediate steps, affecting levels of inbreeding depression and the population structure.

which correlated changes in the ecology and ecophysiology of polyploids may affect the distribution of their sexual systems, and we assess the influence on cytotype distribution of ecological and genetic interactions across ploidy contact zones.

8.15.5.1 Disruption of self-incompatibility and sex determination

It is well established in that genome duplication can directly cause gametophytic self-incompatibility (SI) systems to break down (Lewis, 1960; Stone, 2002), because pollen grains carry multiple alleles at the SI locus (Golz *et al*., 1999). There is also strong evidence that such self-compatible polyploid lineages persist in some families (Ross, 1981). Although the precise mechanism of this loss of function is not fully understood, it has obvious immediate implications for the evolution of the mating system. In a compelling interpretation of the association between polyploidy and dioecy in otherwise self-incompatible (SI) hermaphroditic lineages, Miller and Venable (2000) have suggested that separate sexes have often evolved as a response to selection for outcrossing following the disruption of SI mechanisms through polyploidisation.

The breakdown of sex determination in dioecious lineages has similarly important implications for the mating system if this leads to the expression of both sexes in the same plant. It was an early prediction that such a breakdown would occur (e.g., Muller, 1925), and there is some limited evidence for this in neo-polyploids. For example, artificially induced autopolyploidy in dioecious *Silene* and *Rumex* gave rise to potentially selfing hermaphrodite progeny (reviewed in Westergaard, 1958). Similarly, Durand (1963) found hermaphrodite morphs in the F2 progeny of artificial autotetraploids generated from diploid dioecious individuals of *Mercurialis annua*, with phenotypes similar to those found in natural polyploid populations. Nevertheless, despite general claims for an association between polyploidy and hermaphroditism in otherwise dioecious groups (e.g., Richards, 1997), the only examples from natural populations appear to be *M. annua* and tetraploid *Empetrum hermaphroditum*, derived from dioecious diploid *E. nigrum* (cited in Richards, 1997). More common is the reverse association cited above between polyploidy and dioecy in lineages that were formerly SI hermaphrodites (Miller & Venable, 2000). It is quite possible that dioecy is disrupted by polyploidisation and that subsequent selection on the sexual system re-establishes it over time. This is of course suggested in *M. annua*, where androdioecy

and subdioecy may have evolved from monoecy in hexaploid populations following the earlier breakdown of dioecy. However, as we have hypothesised, there are reasons to believe that monoecy is in fact selectively maintained in *M. annua* over large areas of the species' range.

8.15.5.2 Inbreeding depression

Inbreeding depression denotes the reduced fitness of self-fertilised progeny relative to the fitness of their outcrossed counterparts (Charlesworth & Charlesworth, 1987). In its absence, and under the often realistic assumption that self-fertilisation uses negligible amounts of pollen, Fisher (1941) first pointed out that an outcrossing hermaphroditic population is susceptible to the invasion of self-fertilising mutants because the latter transmit an extra copy of their genome through their seed progeny. This automatic transmission advantage to self-fertilisation is counteracted by levels of inbreeding depression that exceed 0.5 (Charlesworth & Charlesworth, 1987). Populations with a sufficiently high value of inbreeding depression are thus predicted to be maintained as outcrossers unless circumstances cause inbreeding depression to drop. Under continued diploidy, this can only occur if the genetic causes of inbreeding depression are selectively lost, or 'purged', from the population (Barrett & Charlesworth, 1991; Lande & Schemske, 1985). Purging is more efficient under selfing, which causes elevated homozygosity and thus increases the opportunity for the selective loss of deleterious recessive alleles (Crnokrak & Barrett, 2002). Importantly, this process of purging implies that the mating system is stable at either the completely outcrossing or the completely selfing extremes (Lande & Schemske, 1985). On the one hand, selection in selfing populations will continue to remove deleterious genetic load as it arises through mutation, thus maintaining inbreeding depression at low levels. On the other hand, deleterious alleles can continue to accumulate at multiple loci in large outcrossing populations, because they seldom find themselves in a homozygous state when at low frequency (Lande & Schemske, 1985).

Given the importance of inbreeding depression for mating-system evolution and stability, our understanding of the interaction between polyploidy and the sexual system must depend strongly on the fitness effects of inbreeding in lineages experiencing the effects of extensive gene duplication and (potentially) altered patterns of inheritance. It is often claimed that deleterious recessive alleles will be more

effectively hidden from selection in polyploids because duplication results in a smaller proportion of homozygotes after a single round of selfing (e.g., Richards, 1997, p. 382). Indeed, this idea has sometimes been invoked to explain an association between selfing and polyploidy (e.g., Barrett & Shore, 1987). However, the idea of a widespread association between polyploidy and selfing, at least in angiosperms, has been questioned (Ronfort, 1999), and there has been remarkably little research on the subject to substantiate or refute it. Galloway *et al*. (2003) have recently drawn attention to a possible predominance of outcrossing in autopolyploids and of selfing in allopolyploids. If this is so, then it may reflect compromise of morphological outcrossing mechanisms in allopolyploids, and thus a shift to selfing. (This would be analogous to reduced outcrossing upon the breakdown of dioecy or self-incompatibility in polyploids generally.) Alternatively, selfing may be more easily maintained in allopolyploids than autopolyploids as a result of lower inbreeding depression in the former. The few existing theoretical studies concern themselves only with autopolyploids and have only dealt with populations at mutation-selection equilibrium (Bennett, 1976; Lande & Schemske, 1985; Ronfort, 1999), and these provide little consensus on the fundamental question of whether equilibrium values of inbreeding depression should be greater or smaller in tetraploids than diploids.

Of the theoretical work, Lande and Schemske's (1985) conclusion that tetraploids should maintain between half (under complete recessivity) and the same (for partially recessive alleles) inbreeding depression as diploids, has received most attention. However, other workers have reached different conclusions (e.g., Bennett, 1976). In particular, Ronfort (1999) has provided a comprehensive treatment of these issues in autotetraploids and has highlighted the complexities involved. As in the diploid case, predictions depend on such factors as the mutation rate to deleterious alleles, their selection coefficients, the degree of dominance between alleles, and epistatic interactions amongst loci (Ronfort, 1999). For polyploids, the increased number of potential dominance interactions between alleles complicates the expectations substantially. In the special case of complete recessivity, equilibrium inbreeding depression in diploids and autotetraploids is expected be equal. Under partial dominance, however, several of the scenarios investigated by Ronfort (1999) predicted higher inbreeding depression in diploids than autotetraploids, while others predicted the reverse. It appears that without knowing the selection and dominance coefficients associated with genotypes carrying alleles at different dosages (e.g., *AAaa* versus

AAAa), it does not seem possible to make very general predictions about equilibrium values of inbreeding depression, and thus about the evolutionarily stable selfing rate (Ronfort, 1999). It is worth noting that the approach taken in all of these studies requires selection-mutation equilibrium, and it seems unlikely that a neo-polyploid will fulfil this criterion. Indeed, depending on the rate of equilibration relative to the process of diploidisation, selection-mutation equilibrium may never be met in polyploid lineages with polysomic inheritance.

As far as empirical work is concerned, the few relevant comparisons of inbreeding depression between polyploid lineages and their putative diploid progenitors have yielded mixed results. Early work was based largely on polyploid forage and crop plants (e.g., Busbice & Wilsie, 1966; Dewey, 1966). These studies reported increased levels of inbreeding depression in polyploids relative to their diploid counterparts. However, because product yield (rather than fitness) was measured, and because the recent history of the respective polyploid line was not always clear, interpretation of these results is difficult. Of the small number of studies examining inbreeding depression in natural polyploid populations (whether in relation to selfing rates or not), most have not attempted a comparison with their diploid relatives (e.g., Inoue *et al*., 1998; Dudash & Fenster, 2001; Galloway *et al*., 2003). Only Husband and Schemske (1997) and Rosquist (2001) have conducted diploid-polyploid comparisons of inbreeding depression, and both studies found it to be lower in the polyploid lines. We have been unable to find any empirical studies of inbreeding depression that take into account both variation in the selfing rate and the ploidy level. Again, the most convincing, albeit indirect, evidence for the importance of inbreeding depression as a mediating influence in the interaction between polyploidy and the sexual system is the association between polyploidy and dioecy in lineages derived from SI hermaphrodites (Miller & Venable, 2000; Charlesworth, 2001).

There is clearly a great need for further theoretical and empirical work on inbreeding depression in polyploid-diploid comparisons. Here, it will be important to distinguish between inbreeding depression in neo-polyploids, which will influence the evolution of the sexual system immediately following genome duplication, and inbreeding depression at mutation-selection equilibrium following any possible purging. It would thus be particularly instructive to conduct such comparisons between diploid lineages and established polyploids as well as artificially induced polyploid lineages derived from the same putative progenitors.

As yet, there are no estimates of inbreeding depression for *M. annua* populations, but data from this species would be useful because of the existence of both inbred (monoecious) and outbred (androdioecious or subdioecious) polyploid populations in addition to the outbred diploids. Certainly, until we have good estimates of inbreeding depression in the various genetic backgrounds of *M. annua*, it will remain difficult to evaluate the importance that this factor may have had in the evolution and continued maintenance of self-fertile monoecy in the polyploid populations of the species.

8.15.5.3 Reproductive assurance

We have hypothesised that self-fertile monoecy is maintained in M. annua through selection for reproductive assurance under demographic conditions where among-population gene flow is limited and population size fluctuations associated with metapopulation dynamics are more severe (Pannell, 1997a; Pannell, 2000, and Figure 3). Within the distribution of hexaploid M. annua in the Iberian Peninsula, clines between androdioecious and monoecious zones repeatedly coincide with a clear transition in the population structure of the species. In particular, monoecious populations are typically small and geographically isolated from one another, while populations in the corresponding androdioecious zone are more common, less geographically isolated, and orders of magnitude larger (S. M. Eppley and J. R. Pannell, unpubl. data). This trend is also reflected in patterns of both morphological and isozyme diversity, which seem to be higher in androdioecious populations and low in monoecious regions (D. J. Obbard, S. A. Harris and J. R. Pannell, unpubl. data). Higher historic levels of inbreeding in monoecious than in androdioecious regions, estimated indirectly on the basis of patterns in sex allocation (Pannell, unpubl. ms), provide further support for the metapopulation model in M. annua.

Although direct estimates of density-dependent selfing rates and rates of population turnover remain outstanding, there are good indications that metapopulation structure and dynamics do differ between monoecious and androdioecious zones along predicted lines. Because these comparisons can be made independently of ploidy differences, they suggest that selection on the sexual system itself is also likely to have contributed to the maintenance of differences in the sexual system across the two diploid-hexaploid transitions. Nonetheless, it is important to bear in mind that other genetic and ecological differences between ploidy levels may have played, and continue to

play, a substantial role in shaping the geographic distribution and population structures of the two cytotypes. To what factors, for example, can we attribute the fragmented population structure observed in hexaploid relative to diploid zones on either side of the Catalonian contact?

We have already considered the possible importance of inbreeding depression in allowing the evolution of self-fertile hermaphroditism. This, on its own, may have contributed to the ability of self-fertilising monoecious colonisers to expand their range into new or inhospitable environments where small population sizes and mate limitation are prevalent. Such a scenario may also account for the central European distribution of the self-fertilising tetraploid cytotype of the fern *Asplenium* sp., whose more outcrossing diploid progenitors are confined to the more southerly Mediterranean Basin (Vogel *et al*., 1999). It is plausible that high inbreeding depression in diploid populations of *Asplenium* has prevented the evolution of self-fertilisation and thus also the potential for range expansion (Vogel *et al*., 1999). As noted earlier, however, our understanding of inbreeding depression in polyploids remains poor.

Two other issues impinge upon the contrasting population structures and distributional ranges displayed by M. annua diploids and hexaploids. The first concerns possible ecophysiological changes directly associated with polyploidy that may have allowed the species to invade a new environment, irrespective of its sexual system (c.f., Ramsey & Schemske, 2002). The second relates to the post-Pleistocene colonisation history of the respective cytotypes and the possible ongoing genetic and ecological interactions between them across zones of primary or secondary contact (Durand, 1963).

8.15.5.4 Ecophysiological niche differentiation

It has long been thought that polyploidisation may have allowed species to expand their range into novel environments as a result of derived ecophysiological differences relative to diploid progenitors (reviewed in Levin, 2002; Ramsey & Schemske, 2002). Again, it is important here to distinguish between neopolyploids, in which polyploidisation directly causes physiological alterations, and long established polyploids, in which changes may be the result of subsequent adaptive evolution within the new environment that was occupied for other reasons (Ramsey & Schemske, 2002).

The increased DNA content brought about by polyploidisation may lead directly to increased cell size, reduced stomatal density, and to slowed cell division (see Bennett, 1987). These changes may then influence ecophysiological traits such as transpiration and gas exchange (summarised in Levin, 2002), and life history traits such as generation time (Bennett, 1987). Increased gene dosage and disruption of the control of gene expression may cause changes in the level and pattern of gene expression (Guo *et al*., 1996; Comai *et al*., 2000; Wendel, 2000; Osborn, Pires *et al*., 2003). It has also been widely argued that increased levels of heterozygosity may be an important source of increased vigour, ecological tolerance, and thus colonising ability (Bingham, 1980; Stebbins, 1980, 1985; Tomekpe & Lumaret, 1991; Brochmann & Elven, 1992). In the case of allopolyploids, genome duplication has been claimed to represent "a permanent combination of adaptive strategies and ecophysiological capabilities inherited from different diploid progenitor species" (Brochmann & Elven, 1992) which thus allows species range expansions (see also Ainouche *et al*., 2004; Soltis *et al*., 2004). It should be noted that there is little theoretical or empirical evidence to support the notion that increased heterozygosity and gene duplication leads to such increases in fitness.

In *M. annua*, the transition from diploidy to hexaploidy in Catalonia occurs across a gradient of increasing aridity, pointing to possible ecophysiological differences between the cytotypes in this region. However, preliminary results from reciprocal transplant experiments have so far provided no evidence for the hypothesis of ecophysiological local adaptation across the zone of contact between diploid dioecious and hexaploid monoecious population (R. J. A. Buggs and J. R. Pannell, unpubl. data). More generally, gradients in increasing aridity on the Iberian Peninsula and in Morocco repeatedly correspond to transitions from dioecy or androdioecy to monoecy (i.e., from separate to combined sexes). It thus seems possible that transitions in ploidy correlate with shifts in breeding system as a result of selection for reproductive assurance in metapopulations with increasing isolation amongst demes and higher extinction rates (Pannell, 1997a). This would explain observations of increased reproductive failure of females due to mate limitation in the region occupied by monoecious populations in Catalonia (R. J. A. Buggs and J. R. Pannell, unpubl. data).

8.15.5.5 Interaction across zones of contact

The expansion of species ranges generally is thought to be limited by a balance between natural selection, bringing about local adaptation at the range boundary, and gene flow from more central parts of a species range, which impedes local adaptation (Kirkpatrick & Barton, 1997; Lenormand, 2002). The formation of barriers to gene flow between populations may therefore remove this constraint on local adaptation, allowing gene frequencies to change in response to local selection. Polyploidisation may represent an important source of such a barrier if it causes reduced gene flow between diploids at the centre of the range and new polyploid lineages at its extremity.

The sort of genetic isolation between diploids and their polyploid derivatives required by this hypothesis may be either post-zygotic or pre-zygotic, the latter of which may be due to changes in the sexual system. Table 1 lists several examples of pre-zygotic isolation due to changes in the sexual system that have been observed between natural diploid and polyploid populations, including differences in flowering time, pollinator sorting, and shifts from dioecy to monoecy or from sexual reproduction to apomixis. In some cases these changes in the sexual system may be a direct and immediate result of polyploidisation per se, which may contribute to the ability of polyploids to establish under a reduced influence of minority cytotype exclusion (Levin, 1975; Felber, 1991; Husband, 2000; and see below). In other cases, such mechanisms may have evolved gradually as a result of reinforcement selection to avoid the negative consequences of inter-ploidy fertilisations. Although the conditions under which reinforcement may occur are limited (Barton & Hewitt, 1985), in the context of polyploidy it seems most likely to be found in secondary contacts which have a mosaic of patches of differing cytotypes, exposing a high proportion of individuals to selection (Cain, Andreasen & Howard, 1999).

In the examples listed in Table 1 (see also Petit *et al*., 1999), it is not always possible to distinguish between isolating mechanisms as a direct result of polyploidy and those that have arisen in response to reinforcement. In the case of triploid Taraxacum, the chromosome number is directly responsible for the shift to apomixis (see Table 1 for references). Conversely, in the case of *Dactylis glomerata*, selection seems to have acted subsequent to polyploidisation because flowering-time divergence

Table 1. Pre-zygotic reproductive isolating mechanisms involving the sexual system found in wild diploid-polyploid contacts.

Species	Ploidy levels	Pre-zygotic breeding system isolation mechanism	Inter-ploidy hybrid occurrence	References
Anthoxanthum alpinum	2x, 4x	Flowering time difference	Rare	(FelberGirard, Felber & Buttler, 1996)
Mercurialis annua	2x,6x	Dioecy/monoecy?	Common	(Durand, 1963)
Lotus alpinus/ corniculatus	2x,4x	Flowering time difference	Never found	(Gauthier, Lumaret & Bedecarrats, 1999)
Plantago media	2x,4x	Flowering time difference	Very rare	(Van Dijk *et al*., 1992; Van Dijk & Bijlsma, 1994; Van Dijk & Bakx-Schotman, 1997)
Chamerion angustiifolium	2x,3x,4x	Flowering time difference; pollinator sorting	7%	(Husband & Schemske, 1998; Husband 2004)
Arrhenatherum elatius	2x,4x	Flowering time difference; high selfing rate?	1%	(Petit *et al*., 1997; Petit & Thompson, 1997)
Heuchera grossulariifolia	2x,3x,4x	Pollinator sorting	1.4%	(Segraves & Thompson, 1999; Segraves *et al*., 1999)
Carya ovata/ tormentosa	2x,4x	High selfing rate?	?	(McCarthy & Quinn, 1990)
Taxacarum	2x,3x	Sexual/apomictic	Common	(Menken *et al*., 1995)
Claytonia virginica	2x,4x	Flowering time difference	Common	(Lewis, 1976)
Dactylis glomerata	2x,4x	Flowering time difference	No adult found	(Lumaret *et al*., 1987; Lumaret & Barrientos, 1990; Bretagnolle & Lumaret, 1995)

has been found in natural sympatric diploid-tetraploid populations (Lumaret *et al*., 1987; Lumaret & Barrientos, 1990), but not between diploid and neo-tetraploid lineages (Bretagnolle & Lumaret, 1995).

Post-zygotic isolation as a result of low seed viability appears to be particularly common in diploid-polyploid contact zones. In *Arabidopsis thaliana*, this reduction in viability has been attributed to disrupted endosperm development, caused by an imbalance in the ratio of maternal and paternal genomes (Scott *et al*., 1998). Inter-ploidy hybrids may also display reduced fertility, e.g., due to the high incidence of unpaired chromosomes at meiosis (Sybenga, 1975). In mixed-ploidy populations, both of these processes are expected to give rise to frequency-dependent selection, with a fitness disadvantage to the minority cytotype (Levin, 1975; Felber, 1991; Rodriguez, 1996a, 1996b; Husband, 2000). Minority cytotype exclusion places a constraint on the establishment of new polyploids (but see Husband, 2004), and it may also give rise to hybrid 'tension-zone' dynamics at secondary contacts between populations with different ploidy levels.

Tension zones are boundaries of contact between potentially interbreeding populations whose hybrid progeny suffer reduced fitness (Barton & Gale, 1993; Kruuk *et al*., 1999). A particularly important implication of their dynamics is that the relative distribution of the two populations in contact depends largely on the (density-dependent) rates and direction of gene flow between them, and notably not on ecological differentiation or adaptation to different ecological environments (Barton & Hewitt, 1985). Contact zones between diploid and tetraploid *Plantago media* in the Pyrenees (Van Dijk *et al*., 1992; Van Dijk & Bakx-Schotman, 1997), and *Centaurea jacea* in north-east Belgium (Hardy *et al*., 2000, 2001) appear to be good examples of tension zones. In contrast, in diploid and polyploid populations of *Dactylis glomerata* (Lumaret *et al*., 1987; Lumaret & Barrientos, 1990) and *Claytonia virginica* (Lewis, 1976) tension-zone dynamics are mitigated by reproductive isolation due to flowering phenology. Similarly, sympatry is possible for different ploidy levels of *Heuchera grossulariifolia*, partly as a result of differences in flower phenology and pollinator sorting (Segraves & Thompson, 1999).

The position of tension zones is sensitive to differential rates of gene flow from one cytotype to the other, or to 'anisotropy'. This results in the displacement of the cytotype with poorer dispersal by the better disperser (Barton & Hewitt, 1985). Anisotropy can be due in particular to differences in phenology, the mating system or sex allocation.

The best evidence for asymmetry in pre-zygotic isolation between related cytotypes is provided by populations in a zone of sympatry between diploid and tetraploid *Chamerion angustifolium*, where differences in flowering time, pollinator fidelity and pollen competition mean that pollen flow from tetraploids to diploids is much more common than vice versa (Husband, 2000; Husband & Schemske, 2000; Husband *et al*., 2002). Diploid fitness is thus reduced compared with tetraploids in the contact zone. This may ultimately allow tetraploids to expand their range, although the dynamics in this tension zone appear to be dampened by ecological sorting, perenniality, clonal reproduction and variable triploid fitness (Husband, 2004; Husband *et al*., 2002).

The effect of inter-ploidy differences in sex allocation almost certainly play a role in shaping the contact zones between diploid and hexaploid populations of M. annua in north-eastern and north-western Spain. Because of the much greater pollen productivity of males relative to monoecious plants (Pannell, 1997b), pollen flow out of dioecious populations will be substantially higher than out of monoecious ones. Moreover, the fact that males disperse their pollen from erect pedunculate inflorescences, in contrast to the staminate flowers of monoecious individuals that are subsessile in the leaf axils, will no doubt contribute to the anisotropy due to sex allocation (Pannell, 1997c). The degree of anisotropy in M. annua contact zones, and the extent to which self-fertilisation in polyploid populations may mitigate gene flow from diploid individuals, are areas of active research in our lab. Certainly, low levels of gene flow amongst monoecious populations are suggested by higher levels of population differentiation compared with dioecious populations (R.J.A. Buggs, D.J. Obbard, S.A. Harris and J.R. Pannell, unpubl.data).

The effects on fitness of inter-ploidy hybridisation in M. annua appear to be severe. Diploid and hexaploid M. annua do cross-pollinate in several natural sympatric populations in Catalonia, but the resulting tetraploid progeny are sterile (Durand, 1963, and R. J. A. Buggs and J. R. Pannell, unpubl. data). The relative rarity of mixed-ploidy populations in Catalonia and Galicia, as well as the very narrow zone within which populations of either cytotype are found, support a hypothesis of strong tension-zone dynamics in M. annua. Moreover, the contact zone in Catalonia occurs in an area of locally low rainfall along a relatively narrow coastal strip where M. annua populations are rare, a fact that is consistent with tension-zone models that predict maximum steepness in clines to occur in areas of low population density (Barton & Hewitt, 1985).

8.15.6 Conclusions

Despite the clear importance that the sexual system must have in regulating the success of polyploid lineages, and the manifold effects that polyploidy is likely to have on the sexual system in turn, there is remarkably little firm theoretical or empirical research to substantiate the interactions we have discussed in this article. It is safe to conclude that further work aimed at investigating each of the proposed causal links outlined in Figure 1 would be valuable. The poor theoretical understanding and the limited direct empirical support for the relationship between polyploidy and inbreeding depression, both in neo-polyploids and during the course of subsequent genome evolution, are particularly striking and worthy of redress. This is not only because of the general importance of understanding the maintenance of genetic load in populations of plants and animals, but also because of the fundamental role that inbreeding depression is believed to have in regulating the stability of the mating system. The association between polyploidy and dioecy, and its proposed explanation (Miller & Venable, 2000), show convincingly that inbreeding depression is involved in an important way in the interactions between polyploidy and the sexual system (Charlesworth, 2001).

The association between ploidy and the sexual system in M. annua is of course quite the reverse to that exposed more generally by Miller and Venable (2000). Rather than viewing this as an annoying exception, we suggest that it reveals more tellingly the potential importance that selection for reproductive assurance has had in continuing to maintain monoecy in this annual coloniser of ephemeral habitat. The relationship between combined versus separate sexes and population structure, which we have described for hexaploid M. annua independent of ploidy differences, adds weight to our hypothesis. Nevertheless, our studies of M. annua to date have thrown up more questions than answers, and we are still far from understanding the evolutionary significance of the striking variation in ploidy and sexual systems we observe.

We are currently conducting a detailed examination of the ecological and genetic interactions that occur across transitions of ploidy and sexual systems in the species, including the use of transplant experiments to assess the relative importance of reproductive assurance, ecophysiology and tension-zone dynamics in determining

where these transitions occur. We are also analysing patterns of genetic polymorphism in zones represented by contrasting cytotypes and sexual systems to test predictions regarding the demographic history and structure of the respective populations. Although still incomplete, data on patterns of isozyme variation in populations throughout Europe and the Mediterranean Basin strongly implicate Pleistocene refugia in the eastern Mediterranean for diploid M. annua and in southern Iberia or North Africa for tetraploids and hexaploids (D. J. Obbard, S. A. Harris and J. R. Pannell, unpubl. data). This should caution us to consider the hypotheses we have discussed in light of the phylogeographic history of the M. annua species complex. This, of course, is likely to be idiosyncratic in its details, but continued study of M. annua is nevertheless likely to fill important gaps in our general understanding of polyploidy and the sexual system in plants.

8.15.7 Acknowledgements

We thank S. A. Harris for valuable discussion and ongoing assistance; Juan Arroyo, Fernando Ojeda, Xavier Sans and M. Verdú for assistance and hospitality in Spain; the NERC for grant support to JRP; and The Queen's College, Oxford, and the BBSRC for studentships awarded to DJO and RJAB, respectively.

8.15.8 References

Ainouche ML, Baumel A, and Salmom A. 2004. Spartina anglica: a natural model system for studying early evolutionary changes that affect allopolyploid genomes. Biological Journal of the Linnean Society submitted.

Baker HG. 1955. Self-compatibility and establishment after "long-distance" dispersal. Evolution 9: 347-348.

Barrett SCH. 2002. The evolution of plant sexual diversity. Nature Reviews Genetics 3: 274-284.

Barrett SCH, and Charlesworth D. 1991. Effects of a change in the level of inbreeding on the genetic load. Nature 352: 522-524.

Barrett SCH, and Harder LD. 1996. Ecology and evolution of plant mating. Trends in Ecology and Evolution 11: 73-79.

Barrett SCH, and Shore JS. 1987. Variation and evolution of breeding systems in the Turnera ulmifolia complex (Turneraceae). Evolution 41: 340-354.

Barton NH, and Gale KS. 1993. Genetic analysis of hybrid zones. In: Harrison RG, ed. Hybrid zones and the evolutionary process. Oxford: Oxford University Press.

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

Bennett JH. 1976. Expectations for inbreeding depression on self-fertilization of tetraploids. Biometrics 32: 449-452.

Bennett MD. 1987. Variation in genomic form in plants and its ecological implications. New Phytologist 106: 177-200.

Bingham ET. 1980. Maximizing heterozygosity in autopolyploids. In: Lewis WH, ed. Polyploidy: Biological relevance. New York: Plenum Press.

Bretagnolle F, and Lumaret R. 1995. Bilateral polyploidization in Dactylis glomerata L. subsp lusitanica: occurrence, morphological and genetic characteristics of first polyploids. Euphytica 84: 197-207.

Brochmann C, Brysting AK, Also IG, Borgen L, Grundt HH, Scheen AC, and Elven R. 2004. Polyploidy in arctic plants. Biological Journal of the Linnean Society submitted.

Brochmann C, and Elven R. 1992. Ecological and Genetic Consequences of Polyploidy in Arctic Draba-(Brassicaceae). Evolutionary Trends in Plants 6: 111-124.

Busbice TT, and Wilsie CP. 1966. Inbreeding and heterosis in autotetraploids with application to Medicago sativa L. Euphytica 15: 52-67.

Cain ML, Andreasen V, and Howard DJ. 1999. Reinforcing selection is effective under a relatively broad set of conditions in a mosaic hybrid zone. Evolution 53: 1343-1353.Charlesworth D. 1984. Androdioecy and the evolution of dioecy. Biological Journal of the Linnean Society 23: 333-348.

Charlesworth D. 2001. Evolution: An exception that proves the rule. Current Biology 11: R13-R15.

Charlesworth D, and Charlesworth B. 1978. A model for the evolution of dioecy and gynodioecy. American Naturalist 112: 975-997.

Charlesworth D, and Charlesworth B. 1987. Inbreeding depression and its evolutionary consequences. Annual Review of Ecology and Systematics 18: 273-288.

Charnov EL. 1982. The Theory of Sex Allocation. Princeton University Press, Princeton, NJ.

Comai L, Tyagi AP, Winter K, Holmes-Davis R, Reynolds SH, Stevens Y, and Byers B. 2000. Phenotypic instability and rapid gene silencing in newly formed arabidopsis allotetraploids. Plant Cell 12: 1551-1567.

Crnokrak P, and Barrett SCH. 2002. Perspective: Purging the genetic load: A review of the experimental evidence. Evolution 56: 2347-2358.

Dewey DR. 1966. Inbreeding depression in diploid, tetraploid, and hexaploid crested wheatgrass. Crop Science 6.

Dudash MR, and Fenster CB. 2001. The role of breeding system and inbreeding depression in the maintenance of an outcrossing mating strategy in Silene virginica (Caryophyllaceae). American Journal of Botany 88: 1953-1959.

Durand B. 1963. Le complexe *Mercurialis annua* L. s.l.: une étude biosystématique. Annales des Sciences Naturelles, Botanique, Paris 12: 579-736.

Durand B, and Durand R. 1991. Sex determination and reproductive organ differentiation in *Mercurialis*. Plant Science (Limerick) 80: 49-66.

Durand R, and Durand B. 1985. *Mercurialis*. In: Halevy, ed. CRC Handbook of Flowering. 376-387.
Durand R, and Durand B. 1992. Dioecy, monoecy, polyploidy and speciation in annual Mercuries. Bulletin de la Societé Botanique de France, Lettres Botaniques 139: 377-399.
Felber F. 1991. Establishment of a tetraploid cytotype in a diploid population - effect of relative fitness of the cytotypes. Journal of Evolutionary Biology 4: 195-207.
FelberGirard M, Felber F, and Buttler A. 1996. Habitat differentiation in a narrow hybrid zone between diploid and tetraploid Anthoxanthum alpinum. New Phytologist 133: 531-540.
Fisher RA. 1941. Average excess and average effect of a gene substitution. Annals of Eugenics 11: 53-63.
Fritsch P, and Rieseberg LH. 1992. High outcrossing rates maintain male and hermaphrodite individuals in populations of the flowering plant Datisca glomerata. Nature 359: 633-636.
Galloway LF, Etterson JR, and Hamrick JL. 2003. Outcrossing rate and inbreeding depression in the herbaceous autotetraploid, Campanula americana. Heredity 90: 308-315.
Gauthier P, Lumaret R, and Bedecarrats A. 1999. Genetic introgression between tetraploid Dactylis glomerata subspp. reichenbachii and glomerata in the French alps. Insights from morphological and allozyme variation. Plant Systematics and Evolution 214: 219-234.
Guo M, Davis D, and Birchler JA. 1996. Dosage effects on gene expression in a maize ploidy series. Genetics 142: 1349-1355.
Hamilton WD. 1967. Extraordinary sex ratios. Science 156: 477-488.
Hardy OJ, De Loose M, Vekemans X, and Meerts P. 2001. Allozyme segregation and inter-cytotype reproductive barriers in the polyploid complex Centaurea jacea. Heredity 87: 136-145.
Hardy OJ, Vanderhoeven S, De Loose M, and Meerts P. 2000. Ecological, morphological and allozymic differentiation between diploid and tetraploid knapweeds (Centaurea jacea) from a contact zone in the Belgian Ardennes. New Phytologist 146: 281-290.
Holsinger KE. 1991. Inbreeding depression and the evolution of plant mating systems. Trends in Ecology and Evolution 6: 307-308.
Husband BC. 2000. Constraints on polyploid evolution: a test of the minority cytotype exclusion principle. Proceedings of the Royal Society of London Series B-Biological Sciences 267: 217-223.
Husband BC. 2004. The role of triploid hybrids in the evolutionary dynamics of mixed-ploidy populations. Biological Journal of the Linnean Society submitted.
Husband BC, and Schemske DW. 1997. The effect of inbreeding in diploid and tetraploid populations of Epilobium angustifolium (Onagraceae): Implications for the genetic basis of inbreeding depression. Evolution 51: 737-746.
Husband BC, and Schemske DW. 1998. Cytotype distribution at a diploid-tetraploid contact zone in Chamerion (Epilobium) angustifolium (Onagraceae). American Journal of Botany 85: 1688-1694.
Husband BC, and Schemske DW. 2000. Ecological mechanisms of reproductive isolation between diploid and tetraploid Chamerion angustifolium. Journal of Ecology 88: 689-701.
Husband BC, Schemske DW, Burton TL, and Goodwillie C. 2002. Pollen competition as a unilateral reproductive barrier between sympatric Chamerion angustifolium. Proceedings of the Royal Society of London Series B-Biological Sciences 269: 2565-2571.
Inoue K, Masuda M, and Maki M. 1998. Inbreeding depression and outcrossing rate in the endangered autotetraploid plant Aster kantoensis (Asteraceae). Journal of Heredity 89: 559-562.
JF Golz JF, Su V, Clarke AE, and Newbigin E. 1999. Molecular description of mutations affecting the pollen component of the Nicotiana alata S locus. Genetics 152: 1123-1135.
Kirkpatrick M, and Barton NH. 1997. Evolution of a species' range. American Naturalist 150: 1-23.
Krahenbuhl M, Yuan YM, and Kupfer P. 2002. Chromosome and breeding system evolution of the genus *Mercurialis* (Euphorbiaceae): implications of ITS molecular phylogeny. Plant Systematic and Evolution 234: 155-170.
Kruuk LEB, Baird SJE, Gale KS, and Barton NH. 1999. A comparison of multilocus clines maintained by environmental adaptation or by selection against hybrids. Genetics 153: 1959-1971.
Lande R, and Schemske DW. 1985. The evolution of self-fertilization and inbreeding depression in plants: I. Genetic models. Evolution 39: 24-40.
Lande R, Schemske DW, and Schultz ST. 1994. High inbreeding depression, selective interference among loci, and the theshold selfing rate for purging recessive lethal mutations. Evolution 48: 965-978.
Lenormand T. 2002. Gene flow and the limits to natural selection. Trends in Ecology & Evolution 17: 183-189.
Levin DA. 1975. Minority cytotype exclusion in local plant populations. Taxon 24: 35-43.

Levin DA. 2002. The role of chromosomal change in plant evolution. Oxford University Press, New York.
Lewis D. 1960. Genetic control of specificity and activity of the S antigen in plants. Proceedings of the royal Society of London Series B 151: 468-477.
Lewis WH. 1976. Temporal adaptation correlated with ploidy in Claytonia virginica. Systematic Botany 1: 340-347.
Lisci M, and Pacini E. 1997. Fruit and seed structural characteristics and seed dispersal in *Mercurialis annua* L. (Euphorbiaceae). Acta Societatis Botanicorum Poloniae 66: 379-386.
Lloyd D. 1975. The maintenance of gynodioecy and androdioecy in angiosperms. Genetica 45: 325-339.
Lloyd DG. 1983. Evolutionary stable sex ratios and sex allocations. Journal of Theoretical Biology 105: 525-539.
Lloyd DG. 1987. Allocations to pollen, seeds and pollination mechanisms in self-fertilizing plants. Functional Ecology 1: 83-89.
Lumaret R, and Barrientos E. 1990. Phylogenetic relationships and gene flow between sympatric diploid and tetraploid plants of Dactylis glomerata (Gramineae). Plant Systematics and Evolution 169: 81-96.
Lumaret R, Guillerm JL, Delay J, Loutfi AAL, Izco J, and Jay M. 1987. Polyploidy and habitat differentiation in Dactylis glomerata L from Galicia (Spain). Oecologia 73: 436-446.
McCarthy BC, and Quinn JA. 1990. Reproductive ecology of Carya (Juglandaceae): Phenology, pollination, and breeding system of two sympatric tree species. American Journal of Botany 77: 261-273.
Menken SBJ, Smit E, and DenNijs HCM. 1995. Genetical population structure in plants: Gene flow between diploid sexual and triploid asexual dandelions (Taraxacum section Ruderalia). Evolution 49: 1108-1118.
Miller JS, and Venable DL. 2000. Polyploidy and the evolution of gender dimorphism in plants. Science 289: 2335-2338.
Morgan MT, and Schoen DJ. 1997. The role of theory in an emerging new plant reproductive biology. Trends in Ecology and Evolution 12: 231-234.
Muller HJ. 1925. Why polyploidy is rarer in animals than in plants. American Naturalist 59: 346-353.
Osborn TC, Pires JC, Birchler JA, Auger DL, Chen ZJ, Lee HS, Comai L, Madlung A, Doerge RW, Colot V, and Martienssen RA. 2003. Understanding mechanisms of novel gene expression in polyploids. Trends in Genetics 19: 141-147.
Pannell J. 1997a. The maintenance of gynodioecy and androdioecy in a metapopulation. Evolution 51: 10-20.
Pannell J. 1997b. Variation in sex ratios and sex allocation in androdioecious *Mercurialis annua*. Journal of Ecology 85: 57-69.
Pannell J. 1997c. Widespread functional androdioecy in *Mercurialis annua* L. (Euphorbiaceae). Biological Journal of the Linnean Society 61: 95-116.
Pannell JR. 2000. A hypothesis for the evolution of androdioecy: the joint influence of reproductive assurance and local mate competition in a metapopulation. Evolutionary Ecology 14: 195-211.
Pannell JR. 2001. A hypothesis for the evolution of androdioecy: the joint influence of reproductive assurance and local mate competition in a metapopulation. Evolutionary Ecology 14: 195-211.
Pannell JR. 2002. The evolution and maintenance of androdioecy. Annual Review of Ecology and Systematics 33: 397-425.
Pannell JR, and Barrett SCH. 1998. Baker's Law revisited: reproductive assurance in a metapopulation. Evolution 52: 657-668.
Petit C, Bretagnolle F, and Felber F. 1999. Evolutionary consequences of diploid-polyploid hybrid zones in wild species. Trends in Ecology and Evolution 14: 306-311.
Petit C, Lesbros P, Ge XJ, and Thompson JD. 1997. Variation in flowering phenology and selfing rate across a contact zone between diploid and tetraploid Arrhenatherum elatius (Poaceae). Heredity 79: 31-40.
Petit C, and Thompson JD. 1997. Variation in phenotypic response to light availability between diploid and tetraploid populations of the perennial grass Arrhenatherum elatius from open and woodland sites. Journal of Ecology 85: 657-667.
Ramsey J, and Schemske DW. 2002. Neopolyploidy in flowering plants. Annual Review of Ecology and Systematics 33: 589-639.
Richards AJ. 1997. Plant Breeding Systems. Chapman and Hall, London.
Rodriguez DJ. 1996a. A model for the establishment of polyploidy in plants. American Naturalist 147: 33-46.

Rodriguez DJ. 1996b. A model for the establishment of polyploidy in plants: Viable but infertile hybrids, iteroparity, and demographic stochasticity. Journal of Theoretical Biology 180: 189-196.
Ronfort J. 1999. The mutation load under tetrasomic inheritance and its consequences for the evolution of the selfing rate in autotetraploid species. Genetical Research 74: 31-42.
Rosquist G. 2001. Reproductive biology in diploid Anthericum ramosum and tetraploid A. liliago (Anthericaceae). Oikos 92: 143-152.
Ross R. 1981. Chromosome counts, cytology, and reproduction in the Cactaceae. American Journal of Botany 68: 463-470.
Schultz ST. 1999. Can females benefit from selfing avoidance? Genetic associations and the evolution of plant gender. Proceedings of the Royal Society of London B 266: 1967-1973.
Scott RJ, Spielman M, Bailey J, and Dickinson HG. 1998. Parent-of-origin effects on seed development in Arabidopsis thaliana. Development 125: 3329-3341.
Segraves KA, and Thompson JN. 1999. Plant polyploidy and pollination: Floral traits and insect visits to diploid and tetraploid Heuchera grossulariifolia. Evolution 53: 1114-1127.
Segraves KA, Thompson JN, Soltis PS, and Soltis DE. 1999. Multiple origins of polyploidy and the geographic structure of Heuchera grossulariifolia. Molecular Ecology 8: 253-262.
Soltis DE, Soltis PS, Pires JC, Kovarik A, and Tate JA. 2004. Recent and recurrent polyploidy in Tragopogon (Asteraceae): cytogenetic, genomic, and genetic comparisons. Biological Journal of the Linnean Society submitted.
Stebbins GL. 1980. Polyploidy in plants: unsolved problems and prospects. In: Lewis WH, ed. Polyploidy: Biological Relevance. New York: Plenum. 495-520.
Stebbins GL. 1985. Polyploidy, hybridization and the invasion of new habitats. Annals of the Missouri Botanic Garden 72: 824-832.
Stone JL. 2002. Molecular mechanisms underlying the breakdown of gametophytic self-incompatibility. Quarterly Review of Biology 77: 17-32.
Sybenga J. 1975. Meiotic Configurations. Springer Verlag, Berlin.
Tomekpe K, and Lumaret R. 1991. Association between quantitative traits and allozyme heterozygosity in a tetrasomic species - Dactylis glomerata. Evolution 45: 359-370.
Tutin TG, Heywood VH, Burges NA, Valentine DH, Walters SM, and Webb DA. 1964. Flora Europaea. Cambridge University Press, Cambridge.
Van Dijk P, and Bakx-Schotman T. 1997. Chloroplast DNA phylogeography and cytotype geography in autopolyploid Plantago media. Molecular Ecology 6: 345-352.
Van Dijk P, and Bijlsma R. 1994. Simulations of flowering time displacement between two cytotypes that form inviable hybrids. Heredity 72: 522-535.
Van Dijk P, Hartog M, and Vandelden W. 1992. Single cytotype areas in autopolyploid Plantago media L. Biological Journal of the Linnean Society 46: 315-331.
Vogel JC, Rumsey FJ, Schneller JJ, Barrett JA, and Gibby M. 1999. Where are the glacial refugia in Europe? Evidence from pteridophytes. Biological Journal of the Linnean Society 66: 23-37.
Wendel JF. 2000. Genome evolution in polyploids. Plant Molecular Biology 42: 225-249.
Westergaard M. 1958. The mechanism of sex determination in dioecious plants. Advances in Genetics 9: 217-281.
Zhang DY. 2000. Resource allocation and the evolution of self-fertilization in plants. American Naturalist 155: 187-199.