

# University of St Andrews



Full metadata for this thesis is available in  
St Andrews Research Repository  
at:

<http://research-repository.st-andrews.ac.uk/>

This thesis is protected by original copyright

**Molecular analysis of the hybrid origin of  
Oxford Ragwort, *Senecio squalidus* L.,  
and related studies**

**Juliet Katherine Parry James**

**A thesis submitted to the  
University of St. Andrews for  
the degree of Doctor of Philosophy**

**School of Biological and Medical Sciences  
University of St. Andrews**

**December 1999**



In 1586

**This thesis is dedicated to my parents, in thanks for all their support and encouragement over the years.....**

# Abstract

The principal aim of the research presented in this thesis was to investigate the evolutionary history of *Senecio squalidus* L., the Oxford Ragwort. This taxon was introduced to Britain around 300 years ago, and was suspected to be the product of hybridisation between *S. chrysanthemifolius* and *S. aethnensis* on Mt. Etna, Sicily.

Molecular markers to distinguish *S. chrysanthemifolius* from *S. aethnensis* were developed by surveying 'pure' populations of the two taxa for RAPD, ISSR, isozyme and cpDNA RFLP variation.

The molecular markers that differentiate *S. chrysanthemifolius* and *S. aethnensis* were used to confirm the existence of the postulated hybrid zone between the two species, that grow at different altitudes on Mt. Etna. Many markers showed clinal variation, and clines were broadly coincident. Hybrid indices and principal co-ordinate analysis of RAPD and ISSR data provided pictorial representations of the genetic composition of individuals within hybrid zone populations. The nature of the hybrid zone - its origin, maintenance and future - is discussed and an important role for environmental factors is postulated.

RAPD and ISSR data confirm that British *S. squalidus* is the homoploid hybrid derivative of *S. chrysanthemifolius* and *S. aethnensis*. British *S. squalidus* remains morphologically and genetically polymorphic, and in general intermediate between its parents. Nevertheless, it can be considered a new species because it is geographically and ecologically isolated from parental gene-flow and evolving independently. The results suggest that British *S. squalidus* has inherited more of its genome from *S. chrysanthemifolius* than from *S. aethnensis*.

RAPD and ISSR markers were also employed to examine the genomic composition of products of hybridisation between *S. squalidus* and native British *S. vulgaris* var. *vulgaris*, namely *S. vulgaris* var. *vulgaris*, York Radiate Groundsel and *S. cambrensis*. The findings indicate that var. *hibernicus* contains only a very small amount of *S. squalidus* genetic material, while York Radiate Groundsel contains considerably more. Individuals of the allohexaploid species, *S. cambrensis*, exhibited RAPD and ISSR profiles that were to a large extent additive between the parents.

## **Declaration**

I, Juliet Katherine Parry James, hereby certify that this thesis, which is approximately 65,000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

Juliet James  
December 1999.

## Statement

I was admitted as a research student in October, 1995, and as a candidate for the degree of Ph.D. in October 1996; the higher study, for which this is a record was carried out in the University of St. Andrews between 1995 and 1999.

Juliet James  
December 1999.

# Certificate

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Ph.D. in the University of St. Andrews and that the candidate is qualified to submit this thesis in application for that degree.

R. J. Abbott  
December 1999.

# Copyright

In submitting this thesis to the University of St. Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any *bona fide* library or research worker.

Juliet James  
December 1999.

# Acknowledgements

I would like to thank Richard Abbott for his advice and careful reading through of the earlier drafts of this thesis, and Kirsten Wolff for her enthusiasm and encouragement. I am grateful to David Forbes for his patient technical assistance. Heartfelt thanks to Amanda Gillies and Rolf Holderegger for their advice, helpful discussions and support when it was most needed. Thanks to Kirsten, Amanda, and others who taught me molecular and data analysis techniques; and to Richard Milne and my father for accompanying me to Sicily. Finally, thanks to all those who made the Harold Mitchell a good place to work, including Janet, Pam and Urte.

Outside the world of biology, I would like to thank my friends who've been so supportive. In particular, I would like to thank Kevin for his friendship throughout the last few years. Thanks to Phil and Graham (and Laura!) for their excellent company, which made the last few years in St. Andrews much more pleasant. Also to Caroline, and others, for listening to my moaning!

Last, and by no means least, thanks to my family. None of this would have been possible without the support and encouragement (and financial assistance!) of my parents and sisters. Special thanks to Jo for helping me put this thesis together, and not losing her cool.

The financial assistance of a Natural Environment Research Council research studentship is gratefully acknowledged.

# Contents

## **Chapter 1 General Introduction**

1.1	Introduction to Hybridisation .....	1
1.1.1	Definitions	1
1.1.2	Hybridisation and species concept	3
1.1.3	The frequency of plant hybridisation	3
1.2	Hybrid Zones.....	5
1.2.1	The origin of hybrid zones – primary versus secondary intergradation	5
1.2.2	Hybrid zone models	8
1.2.3	The fate of hybrid zones	13
1.3	The evolutionary consequences of hybridisation and introgression in plants.....	14
1.3.1	An increase in genetic diversity	14
1.3.2	The transfer of adaptations	14
1.3.3	The origin of new adaptations	15
1.3.4	The breakdown or reinforcement of reproductive barriers	16
1.3.5	A dispersal mechanism	17
1.3.6	The origin of new homoploid types	18
1.3.7	The origin of new allopolyploids	24
1.4	Molecular markers and the study of hybrids and hybrid zones.....	27
1.4.1	Isozyme analysis	27
1.4.2	Restriction fragment length polymorphism (RFLP) analysis	28
1.4.3	RAPDs	32
1.4.4	Inter simple sequence repeat markers (ISSRs)	41
1.4.5	Amplified fragment length polymorphisms (AFLPs)	43
1.4.6	Simple sequence repeat (SSR) or microsatellite analysis	44
1.4.7	DNA sequencing	45
1.5	Aims and Objectives.....	47

## **Chapter 2 Materials and Methods**

2.1	Origin of plant material.....	48
2.2	Cultivation of plant material.....	48
2.3	DNA extraction.....	52
2.4	Running agarose gels.....	53
2.5	Quantification of the DNA.....	54
2.6	Additional DNA purification .....	55
2.7	Restriction fragment length polymorphism (RFLP) analysis.....	55
2.8	RAPD and ISSR procedure.....	59
2.9	Testing the homology of RAPD and ISSR bands.....	61
2.10	Allozyme analysis.....	63
2.11	Data analysis.....	65
2.11.1	Recording RAPD and ISSR results	65
2.11.2	Similarity coefficients: Nei and Li, and Jaccard	65
2.11.3	Principle coordinate analysis (PCO)	66
2.11.4	Analysis of molecular variance (AMOVA)	68
2.11.5	Cluster analyses	69
2.11.6	Shannon's Diversity Index	70
2.11.7	General Statistics	71
2.11.8	<i>F</i> -statistics	71
2.11.9	Nei's unbiased genetic distance	72

## Chapter 3 The Development of Taxon-specific Molecular Markers

3.1	Introduction.....	73
3.1.1	What is a taxon-specific molecular marker?	74
3.1.2	Choice of molecular marker methods	75
3.1.3	Aim	76
3.2	Materials and Methods.....	77
3.2.1	RAPD and ISSR survey of variation within and between <i>S. aethnensis</i> and <i>S. chrysanthemifolius</i>	77
3.2.2	Survey of isozyme variation within and between <i>S. chrysanthemifolius</i> and <i>S. aethnensis</i>	80
3.2.3	Survey of chloroplast DNA RFLP variation within and between <i>S. chrysanthemifolius</i> and <i>S. aethnensis</i>	81
3.3	Results.....	82
3.3.1	Results of the survey of RAPD and ISSR variation within and between <i>S. chrysanthemifolius</i> and <i>S. aethnensis</i>	82
3.3.2	Results of the survey of isozyme variation within and between <i>S. chrysanthemifolius</i> and <i>S. aethnensis</i>	90
3.3.3	Results of the survey of chloroplast DNA RFLP variation within and between <i>S. chrysanthemifolius</i> and <i>S. aethnensis</i>	92
3.4	Discussion.....	95
3.4.1	The effectiveness of the survey of RAPD and ISSR variation within and between <i>S. chrysanthemifolius</i> and <i>S. aethnensis</i>	95
3.4.2	The survey of isozyme variation between <i>S. chrysanthemifolius</i> and <i>S. aethnensis</i>	97
3.4.3	The survey of chloroplast DNA RFLP variation between <i>S. chrysanthemifolius</i> and <i>S. aethnensis</i>	97
3.4.4	The taxonomic status of <i>S. chrysanthemifolius</i> and <i>S. aethnensis</i>	98
3.4.5	Assumptions made and improvements suggested	98
3.4.6	The potential of other molecular marker systems to produce taxon-specific bands for <i>S. chrysanthemifolius</i> and <i>S. aethnensis</i>	101
3.4.7	Summary	102

## Chapter 4 Analysis of the Hybrid Zone Between *Senecio chrysanthemifolius* and *Senecio aethnensis* on Mt. Etna, Sicily, Using Molecular Markers

4.1	Introduction.....	103
4.1.1	Early work on the postulated hybrid zone between <i>S. chrysanthemifolius</i> and <i>S. aethnensis</i> on Mt. Etna, Sicily	103
4.1.2	The use of molecular markers	105
4.1.3	Aim	107
4.2	Materials and Methods.....	108
4.2.1	Choice of transect and populations	108
4.2.2	Survey of RAPD and ISSR variation along transect S, through the postulated hybrid zone between <i>S. chrysanthemifolius</i> and <i>S. aethnensis</i> , on Mt. Etna	109
4.2.3	Survey of isozyme variation along transect S	110
4.2.4	Survey of chloroplast DNA RFLP variation along transect S	110
4.2.5	Data analysis	111
4.3	Results.....	114
4.3.1	Description of the postulated hybrid zone between <i>S. chrysanthemifolius</i> and <i>S. aethnensis</i> on Mt. Etna, Sicily	114
4.3.2	RAPD and ISSR variation along transect S, through the postulated hybrid zone between <i>S. chrysanthemifolius</i> and <i>S. aethnensis</i> , on Mt. Etna	116
4.3.3	Isozyme variation along transect S, through the postulated hybrid zone between <i>S. chrysanthemifolius</i> and <i>S. aethnensis</i> , on Mt. Etna	124
4.3.4	Chloroplast DNA RFLP variation along transect S, through the postulated hybrid zone between <i>S. chrysanthemifolius</i> and <i>S. aethnensis</i> , on Mt. Etna	129

4.4	Discussion.....	130
4.4.1	Evidence for a hybrid zone between <i>S. chrysanthemifolius</i> and <i>S. aethnensis</i> on Mt. Etna	130
4.4.2	Primary or secondary intergradation	130
4.4.3	Ancestral polymorphism	131
4.4.4	Single locus variation	132
4.4.5	Hybrid indices and principle coordinate analysis	134
4.4.6	Population level analyses	136
4.4.7	Maintenance of the hybrid zone	138
4.4.8	The future of the hybrid zone	144
4.4.9	Improvements to method of analysis and future work	145

## Chapter 5 Molecular Analysis of the Hybrid Origin of British *Senecio squalidus*

5.1	Introduction.....	147
5.1.1	Early history - introduction and spread of British <i>S. squalidus</i>	147
5.1.2	Current distribution	148
5.1.3	The hybrid origin of <i>S. squalidus</i>	148
5.1.4	Aim	150
5.2	Materials and Methods.....	151
5.2.1	Survey of RAPD and ISSR variation in British <i>S. squalidus</i> populations	151
5.2.2	Survey of chloroplast and isozyme variation	153
5.3	Results.....	154
5.3.1	Survey of RAPD and ISSR variation in British <i>S. squalidus</i> populations	154
5.3.2	Hybrid index analysis	157
5.3.3	Principle coordinate analysis	157
5.3.4	Analysis of molecular variance	159
5.3.5	Shannon's diversity index	160
5.4	Discussion.....	162
5.4.1	The identification of homoploid hybrid species	162
5.4.2	Hybrid origin of British <i>S. squalidus</i>	166
5.4.3	Comparison of modes of homoploid hybrid speciation	168
5.4.4	Proportion of the genome of British <i>S. squalidus</i> attributable to <i>S. chrysanthemifolius</i> and <i>S. aethnensis</i> .	170
5.4.5	The genetic structure of British <i>S. squalidus</i>	172
5.4.6	The success of British <i>S. squalidus</i>	174
5.4.7	<i>Senecio rupestris</i> - a confounding factor	177

## Chapter 6 A Preliminary Study of the Genomic Composition of Hybrid Derivatives of *Senecio squalidus* and *Senecio vulgaris* var. *vulgaris*

6.1	Introduction.....	180
6.1.1	Hybridisation between <i>S. squalidus</i> and <i>S. vulgaris</i> var. <i>vulgaris</i> , common groundsel	180
6.1.2	F <sub>1</sub> hybrids between <i>S. squalidus</i> and <i>S. vulgaris</i> var. <i>vulgaris</i>	180
6.1.3	Origin of <i>S. vulgaris</i> var. <i>hibernicus</i> , radiate groundsel	181
6.1.4	Origin of York Radiate Groundsel	183
6.1.5	Origin of <i>S. cambrensis</i> , the Welsh groundsel	184
6.1.6	RAPD variation in the hybrid derivatives of <i>S. squalidus</i> and <i>S. vulgaris</i> var. <i>vulgaris</i>	185
6.1.7	Aim	185
6.2	Materials and Methods.....	186
6.2.1	RAPD and ISSR survey of <i>S. vulgaris</i> var. <i>vulgaris</i> , <i>S. vulgaris</i> var. <i>hibernicus</i> , <i>S. cambrensis</i> and York Radiate Groundsel	186
6.3	Results.....	189
6.3.1	Screen of isolated <i>S. vulgaris</i> var. <i>vulgaris</i>	189

6.3.2	Presence/absence of <i>S. chrysanthemifolius</i> and <i>S. aethnensis</i> marker bands in the hybrid derivatives	189
6.3.3	Presence/absence of var. <i>vulgaris</i> and <i>S. squalidus</i> marker bands in the hybrid derivatives	191
6.4	Discussion.....	196
6.4.1	Screen of isolated <i>S. vulgaris</i> var. <i>vulgaris</i>	196
6.4.2	Presence/absence of <i>S. chrysanthemifolius</i> , <i>S. aethnensis</i> , " <i>S. squalidus</i> " and " <i>S. vulgaris</i> " bands in the hybrid derivatives of <i>S. squalidus</i> and <i>S. vulgaris</i>	197
6.4.3	<i>S. vulgaris</i> var. <i>hibernicus</i>	198
6.4.4	Non-isolated <i>S. vulgaris</i> var. <i>vulgaris</i>	199
6.4.5	York Radiate Groundsel	200
6.4.6	<i>S. cambrensis</i>	201
6.4.7	Improvements to the study and future work	203

## Chapter 7 General Conclusions

7.1	Conclusions.....	205
-----	------------------	-----

# **Chapter 1**

## **General Introduction**

## **General introduction**

### **1.1 Introduction to hybridisation**

The concept of hybridisation has intrigued man through the ages. Early scholars often explained the morphologies of bizarre creatures, both real and mythological, by invoking crossing between discrete species (Zirkle, 1941; in Harrison, 1993). However, it was not until the mid 1700s that a more scientific approach was taken by Linnaeus and his contemporaries, followed in the late 1800s by the likes of Darwin and Wallace (Harrison, 1993). This century, the study of hybridisation has flourished and in the last three decades the subject has been revolutionised by the development of molecular markers.

Today, hybridisation is no longer merely the subject of academic interest but a topic that is frequently in the news (Brookes, 1999). Wildlife experts in Britain are arguing about the environmental impact of hybridisation between species; for example between the endangered white-headed duck and the American ruddy duck (Brookes, 1999), between native red deer and Japanese sika deer (Abernethy, 1994), and between our native flora and plants from wildflower mixes of international provenance. In America, the conservation status of hybrid taxa such as the red wolf (a wolf-coyote hybrid) and the Florida panther has yet to be resolved (Rhymer and Simberloff, 1996). Even more controversial are the debates surrounding potential hybridisation between genetically modified crops and wild relatives. Biotechnology firms may reassure, but fears about "super-weeds" and damage to soil microfauna and flora, insect and bird populations remain (Abbott, 1994; Hails, 2000). The continuing interest in issues such as these suggests that it has never been more important to understand the theoretical concepts and practical realities of hybridisation.

#### **1.1.1 Definitions**

##### *Hybridisation*

In scientific literature, various definitions of the term hybridisation have been used (Harrison, 1993). At one extreme, hybridisation may be viewed exclusively as the crossing of individuals from different species. At the other, it may be viewed as crossing between genetically distinct individuals. Harrison took the middle ground and proposed that:

*"hybridisation is the interbreeding of individuals from two populations, or groups of populations, which are distinguishable on the basis of one or more heritable characters"* (Harrison, 1990, 1993). Harrison's definition is widely accepted and has four advantages. First, it does not rely upon the acceptance of a particular species concept. Second, the arbitrary assignment of populations to particular taxonomic categories (e.g. subspecies, races) is avoided. Third, it is unnecessary to judge the relative fitness of hybrids or the differences in adaptive norms between parental types (Harrison, 1993); and finally, the definition is easy to use (Arnold, 1997).

### *Hybrid*

*"A... hybrid individual derives from crosses...between individuals from two populations, or groups of populations, which are distinguishable on the basis of one or more heritable characters"* (Arnold, 1997). The term 'hybrid' is applied (unless otherwise stated) not only to F<sub>1</sub> offspring but to all individuals of mixed ancestry including backcrosses, F<sub>2</sub> individuals and so on (Harrison, 1993).

### *Hybrid zone*

Many definitions of hybrid zones have been suggested (Harrison, 1990). Some restrict the term to regions of secondary contact (see below), while others contain explicit assumptions about the dynamics of the zone (Harrison, 1993). However, as it is often difficult to infer the origin and dynamics of a hybrid zone, it is better to use a broad definition, e.g.: *"hybrid zones occur when genetically distinct groups of individuals meet and mate, resulting in at least some offspring of mixed ancestry"* (Harrison, 1993). Hybrid zones are frequently identified by patterns of clinal variation; that is by *"a gradient or set of gradients in morphology or gene frequency, at one or more loci"* (Barton and Hewitt, 1985).

### *Introgression*

Hybridisation, followed by back-crossing to one or both parental types can lead to introgression (Harrison, 1993). Introgression is *"the permanent incorporation of genes from one set of differentiated populations into another, i.e. the incorporation of alien alleles into a new, reproductively integrated population system"* (Rieseberg and Wendel, 1993).

### 1.1.2 Hybridisation and the species concept

Several species concepts have been proposed and all have their strengths and weaknesses. The most well-known and popular is the biological species concept (BSC), according to which species are “*groups of actually or potentially interbreeding natural populations which are reproductively isolated from other such groups*” (Mayr, 1963; in Barton and Hewitt, 1985). It is clear that hybridisation between distinct taxa, that might otherwise be identified as separate species, seriously challenges the validity of the BSC. Even a revised version of the BSC (Mayr, 1982; in Barton and Hewitt, 1985) does not satisfactorily deal the criticisms levelled at it (Barton and Hewitt, 1985). Hybridisation also poses problems for other species concepts, including recognition, cohesion and phylogenetic (Arnold, 1997). Mallet (1995a, 1995b) suggests that these “*assumption-laden*” species concepts should be abandoned. Instead, he advocates the “*genotypic cluster definition*”, which states that “*species are groups that remain recognisable in sympatry because of the morphological and genotypic gaps between them*”. Mallet's definition permits species to “*hybridise in theory as well as practice*” and to evolve in a reticulate manner. Above all, the phenomenon of hybridisation is a timely reminder that the search for a single, unequivocal species definition may prove to be futile, not least because species are defined for our convenience and do not necessarily reflect biological truths (Mallet, 1995b).

### 1.1.3 The frequency of plant hybridisation

Various attempts have been made to assess the extent of hybridisation in plants. One approach has been to estimate the number of hybrid plant taxa. For instance, Knobloch (1972; in Rieseberg and Ellstrand, 1993) compiled a list of over 23000 possible interspecific and intergeneric hybrids, while Stace (1991; in Rieseberg and Ellstrand, 1993) calculated that around 20% of the British flora were of hybrid origin. Other authors have examined introgression. For example, Rieseberg and Wendel (1993) evaluated 165 examples of putative introgression: in 65 cases there was good evidence of introgression and in a further 94 cases hybrid swarm formation was established. However, hybridisation does not appear to be randomly distributed among taxa and instead may be concentrated in certain families and genera (Ellstrand *et al.*, 1996). An alternative approach has been to estimate the number of polyploid species, leading to suggestions that between 50-70% of angiosperms, and an even greater percentage of pteridophytes, are of hybrid origin (Grant,

1981; Arnold, 1997). This method assumes that chromosome numbers are a reliable indicator of polyploid status and that most polyploids are allopolyploids rather than autopolyploids (Rieseberg and Ellstrand, 1993). Nonetheless, both these assumption are believed to be reasonable (Arnold, 1997). Overall, therefore, the data indicate that hybridisation is widespread in plant groups.

## 1.2 Hybrid Zones

### 1.2.1 The origin of hybrid zones - primary versus secondary intergradation

Hybrid zones may be created by two processes. Primary hybrid zones are formed by primary intergradation (primary contact), i.e. the differentiation, *in situ*, of more or less continuously distributed populations as a result of spatially varying selection, along an environmental gradient (Futuyma, 1998). Secondary hybrid zones are formed by secondary intergradation (secondary contact), i.e. the meeting of populations that have differentiated in allopatry (Harrison, 1993).

The issue of hybrid zone origins, i.e. primary versus secondary intergradation, can be partitioned into two discrete questions. Firstly, "*where and how did existing variants arise (in what geographic context and as a result of what evolutionary forces)?*" (Harrison, 1990). Secondly, "*what historical and ecological factors account for the current distribution of these variants (i.e. the current geographic extent of the hybrid zone and patterns of variation within it)?*" (Harrison, 1990). Harrison (1990) argues that one cannot obtain answers to the first question simply by observing patterns of variation in hybrid zones because the process of differentiation (speciation) is often obscured by ensuing events (e.g. dispersal, colonisation and local extinction). In other words, "*attempts to resolve hybrid zone origins must focus on factors determining the current distribution of the hybridising taxa and not on the geographic context in which differences initially arose*" (Harrison and Rand, 1989). For example, a non-allopatric speciation event may produce species that subsequently colonise new habitats, in which the species meet and form a zone of secondary contact (Harrison and Rand, 1989). Thus, two species may arise by means of primary intergradation and later meet in a zone of secondary contact.

#### 1.2.1.1 Distinguishing between primary and secondary intergradation

It can be extremely difficult to distinguish primary and secondary intergradation; both processes give rise to very similar patterns of variation. The various lines of evidence that may be informative are discussed below:

### *Historical data*

Occasionally, historical evidence provides conclusive proof of hybrid zone formation by secondary intergradation. For example, a hybrid zone between two species of fire ant was known to have been formed by secondary contact because both ant species were introduced (Shoemaker *et al.*, 1994). Likewise, a hybrid zone between two species of thistle was known to be a zone of secondary intergradation because both thistle species were introduced (Warwick *et al.*, 1989).

Without historical evidence, distinguishing between primary and secondary intergradation is difficult, and researchers must rely upon clinal, phylogenetic, biogeographical and ecological data.

### *Evidence from clines*

If a hybrid zone is formed by primary intergradation one would expect non-coincident clines because "*different loci might be differentially affected by the environmental gradient*" (Futuyma, 1998). Unfortunately, the situation is not as simple as that. First, selection could act independently on many loci to produce coincident clines in a primary zone. Second, coincident clines, in a primary zone, could result from widespread co-adaptation, if a majority of newly emerging alleles are beneficial in only one genetic background (Harrison, 1990). Although these two scenarios are not impossible, their likelihood decreases as the number of coincident clines increases. Consequently, "*many investigators take secondary contact to be the most likely explanation of hybrid zones when many characters show coincident clines*" (Futuyma, 1998).

It does not follow that a lack of coincident clines is necessarily indicative of primary intergradation. A lack of coincident clines could also result from spatially varying selection in a zone of secondary contact with an environmental gradient (Barton and Hewitt, 1985). Overall, therefore, clinal patterns of variation must be interpreted cautiously and the evidence is frequently inconclusive.

### *Evidence from phylogenetics*

Phylogenetics can be a valuable means of distinguishing primary from secondary intergradation. "*In general, hybridisation between non-sister taxa would seem to signal secondary contact*" (Harrison, 1990). However, the converse is not true, that is

hybridisation between sister-taxa does not necessarily indicate primary contact (Rieseberg and Ellstrand, 1993).

Harrison (1990) details a phylogenetic procedure for distinguishing between primary and secondary hybrid zones. In this method, a set of phylogenies is constructed using samples from many populations on either side of a linear hybrid zone (or within a mosaic hybrid zone). Dendrograms in which all populations from one side (or one patch type) group together could denote widespread co-adaptation, current ecology or recent common ancestry. If many independent phylogenies (constructed from different character sets) display the same pattern, the latter explanation is more probable; that is the two hybridising lineages presumably existed before the hybrid zone was formed and therefore the zone is one of secondary contact. An alternative phylogenetic approach was developed by Thorpe (1984). In this method, phylogenetic analysis is employed to postulate the range expansion of a species. The origin (primary or secondary) of a hybrid zone can then be deduced by comparing the geographical location of the hybrid zone with the conjectured range expansion. Nevertheless, there are many problems associated with using phylogenetics to ascertain hybrid zone origins, for example lineage sorting (Harrison, 1990).

#### *Evidence from biogeography*

Many authors conclude that the hybrid zones they have studied are of secondary origin on the basis of biogeographical data, such as large-scale patterns of climate and vegetation change (e.g. Perron and Bousquet, 1997). For example, many hybrid zones in northern temperate regions are thought to be the result of post-Pleistocene secondary contact (Hewitt, 1989). It is assumed that populations differentiated in isolated Pleistocene refugia and re-established contact as a result of range expansion following glacial retreat. Although biogeographical evidence is often persuasive it does not, in itself, rule out the possibility that a hybrid zone may have been formed by primary contact.

#### *Evidence from ecology*

The absence of an obvious environmental gradient has been argued as an indicator of secondary contact (Millar, 1983). Yet, without extensive work, it is hard to prove the lack of an environmental gradient. Furthermore, the presence of an environmental gradient does not, of course, indicate that the hybrid zone is of primary origin as secondary intergradation may occur along an environmental gradient.

Habitat alteration or fragmentation can lead to the breakdown of ecological barriers between taxa or provide corridors of dispersal that allow formerly isolated populations to come into contact (Young, 1996). In addition, human disturbance may produce new habitats where hybrids can thrive. For these reasons, evidence of human disturbance may provide weak support for a hypothesis of secondary intergradation.

### **1.2.2 Hybrid zone models**

Various models have been proposed to account for stable hybrid zones. Some models explain the persistence of such zones in terms of selection alone, others invoke a balance between selection and dispersal. Some models are environment-dependent, others environment-independent. Models also differ in the assumed fitness of hybrid genotypes, relative to parental taxa.

#### ***1.2.2.1 Bounded hybrid superiority model***

The bounded hybrid superiority model was put forward by Moore (1977). This model is dispersal-independent; dispersal is assumed to be negligible and a stable hybrid zone is maintained because hybrids have higher fitness than parental taxa within a restricted geographic area. The chief determinant of relative fitness is considered to be the environment, i.e. environment-dependent natural selection favours hybrids in certain habitats. Moore believed that hybrids would tend to be formed in ecotones; in marginal, transitional and disturbed areas.

#### ***1.2.2.2 Environment-dependent selection and dispersal model***

In this model, a stable hybrid zone is maintained by a balance between dispersal and natural selection acting to promote adaptation to the external environment: "*different alleles are favoured in different places*" (Barton and Gale, 1993). Therefore, a hybrid zone of this type is associated with an environmental gradient or ecotone and is stationary, trapped at a certain point along the gradient or ecotone.

#### ***1.2.2.3 Tension zone model***

This model was developed by Barton and Hewitt (1985) for stable hybrid zones that are maintained by a balance between dispersal and intrinsic selection against hybrids. Selection against hybrids may be direct, due to epistatic interactions or stem from frequency-

dependent selection against rare alleles (Harrison, 1990). This type of hybrid zone is potentially mobile, because its maintenance is unconnected to local environmental conditions. Such hybrid zones move to minimise their length, hence the name "tension zone". Tension zones are inclined to come to rest where there is restricted gene flow, at barriers to dispersal or in regions of low population density.

Barton and Hewitt (1985) suggest that a substantial proportion, if not a majority, of hybrid zones fit the tension zone model, rather than the bounded hybrid superiority model or the environment-dependent selection/dispersal model.

They criticise the bounded hybrid superiority model on three counts (Barton and Hewitt, 1985): First, in a dispersal-independent cline, the width of the cline must be substantially greater than the characteristic scale of selection (the ratio of dispersal distance to the square root of the selection coefficient). In other words, the cline must be much wider than the dispersal distance of the organism and selection must be moderate to strong. However, hybrid zone clines are frequently much narrower than probable environmental gradients, and have widths that approximate to the individual dispersal range (Barton and Gale, 1993). Second, if cline shape is a direct response to local selection pressures one would expect considerable variation from location to location. Indeed, Barton and Hewitt (1985) suggest that "*hybrid superiority should give a broken, patchy distribution*". However, they argue that many clines have a consistent width and shape across different transects. Finally, if clines for each phenotypic trait or for each locus are determined directly by the environment one would not expect them to be centred in the same position (coincident) and equal in width (concordant). However, according to Barton and Hewitt (1985), most hybrid zones comprise a number of coincident and concordant clines, including those of functionally unrelated characters. Barton and Hewitt (1985) and Barton and Gale (1993) admit that there are exceptions to their arguments. For example, the hybrid zone of *Thomomys bottae* (the pocket gopher) is so wide that, in effect, dispersal must be negligible. Nevertheless, the judgement is made that "*both geographic and genotypic patterns fit with a balance between dispersal and selection, and not with any form of hybrid superiority*" (Barton and Hewitt, 1985).

Single locus models indicate that a cline maintained by a balance between selection against hybrids and dispersal will be difficult to distinguish from a cline maintained by a balance

between environment-dependent selection and dispersal. However, Barton and Hewitt (1985) present a number of reasons for their belief in the prevalence of intrinsic selection against hybrids over environment-dependent selection: First, they believe that direct evidence indicates that hybrids are often less fit than parental types. Second, Barton and Hewitt find it difficult to envisage how selection could act independently on many loci, which must occur if the hybrid zone is based upon environment-dependent selection. Even if the initial genetic divergence was in response to environmental differences, differential co-adaptation is likely to evolve (Harrison, 1990). Third, Barton and Hewitt argue that the geographic patterns do not fit environment-dependent selection. They suggest that a more broken pattern of clinal variation would be expected with environment-dependent selection. Moreover, although hybrid zones are frequently correlated with environmental gradients this does not necessarily mean that they are causally related. Tension zones can become associated with environmental gradients because of their tendency to come to rest where the two forms have equal fitness or because the meeting of two populations in secondary contact happens to coincide with an environmental gradient. Furthermore, hybrid zones are sometimes located at local physical barriers (e.g. rivers or scree), which is not consistent with environmental-dependent selection. Fourth, many characters show coincidence. If selection is environment-dependent, Barton and Hewitt suggest that the clines should be more scattered, because it is unlikely that all the environmental variables will be aggregated. In contrast, coincidence of clines in a zone maintained by intrinsic selection against hybrids may be brought about by: (i) the influence of population structure, e.g. density troughs or range constrictions may 'trap' clines in the same place; (ii) the attraction of overlapping clines; or (iii) it may reflect the first point of secondary contact between two expanding populations (Barton and Hewitt, 1985; Butlin *et al.*, 1991).

Barton and Hewitt (1985) conclude: "*The characteristic spatial configuration of most hybrid zones and the wide range of genotypes found within them persuade us that the great majority are maintained in a stable balance between dispersal and selection. It is harder to distinguish whether parapatrically distributed forms remain distinct because they are adapted to different environments, or because hybrids between them are less fit. However, both the direct evidence of hybrid unfitness and the indirect evidence of the close concordance of different characters lead us to believe that the latter is more likely, and that most hybrid zones are in fact "tension zones."*"

#### ***1.2.2.4 Mosaic hybrid zone model***

All the models described above depict hybrid zones as areas of (relatively) smooth transition from one form to another which can be represented by simple monotonic clines. However, it has become apparent that not all hybrid zones can be depicted in this way and, therefore, the mosaic model was formulated (Harrison, 1986; Harrison and Rand, 1989).

In this model, the environment is a mosaic of patches, representing different physical environmental factors, resources or habitats. The transition from one environment to another may not comprise a succession of intermediate environments; instead it may reflect a change in the relative proportion of patch types (Harrison, 1990). The hybridising taxa differ in their fitness in or preference for patch types and interact at the boundaries between patches. The genetic structure of the hybrid zone is a reflection of this underlying spatial structure. Thus a mosaic hybrid zone is characterised by a patchy distribution of different genotypes (Arnold, 1997) and a linear transect through the zone will reveal major reversals in marker frequencies.

Harrison and Rand (1989) suggest that the applicability of the mosaic model depends upon the dispersal ability of the taxa involved, relative to average patch size. If dispersal distance (of the organisms involved) is low in comparison to patch size, the hybrid zone will be a mosaic. However, if dispersal distance is high in comparison to patch size, the hybrid zone will be observed as a simple monotonic cline rather than as a mosaic. The scale at which the hybrid zone is observed can be very important. If dispersal distance is much smaller than patch size, patch boundaries may be environmental gradients that can be represented by monotonic clines, although the hybrid zone remains a mosaic on a larger scale.

The dynamics of mosaic zones are rather different from zones based on simple clines. In a mosaic zone, there may be many local interactions between the hybridising species and each contact may effectively be independent. The outcome of any particular interaction will depend upon the relative abundance of the hybridising taxa, local selection pressures, random drift and the level of inter-patch genetic exchange (Harrison, 1990). These factors may vary from site to site within the hybrid zone. Consequently, local interactions cannot be used to predict the global outcome of the hybrid zone (Harrison and Rand, 1989) and global extinction of an allele or species is unlikely (Harrison, 1990).

### ***1.2.2.5 Arnold's evolutionary novelty model***

Arnold (1997) argues that the bounded hybrid superiority model and the mosaic model are unsatisfactory for explaining hybrid zones. However, he reserves his greatest disapproval for the tension zone model. Firstly, Arnold points out that many of the expectations of this model (e.g. linkage disequilibria and parallel clines) are not unique to tension zones. Therefore, hybrid zones which exhibit these characteristics should not be automatically assumed to be tension zones. Arnold proceeds to reconsider data from three putative animal tension zones and one plant hybrid zone, concluding that none of these four zones consistently demonstrate the genetic and ecological patterns predicted by the tension zone model. In particular, Arnold believes there is evidence of environment-dependent selection and hybrid genotypes that are fitter than their parents in certain environments. On a broader scale, Arnold draws attention to 37 studies in which the relative fitness of hybrids and parents has been estimated (Arnold and Hodges, 1995; Arnold, 1997). There is no evidence to suggest that hybrids are uniformly unfit. Instead, they fall into a range of fitness categories and, overall, most are as fit, or fitter, than the fittest parental genotype.

Based upon these observations and opinions, Arnold proposes a new hybrid zone model. The "evolutionary novelty" model can be summarised as follows: (i) The formation and establishment of F<sub>1</sub> hybrids, which may be a rare event; (ii) endogenous selection against certain hybrid genotypes, and exogenous selection for or against different hybrid genotypes; (iii) the formation of backcross and filial generations; (iv) further endogenous and exogenous selection; (v) the occupation by hybrid derivatives of novel habitats, or the replacement of parental taxa by more fit hybrid genotypes; and finally (vi) the prediction that evolutionary stable lineages will be established.

The evolutionary novelty model has been challenged by a number of authors. Ritchie (1998) refers to the model as "*more of an amalgam of... previous theories than a radically new viewpoint*". He states that "*I was not convinced that there are enough data clearly supporting Arnold's viewpoint*"; for example, he queries Arnold's evidence relating to the comparative fitness of hybrids and parents. Moreover, Eckenwalder (1998) criticises the narrative form of the model and its lack of rigorously testable predictions.

### ***1.2.2.6 Conclusion and examples***

All the models described above have advantages and disadvantages. No model is applicable to all situations, indeed it could be argued that no model can really provide an adequate explanation for the complex patterns observed in nature. Nevertheless, when studying a hybrid zone it is often possible to select the most appropriate model for that instance. For example, Wang *et al.* (1997; 1999) concluded that the hybrid zone between two subspecies of *Artemisia* could best be represented by the bounded hybrid superiority model. The mosaic model has been used to describe the hybrid zone between two spruce species (Perron and Bousquet, 1997) and the hybrid zone between two oak species (Howard *et al.*, 1997). Perhaps unsurprisingly, Arnold (1997) implies that the hybrid zone between *Iris hexagona*, *I. fulva* and *I. brevicaulis* (e.g. Arnold and Bennett, 1993) is a good example of his evolutionary novelty model! Used wisely, a model may provide helpful insights into the structure and dynamics of a stable hybrid zone, as well as generating hypotheses for future study.

### **1.2.3 The fate of hybrid zones**

Various fates may befall a hybrid zone (Harrison, 1993, Futuyma, 1998): (i) The hybrid zone may persist indefinitely (see above). However, it will not necessarily remain unaltered; for example, a tension zone can move. (ii) The hybrid zone may represent the "wave of advance" of a superior competitor, ultimately leading to the extinction of one of the hybridising taxa. (iii) The hybrid zone may represent neutral diffusion, resulting in the fusion of hybridising taxa. Harrison (1993) points out that fusion and extinction are not mutually exclusive, because both may involve random or non-random loss of alleles from each parent taxon. (iv) Reinforcement may take place (section 1.3.4); or, conversely, (v) the barriers to gene exchange may breakdown (section 1.3.4). (v) Finally, some of the hybrids in the zone may become reproductively isolated from the parental types, creating a new taxon (sections 1.3.6 and 1.3.7).

### **1.3 The evolutionary consequences of hybridisation and introgression in plants**

Research has highlighted several possible consequences of hybridisation and introgression with evolutionary significance in plants. These are: an increase in genetic diversity, the transfer of adaptations, the origin of new adaptations, the breakdown or reinforcement of reproductive barriers, a dispersal mechanism and the origin of new types (Rieseberg and Wendel, 1993). Each of these consequences, which are not mutually exclusive, will be considered in turn.

#### **1.3.1 An increase in genetic diversity**

Early researchers believed that hybridisation and introgression were important sources of genetic variation. Hybrids would possess alleles from both parents and exhibit new single and multilocus genotypes and, consequently, could have greater evolutionary potential than their parents (Rieseberg and Wendel, 1993). An increase in genetic variation has indeed been observed in many hybrid plant populations. For example, isozyme analysis indicated that hybrid populations of three *Aesculus* species were more genetically diverse than parental populations (dePamphilis and Wyatt, 1990). Likewise, populations from the hybrid zone between two races of *Castanea sativa* in Turkey displayed higher levels of genetic polymorphism (in terms of expected heterozygosities of allozymes) than populations outside the hybrid zone (Villani *et al.*, 1999).

However, Rieseberg and Wendel (1993) argue that we must distinguish between recently hybridised or introgressant populations (those that have experienced recent biparental gene-flow) and stabilised introgressants (those that have not experienced recent biparental gene-flow). Stabilised introgressants may be less genetically diverse than parental progenitors, due to the possibility of a single origin, subsequent founder effects and genetic drift in small populations, etc..

#### **1.3.2 The transfer of adaptations**

It has been proposed that adaptations may be transferred from one taxon to another via introgression. Rieseberg and Wendel (1993) point out that this is a difficult process to verify experimentally. Although it may be possible to establish, with reasonable certainty, that introgression has taken place, it does not necessarily follow that genes of adaptive

importance will have been transferred. It could be argued that introgression is synonymous with the transfer of adaptations, because beneficial alleles are predicted to cross a hybrid zone more easily than neutral molecular markers (used to document introgression). However, Rieseberg and Wendel consider this assumption premature because 'neutral' molecular markers vastly outnumber advantageous alleles. If introgression between two taxa (A and B) is confirmed and taxon A acquires an adaptation that is typical of taxon B, it is tempting to assume this proves that transfer of an adaptation has taken place. However, the adaptation could have been obtained by taxon A independently of gene flow from taxon B; for instance, hybridising taxa may gain analogous adaptations via convergent evolution, if both are subject to similar selection pressures.

Several examples of postulated inter-specific transfer of adaptations can be found in the literature. For example, Heiser (1951) suggested that *H. annuus* was able to colonise eastern Texas due to the introgression of beneficial alleles from *H. debilis* ssp. *cucumerifolius*, an adapted native species. Molecular work (cpDNA, rDNA) provided support for the hypothesised introgressive origin of Texan *H. annuus* (ssp. *texanus*) but Rieseberg *et al.* (1990a) pointed out that "*evidence for introgression does not necessarily prove that the introgression of H. debilis ssp. cucumerifolius into H. annuus was in any way adaptive*". Keim *et al.* (1989) and Paige *et al.* (1991) suggested that the introgression of nuclear and mitochondrial genes from *Populus fremontii* (a low elevation species) into *P. angustifolia* (a high elevation species) was enabling *P. angustifolia* to become better adapted to lower elevations.

### **1.3.3 The origin of new adaptations**

Rieseberg and Ellstrand (1993) drew attention to the common occurrence of novel characters in plant hybrids. Novel characters may be the product of (i) new multilocus genotypes; (ii) new alleles; or (iii) reduced developmental stability (Rieseberg and Ellstrand, 1993). New multilocus genotypes may result from the complementary action of new combinations of normal alleles, unexpressed (or expressed) alleles being placed under a new pattern of regulation or the fixation (e.g. through forced selfing) of recessive alleles present in the heterozygous condition in the parents (Rieseberg, 1995). New alleles have been attributed to an increased mutation rate in hybrids, reduced selection or intragenic recombination between different alleles of the parental taxa (Golding and Strobeck, 1983;

Rieseberg and Wendel, 1993). It is likely that at least some of these novel characters will be adaptive (Rieseberg and Wendel, 1993, Rieseberg *et al.*, 1999).

The origin of new adaptations in plant hybrids has not been experimentally verified but circumstantial evidence does exist (Rieseberg and Wendel, 1993). One example is provided by Rieseberg's (1991) work on *Helianthus* hybrid species. *Helianthus anomalus*, *H. deserticola* and *H. paradoxus* are found in markedly different habitats to their parental species, *H. annuus* and *H. petiolaris*. *H. anomalus* and *H. deserticola* are adapted to xeric conditions and *H. paradoxus* to salty wetlands, while the parents occupy environments that are more moderate. Neuffer *et al.* (1999) proposed that introgression from *Viola reichenbachiana* into *V. riviniana* has resulted in novel genotypes that can invade polluted pine forests in the Dübener Heide (Germany).

Evolution by natural selection relies on plentiful genetic variation but, under normal circumstances, mutation creates genetic variation rather slowly. In contrast, hybridisation and introgression may lead to a rapid increase in genetic diversity, the transfer of adaptations and the origin of new adaptations. Thus, hybridisation and introgression may play a vital role in the plant kingdom as "*a source of variability upon which selection can act*" (Rieseberg, 1995).

#### **1.3.4 The breakdown or reinforcement of reproductive barriers**

The theory of reinforcement was first proposed by Dobzhansky in the 1940's and has been a matter of controversy ever since (Howard, 1993). Reinforcement may be defined as "*the evolution of prezygotic isolating barriers in zones of overlap or hybridisation (or both) as a response to selection against hybridisation*" (Howard, 1993). The predicted outcome of reinforcement is reproductive character displacement: "*a pattern of greater divergence of an isolating trait in areas of sympatry between closely related taxa than in areas of allopatry*" (Howard, 1993). In the extreme, reinforcement should result in the cessation of hybridisation (Rieseberg and Wendel, 1993).

The hypothesis of reinforcement has been rejected by many because of theoretical difficulties (Phelan and Baker, 1987). Howard (1993) refuted many of the criticisms, and argued that the hypothesis could not be rejected on theoretical grounds alone. Instead, he advocated studying natural populations for evidence of reinforcement, for example reproductive character displacement and positive assortative mating in hybrid zones. Some

authors (e.g. Phelan and Baker, 1987) have suggested that such evidence is lacking. However, a literature review by Howard showed that reproductive character displacement was a common phenomenon in nature, and that deviations from random mating expectations in hybrid zones were frequent (indicative of positive assortative mating). Nevertheless, Howard points out that "*we still lack... the detailed studies of hybrid zones and of individual cases of reproductive character displacement necessary to determine whether the patterns suggestive of the operation of reinforcement can indeed best be explained by this process*".

Most of the papers reviewed by Howard were zoological, and only three potential cases of reproductive character displacement in plants were identified (out of a total of 48). The most convincing study was of *Phlox pilosa* and *P. glaberrima* (Levin and Kerster, 1967); in which reproductive character displacement (corolla colour), inter-specific matings and selection against hybridisation were documented.

In conclusion, it seems sensible to support Howard's view that both the case for reinforcement and the case against reinforcement remain not proven. Howard (1993) believed that much more data were needed and that "*the eventual fate of the model will be determined by some difficult field and laboratory studies*".

Hybridisation may lead to the breakdown of reproductive isolating barriers and to the eventual merger of previously isolated taxa. Breakdown occurs when selection weakens the barriers to gene exchange, by favouring those variants that display the smallest reduction in viability and fertility when crossed with either parental type (Harrison, 1993). Although this hypothesis is uncontroversial, there are only a few examples in the literature (Rieseberg and Wendel, 1993). Hybridisation between *Clarkia nitens* and *C. speciosa* ssp. *polyantha*, two chromosomally differentiated taxa, has led to the production and establishment of new chromosome arrangements that genetically link the parental species. These new chromosome arrangements have become distributed throughout the hybrid zone and allow inter-specific gene-flow with little or no loss of fertility (Hauber and Bloom, 1983).

### **1.3.5 A dispersal mechanism**

Potts and Reid proposed that hybridisation through pollen dispersal may be of evolutionary significance as a means of gene dispersal where a taxon is limited by restricted seed

dispersal (Potts and Reid, 1988). Potts and Reid's (1988, 1990) hypothesis was based upon their study of two *Eucalyptus* species. They discovered that *Eucalyptus risdonii*, a localised endemic, was expanding into the range of *E. amygdalina*, a more broadly distributed species. Dispersal of *E. risdonii* seed was limited, but the flow of *E. risdonii* genes into the range of *E. amygdalina* by pollen dispersal and subsequent inter-specific hybridisation was widespread. Therefore, Potts and Reid suggested that *E. risdonii* was invading favourable habitat islands via pollen-mediated gene-flow through *E. amygdalina*. Potts and Reid's theory was attacked by Schemske and Morgan (1990) on both methodological and theoretical grounds. Schemske and Morgan argued that hybridisation between the *Eucalyptus* taxa "*might be better viewed as an accidental consequence of the evolutionary history of the two species, rather than as an 'evolutionary advantage' arising from 'increased dispersal potential'*".

### **1.3.6 The origin of new homoploid types**

The origin of new homoploid types (i.e. at same ploidy level as the parents) is one of the most frequently documented evolutionary consequences of hybridisation and introgression (Rieseberg and Wendel, 1993). Homoploid taxa derived via hybridisation may be new species, sub-species, races or ecotypes and are commonly referred to as stabilised introgressants, that is "*populations that breed true for an alien allele (or alleles)*" (Rieseberg and Wendel, 1993).

The genetic composition of stabilised introgressants can vary considerably. For example, the type of cross made by early hybrid generations may be important (Abbott, 1992). If hybrids backcross to one parental taxon, this may result in stabilised introgressants that are genetically, and phenotypically, referable to the backcross taxon but with a character or characters inherited from the other parent (Abbott, 1992). In contrast, if hybrids inbreed or cross with their siblings, this may produce stabilised introgressants in which the genetic contribution of each parent is more evenly balanced. Such hybrids may have a phenotype intermediate between the parents or resemble an extreme variant form of one parent (Abbott, 1992).

The maintenance of a stabilised introgressant depends upon the stability of its hybrid gene complexes, which in turn depends upon prevention of recombination and segregation (Abbott, 1992). In other words, gene flow between the stabilised introgressant and its parental taxa must be restricted. To a large extent, it is the level of parent-hybrid gene

exchange that determines whether the stabilised introgressant is referred to as a new species, sub-species, race, or merely an ecotype - to recognise a new species one would obviously expect greater isolation than would be necessary to recognise a new sub-species, ecotype or race (Abbott, 1992).

Of the many examples of putative stabilised introgressants, only a few have been validated using molecular markers. Intra-specific taxa that have been confirmed in this way include *Helianthus annuus* ssp. *texanus* and *Senecio vulgaris* var. *hibernicus*. *Helianthus annuus* ssp. *texanus* is the product of hybridisation between *H. annuus* and *H. debilis* ssp. *cucumerifolius* (Rieseberg *et al.*, 1990a; see above). The radiate groundsel, *Senecio vulgaris* var. *hibernicus*, is the result of hybridisation between non-radiate *S. vulgaris* and radiate *S. squalidus* (Abbott *et al.*, 1992a, see chapter 6). Examples of homoploid hybrid species are noted below.

#### ***1.3.6.1 Homoploid hybrid speciation***

Homoploid hybrid speciation is a form of sympatric speciation because the parent species must geographically co-occur to hybridise. It has been proposed that homoploid hybrid species may arise, i.e. become isolated from parental gene flow, in one of two ways – speciation with external barriers or recombinational speciation (Grant, 1981).

##### *Homoploid hybrid speciation with external barriers to gene exchange*

This is "the formation of a new recombination type in the progeny of a hybrid which is separated from the parental species by external isolating mechanisms" (Grant, 1981). The scenario can be outlined as follows: (i) Hybridisation between two interfertile plant species leads to the formation of diverse recombination types for character differences between the parent species; (ii) some of the new character combinations bring about external reproductive isolation, of one sort or another, between the hybrid derivatives and the parents; and (iii) the externally isolated recombinants are maintained if well-adapted to an available environment (Grant, 1981). The external reproductive isolation might be ecological, phenological, mechanical, ethological, or a combination of these (Grant, 1981). An alternative scenario involves the chance colonisation by recombinant individuals of a favourable allopatric location, completely removing hybrid individuals from parental gene-flow and competition. Such spatial separation eliminates the need for ecological,

ethological, etc. isolation between the hybrid species and its parents, as long as they remain apart (Rieseberg, 1997).

*Recombinational homoploid hybrid speciation*

In contrast, recombinational speciation requires the development of chromosomal and/or genic sterility barriers between the neospecies and its parents (Rieseberg, 1997). The basic model can be summarised as follows: (i) The two parental species are distinguished by a chromosomal sterility barrier composed of two or more separable chromosome rearrangements, and/or by a genic sterility barrier; (ii) their partially sterile hybrid gives rise, via segregation and recombination, to new homozygous (or true breeding) recombinant types for the chromosome rearrangements and/or the genic sterility factors; and (iii) the recombinant types are fertile within the line but at least partially sterile with parental taxa (Grant, 1981; Rieseberg, 1997). Although this model stresses the importance of postmating barriers to speciation, it would be wrong to assume that premating barriers are unimportant. Ecological divergence between a recombinational hybrid species and its parents increases the likelihood of hybrid establishment and persistence, because it may enable the hybrid to coexist with one or both parents or occupy a new habitat (Charlesworth, 1995; Rieseberg, 1997). Without such divergence, or spatial isolation between hybrid and parents, a fitness advantage of the hybrid over its progenitors may be critical (Morrell and Rieseberg, 1998).

McCarthy *et al.* (1995) used a computer simulation to examine the feasibility and dynamics of recombinational speciation, and their findings were generally consistent with expectation. Recombinational speciation was facilitated by a great selective advantage for the hybrid type, a long hybrid zone interface (which boosts the number of hybrid matings), increased  $F_1$  fertility, a high level of selfing (which increases the frequency of the union of balanced gametes) and a low number of chromosomal differences between the parental species (because the chance emergence of the optimal recombinant karyotype is more likely). The simulation also suggested that recombinational speciation dynamics might be characterised by long-term hybrid zone stasis followed by an abrupt transition to the optimal recombinant type. Such a transition would be favoured by a high number, and aggregated distribution, of the optimal type.

The recombinational model has been experimentally verified by the artificial synthesis of several new "hybrid" species, under conditions of both weak and strong sterility barriers (Morrell and Rieseberg, 1998). One of the most elegant studies was conducted by Grant (1966a; 1966b) on *Gilia modocensis* and *G. malior*. F<sub>1</sub> hybrids between the two species were highly sterile, with low chromosome pairing. However, several generations of selfing, with artificial selection for vigour and fertility, led to the production of a fully fertile homoploid hybrid derivative. The hybrid derivative was vigorous, had a distinctive combination of morphological characters and was intersterile with both parent species. More recently, Rieseberg *et al.* (1996) investigated the role of gene interactions in recombinational speciation. Concordance in genomic composition among three experimentally synthesised hybrid lineages (*Helianthus annuus* x *H. petiolaris*), suggests that hybrid species formation is largely governed by selection rather than chance. In these *Helianthus* lineages, selection was acting primarily to increase fertility (the hybrids were produced under glass and there was a rapid rise in fertility from F<sub>1</sub>s to fifth generation hybrids). Furthermore, the genomic composition of the synthetic hybrids was similar to that of the ancient hybrid species, *H. anomalus* (itself the product of hybridisation between *H. annuus* and *H. petiolaris*). This congruence implies that the genomic structure and composition of recombinational hybrid species may be stabilised within a few generations and remain fairly constant thereafter. Rieseberg *et al.* (1996) also analysed parental marker segregation in the synthesised hybrid lineages. In each lineage, non-random rates of introgression and significant associations among unlinked markers were noted. This indicates that the genomic composition of hybrid species may be constrained by interactions between the coadapted genes of the parental species. While a majority of interspecific gene interactions may be negative or neutral, a small number may be favourable and provide the raw material for adaptive evolution in hybrid taxa.

Burke *et al.* (1998a) used RAPDs to study genetic interactions in F<sub>2</sub> hybrids between *Iris fulva* and *I. brevicaulis*. Their results suggested that heterospecific nuclear and cytonuclear epistatic interactions were important for determining F<sub>2</sub> hybrid viability. Like Rieseberg *et al.* (1996) above, Burke *et al.* (1998a) concluded that "*the traditional view that interactions between divergent genomes are always deleterious is an oversimplification. Rather, it seems likely that crosses between divergent lineages can lead to the production of both fit and unfit hybrid genotypes*".

*Examples of homoploid hybrid species*

The frequency of homoploid hybrid speciation in nature remains unclear; partly due to the difficulty of unambiguous documentation (section 5.4.1). The literature contains over 50 putative homoploid hybrids but only a minority have been rigorously tested with molecular markers (Rieseberg, 1997).

Confirmed recombinational homoploid hybrid species include: *Helianthus anomalus* (Rieseberg, 1991), *H. deserticola* (Rieseberg, 1991) and *H. paradoxus* (Rieseberg *et al.* 1990b; Rieseberg, 1991), *Iris nelsonii* (Arnold, 1993) and *Stephanomeria diegensis* (Gallez and Gottlieb, 1982). The three *Helianthus* species are all derived from hybridisation between *H. annuus* and *H. petiolaris*, demonstrating that segregation of sterility factors from one set of parents can yield multiple reproductively isolated derivatives (Morrell and Rieseberg, 1998). *Helianthus anomalus*, perhaps the best studied of any recombinational species, is semisterile with its parents due to meiotic difficulties, apparently caused by chromosomal structural differences. Genetic linkage maps (constructed using RAPDs) for *H. anomalus*, *H. annuus* and *H. petiolaris*, showed that the hybrid species had undergone extensive genome reorganisation relative to its parents, resulting from the merger of pre-existing structural differences between the parents and chromosomal rearrangements seemingly induced by recombination (Rieseberg *et al.*, 1995). Furthermore, the data indicated that three large linkage groups from *H. annuus* had been transmitted intact into *H. anomalus*. Rieseberg *et al.* (1995) speculated that structural differences in the parental genomes protected these three blocks of markers from recombination during speciation.

Homoploid hybrid species that are isolated from their progenitors by external barriers to gene exchange include: *Paeonia emodi* and other *Paeonia* species (Sang *et al.*, 1995), *Pinus densata* (Wang and Szmidt, 1990; Wang *et al.*, 1990; Wang and Szmidt, 1994), *Encelia virginensis* (Allan *et al.*, 1997) and *Penstemon clevelandii* (Wolfe *et al.*, 1998). For example, Allan *et al.* (1997) suggested that *E. virginensis* is isolated from its parents by strong ecological selection against backcross progeny, although the precise nature of this selection is unknown (Kyhos *et al.*, 1981).

*The biology of homoploid hybrid species*

A number of similarities and differences among the homoploid hybrid species detailed above can be highlighted (Morrell and Rieseberg, 1998). First, all the listed species have diverged ecologically from their parents. The ubiquity of ecological divergence indicates that separating species into recombinational and non-recombinational categories may be rather arbitrary. For the non-recombinational species, ecological divergence is the primary isolating factor between hybrid and parental taxa, while in the case of both non-recombinational and recombinational species ecological divergence reduces competition between hybrid and parental taxa by allowing niche/habitat differentiation. Interestingly, several of the listed species occupy habitats that are novel or extreme, rather than intermediate between the parents. For instance, the *Helianthus* hybrid species occur in more xeric or marshy habitats than either parent (Rieseberg, 1991) and *Pinus densata* grows at higher altitude than its parents (Wang and Szmidt, 1990).

Some of the listed hybrid species (e.g. the *Helianthus* hybrid species; Rieseberg, 1991) appear to have undergone transgressive segregation, i.e. they possess phenotypic characters that are novel or extreme relative to parental characters (section 1.3.3). Rieseberg *et al.* (1999) speculate that transgressive segregation may be an important mechanism for hybrid adaptation, and thereby responsible for the niche and habitat divergence described above.

All the listed species are outcrossers, confounding the theoretical expectation that rates of hybrid speciation should increase with selfing (McCarthy *et al.*, 1995). This could be an artefact of the small sample size. A second possible explanation is that although selfing may assist hybrid establishment, this is counterbalanced by a lower level of hybridisation among selfing taxa (Rieseberg *et al.*, 1998). Alternatively, conditions in hybrid zones might promote selfing, even in normally outcrossing species. Rieseberg *et al.* (1998) recorded increased selfing in three hybrid zones of (normally outcrossing) *Helianthus* species. However, the observed increase in selfing was in the parental-like fraction of the populations and was, therefore, unlikely to facilitate hybrid speciation (semisterility prevented selfing in the critical hybrid fraction of the populations).

A fair proportion of the listed species are annuals, which is surprising because annual hybrids are less likely to be found and identified than perennial hybrids (Ellstrand *et al.*, 1996).

With regards to genetic diversity, there is no overall trend - some of the listed hybrid species were more diverse, some equivalent and some less diverse than their parents (section 5.4.5). Contemporary hybridisation has been recorded between the parents of a number of the listed species, e.g. among the parents of *Iris nelsonii* but not among the parents of *Paeonia emodi*.

Finally, the listed species differ in their distribution, some are fairly rare and restricted, e.g. the *Helianthus* hybrids, others are more abundant and widespread, e.g. *Pinus densata*. Morrell and Rieseberg (1998) suggest the differences in geographical range might reflect the availability of suitable habitat and/or time since origin.

### **1.3.7 The origin of new allopolyploids**

New allopolyploids are formed by a combination of hybridisation and polyploidization (chromosome doubling - via somatic chromosome doubling or the fusion of 2x gametes). It has been suggested that a majority of plant taxa are of polyploid ancestry (Arnold, 1997) and that most polyploids are allopolyploids (autopolyploids may be subject to fertility problems caused by multivalent formation at meiosis). This indicates that the origin of allopolyploids is one of the most significant consequences of hybridisation.

Polyploidization is important as a means of circumventing hybrid sterility and hybrid breakdown (Grant, 1966b). Hybrids are often sterile because chromosomes from the two parents are too dissimilar to form bivalents during meiosis. This sterility is overcome by polyploidization, because each chromosome can pair with its newly created double. Furthermore, allopolyploidy is one-step sympatric speciation, instantaneously leading to reproductively isolated species (Novak *et al.*, 1991). Crosses between allopolyploids and their parental taxa are generally unsuccessful because the progeny tend to have unbalanced chromosome sets (Thompson and Lumaret, 1992).

In the past, allopolyploids were believed to be genetically depauperate. This view was based upon the assumption that each polyploid species arose only once, resulting in a genetically uniform species, and secondly, that the "buffering" effects of multiple genomes would retard the action of mutation and recombination (Stebbins, 1971; in Cook *et al.*, 1998). In contrast, recent research suggests that allopolyploid species may be genetically diverse. Multiple origins of polyploid species are the rule rather than the exception (Soltis

and Soltis, 1993; Soltis and Soltis, 1999) and this phenomenon may greatly increase a taxon's gene pool (if the parent populations are differentiated). If polyploids of independent origin come into contact and hybridise, subsequent segregation and recombination will generate yet more genetic variation (Soltis and Soltis, 1999). Genome evolution may also provide an important source of genetic diversity in polyploids. Intra- and intergenomic chromosomal rearrangements can occur and such changes may be both extensive and rapid (Bailey *et al.*, 1993; Song *et al.*, 1995). In addition, there may be gene silencing, which can lead to diploidization, and the regulatory or functional divergence of duplicate genes (Soltis and Soltis, 1993).

Allopolyploid species are often successful, they may be able to out-compete their parental progenitors, withstand broader environmental conditions than their parental progenitors and even exploit entirely new habitats (Thompson and Lumaret, 1992). The successful establishment and persistence of allopolyploid species can be attributed to a number of factors including: high levels of genetic diversity and novel gene combinations (see above); fixed heterozygosity which may promote hybrid vigour (Thompson and Lumaret, 1992); enzyme multiplicity, novel heteromeric enzymes and increased levels of enzyme activities (Roose and Gottlieb, 1976); greater reproductive flexibility e.g. the frequent breakdown of self-incompatibility (Thompson and Lumaret, 1992); and wider physiological tolerance (Thompson and Lumaret, 1992).

Many putative allopolyploids are believed to be the product of ancient hybridisation and polyploidization events (Soltis *et al.*, 1993) and consequently difficult to investigate. The most extensively studied allopolyploid species are of recent origin - and include two tetraploids, *Tragopogon mirus* and *T. miscellus*. These species were first described in 1950 by Ownbey, following their discovery in Washington and Idaho, USA (Ownbey, 1950; in Novak *et al.*, 1991). Using morphological, cytological and other procedures, Ownbey demonstrated that the diploid progenitors of *T. mirus* were *T. dubius* and *T. porrifolius*, and the diploid progenitors of *T. miscellus* were *T. dubius* and *T. pratensis*. The three parental species were introduced to America in the early 20<sup>th</sup> century, indicating that *T. mirus* and *T. miscellus* originated in the last 60-70 years. Molecular marker evidence (isozymes, cpDNA RFLP, rDNA RFLP) provided support for Ownbey's conclusions, and additional information about pathways of origin of the species (Roose and Gottlieb, 1976; Soltis and

Soltis, 1989;). Both allotetraploids have increased in number and geographical range since the 1950's (Novak *et al.*, 1991) and molecular markers have confirmed several independent origins of the species (Soltis and Soltis, 1991; Cook *et al.*, 1998).

## 1.4 Molecular markers and the study of hybrids and hybrid zones

### 1.4.1 Isozyme analysis

Isozyme and allozyme analysis are based upon enzyme polymorphism. Isozymes are "*distinct forms of an enzyme that have identical or nearly identical chemical properties but are encoded by different loci*", whereas allozymes are "*distinct forms of an enzyme encoded by different alleles at a single locus*" (Li and Graur, 1991). The terms 'isozyme' and 'allozyme' are often used interchangeably.

Isozyme and allozyme variation are surveyed by protein electrophoresis. Crude protein extracts are prepared and loaded onto a gel (typically starch or agarose), through which an electric current is passed. The electric current causes proteins to migrate, according to their overall charge, size and shape (Avisé, 1994). Following electrophoresis, the gel is incubated with a histochemical stain specific for a certain enzyme. Discrete bands are formed at the positions in the gel to which the enzyme has migrated. The number of bands resolved, per individual, depends upon: the number of loci present, whether each locus is homo- or hetero-zygous, the ploidy level of the individual, and the molecular configuration of the enzyme (Weising *et al.*, 1995).

Isozyme analysis was one of the first molecular marker techniques to be developed and it is still in widespread use because of its many advantages. It is cheap, comparatively easy to perform, quick, allows a high through-put of samples and is reasonably reliable and robust. Markers are codominant, allowing heterozygotes and homozygotes to be distinguished unambiguously and genotype frequencies to be accurately assessed. The genetic basis of marker inheritance is clearly understood (and has been frequently tested) and methods of data analysis are well established (e.g. Avisé, 1994).

Nevertheless, there are complications in the interpretation of banding patterns (e.g. cryptic differences in the migration of alleles (Schaal *et al.*, 1991), null alleles and difficulties in confirming the homology of co-migrating bands between more distantly related species). More importantly, compared to sampling at the level of the DNA, isozyme analysis has a number of disadvantages (Bachmann, 1994). Firstly, an insufficient number of isozyme markers may be obtained for the species under study (DNA molecular markers are

normally found in much higher abundance than isozyme markers). Isozyme analysis is often less sensitive than DNA-based techniques (for example, the latter may reveal silent mutations and some amino acid substitutions that protein electrophoresis cannot resolve). Isozyme markers sample only coding regions of the genome whereas DNA molecular markers enable both coding and non-coding regions of the genome to be sampled (markers in non-coding regions of the genome are more likely to be selectively neutral and may evolve more rapidly, as genetic change is less constrained in non-coding areas). Finally, isozyme expression may be tissue-specific and affected by the environment or the developmental stage of the organism (Peakall *et al.*, 1995).

Allozymes were used to study hybridisation in *Aesculus* (dePamphilis and Wyatt, 1989; 1990). Three parapatric *Aesculus* species formed a broad hybrid zone, which was apparently maintained by long-distance pollen dispersal via the movement of humming birds. Gallez and Gottlieb (1982) employed allozymes to investigate the putative homoploid hybrid origin of *Stephanomeria diegensis*. The taxon was a genetic composite of its presumed parental species, confirming its hybrid status.

#### **1.4.2 Restriction fragment length polymorphism (RFLP) analysis**

In a traditional RFLP analysis, restriction enzymes are employed to cut DNA into fragments. The resultant fragments are separated by gel electrophoresis, transferred to a membrane via Southern blot hybridisation and visualised using fluorescently or radioactively labelled probes (Avisé, 1994). The probe may be homologous (i.e. cloned from the species itself) or heterologous (i.e. cloned from another, related, species) (Bachmann, 1992). In RFLP analysis of chloroplast DNA (cpDNA), the probe is usually a known section of chloroplast DNA and in RFLP analysis of ribosomal DNA (rDNA) the probe is a fragment of the ribosomal DNA.

RFLP polymorphism can be attributed to restriction site mutations, the loss or gain of a restriction site resulting from a point mutation (e.g. the loss of a restriction site will result in the replacement of two restriction fragments with one of their combined length; Bachmann, 1992), or length mutations, the insertion, deletion or inversion of DNA (e.g. the insertion of a small section of DNA will result in the replacement of one restriction fragment with another larger fragment). Restriction site polymorphisms are thought to be

more reliable than length polymorphisms (because the allelic homologies of length mutations may be difficult to determine; Schaal *et al.*, 1991; Hong *et al.*, 1993).

#### ***1.4.2.1 Chloroplast DNA (cpDNA) and RFLP analysis***

Chloroplast DNA molecular markers may provide very different information from many nuclear DNA molecular markers. Chloroplast DNA has a conservative rate of evolution, and chloroplasts are inherited uniparentally (in angiosperms, generally maternally) and asexually, i.e. there is no sexual recombination (McCauley, 1995). The effective population size of chloroplast genes is approximately one-quarter that of nuclear genes (as a result of uniparental and asexual transmission), causing an increase in the rate of gene fixation via drift and a decrease in expected gene diversity (Rieseberg and Soltis, 1991). In general, intraspecific cpDNA variation is fairly low but spatial structuring may be pronounced, especially in angiosperms (Hong *et al.*, 1993; McCauley, 1995). Chloroplast DNA is not always selectively neutral and chloroplast genomes may undergo periodic selection (Hong *et al.*, 1993).

Hybrids are normally recognised by an additive pattern of parental molecular markers. Consequently, chloroplast DNA markers, on their own, cannot be used to identify hybrids (because of the uniparental and asexual inheritance of chloroplasts). Even at the population level there is a much lower probability of detecting and correctly diagnosing a case of cytoplasmic introgression, than a case of nuclear gene introgression (Rieseberg and Soltis, 1991); the likelihood of maintaining two different chloroplast genomes (e.g. alien and native) in a population, for any length of time, is less than the likelihood of maintaining nuclear gene diversity (due to the lower effective population size of the former). Indeed, cytoplasmic gene flow may even result in the complete replacement of a native cytoplasm by an alien one, within a lineage, making it very difficult to detect cytoplasmic introgression. The probability of complete replacement increases with time elapsed since the transfer of the cytoplasm, and decreases with greater geographical range of the taxon.

It is clear, therefore, that to identify hybrid individuals or populations cpDNA markers must be used in conjunction with nuclear markers (Rieseberg and Soltis, 1991). However, employed in this manner, chloroplast markers may be very valuable for a number of reasons: First, slow evolution and lack of recombination in the chloroplast genome may

reveal cases of ancient introgression (Schaal *et al.*, 1991). Second, looking at the chloroplast genome may help to confirm the polarity of a putative hybrid - putative parent relationship (i.e. confirm that a putative parent is ancestral to a putative hybrid; Rieseberg *et al.*, 1990b). Third, most angiosperms inherit chloroplast DNA maternally, so a combined study of cpDNA and nuclear markers is a means of comparing seed-mediated (cytoplasmic plus nuclear) gene flow to pollen-mediated gene flow (nuclear only) (McCauley, 1995). Finally, there is a suggestion that cytoplasmic gene flow between taxa occurs more frequently and readily than nuclear gene flow, and indeed may occur in the absence of nuclear gene flow (this phenomenon is known as 'chloroplast capture'; Rieseberg and Soltis, 1991). If this is the case, hybridisation and introgression events would go undetected if chloroplast markers were ignored. Chloroplast capture is well-documented and Rieseberg and Soltis (1991) discuss the various explanations for the process (e.g. the effects of selection and linkage, cytoplasmic male sterility, founder effects, semigamy etc.).

Chloroplast RFLP markers are co-dominant, highly reliable and reproducible and scattered throughout the chloroplast genome. However, although few workers find it necessary to clone their own probes (due to the wide availability of heterologous probes), traditional cpDNA RFLP remains one of the more laborious and time-consuming techniques and comparatively large quantities of pure DNA are required.

Whittemore and Schaal (1991) used cpDNA RFLP to investigate hybridisation among five sympatric white oak (*Quercus*) species in the eastern USA. Cytoplasmic introgression between the oak taxa was considerable, in that the distribution of chloroplast genotypes was concordant with the geographical location of populations, rather than species boundaries defined by morphology and nuclear rDNA. Smith and Sytsma (1990) employed cpDNA RFLP to study the origin of *Populus nigra* (black poplar). The chloroplast genome of *P. nigra* was very similar to that of *P. alba*, although the two species are in different taxonomic sections and have distinct nuclear genomes (based on nuclear rDNA RFLP). *Populus nigra* is probably the product of an ancient hybridisation event between *P. alba* or its ancestor/close relative (the maternal parent) and the immediate ancestor *P. nigra*, followed by backcrossing to the immediate ancestor of *P. nigra*.

#### **1.4.2.2 Ribosomal DNA (rDNA) and RFLP analysis**

Ribosomal DNA can be a useful nuclear marker. Polymorphisms typically occur in the non coding rDNA regions of the intergenic spacer (which consists of tandem repeats of simple sequences), whereas coding regions of the rDNA are generally conserved (Bachmann, 1992). The results of a rDNA analysis must be interpreted carefully (Weising *et al.*, 1995) because rDNA is repetitive and is therefore prone to concerted evolution (i.e. the homogenisation of rDNA sequences within a species) via unequal crossing over or gene conversion (Li and Graur, 1991).

Ribosomal DNA probes are usually heterologous in origin, due to the conserved nature of rDNA coding sequences across species, and polymorphisms are easy to detect because ribosomal DNA is abundant. However, like cpDNA RFLP, rDNA RFLP is time-consuming.

Arnold *et al.* (1990) used RFLP analysis of rDNA to investigate the pattern of interspecific introgressive hybridisation in allopatric and parapatric populations of *Iris fulva* and *I. hexagona*. Rieseberg (1991) employed RFLP analysis of rDNA to confirm the homoploid hybrid origin of three *Helianthus* species. Traditionally, hybrids have been expected to exhibit additive rDNA phenotypes, as was the case in the two examples above. However, it is now recognised that repeated backcrossing of a hybrid type to one parent or concerted evolution may lead to a hybrid derivative possessing the rDNA of only one of its putative parents, thus masking a hybrid origin (Fuertes Aguilar *et al.*, 1999). Concerted evolution may be slowed or prevented by a reduced rate of sexual recombination (e.g. vegetative reproduction or a long generation time) and presence of the rDNA on non-homologous chromosomes (Fuertes Aguilar *et al.*, 1999).

#### **1.4.2.3 PCR-RFLP or cleaved amplified polymorphic sequence (CAPS)**

PCR-RFLP or CAPS involves PCR amplification of a fragment of DNA, which is then digested by a range of restriction enzymes (Rafalski and Tingey, 1993). The restriction products are run out on a gel (e.g. polyacrylamide) and stained (e.g. with ethidium bromide). Like traditional RFLP, this technique produces codominant markers. However, in contrast to traditional RFLP analysis, it is faster (probes and Southern blotting are not required), allowing more enzymes to be investigated, and it uses less DNA. When studying

rDNA or cpDNA universal primers can be employed (e.g. in King and Ferris, 1998) but for the analysis of non-ribosomal nuclear DNA, primers often need to be constructed for each species (e.g. by random sequencing of *Pst*I genomic clones). Obviously, primer construction increases start-up costs and time required for development.

Ferris *et al.* (1993) used PCR-RFLP in a survey of cpDNA variation to investigate glacial refugia and postglacial migration in European *Quercus* species. Their results suggested that ancient cytoplasmic introgression had occurred between *Q. robur* and *Q. petraea* in a south eastern glacial refugium, and more recent cytoplasmic introgression had occurred between oaks from the south eastern and south western refugia.

### **1.4.3 RAPDs**

Multiple arbitrary amplicon profiling (MAAP; Caetano-Anollés, 1993) is the name given to a group of protocols that include randomly amplified polymorphic DNAs (RAPDs), arbitrary primed PCR (AP-PCR) and DNA amplification fingerprinting (DAF). These techniques produce arbitrary fragment length polymorphisms and are based upon a modified polymerase chain reaction (PCR) methodology, in that usually only one primer, with an arbitrary nucleotide sequence, is employed (Weising *et al.*, 1995). Whilst AP-PCR (Welsh and McClelland, 1990) and DAF (Caetano-Anollés *et al.*, 1991; in Caetano-Anollés *et al.*, 1992) are rarely practised, the use of RAPDs is extremely widespread (Weising *et al.*, 1995).

The popularity of RAPDs is due, in part, to the comparative simplicity of the technique. The RAPD procedure (Williams *et al.*, 1990) uses single, short - 10-bp - arbitrary oligonucleotide primers with a G-C content greater than 50%. Amplified products are separated by electrophoresis and visualised, typically by means of an agarose gel and ethidium bromide staining. On a gel, each individual is represented by a number of bands, which may range in size from approximately 200 bp to 2000 bp.

#### ***1.4.3.1 Molecular mechanisms underlying RAPD analysis***

For amplification to occur, a primer must anneal to complementary strands of the template DNA (Black, 1993). The distance between primer annealing sites must not be greater than a few kilobases (around 3000 bp), as PCR cannot normally amplify fragments larger than

this. Furthermore, the 3' ends of the annealed primers must face one other, indicating that annealing sites are inverted repeats (Black, 1993). Priming sites may be distributed throughout the genome, in repetitive DNA and single-copy DNA (Williams *et al.*, 1993), and in both non-coding and coding regions (Harris, 1999).

Not all amplifications result from a perfect primer to priming site match. The theoretical number of fragments that a single primer would be expected to produce, annealing with 100% homology, can be estimated from the primer length and target genome size (a large genome should, by chance, contain more suitable priming sites than a small genome). In reality, little correlation is observed between target genome size and the number of fragments per primer (Weising *et al.*, 1995). This lack of correlation, together with evidence from experiments by Williams *et al.* (1993), demonstrates the importance of competition for priming sites. Primers will preferentially anneal to sites with which they are most homologous. If many priming sites of 100% homology are present, as in a large target genome, primers are unlikely to anneal to sites of lower homology. In contrast, when fewer sites of high homology are available, as in a smaller target genome, primers may anneal to sites with a certain degree of nucleotide mismatch; this being permitted by the low annealing temperatures used in the RAPD procedure. In other words, the presence or absence of a RAPD band does not depend solely on the presence or absence of a particular DNA sequence.

Caetano-Anollés (1993) proposed a model to explain the interactions of molecular species formed during RAPD reactions. From this model, it is clear that the RAPD reaction process is both complex and stochastic.

### *RAPD polymorphism*

Two types of polymorphism are observed on RAPD gels: bands may be present or absent and/or they may vary in intensity. The presence or absence polymorphism may be caused by: (1) a large insertion between two primer annealing sites (making the fragment too large to be amplified); the creation or destruction of a priming site through: (2) insertion, deletion or inversion of a priming site; or (3) nucleotide substitutions in primer annealing sites (substitutions at the 3' end of the priming site having more effect than those at the 5' end; Stammers *et al.*, 1995). Explanations for intensity polymorphism include: heterozygosity, comigration of non-homologous fragments, product copy number

differences, competition between PCR products or degree of mismatching of primer sites (Venugopal *et al.*, 1993; in Harris, 1999).

RAPD data is usually scored as band presence versus band absence. It is inadvisable to score intensity differences between bands because the basis of such intensity variation is unclear (and cannot be guaranteed to have a genetic source; Harris, 1999).

#### ***1.4.3.2 Assumptions and problems***

##### ***Reproducibility of RAPD fragments***

The potential lack of reproducibility of RAPD profiles is considered to be one serious disadvantage of the technique. RAPD reproducibility may be influenced by numerous factors including: (1) the type of DNA polymerase used; (2) magnesium ion concentration; (3) dNTP concentration; (4) choice of primer and primer concentration; (5) template DNA concentration; (6) type of thermocycler and (7) temperature profile (Black, 1993; Pammi *et al.*, 1994 and Weising *et al.*, 1995). Nevertheless, although replication of RAPD results in different laboratories may be difficult (e.g. due to the use of different thermocyclers), within a single laboratory reproducibility of an acceptable standard can be achieved with care (Weising *et al.*, 1995 but see Pérez *et al.* (1998) for an opposing view). Reaction conditions must be optimised at the start of a study and the conditions strictly adhered to. To ensure that a reasonable level of reproducibility is achieved, appropriate standards should be included in each RAPD reaction and, ideally, both DNA extractions (Russell *et al.*, 1993) and RAPD reactions should be repeated. Conservative scoring of RAPD gels is recommended, for example scoring only high intensity fragments and those in the middle molecular weight range (Bachmann, 1994; Stewart and Excoffier, 1996). Finally, some suggest that gels should be independently scored by a number of people and the results compared, to reduce scoring errors and bias (Gomes *et al.*, 1998).

Contamination, from the laboratory or from the material sampled (e.g. commensal micro-organisms), poses another problem for users of RAPDs. Contamination can be reduced to an acceptable level through prudent laboratory working practices (Weising *et al.*, 1995) and care in the growing and harvesting of plant material (Bachmann, 1994). A negative control, which includes all reaction ingredients bar the template DNA, can also be used (Haig *et al.*, 1994).

*Allelism and dominance*

Most RAPD polymorphism is scored as band presence versus band absence. It is assumed that all fragments of a certain size are referable to a unique locus and that each locus can be treated as a bi-allelic system (Lynch and Milligan, 1994). One allele (the marker allele) is amplifiable and produces a fragment, while the alternative allele is a "null" allele and does not amplify. Absence of a band represents the homozygous recessive state (two null alleles). Presence of a band represents both the homozygous dominant state (two marker alleles) and heterozygote state (one marker and one null allele) - as the marker allele is "dominant" to a null allele on a gel; thus RAPD fragments segregate as dominant markers.

The assumption that each locus has one amplifiable allele is not always fulfilled. Occasionally, multiple amplifiable alleles, in the form of different length variants, exist at a single locus. These result from small insertions and deletions in the region between two priming sites and can be scored as codominant markers (Lynch and Milligan, 1994). Codominant markers may accidentally be scored as if they are dominant, because few workers carry out the extensive crossing programmes needed to test marker segregation patterns (Harris, 1999). However, studies have shown that the number of codominant RAPD markers is very low. For example, Williams *et al.* (1993) and Fritsch and Rieseberg (1992) found that less than 5% of RAPD fragments behaved as codominant markers and Echt *et al.* (1992) identified no codominant RAPD markers at all. Therefore, the presence of unidentified codominant markers will not necessarily introduce significant errors into a data set.

The assumption that there is only one null allele per locus may also be incorrect, instead there may be a pool of heterogeneous unamplifiable alleles, each resulting from a different mutation. One null allele per locus is more likely to be a valid assumption when considering closely related individuals (e.g. from the same population) than at higher taxonomic levels (e.g. between species; Black, 1993). It is not possible to easily test the hypothesis of multiple null alleles at a locus and, consequently, the absence of a RAPD band should not normally be used as a character or marker.

### *Homology of RAPD fragments*

It is presumed that comigrating RAPD fragments, i.e. fragments of apparently the same size, are homologous (i.e. have the same sequence and origin). This is not necessarily the case; non-homologous fragments may comigrate due to inadequate gel resolution or chance size convergence of unrelated fragments. Homology can be tested using a Southern blot technique or restriction digest analysis (Rieseberg, 1996; see section 2.9). Rieseberg (1996) found 91% homology between 220 comigrating RAPD fragments. However, linkage mapping data suggested that around 13% of the homologous fragments were of paralogous rather than orthologous origin (i.e. similar through gene duplication rather than by descent, and possibly associated with repetitive DNA).

It is presumed that errors due to non-homology of comigrating fragments will increase with increasing genetic distance between the individuals being compared. Thus, the use of RAPDs for investigating intraspecific populations and closely related species is more satisfactory than their use for higher level systematics (Harris, 1999). Whenever possible the homology of RAPD fragments should be tested rather than assumed (whilst remembering that neither technique for homology testing is very sensitive and that the methods do not test for orthology). In addition, improving gel resolution will reduce the number of misclassified fragments. This may be achieved by running gels over a longer time period or switching to a higher resolution gel system e.g. polyacrylamide gels with silver-staining (Rieseberg, 1996).

### *Genome sampling*

RAPD fragments of cytoplasmic origin are likely to diverge at different rates from RAPD fragments of nuclear origin because nuclear DNA has a greater mutation rate than chloroplast or plant mitochondrial DNA (Li and Graur, 1991). In addition, cytoplasmic RAPDs will tend to be inherited in a uni-parental manner unlike nuclear RAPDs, which are inherited bi-parentally. However, the small size of cytoplasmic genomes, in comparison to the nuclear genome, leads to the theoretical expectation that the source of most RAPD markers will be the nuclear genome (Weising *et al.*, 1995). This expectation appears to be broadly met, as a number of studies have shown that only a small percentage of bands may originate in the chloroplast or mitochondrial genomes. For example, it has been demonstrated that the mitochondrial and chloroplast genomes make a small contribution to

the RAPD phenotypes of *Brassica* species and *Beta vulgaris* (Lorenz *et al.*, 1994; Thormann *et al.*, 1994). Occasionally, larger numbers of cytoplasmic bands have been discovered, for example 45% of the RAPD fragments scored in the Douglas fir were mitochondrial in origin (Aagaard *et al.*; 1995) but this is believed to be the exception rather than the rule.

It is supposed that sampling is random within a genome, but this may not be true. The high G-C content of RAPD primers means that G-C rich regions of the genome may be preferentially screened (Harris, 1999). Rapidly evolving, repetitive DNA may also be preferentially screened because this class of DNA contains an especially high number of inverted repeats, that are thought to act as primer annealing sites (Black, 1993).

### *Inheritance*

RAPD markers are assumed to show Mendelian inheritance. The inheritance of RAPD markers has been tested and confirmed as Mendelian on a number of occasions (e.g. Dawson *et al.*, 1993). Non-Mendelian inheritance of RAPD fragments may be explained by the origin of fragments in uniparentally inherited cytoplasmic DNA or artefactual bands.

Ideally, the inheritance of RAPD fragments should be tested, using an F<sub>2</sub> or back-cross generation, in all studies. This is often not done due to time constraints, especially if dealing with a species with a long generation time.

### *Independence of RAPD fragments and artefactual bands*

Different primers may anneal at the same (with a degree of nucleotide mismatch) or overlapping priming sites and this will result in production of identical or interdependent fragments. Within a single DNA amplification reaction, the non-independence of RAPD fragments can be ascribed to a number of processes. First, product competition may occur, meaning that the amplification of certain bands may interfere with, or prevent, the amplification of other bands (Black, 1993). Second, there is the phenomenon of nested priming sites, in which priming sites are present within the amplified copies of genomic fragments (Rabouam *et al.*, 1999). Other sources of non-independent RAPD fragments include: cryptic codominance of markers, fragments associated with repetitive DNA sequences, heteroduplex formation and non-genetic artefacts. It is possible to confirm the

independence of RAPD fragments by southern blotting and/or segregation tests; unfortunately such tests require considerable effort and are rarely carried out. Simple ways to reduce the number of interdependent fragments are to score only a small number of bands (1-2) per primer and to use a large number of primers.

A related problem is that of artefactual bands, since many non-independent fragments may also be artefactual. Rabouam *et al.* (1999) investigated the origin of RAPD fragments in samples of *Calonectris diomedea* (Cory's shearwater) and *Haemonchus contortus* (Nematoda) and discovered that many of the RAPD bands were artefacts. Some artefactual bands appeared to be attributable to amplification of DNA from commensal micro-organisms. Other artefacts were attributed to interactions between molecular species formed during DNA amplification; including heteroduplex formation, intrastrand interactions, interstrand interactions and nested primer annealing.

#### *Analysis of RAPD data*

As the majority of RAPD markers are dominant, RAPD data cannot be used to calculate allele frequencies directly. This is a significant difficulty, as traditional population genetic methods (e.g. for estimating and partitioning genetic diversity) rely on the use of allele frequencies. Some have tried to overcome this complication (e.g. Haig *et al.*, 1994; Aagaard *et al.*, 1995) by using the Hardy-Weinberg equation to estimate allele frequencies, and other population genetic parameters, indirectly from phenotype frequencies. Lynch and Milligan (1994) mention several precautions that should be taken to increase the likelihood of obtaining a credible and unbiased estimate of allele frequencies from RAPD data: in the order of two to ten times more individuals per locus must be sampled compared to codominant markers; the marker alleles for a majority of the loci should be at relatively low frequency (1-3/no. of individuals); and more loci need to be sampled (compared to codominant markers). To estimate other population genetic parameters, such as within and between population genetic diversity, Lynch and Milligan recommend a minimum of 100 individuals per population and that sample sizes across populations should be roughly equal. Even if the stringent conditions of Lynch and Milligan are met, which is rare, the approach may still be of dubious value because the implicit assumption of one amplifiable allele and one non-amplifiable allele per locus may not always be met. Although the problem of multiple amplifiable alleles can be dealt with to some extent, the problem of

multiple null alleles cannot (Lynch and Milligan, 1994). In addition, some populations will not meet the assumptions of the Hardy-Weinberg equation (very large population size, random-mating etc.; Stewart and Excoffier, 1996). Due to these difficulties, some have concluded that estimation of allele frequencies from RAPD data is a flawed process (Haig *et al.*, 1994). Instead, the direct analysis of RAPD phenotype frequencies is recommended (Haig *et al.*, 1994), e.g. Shannon's index of phenotypic diversity and analysis of molecular variance (see chapter 2).

For determining genetic relationships between OTUs (operational taxonomic units), using RAPDs, there are two choices. Firstly, RAPD data can be treated as distance data. In this case, a binary matrix of RAPD band presence/absence is converted into pairwise genetic distances between OTUs. The genetic distances can then be used in a cluster analysis (e.g. UPGMA) or in an ordination procedure such as principal coordinate analysis (see chapter 2). Distance approaches are the mainstay of phenetics, in which OTUs are grouped together on the basis of overall similarity. Alternatively, RAPD data could be treated as discrete or character-state data. In this case, the binary matrix would be utilised directly e.g. in a parsimony or maximum likelihood analysis. Character-state approaches are employed in cladistics, when the grouping of taxa reflects inferred evolutionary relationships (based on the presence of shared-derived characters). Character-state methods are often preferred to distance methods because the latter are thought to result in a loss of information from the data. However, it is now accepted that RAPD data should not generally be treated as character-state data (Harris, 1999) because RAPD characters will frequently not meet the strict independence and homology criteria necessary (Harris, 1999). Even a few non-homologous, i.e. homoplastic, characters can have a dramatic effect on the outcome of a character-state analysis (Adams and Demeke, 1993).

### *Summary*

The assumptions of, and problems associated with, the RAPD technique might lead one to imagine that RAPDs are of little use. This is definitely not the case. RAPDs have a comparatively long history of successful use in studies of population and evolutionary biology, and continue to be employed today. Often, the problems associated with the RAPD technique can be reduced to an acceptable level. For example, poor reproducibility can be minimised by careful laboratory practice and conservative gel scoring, and the

dominance of RAPD markers can be dealt with by using methods of analysis based upon phenotype frequencies (rather than allele frequencies). Likewise, the various assumptions, e.g. the homology of co-migrating bands, need not be the cause of undue concern. Many assumptions may be valid, especially in studies within and between closely related species (rather than deeper level systematic studies; Harris, 1999) and assumptions can be tested e.g. tests for homology, Mendelian inheritance and the independence of bands (although these tests are frequently not carried out). Finally, the impact of the assumptions can be negated by acknowledging them (and likelihood of their being met) and interpreting results in the light of this information.

Moreover, it is important to remember that all molecular markers have their disadvantages and are based upon assumptions that may not always be met. Indeed, as a comparatively mature molecular marker technique, critical analysis of RAPDs is to be expected, whereas other, more recently developed techniques (e.g. AFLPs) have yet to be fully appraised with respect to their assumptions and disadvantages.

#### ***1.4.3.3 Advantages of RAPDs***

The practical advantages of RAPDs are numerous: the technique is easy, non-hazardous, quick to carry out, allows a high through-put of samples, requires no developmental work or sequence knowledge, can be comparatively cheap (low start-up and running costs) and uses only very small quantities of DNA (Hadrys *et al.*, 1992). Furthermore, the RAPD technique normally produces reasonable numbers of polymorphic markers and it is believed that a large proportion of the genome is sampled (Harris, 1999).

Hybridisation in *Cardamine* (Brassicaceae) was studied using RAPDs (Neuffer and Jahncke, 1997). It was established that two diploid *Cardamine* species had crossed and produced a triploid hybrid derivative (which had an additive profile of parental RAPD markers). Furthermore, polyploidization of the triploid led to the creation of a new allohexaploid species. Padgett *et al.* (1998) used RAPDs to confirm the hybrid origin of *Nuphar x rubrodisca*, and De Greef and Triest (1999) employed RAPDs to examine interspecific hybridisation in *Scirpus* (Cyperaceae).

#### **1.4.3.4 RAPD primer-pairs**

RAPD primers can be employed not only individually, but also in pair-wise combinations (Micheli *et al.*, 1993; Gomes *et al.*, 1998). RAPD profiles obtained using two primers simultaneously are distinct from those obtained with each single primer (Weising *et al.*, 1995). In general, most of the fragments generated using two primers at the same time are different from those produced by the two primers used separately (Weising *et al.*, 1995). This observation has two possible explanations. Firstly, smaller products generated by two primers together (because the average distance between two suitable priming sites is less) out-compete the longer products produced by either primer used singly, during the extension period of the RAPD reaction (Welsh and McClelland, 1991). The second explanation involves the competitive nature of primer target site selection (Weising *et al.*, 1995). Fragments generated by a single primer can form hairpin structures (since they have palindromic ends) and therefore the primer must compete with hairpin formation. In contrast, fragments produced by a pair of primers cannot form hairpin structures (their ends are non-palindromic) and so the primers do not have to compete with hairpin structure formation.

Using RAPD primer-pairs can increase the number of polymorphisms derived from a restricted number of RAPD primers. However, Weising *et al.* (1995) argue that bands generated by primers used singly and in combination must not be included in the same data-set, because an identical fragment might appear twice. Weising *et al.* conclude that using primer-pairs will normally be a less effective method of acquiring new information than employing additional single primers.

#### **1.4.4 Inter simple sequence repeat markers (ISSRs)**

ISSR analysis (Zietkiewicz *et al.*, 1994) is closely related to RAPD analysis. However, whereas RAPDs uses primers of arbitrary sequence, ISSR analysis uses primers designed from microsatellite regions. A typical ISSR primer is longer than a RAPD primer (around 17-18 bp long), consists of short (di-, tri- or tetra-) nucleotide repeat motifs and has an anchoring sequence of one to three nucleotides at the 5' or 3' end (this anchoring sequence is generally random). The addition of the anchoring sequence prevents strand-slippage artefacts that might otherwise occur (Wolfe *et al.*, 1998). Unique banding patterns are produced if different anchoring sequences are added to a common repeat motif (Wolfe *et*

*al.*, 1998), although this may not apply if the anchoring sequences include redundant bases (e.g. (CA)<sub>8</sub>-RY and (CA)<sub>8</sub>-AT may yield similar bands). ISSR analysis often employs higher annealing temperatures than RAPD analysis; for example, Charters *et al.* (1996) used an annealing temperature of 55°C (more than ten degrees higher than the annealing temperature in a typical RAPD reaction). ISSR bands, are generated in an analogous manner to RAPD bands (i.e. the region between two simple sequence repeat priming sites will be amplified if the sites are on complementary strands, correctly orientated and not too far apart). Likewise, ISSR polymorphisms result from similar processes to those causing RAPD polymorphisms.

ISSR markers share many of the advantages (e.g. easy to use, cheap), disadvantages (e.g. dominance) and assumptions (e.g. homology of co-migrating bands) associated with RAPD markers. Some workers believe that ISSR markers are more reliable and reproducible than RAPD markers, which could be attributed to the higher annealing temperatures (reducing template-primer mismatch artefacts) and longer primers (Charters *et al.*, 1996; Wolfe *et al.*, 1998). Other workers have found little or no difference in reliability and reproducibility between RAPDs and ISSRs (Hollingsworth *et al.*, 1998). It has also been suggested that ISSRs generate more polymorphic markers per primer than RAPDs (Zietkiewicz *et al.*, 1994; Esselman *et al.*, 1999). However, Hollingsworth *et al.* (1998) found no such difference between the two methods. Hollingsworth *et al.* point out that the sensitivity of RAPDs, ISSRs and similar techniques may, to large extent, be dependent on the resolution of the gel separation system employed. In other words, the choice of high-resolution polyacrylamide or low-resolution agarose gels can have a greater influence on the number of polymorphic fragments per primer than the choice of ISSRs or RAPDs.

Wolfe *et al.* (1998) used ISSRs to study hybridisation in populations of *Penstemon*. The ISSRs provided a higher degree of resolution than other molecular methods (allozymes, rDNA and cpDNA RFLP) and enabled Wolfe *et al.* to confirm the hybrid status of *P. clevelandii* (which had an additive profile of parental ISSR markers) and reject a hybrid origin for *P. spectabilis* (which did not have an additive profile). Hollingsworth *et al.* (1998) employed ISSRs to investigate clonal diversity and genotype inter-relationships among Japanese knotweed (*Fallopia japonica*), giant knotweed (*F. sachalinensis*) and their hybrid *F. x bohemica*.

### 1.4.5 Amplified fragment length polymorphisms (AFLPs)

Amplified fragment length polymorphisms (AFLPs) are produced by a combination of restriction digestion and PCR procedures. AFLP markers are typically developed as follows (Ridout and Donini, 1999): The genomic DNA is restricted with two restriction enzymes (one rare and one frequent cutting) and adapters are ligated to the restriction fragments generated. Non-selective primers are used in a first PCR step. An aliquot of the non-selective PCR product is then used in a second PCR step. The second PCR step is selective (uses primers with selective nucleotides) and ensures that only a subset of the original fragments are amplified. Amplified fragments are separated on a sequencing gel and visualised by means of radioactivity, silver-staining or, more commonly, an automatic sequencer. AFLP markers are thought to be distributed throughout the genome (Ridout and Donini, 1999).

The advantages of AFLPs have been heavily promoted by some researchers (e.g. Palacios *et al.*, 1999). The AFLP technique has a high effective multiplex ratio (i.e. a large number of polymorphic loci can be analysed simultaneously) and so a large number of molecular markers can be rapidly produced. Moreover, AFLPs are thought to be more reliable and reproducible than, for example, RAPDs (Palacios *et al.*, 1999).

However, AFLPs do have disadvantages. The start-up costs for AFLPs can be considerable. The technique can be difficult to master initially (Ridout and Donini, 1999). In particular, the visualisation of AFLP bands can be problematic; radioactivity is hazardous and does not always produce clear bands, silver-staining can be unpredictable and automatic sequencers must be used carefully (Escaravage *et al.*, 1998). Furthermore, AFLPs share many of the assumptions and problems associated with RAPDs (Harris, 1999): the markers are dominant, the origin of markers is unknown, locus and allele designations are unclear, the independence of bands cannot be guaranteed and the homology of comigrating bands has to be assumed (and is harder to test than in the case of RAPDs). Finally, the claim that AFLPs are more reliable and reproducible than RAPDs has yet to be fully substantiated.

O'Hanlon *et al.* (1999) employed AFLPs to study hybridisation in invasive *Onopordum* thistles. AFLPs have also been used to identify *Salix* hybrids (Beismann *et al.*, 1997) and *Populus* hybrids (Arens *et al.*, 1998).

### 1.4.6 Simple sequence repeat (SSR) or microsatellite analysis

Simple sequence repeat or microsatellite DNA consists of tandemly repeated sequences whose unit of repetition is between one and five base pairs (Jarne and Lagoda, 1996). A SSR or microsatellite marker is developed by designing primers for the flanking regions of a stretch of microsatellite DNA. PCR is then used to amplify the microsatellite DNA and the amplification products are separated on polyacrylamide (or high resolution agarose) gels and visualised by silver-staining or autoradiography.

An amplified microsatellite represents a single locus. Microsatellite markers behave in a codominant fashion and therefore one or two bands will be observed at a locus, depending on whether the locus is homo- or heterozygous. For each locus a large number of alleles may exist in a population and these are characterised by polymorphism in size i.e. variation in the number of repeat units. Microsatellite alleles evolve at a rapid rate and the mutation of the alleles is attributed to polymerase slippage at DNA replication and unequal crossing-over (Jarne and Lagoda, 1996). Mutation models are discussed by Jarne and Lagoda (1996).

Microsatellites have many advantages as molecular markers; they are codominant, highly reproducible, Mendelian, neutral and dispersed throughout the genome. For the population geneticist, microsatellites are often the molecular marker of choice.

However, microsatellites are perhaps not the optimal molecular markers that some have suggested. The process of developing microsatellites is often extremely laborious and may involve the cloning and enrichment of a genomic library (Weising *et al.*, 1995). Compared to RAPDs or AFLPs, fewer loci may be resolved in a particular species and only one locus can be characterised at a time. Visualising the microsatellites can be prone to difficulties as either radioactivity or silver-staining is required and slippage artefacts may occur. Microsatellites are perhaps one of the more expensive molecular marker techniques (Ridout and Donini, 1999). In addition to the technical problems, care must be taken when scoring and interpreting microsatellite data. For example, one must allow for the possibility of null alleles (Jarne and Lagoda, 1996) and potential homoplasy in microsatellite data (alleles must be identical by descent and not merely the same size and sequence; Jarne and Lagoda, 1996). In particular, microsatellites may be less than ideal for comparing species or widely diverged populations because of the increased occurrence of homoplasy (Harris, 1999) and other problems such as ascertainment bias (Jarne and Lagoda, 1996) and non-amplification of homologous loci in related species (Burke and Arnold, 1999).

Burke and Arnold (1999) discuss the development of five microsatellite loci cloned from *Iris fulva* and *I. brevicaulis*. The authors suggest that these highly polymorphic loci will enable them to carry out paternity studies in hybrid populations of the two *Iris* species. Bucci *et al.* (1998) used chloroplast microsatellite markers to examine hybridisation between two sympatric *Pinus* species in Turkey, and found good evidence of unidirectional introgression. Overall, microsatellites have not been widely used in studies of plant hybridisation. This may be due to a number of the factors discussed above; in particular the development work required and the problems associated with using microsatellites at the species level.

#### **1.4.7 DNA Sequencing**

The molecular marker techniques described above are indirect methods of studying polymorphisms at the DNA level. The most direct approach is DNA sequencing - to determine the nucleotide sequence of a defined DNA region in the genome of one individual and to compare this to the correctly aligned sequence of the orthologous region in the genome of a more or less related individual (Weising *et al.*, 1995). The popularity of sequencing has increased as the technique has become easier, through the advent of PCR and automated sequencers, and sequence data can be highly reproducible and informative. Disadvantages of sequencing include potential difficulties in sequence alignment, high cost and low through-put of samples. Indeed, sequencing may provide high resolution at the expense of examining a broad range of loci and individuals. Moreover, sequencing depends upon the availability of suitable primers; i.e. primers that can amplify a region of DNA that exhibits an appropriate level of variation for the study being undertaken. Many universal primers amplify fairly conserved regions of the genome e.g. the *rbcL* gene of the chloroplast (Clegg, 1993). For this reason (together with other considerations e.g. cost and sample through-put) sequencing is most commonly used in studies at the species level and above, i.e. phylogenetics rather than population genetics.

Fuertes Aguilar *et al.* (1999) sequenced the nuclear ribosomal internal transcribed spacers (ITS) of *Armeria villosa* ssp. *longiaristata*, *A. colorata*, and F<sub>1</sub>, F<sub>2</sub> and backcross hybrids between the two species. They found evidence of rapid homogenisation of ITS sequences in the hybrids, due to concerted evolution and/or backcrossing. F<sub>2</sub> individuals had ITS

sequences similar to those of *A. colorata*, backcross individuals had ITS sequences similar to the recurrent parent. The findings were consistent with the hypothesis of a hybrid origin of *A. villosa* ssp. *carrtracensis*, a putative hybrid between *Armeria villosa* ssp. *longiaristata* and *A. colorata*, which possesses the rDNA of *A. colorata*.

## 1.5 Aims and objectives

*Senecio squalidus* L., the Oxford Ragwort, is a diploid member of the Asteraceae. The species was introduced to Britain approximately 300 years ago, when a population was established in the Oxford Botanic Garden. After a century or so of cultivation, *S. squalidus* escaped from the garden and began to spread across Britain. It is now found in most non-rural areas of the country (Crisp, 1972). During its range expansion, *S. squalidus* has come into contact, on numerous occasions, with native *S. vulgaris* L. var. *vulgaris* (common groundsel). The two species hybridise, leading to the formation of three new taxa (chapter 6) - *S. cambrensis* Rosser, *S. vulgaris* var. *hibernicus* Syme and York Radiate Groundsel (Abbott and Lowe, 1996).

These evolutionary events generated interest in the origins of *S. squalidus*. Tradition held that the taxon had been introduced to the British Isles from Mt. Etna, Sicily. To investigate this possibility, Crisp (1972) travelled to Sicily and attempted to locate the source of the British material. However, Crisp did not find one species resembling British *S. squalidus* on Mt. Etna. Instead, he recorded three *Senecio* taxa similar to *S. squalidus*: low-altitude *S. chrysanthemifolius* Poir., mid-altitude *S. incisus* Presl. and high-altitude *S. aethnensis* Jan.. Crisp suggested that *S. chrysanthemifolius* and *S. aethnensis* were hybridising and that *S. incisus* was an array of hybrid swarms (chapter 4). Furthermore, he proposed that British *S. squalidus* was derived from individuals sampled from this hybrid zone (chapter 5). More recently, the putative hybrid origin of British *S. squalidus* has been investigated by Abbott (Abbott *et al.*, 1995; Abbott and Milne, 1995) using isozymes and cpDNA.

The aim of the present study, was to examine aspects of the evolutionary history of *S. squalidus* in greater detail. The first objective was to develop taxon-specific molecular markers for *S. chrysanthemifolius* and *S. aethnensis*. Second, it was intended to use these taxon-specific markers to study the putative hybrid zone between *S. chrysanthemifolius* and *S. aethnensis* on Mt. Etna. Third, the taxon-specific *S. chrysanthemifolius* and *S. aethnensis* markers were to be employed to confirm or refute the hypothesis of a Sicilian homoploid hybrid origin of British *S. squalidus*. If the hybrid origin of British *S. squalidus* was established, it was hoped that an estimate could be made of the proportion of its genome attributable to each parental species. Finally, the taxon-specific markers would be used to examine the products of hybridisation between *S. squalidus* and *S. vulgaris* var. *vulgaris*.

# **Chapter 2**

## **Materials and Methods**

## **Materials and Methods**

### **2.1 Origin of plant material**

Seed material of *Senecio chrysanthemifolius*, *S. aethnensis* and their hybrid was collected by the author on two field trips to the Mt. Etna area of Sicily (table 2.1). Seed material of British *S. squalidus*, *S. vulgaris*, *S. cambrensis* and the York Radiate Groundsel was taken from stocks held in the Laboratory of Plant Sciences, University of St. Andrews (table 2.2).

### **2.2 Cultivation of plant material**

Leaf material for DNA extraction was collected from plants grown from seed. Seeds were sown on damp filter paper in partitioned transparent plastic boxes and germination took place in a growth cabinet maintained at  $20^{\circ}\text{C} \pm 3^{\circ}\text{C}$ , with a 16 h photoperiod. Alternatively, seeds were sown in pots of damp compost covered with plastic sheeting and germination took place in the greenhouse. Young seedlings were potted on into pots containing a 3:1 compost and gravel mixture and kept in the greenhouse. Greenhouse illumination was supplied by natural daylight and 400-W metal halide lights set to a 16 h photoperiod and the temperature was maintained at  $20^{\circ}\text{C} \pm 3^{\circ}\text{C}$ . Plants were watered, fertilised and protected from insects and disease, as necessary.

**Table 2.1** Geographic locations, coordinates (latitude and longitude), altitude and date collected of populations of *S. chrysanthemifolius*, *S. aethnensis* and putative hybrids sampled from Mt. Etna, Sicily, and used in the present study. All samples were collected by the author.

Species or transect	Popn. Code	Locality	Lat. 00° 00' N	Long. 00° 00' E	Alt. m	Date
<b><u><i>S. aethnensis</i></u></b>						
	BB	below Bocca Superiore	37 44	14 59	2525	02.08.97
	VB	nr. Cisternazza	37 44	15 01	2600	01.08.97
	TJ	between VB and BB	37 44	15 00	2590	02.08.97
	TC	below Monte Dagalotto	37 47	15 01	2550	05.08.97
	UJ/VO	nr. observatory of vulcanology	37 46	15 01	2800	05.08.97
	S13	1/2 way between S14 and Rifúgio Sapienza	37 43	15 00	2200	19.06.96
	S14	cable car station, Píccola Rifúgio	37 43	15 00	2500	19.06.96
<b><u><i>S. chrysanthemifolius</i></u></b>						
	C0	northern Catania	37 32	15 05	150	13.06.96
	C1	nr. sports stadium, Pedara	37 37	15 04	600	13.06.96
	C9	Randazzo	37 53	14 57	750	16.06.96
<b><u>Transect S populations</u></b> (includes hybrids)						
	S2	minor road nr. Casa Maugen	37 38	15 00	950	14.06.96
	S5	restaurant car-park, just below road to Monte Rinazzelli	37 40	15 00	1300	14.06.96
	S6	nr. Monte Manfre	37 40	14 59	1400	15.06.96
	S8	~ ¼ km above Casa Costarelli	37 41	14 59	1500	18.06.96
	S9	~ 1 km below road to Grande Albergo del Parco	37 41	14 59	1600	18.06.96
	S10	just after road to Grande Albergo del Parco	37 41	15 00	1650	18.06.96
	S11	~ 1 ½ km below Rifúgio Sapienza	37 42	14 59	1800	02.08.97
	RD	above ski lift change-over, Capannina to Monte Silvestri	37 43	15 00	2175	02.08.97

**Table 2.2** Geographic locations, site descriptions, coordinates (latitude and longitude), altitude, collector and date collected of populations of *S. squalidus*, *S. vulgaris* var. *vulgaris*, *S. vulgaris* var. *hibernicus*, York Radiate Groundsel and *S. cambrensis* from the British Isles, used in the present study.

Species	Popn. Code	Locality	Lat. 00° 00'	Long. 00° 00'	Alt. ~ m	Collector (date)
<i>S. squalidus</i>	PB	HMS Osprey, <b>Portland Bill</b> , Dorset, England	50 28 N	02 26 W	< 50	JAI (03.09.90)
	OX	waste-ground, Osney St./Hollybush Rd., <b>Oxford</b> , England	51 45 N	01 09 W	< 50	RJA (22.05.91)
	CA	Barry Docks, <b>Cardiff</b> , Wales	51 24 N	03 08 W	< 50	JAI (07.09.90)
	NO	Riverside/Carran Rd., <b>Norwich</b> , England	52 40 N	01 20 E	< 50	PAA (07.10.90)
	HU	Myton St. car park, <b>Hull</b> , England	53 40 N	00 18 W	< 50	JAI (08.09.90)
	KY	car park/Coal Wynd, <b>Kirkcaldy</b> , Fife, Scotland	56 08 N	03 10 W	< 50	RJA (90)
<i>S. vulgaris</i> var. <i>vulgaris</i>	AF	<b>Aberffraw</b> , SW Anglesey, Gwynedd, Wales	53 09 N	04 34 W	< 50	RJA (87)
	PI	<b>Puffin Island</b> , nr. Beaumaris, E Anglesey, Gwynedd, Wales	53 13 N	04 09 W	< 50	RJA (87)
	GL	<b>Glenluce</b> , nr. Stranraer, Dumfries and Galloway, Scotland	54 45 N	04 44 W	< 100	DGF (23.07.93)
	LA	<b>Letham</b> , nr. Forfar, <b>Angus</b> , Scotland	56 36 N	02 42 W	< 200	DGF (06.93)
	TO	<b>Tomintoul</b> , Moray, Scotland	57 14 N	03 29 W	300	DGF (07.09.97)
	BV	Botanic Garden/Jamaica St.?, <b>Bristol</b> , England	51 24 N	02 36 W	< 50	AJL (23.06.93)?
	CV	<b>Cardiff</b> , Wales	51 25 N	03 09 W	< 50	RJA (01.06.87)?
	IV	<b>Birmingham</b> , England	52 25 N	01 54 W	< 50	DFM (87)?
	SV	<b>St. Helens</b> , Merseyside, England	53 23 N	02 45 W	< 50	PAA ?
	RV	<b>Brymbo</b> , nr. Wrexham, Wales	53 04 N	03 03 W	< 200	PAA (01.06.87 or 01.09.87)?

Table 2.2 continued.

Species	Popn. Code	Locality	Lat. 00° 00'	Long. 00° 00'	Alt. ~ m	Collector (date)
	OV	Cork and Passage West (St. Mary's and basket ball court), County Cork, Eire	51 55 N	08 27 W	< 50	AJL (16.06.93)
	YV	York, England	53 53 N	01 04 W	<100	AJL (30.06.93)
	EV	Bath Road, Leith, Edinburgh, Scotland	55 58 N	03 07 W	< 50	AJL (29.04.93)
<i>S. vulgaris</i> var. <i>hibernicus</i>	BH	Botanic Garden?, Bristol, England	51 24 N	02 36 W	< 50	AJL (23.06.93)
	CH	Cardiff, Wales	51 25 N	03 09 W	< 50	RJA (01.06.87)?
	IH	Birmingham, England	52.25 N	01 54 W	<50	DFM (87)?
	SH	St. Helens, Merseyside, England	53 23 N	02 45 W	< 50	PAA (01.09.87)?
	RH	Brymbo, nr. Wrexham, Wales	53 04 N	03 03 W	< 200	PAA (01.06.87 or 01.09.87)?
	OH	Cork, County Cork, Eire	51 55 N	08 27 W	< 50	AJL (16.06.93)?
	EH	Bath Road, Leith, Edinburgh, Scotland	55 58 N	03 07 W	< 50	AJL (29.04.93)
York Radiate Groundsel	YR	Dalton Terrace, York, England	53 53 N	01 04 W	<100	AJL (30.06.93)?
<i>S. cambrensis</i>	FC	Ffrith, nr. Wrexham, Wales	53 05 N	03 04 W	< 150	PAA (86 or 87)?
	SC	Southsea, Wrexham, Wales	53 03 N	03 02 W	<100	PAA (86, 87 or 90)?
	WC	Wrexham, Wales	53 02 N	03 00 W	< 100	PAA (86 or 87)?
	MC	Mochdre, nr. Colwyn Bay, Conwy, Wales	53 15 N	03 47 W	< 50	PAA (86 or 87)?
	EC	Leith, Edinburgh, Scotland	55 58 N	03 07W	< 50	PAA (86, 87, 90 or 93)?

Collectors: RJA (RJ Abbott), DGF (DG Forbes), AJL (AJ Lowe), JAI (JA Irwin), PAA (PA Ashton), DFM (DF Marshall).

### 2.3 DNA extraction

A number of problems are associated with the isolation and purification of genomic DNA from plants (Weising *et al.*, 1995). The extraction of DNA from *Senecio* material was difficult due to high levels of viscous polysaccharides and the presence of polyphenols and quinonic compounds. To obtain DNA of reasonable quality and quantity a large scale CTAB procedure, modified by Wolff (personal communication) from Weising *et al.* (1995), was used.

The large scale CTAB method is time-consuming and therefore two small scale extraction methods were investigated - a minipreparation method (based upon that of Weising *et al.* (1995) and a scaled down version of the CTAB procedure, to fit in a 2 ml microcentrifuge tube. Unfortunately, these small scale extraction methods produced very poor quality DNA and, consequently, were not employed.

#### *CTAB procedure*

Approximately 2 g of fresh leaf material was harvested from a plant, placed in a plastic bag and left in a refrigerator overnight to de-starch. The following morning, leaf material was flash frozen with liquid nitrogen and ground with a pestle and mortar. A small spatula of sterile alumina was added to aid grinding. The addition of liquid nitrogen and subsequent grinding was repeated once or twice until the leaf material was a fine powder. Once the powder had thawed slightly, 9 ml of CTAB extraction buffer with 2% 2-mercaptoethanol (appendix A), pre-warmed to 55°C, was added to the mortar. The powder and buffer were thoroughly mixed and the resulting suspension was transferred to an oak-ridge centrifuge tube. The tube was incubated at 55°C (in a water-bath) for 15 min and then cooled (in a cold water-bath) for 5 min. Half the volume, 4.5 ml, of wet chloroform (appendix A) was added to the tube and it was vortexed briefly. The tube was then centrifuged for 10 min at 8000 rpm, in a Sorvall<sup>®</sup> RC5C centrifuge (Sorvall Instruments). The aqueous supernatant was brought into a clean centrifuge tube, using a sterile wide-bore pasteur pipette. Another 4.5 ml of wet chloroform was added and the centrifugation repeated, as above. Following the second centrifugation, the aqueous supernatant was transferred to a sterile 50 ml plastic tube, using a sterile wide-bore pasteur pipette. Approximately twice the volume (~30 ml) of cold, -20°C, ethanol was added to precipitate the DNA. Using a sterile glass hook, the DNA was spooled from the interface of the supernatant and ethanol. The glass hook, with

the DNA attached, was placed in a 15 ml sterile tube containing 3-4 ml of sodium acetate wash buffer (appendix A). After 10 min, the hook was placed in a fresh tube of sodium acetate wash buffer and left to soak for 50 min. The hook was then removed, dipped briefly in ammonium acetate wash buffer (appendix A) and the end of the hook broken off into a 1.5 ml microtube containing 1 ml of TE (appendix A). To ensure that the DNA dissolved completely, the microtube was placed in a shaking incubator at 37°C for 2 h and then left overnight in a refrigerator.

#### *Cleaning the DNA*

An RNase step and ammonium acetate step were performed to remove RNA and polysaccharides, respectively. To the microtube containing the dissolved DNA, 20 µl RNase solution (appendix A) was added. After 1 h on a rotary shaker, half the volume, 500 µl, of cold (4°C) 7.5 M ammonium acetate was added to the microtube. The microtube was cooled at 4°C for 20 min and then centrifuged at 13000 rpm for 15 min in a bench-top microfuge (*Biofuge pico*, Heraeus Instruments). The resulting supernatant was divided between two new 1.5 ml microtubes and sufficient isopropanol was added to fill each microtube. Microtubes were mixed gently but thoroughly before placing at -20°C. After 30 min, the microtubes were removed from the freezer and centrifuged for 15 min at 13000 rpm. The supernatant was discarded, leaving a pellet of DNA in each microtube to which 200µl of cold (-20°C) 70% ethanol was added. The microtubes were flicked to wash the pellets and the microtubes were spun again for 15 min at 13000 rpm. The supernatant in each microtube was discarded and the microtubes inverted to ensure that most of the remaining ethanol was removed from the DNA pellets. The microtubes were left at room temperature for approximately 20 min to dry the pellets. Pellets were re-suspended in 150 µl of TE and the microtubes placed in a shaking incubator, at 37°C, for 2 h and then overnight in a refrigerator to dissolve the DNA. DNA was stored at 4°C or -20°C until needed.

#### **2.4 Running agarose gels**

For RFLP analysis 0.8% (w/v) agarose gels were used, for quantification of DNA 1% agarose gels were employed and for RAPD analysis 1.4% (w/v) gels were used. Large gels (300 ml volume) were employed for most procedures, but small gels (50 ml volume) were

occasionally employed for DNA quantification and RAPD analysis. The gel rigs used were Gallenkamp maxicell<sup>®</sup>s for large gels and Gallenkamp minicell<sup>®</sup>s for small gels.

The appropriate amount of agarose was added to the appropriate volume of 0.5X TBE (appendix A) buffer in a conical flask. The flask was heated in an 850-W microwave, with regular vigorous swirling. Large gels required 3 min 20 s of heating, small gels about 1 min 30 s of heating. The hot flask of agarose was placed on a rotary shaker and left to cool until it was "hand hot" (~ 60°C). Ethidium bromide, 1 mg/ml, was added to the gel solution as required. The gel solution was poured into a level plastic gel mould, which had been sealed at both ends with masking tape. Immediately, a comb or combs were added, parallel with the ends of the mould. Once the gel had set (approximately 1 h at room temperature), the tape and comb(s) were removed and the gel was placed in a gel rig. The gel was immersed in 0.5X TBE buffer and a Gilson Pipetteman was used to load DNA samples into the wells formed by the gel comb. Ethidium bromide, 1 mg/ml, was added to the tank buffer as required and the gel was run at the correct voltage.

## **2.5 Quantification of the DNA**

Quantification of DNA samples was carried out using the ethidium bromide staining method (Weising *et al.*, 1995). This method also provides a means of checking that the DNA is not degraded and that the level of RNA contamination is low.

Known volumes of DNA samples (e.g. 2-5  $\mu$ l) were mixed with (~9  $\mu$ l) loading buffer (appendix A). The samples were electrophoresed on a 1% agarose gel (containing 25  $\mu$ l of ethidium bromide if a large gel, 8  $\mu$ l if a small gel). The gel was run at 90 V until the dye front had moved 2 cm from the wells. Standard bands of 100 ng and 200 ng (100 ng/ $\mu$ l  $\lambda$  DNA, GIBCO BRL) were also included on the gel. The gel was photographed under UV light using a gel image analysis system (Herolab, E.A.S.Y. Store software). The photograph was used to calculate the concentration of each DNA sample; by comparing the intensity of the DNA band of the sample with the intensity of the standard bands. The comparison was made by eye, although this was not believed to very reliable.

## 2.6 Additional DNA purification

DNA samples that amplified poorly in PCR reactions were cleaned using the GENE CLEAN<sup>®</sup> III kit (BIO 101). The kit uses a silica matrix (EZ-GLASSMILK<sup>®</sup>) which binds DNA but does not bind DNA contaminants.

A modified version of the manufacturer's protocol was employed as follows:

An aliquot of the DNA sample to be cleaned was placed in a new 1.5 ml microtube. Three volumes of sodium iodide (NaI) and 10% of the volume of TBE Modifier<sup>®</sup> were added to the microtube. Next, EZ-GLASSMILK<sup>®</sup> suspension was added, 5  $\mu$ l for the first 5  $\mu$ g of DNA and 1  $\mu$ l for each additional 0.5  $\mu$ g. After adding the EZ-GLASSMILK<sup>®</sup>, the solution was mixed and left at room temperature for 15 min, mixing every 2-3 min. The microtube was spun in a microfuge for 5 s and the NaI supernatant discarded. The microtube was spun again and the remaining NaI removed with a small bore pipette, leaving a pellet of EZ-GLASSMILK<sup>®</sup> and bound DNA. The pellet was washed with 700  $\mu$ l of NEW wash for 5 min, during which time the pellet was very gently re-suspended. After 5 min, the microtube was spun for 5 s and the supernatant discarded. This washing step was repeated two more times. Having removed the supernatant following the third washing step, the microtube was spun again for 5 s and the remaining NEW Wash removed with a small bore pipette. The pellet was left to air dry for 10 min.

Elution of the DNA from the EZ-GLASSMILK<sup>®</sup> was carried out as follows: 15  $\mu$ l TE was added to the pellet and the pellet was gently re-suspended. The microtube was placed in a 55°C water-bath for 5 min and then spun in a microfuge for 30 s to make a solid pellet. The supernatant, containing the eluted DNA, was removed and placed in a fresh microtube. A second elution step was then carried out by adding another 15  $\mu$ l of TE to the microtube containing the pellet and repeating as above.

After cleaning, amplification of the DNA samples was less problematic.

## 2.7 Restriction fragment length polymorphism (RFLP) analysis

### *Restriction digestion*

Restriction digests were set up in 0.6 ml microtubes. Each microtube contained 500 ng of DNA to which the mega-mix was added. Quantities of the mega-mix were calculated according to the number of restriction digests, with an additional amount to allow for pipetting errors. For each DNA sample the mega-mix contained: 3  $\mu$ l of the appropriate

10X enzyme buffer, 0.3  $\mu$ l 10 mg/ml BSA [bovine serum albumin], 1.2  $\mu$ l 10 mg/ml spermidine, 3U of the restriction enzyme and enough sterile distilled water to bring the total reaction volume (including DNA) up to 30  $\mu$ l. The mega-mix was vortexed and centrifuged before the correct amount was added to each DNA sample. Once the mega-mix had been added, the restriction digest tubes were mixed, centrifuged and left over-night in an incubator set at the correct temperature. The following morning, the restriction reactions were stopped by adding 6  $\mu$ l of RFLP loading buffer (appendix A) to each microtube. Tubes were stored in the refrigerator or freezer until needed.

In the present study the enzyme *Cla* I was used (table 2.3; Abbott *et al.*, 1995).

### *Southern blotting*

A large 0.8% gel, containing 100  $\mu$ l of ethidium bromide, was made. At the top of the gel was one row of 28 wells. The restriction digest samples were loaded into wells along with two  $\lambda$  DNA *Hind* III size markers (appendix A). The gel was run at 60 V overnight and until the dye front was approximately 1 cm from the end of the gel. Once the gel had finished running it was photographed over UV light and measurements were made of the bands in the marker lanes to enable the size of the restriction fragments to be calculated. The gel was trimmed (the portion above the wells and empty lanes were removed), measured and immersed in denaturation buffer (appendix A) for 30 min. The gel was washed twice in distilled water and immersed in neutralisation buffer (appendix A) for 30 min.

The Southern blot apparatus was constructed as follows: A 26 cm x 26 cm glass tray was filled, to one third of its volume, with 20X SSC (appendix A). A sheet of perspex (longer but narrower than the tray) was lain over the tray. Three pieces of 19 cm x 46 cm chromatography paper were cut, wetted with 20X SSC and placed over the perspex so that the ends of the paper (at least 8 cm on each side) were in contact with the SSC in the tray, forming a wick for the liquid. The gel was inverted onto the wick and any air bubbles gently removed. A piece of nylon blotting membrane (Electran<sup>®</sup>, BDH), the same size as the gel, was wetted with SSC and placed carefully onto the gel. A piece of chromatography paper (length and width 6 mm less than that of the gel) was briefly wetted in SSC and placed on the membrane. A glass pipette was rolled across the paper to remove any air bubbles. Cling film was positioned around the gel to ensure that the upward flow of 20X

SSC could only take place through the gel. Two pieces of dry chromatography paper (cut to the same size as the wetted paper) were placed on top of the wetted paper, followed by a stack of paper towels, a glass plate and a weight. The apparatus was left overnight to allow the DNA fragments from the gel to be transferred to the nylon membrane.

The following morning, the apparatus was dismantled. The membrane was removed, rinsed in 2X SSC for 5 min on a rotary shaker and left to dry. The DNA fragments were cross-linked (bound) to the membrane by exposing the membrane to UV light for 30 s and baking at 80°C for 2 h. Care was taken to ensure the membrane was labelled correctly and a corner was cut off to indicate which side the DNA was on. The membrane was stored at room temperature until needed.

#### *Probing the membranes*

DNA fragments were visualised by means of DNA-DNA hybridisation with pLsC (*Lactuca sativa* L.) cpDNA probes (cloned into the plasmid vector pUC18). The probes were a gift from Dr. J. D. Palmer and Dr. R. K. Jansen (Indiana University) and were maintained by Dr. S. A. Harris and Dr. K. Wolff. In the present study, the probe pLsC6 (Abbott *et al.*, 1995) was used (pLsC6 is synonymous with probe number 3 in Jansen and Palmer (1987); see Milne, 1997 p 32). The pLsC6 probe is 14.7 kb long and comes from the Large Single Copy region of the cpDNA molecule (Jansen and Palmer, 1987).

Membranes were probed using the non-radioactive DIG labelling method (which is faster and safer than radioactive labelling). The protocol was developed by Dr. K. Wolff (personal communication) from the protocols of Boehringer Mannheim and the CIMMYT Laboratory in Mexico (González-de-León *et al.*, 1995).

The principles of DIG labelling are as follows: A DNA probe is labelled with digoxigenin (DIG) coupled to dUTP (i.e. digoxigenin-11-dUTP) using the random-primed (hexamer) labelling method. The DIG-labelled probe is brought into contact with the immobilised (membrane blotted) target DNA and hybridises to homologous regions of the target DNA. The hybridised probe is immunodetected using anti-digoxigenin Fab fragments conjugated to alkaline phosphatase (the anti-DIG antibodies bind to the digoxigenin of the probe). When the chemiluminescent substrate CSPD<sup>®</sup> is added, the alkaline phosphatase causes the

enzymatic dephosphorylation of CSPD<sup>®</sup>. This leads to an emission of light (at a maximum wavelength of 477 nm) which is recorded on X-ray film. Consequently, the regions of the membrane to which the probe has bound, are visualised.

#### *Labelling of the probe*

Sterile distilled water was added to the probe DNA to achieve a total volume of 13.5 µl. Approximately 120 ng of probe DNA was required for the first membrane, with an extra 60 ng for each additional membrane. The DNA probe solution was boiled for 10 min and immediately quenched on ice to prevent re-annealing. The following solutions were then added to the probe DNA to give a total volume of 20 µl: 2 µl of 10X Klenow buffer (100 mM tris HCl pH 8.5, 100 mM MgCl<sub>2</sub>), 2 µl of 10X hexanucleotide mix (Boehringer Mannheim), 2 µl of 10X DIG DNA labelling mix (Boehringer Mannheim) and 0.5 µl of *Klenow* enzyme (Boehringer Mannheim). The mixture was shaken, spun in a microfuge and incubated overnight at 37°C. The following morning, 2 µl of 0.2 M EDTA was added to the probe solution to end the reaction.

The probe was purified as follows: The probe DNA was precipitated with 2.5 µl 4 M LiCl and 75 µl 96% ethanol and the tube left at -20°C for 2 h. The tube was then spun at 1300 rpm, in a microfuge, for 10 min. The supernatant was discarded and the pellet of probe DNA was washed with 100 µl of 70% ethanol. The tube was spun again (as above) and the supernatant discarded. The pellet of probe DNA was dried for 10-15 min in a 37°C incubator, re-suspended in 100 µl TE and left for 30 min in a 37°C incubator to dissolve. Labelled probe was stored at -20°C until needed.

#### *Pre-hybridisation and hybridisation*

Up to four nylon membranes were placed in a hybridisation cylinder (*Techne*), each separated by a nylon mesh. Pre-hybridisation buffer (appendix A) was added to the cylinder, 20 ml for the first membrane and 2 ml for each additional membrane. Pre-hybridisation was carried out at 65°C for 2 h, in a hybridisation oven (*Techne*).

The purified probe was boiled for 5 min, quenched immediately on ice and added to hybridisation buffer (appendix A), pre-warmed to 65°C. The pre-hybridisation buffer was poured out of the cylinder and the hybridisation solution (containing the probe) poured into

the cylinder, 10 ml for the first membrane and 1 ml for each additional membrane. Hybridisation was carried out at 65°C overnight, in the hybridisation oven.

#### *Post-hybridisation washes and detection*

After hybridisation, the membranes were removed from the cylinder and washed in low stringency solution (appendix A) at room temperature; two washes of 5 min. The membranes were then washed in medium stringency solution at 65°C, under constant agitation; two washes of 15 min. After rinsing in buffer 1 (appendix A) and 0.3% Tween 20 [polyethylene sorbitol monolaurate], the membranes were incubated at room temperature in buffer 2 (appendix A), under constant agitation. For each membrane, 25 ml of buffer 2 was required. The buffer 2 was removed after 30 min and replaced with fresh buffer 2, 25 ml of buffer 2 per membrane plus 1.7 µl of anti-DIG Fab fragments (Boehringer Mannheim) per membrane. The remaining washing steps were carried out at room temperature and under constant agitation on a rotary shaker. Two washes of 15 min were carried out, in buffer 1 (250 ml per membrane) and 0.3% Tween 20 and these were followed by a 5 min wash in buffer 3 (at least 20 ml per membrane) (appendix A). Finally, the membranes were incubated in 25 ml of CSPD<sup>®</sup> solution (appendix A). The incubation step was carried out at room temperature, on a rotary shaker, and lasted 5 min.

#### *Luminography*

Membranes were drained of CSPD solution, sealed in transparent plastic bags and incubated at 37°C for 15 min to enhance the luminescent reaction. Each bag was placed in an autoradiograph cassette and covered with a sheet of X-ray film. The film was exposed for several hours.

## **2.8 RAPD and ISSR procedure**

### *Setting up RAPD and ISSR reactions*

RAPD and ISSR reactions were set up in a thermowell™ 96 (12 x 8) well plates (Costar) or, occasionally, in 0.6 ml microtubes. To each well or microtube, 1 µl of template DNA was added (0.25 ng or 0.5 ng DNA), followed by 24 µl of freshly prepared, thoroughly shaken mega-mix (making a total reaction volume of 25 µl). Quantities of the mega-mix were calculated according to the number of RAPD or ISSR reactions, with an additional amount to allow for pipetting errors. For each DNA sample the mega-mix contained: 18.55

$\mu\text{l}$  sterile distilled water, 0.5  $\mu\text{l}$  of 2 mM  $\text{MgCl}_2$ , 2  $\mu\text{l}$  of dNTP mixture (containing 2 mM dATPs, 2 mM dCTPs, 2 mM dGTPs and 2 mM dTTPs; HT Biotechnology), 0.2  $\mu\text{l}$  of 25 pmol/ $\mu\text{l}$  RAPD (Operon Technologies) or ISSR (University of British Columbia) primer, 2.5  $\mu\text{l}$  of 10X *Dynazyme* buffer and 0.25  $\mu\text{l}$  (= 0.5U) of *Dynazyme* polymerase (F-500L, Flowgen). The mega-mix was overlaid with approximately 40  $\mu\text{l}$  of mineral oil, to prevent evaporation, and plates were sealed with pressure sensitive plate sealers (Costar).

#### *RAPD primer-pairs*

RAPD reactions were also performed using RAPD primer-pairs. In this case, 0.1  $\mu\text{l}$  of each RAPD primer (25 pmol/ $\mu\text{l}$ ) was used per RAPD reaction.

#### *PCR programmes*

A PTC-100<sup>TM</sup> Programmable Thermal Controller (MJ Research) PCR machine was used for the PCR. For RAPD reactions the PCR machine was programmed as follows: 3 min of denaturation at 94°C; 45 cycles of 30 s denaturation at 94°C, 45 s of annealing at 40°C and (with a ramp rate of 0.4°C per s) 1.5 min of extension at 72°C; 4 min at 72°C and 5°C for 1 min. For ISSR reactions the PCR machine was programmed as follows: 3 min of denaturation at 94°C; 45 cycles of 30 s denaturation at 94°C, 45 s of annealing at 52°C and 2 min of extension at 72°C; 4 min at 72°C and 5°C for 1 min. Amplification products were stored at 4°C or at -20°C for longer periods.

#### *Running out the amplification products*

Loading buffer was added to the PCR products (9  $\mu\text{l}$  per sample) before they were loaded into the wells of a 1.4% gel (made with 25 $\mu\text{l}$  of ethidium bromide for a large gel, 8 $\mu\text{l}$  for a small gel). Normally, each large gel had two rows of 27 wells. A 1 kb ladder (Promega) was used as a marker, two markers were run per row. Ethidium bromide (100  $\mu\text{l}$  for a large gel rig, 25  $\mu\text{l}$  for a small gel rig) was added to the tank buffer. The gel was run at 40 V until the dye front left the wells and then at 90 V for approximately 4 h, until the dye front from the lower row of wells had travelled to within 1 cm or so of the end of the gel. Gels were photographed under UV light using a gel image analysis system and normal photographic film (with a UV filter). RAPD and ISSR markers were scored directly from the gels, whenever possible, and from the photographic prints. Gels were scored by eye because this

was found to be less time-consuming and just as objective as scoring the gels using a gel analysis package (Phoretix 1D, version 3.0; Phoretix International).

## **2.9 Testing the homology of RAPD and ISSR bands**

*"The use of RAPDs [or ISSRs] for comparative purposes relies on the assumption that similarity of fragment size is a dependable indicator of homology"* (Rieseberg, 1996). There are two methods for testing the homology of co-migrating fragments: Southern hybridisation and restriction digestion. In the former, a RAPD or ISSR fragment is transferred to a nylon membrane and probed with a gel-isolated, labelled fragment that is putatively homologous. If the fragments are homologous the probe will hybridise to the fragment on the membrane. However, the Southern hybridisation approach is *"extremely time-consuming"* (Rieseberg, 1996) and, therefore, restriction digestion was used in the present study.

### *Restriction digestion method for testing RAPD and ISSR band homology*

A band to be tested was excised from the agarose gel (under UV light) using a narrow bore pasteur pipette, placed in a microtube with 50  $\mu$ l of sterile distilled water and stored at -20°C. Following removal from the freezer, the microtube was incubated at 37°C for 2-3 h to ensure that the DNA from the band had dissolved into the water. The band was re-amplified using a small volume of water, the original primer and standard RAPD or ISSR methodology. A small sub-sample of the re-amplified DNA was run out on a 1.4% agarose gel to ensure that only the desired band was visible. If only the desired band was visible, the remaining re-amplified DNA was split into three aliquots. Two aliquots were digested with different 4 bp recognition site restriction enzymes, used singly or in combination (table 2.3), using the standard restriction digest procedure (section 2.7). The digested samples were run out on 1.4% agarose gels (50  $\mu$ l of ethidium bromide for a large gel, plus 200  $\mu$ l in the tank buffer) and photographed under UV light. RAPD or ISSR bands with congruent restriction profiles for both enzymes were regarded as homologous (Rieseberg, 1996). For two profiles to be considered congruent, they had to contain the same number of fragments of the same size, with the sum of fragment sizes in each profile equal to the size of the original (pre-digestion) band). However, when congruent profiles were detected for only one of the two enzymes, the third aliquot of each band was digested with a third

enzyme. If congruent profiles were observed for two of the three enzymes this was considered to be evidence for homology (Rieseberg, 1996).

When a sub-sample of a re-amplified RAPD or ISSR band was electrophoresed, other bands were sometimes present, along with the desired band. If this was the case the following procedure was carried out: The remaining re-amplified DNA was run out on a 0.8% TAE (appendix A) agarose gel (100 µl of ethidium bromide for a large gel) immersed in 1X TAE tank buffer (100 µl of ethidium bromide in the tank buffer) for several hours. The correct band was cut out using a scalpel, under UV light. The band was placed in a microtube with three volumes of NaI and 10% TBE Modifier<sup>®</sup> and incubated at 55°C in a water-bath for 5 min, mixing once. After 5 min, the solution was observed to ensure that the agarose had melted. EZ-GLASSMILK<sup>®</sup> was then added and the DNA from the band was isolated following the steps outlined above for DNA purification (section 2.6). The purified DNA was then digested as above.

The presence of unwanted bands could be attributed to 1) the agarose of the excised band having contained additional DNA molecules due to the idiosyncrasies of gel electrophoresis (Weising *et al.*, 1995) or 2) internal priming sites within the chosen fragment.

**Table 2.3** Restriction enzymes used in the present study.

Restriction enzyme	Recognition/cutting site 5'-3'	Incubation temperature °C	Supplier
<i>Alu</i> I	AG↓CT	37	Promega
<i>Cfo</i> I	GCG↓C	37	Promega
<i>Hae</i> III	GG↓CC	37	Promega
<i>Hinf</i> I	G↓ANTC	37	Promega
<i>Hpa</i> II	C↓CGG	37	Promega
<i>Rsa</i> I	GT↓AC	37	Promega
<i>Taq</i> I	T↓CGA	65	Promega
<i>Cla</i> I*	AT↓CGAT	37	Promega

\* 6 bp cutter used for RFLP analysis, all other enzymes 4 bp cutters used to check homology of RAPD fragments

↓ Indicates the cutting site; N = A, T, C or G

## 2.10 Allozyme analysis

Six enzymes were assayed for allozyme variation: aconitase (*Aco*), acid phosphatase (*Acp*), aspartate aminotransferase (*Aat*), isocitrate dehydrogenase, phosphoglucoisomerase (*Pgi*) and phosphoglucomutase (*Pgm*). Buffer systems and stains for these enzymes are described in appendix A.

### *Starch gel electrophoresis*

#### *Preparing a gel*

36 g of hydrolysed potato starch was added to 300 ml of gel buffer (appendix A) in a 1 l Buchner flask. The flask was sealed with cling-film and swirled vigorously to help the starch dissolve. The flask was heated over a Bunsen burner for approximately 5 min; during this period the flask was swirled vigorously. Once the fluid was of the right colour and consistency (clear and thick) a vacuum pump was used to evacuate the flask, thus removing some of the air bubbles that would adulterate the gel. The solution was quickly poured into a pre-prepared gel mould. A glass plate was carefully lowered onto the top of the gel mould, to spread the gel solution evenly across the mould and seal it. Care was taken to avoid or remove any bubbles. The gel was left to set; cooling it first by an open window and then in a refrigerator.

Once the gel had cooled, the glass plate was removed and the mould was cleaned of excess gel. A plastic comb was used to sink wells, normally 34, into the gel at the origin line (5 cm from the top edge of the gel). The fully prepared gel was covered in cling-film to keep it moist.

#### *Collecting and loading the samples*

From each of the plants being tested, a small flower-bud (for highest enzyme activity) was removed and placed in a numbered cavity of a microtitre plate. From this stage onwards, it was essential to ensure that the samples were kept cool to avoid enzyme degradation; for example, by placing the microtitre plate on ice. To each cavity a drop of extraction buffer (appendix A) was added and the sample crushed to a homogeneous solution using a glass rod. Cross-contamination was avoided by washing the glass rod thoroughly between samples. Once all the samples had been crushed, blotting paper wicks (3 mm x 5 mm) were added to the cavities to soak up the plant extract. Each soaked wick was carefully placed

into a different well of the prepared gel. Marker dye was added to the outer wells and the cling film replaced.

#### *Running, slicing and staining the gel*

The gel was loaded onto a gel-rig, containing appropriate electrode buffer (appendix A), covered with an ice-bag and placed in the refrigerator. The gel was run at approximately 250 V for 3-4 h until the tracker dye had reached or passed a mark 13 cm from the origin.

Using fine forceps, the wicks were removed from the wells and the gel loosened from the mould. The gel was cut along the origin line and just above the tracker dye line. The centre gel piece was carefully removed from the mould and placed on a glass sheet, having removed any excess moisture. Two plastic spacers of different thicknesses were placed on each side of the gel (i.e. parallel to the origin line) and a glass plate placed over the top. Nylon fishing line was then pulled between the upper glass plate and upper spacer, to produce a thin slice of gel; this slice was discarded. By re-arranging the plastic spacers three more slices were produced and these were placed in separate staining dishes.

Stain ingredients (appendix A) were prepared in advance but not mixed until each stain was needed. The stains were added to appropriate slices and incubated at 37°C. Fifteen minutes to several hours later the banding patterns appeared (*Pgi/Aat* over-stained easily, whilst *Pgm* took a long time to develop). Once banding patterns were clear, gel slices were washed to remove excess stain and covered in 50% glycerol to preserve them.

Banding patterns were recorded by means of diagrams and colour photographs. The genetics of electrophoretic variants of *Aat* and *Pgi* were inferred from previous inheritance studies (Abbott *et al.*, 1992a, b; Ashton and Abbott, 1992b). Genetic interpretation of the remaining enzymes (*Acp*, *Aco*, *Idh*, *Pgm*) was based upon the assumptions of Abbott *et al.* (2000) and Forbes (personal communication). Three or four standards were run per gel, to aid in the correct identification of alleles. Isozymes were named numerically and allozymes alphabetically, according to declining migration rate and following the convention of Abbott *et al.* (2000) and Forbes (personal communication).

## 2.11 Data analysis

### RAPD and ISSR data

#### *2.11.1 Recording RAPD and ISSR results*

RAPD and ISSR bands were scored as present (1) or absent (0) in each individual. RAPDs and ISSRs are primarily dominant markers which means that allele frequencies cannot be calculated directly from RAPD and ISSR band data (see section 1.4.3.2). Therefore, direct analysis of RAPD and ISSR phenotypes (i.e. band presence or absence) was undertaken. This means that when referring to the frequency of a RAPD or ISSR marker we are referring to a phenotype frequency (the number of individuals containing a band), not an allele frequency.

#### *2.11.2 Similarity coefficients: Nei and Li, and Jaccard*

A pair-wise similarity matrix can be constructed from a binary data matrix using a similarity coefficient. In the present study Jaccard's (1908; cited in Sneath and Sokal, 1973) coefficient and Nei and Li's (1979; cited in Sneath and Sokal, 1973) coefficient were employed.

##### *Jaccard's coefficient*

The similarity (S) between two individuals (x and y) can be calculated as follows:

$$S = \frac{n_{xy}}{(n-n_0)}$$

where,

$n_{xy}$  = number of bands present in both x and y (number of shared product presences)

$n$  = total number of bands scored

$n_0$  = number of bands in neither x nor y.

*Nei and Li's coefficient* (equivalent to the coefficient of Dice)

The similarity (S) between two individuals (x and y) can be calculated as follows:

$$S = \frac{2 n_{xy}}{(n_x + n_y)}$$

where,

$n_{xy}$  = number of bands present in both x and y (number of shared product presences)

$n_x$  = number of bands present in x

$n_y$  = number of bands present in y.

In Jaccard's and Nei and Li's coefficient, negative (0,0) matches are excluded from the numerator and, therefore, only shared presence of a fragment is counted (Harris, 1999). This is advisable because the absence of a band in two individuals cannot be assumed to have arisen from an identical mutation (i.e. it is possible that two recessive alleles at a locus are not identical in state; Black, 1993).

However, with Jaccard's coefficient matched (1,1) and unmatched pairs (1,0; 0,1) are equally weighted, whereas with Nei and Li's coefficient matched pairs carry twice the weight of unmatched pairs. Harris (1999) recommends Nei and Li's coefficient rather than Jaccard's coefficient because the former has a lower percentage bias (the percentage difference between 'true' and estimated similarity values due to bias introduced by false positives and false negatives) than the latter.

A pair-wise similarity matrix can be converted into a pair-wise dissimilarity or distance matrix using the following equation:

$$D \text{ (pair-wise dissimilarity/distance)} = 1 - (S)$$

In the present study, all the above calculations were made using the programme RAPDistance, version 1.04 (Armstrong *et al.*, 1994). The authors supplied a modified version of the programme that could deal with sample sizes of up to 200 individuals.

### ***2.11.3 Principal coordinate analysis (PCO)***

The aim of ordination is to display the relationships between OTUs (operational taxonomic units) in a low-dimensional space, thereby enabling direct visual examination of the relative positions of the OTUs (Dunn and Everitt, 1982). The relationships between OTUs

can be based upon the values of a large number of variables measured on each OTU, as in principal components analysis (PCA; Manly, 1994). Alternatively, the relationships between OTUs can be based upon a matrix of similarity or dissimilarity/distance, as in principal coordinate analysis (Gower, 1966). In summary, therefore, PCO "*is a technique for providing a geometrical representation of the distance or association between individuals*" (Reyment *et al*, 1984).

PCO is apparently robust to homology errors in data, which may be particularly useful if employing RAPDs or ISSRs (see section 1.4.3.2). Whereas a few non-homologous characters may dramatically alter the results obtained in a cladistic analysis, random errors of homology can be accounted for as noise in a PCO (Adams and Demeke, 1993).

Principal coordinate analysis can be split into three stages (Reyment *et al*, 1984). The first step is to form a matrix which represents the similarity between OTUs (a dissimilarity matrix can also be used with a slight adjustment). The second step, is to double-centre the matrix, so that all columns and rows sum to zero. The third step is to perform an eigen analysis on the transformed matrix; the matrix is factored to compute its eigenvalues and eigenvectors. Each eigenvector is normalised so that the sum of squares of its components is equal to the corresponding eigenvalue.

The elements of the normalised eigenvectors are the coordinates of the OTUs, providing an exact representation of the distances between them in a number of dimensions. If there is an exact solution, the maximum number of dimensions will be, at the most, one less than the number of individuals. Moreover, the coordinates are arranged relative to their orthogonal (perpendicular and independent) principal axes. This means that the first dimension of the solution explains the greatest amount of variance in the data, the second dimension explains the second largest amount of variance in the data that is uncorrelated with the variance accounted for by the first dimension, and so on, with each subsequent dimension explaining progressively less of the variance. The percentage variation explained by each dimension can be ascertained by looking at the relative magnitude of the associated eigenvalues. When much of the variation is contained in the first few (typically two or three) dimensions, the coordinates given by the corresponding normalised eigenvectors can be plotted. In the resulting PCO plot(s) the Euclidean distance between two points reflects as accurately as possible the relationship between the corresponding OTUs, as implied by their observed proximities (i.e. points representing similar individuals

will be located close together and points representing dissimilar individuals will be located further apart).

PCO was originally developed for use with Euclidean distance and association coefficients (Sneath and Sokal, 1973). However, it is valid to employ non-Euclidean coefficients, as long as there are no large negative eigenvalues (Sneath and Sokal, 1973). For example, both Nei and Li's similarity coefficient (e.g. Liao and Hsiao, 1998) and Jaccard's similarity coefficient (e.g. Gabrielsen and Brochmann, 1998) have been utilised in PCO. Nei and Li's coefficient and Jaccard's coefficient are well-suited for RAPD and ISSR data, because they only take into account the shared presence of bands (2.11.2), whereas Euclidean coefficients treat shared presence *and* shared absence of bands alike. A Euclidean coefficient (e.g. the Euclidean distance coefficient) can be employed as a precaution, to ensure that the results generated using Nei and Li's and Jaccard's coefficients are real and not artefacts of using a non-Euclidean coefficient.

In the present study, the programme NTSYS-pc version 2.0 (Rohlf, 1998) was used to perform PCO analyses in two ways:

In the first method, a binary data matrix was used as the input for NTSYS and PCO analysis was conducted using the following commands: standardise, simint, dcenter, eigen and output. The coefficient chosen was Euclidean distance.

In the second method, a distance matrix (derived from Nei and Li's coefficient or Jaccard's coefficient) was used as the input for NTSYS and PCO analysis was conducted using the following commands: dcenter, eigen and output.

#### **2.11.4 Analysis of molecular variance (AMOVA)**

The analysis of molecular variance (AMOVA) procedure can be used to partition genetic variation at different hierarchical levels, such as among individuals within populations and among populations within regions.

AMOVA was developed by Excoffier *et al.* (1992) for RFLP haplotypes. In an AMOVA, a matrix of squared Euclidean distances among all pairs of haplotypes is produced. From this distance matrix, the total sum of squared deviations (equivalent to the sum of squares) is calculated. The total sum of squared deviations is then used in a standard analysis of

variance, so that the total sum of squared deviations is apportioned into sum of squared deviations at various hierarchical levels (e.g. among individuals within populations, among populations within regional groups and among regional groups). Mean squared deviations are determined by dividing each sum of squared deviations by the correct degrees of freedom. By equating the mean squared deviations to their expectations, the variance components of each hierarchical level are estimated. *F*-statistic analogues,  $\Phi$ -statistics, are estimated from the variance components. A permutational approach is used to test the significance of the variance components and  $\Phi$ -statistics (thus avoiding the normality assumption that may be inappropriate for molecular data).

Although the AMOVA procedure was developed for haplotypic data it has also been used successfully with RAPD phenotype data (Huff *et al.*, 1993; Huff *et al.*, 1998; Kölliker *et al.*, 1998). In a RAPD-based AMOVA, the pair-wise distance between two RAPD phenotypes is estimated using Escoffier's Euclidean distance metric, defined for RAPD markers by Huff *et al.* (1993) as:

$$E = \{\varepsilon_{xy}^2\} = n [1 - 2n_{xy}/2n]$$

where,

$n_{xy}$  = the number of bands shared by individuals x and y

$n$  = the total number of polymorphic bands.

In the present study, RAPDistance was used to produce a pair-wise distance matrix, based upon the Euclidean metric of Huff *et al.* (1993, as above) from a binary data matrix. This matrix was used as the input for the WINAMOVA programme, (version 1.53; Excoffier, 1993) which performs the AMOVA. The number of permutations, for testing the significance of the variance components and  $\Phi$ -statistics, was set at 1000 for all analyses.

### **2.11.5 Cluster analyses**

Phenograms were generated using two clustering methods - the UPGMA (Unweighted Pair Group Method using Arithmetic Averages) and the Neighbour-Joining procedures. The UPGMA is one of the simplest algorithms for tree building, with the most similar OTUs clustered together sequentially. UPGMA joins one cluster to another on the basis of the average distance between members of one cluster and members of the other cluster;

distances between all possible pair-wise combinations of OTUs in the two clusters are calculated, summed and divided by the number of combinations (Quicke, 1993).

In neighbour-joining clustering, there is sequential identification of neighbour pairs that minimise the total length of the tree (Li and Graur, 1991).

UPGMA assumes the operation of a universal molecular clock. i.e. that all lineages will diverge at the same speed (Glover *et al.*, 1994). However, the inherent assumption of the UPGMA method, of constant change at the molecular level, may be invalid for RAPD data (Glover *et al.*, 1994). This is because RAPD markers are random, they may be derived from coding as well as non-coding regions and from organelle genomes as well as the nuclear genome, and these different categories of DNA may evolve at different rates (for example, coding regions change more slowly than non-coding regions). Consequently, it has been suggested that the neighbour-joining method (which does not assume the operation of a universal molecular clock) should be used to cluster RAPD data (Glover *et al.*, 1994).

In the present study, distance matrices were used as the input for the programme PHYLIP for windows version 3.5C (Felsenstein, 1993). The data were clustered by the UPGMA and neighbour-joining methods using the sub-programme NEIGHBOUR. The tree-viewing programme TREEVIEW version 1.4 (Page, 1996) was used to produce dendrograms from the output of NEIGHBOUR.

#### ***2.11.6 Shannon's Diversity Index***

Shannon's diversity index (King and Schaal, 1989) can be used to quantify the level of genetic diversity present within each population of a sample and to partition the total genetic diversity (in the sample as a whole) into within and between population components. Shannon's diversity index has been used successfully with RAPD data on several occasions (e.g. Russell *et al.*, 1993, Gillies *et al.*, 1997 and Wolff *et al.*, 1997). Shannon's diversity index is thought to be well-suited to RAPD data analysis, because it is comparatively insensitive to the potential bias in the data resulting from an inability to spot heterozygous individuals (Gillies *et al.*, 1997).

Genetic diversity within a single population ( $H_0$ ) can be calculated as follows:

$$H_0 = -\sum p_i \log_2 p_i$$

where,

$p_i$  = the frequency of a RAPD or ISSR phenotype in that population.

The total genetic diversity (i.e. the genetic diversity in the sample as a whole, assuming no sub-structuring) can be calculated as follows:

$$H_{sp} = -\sum p_i \log_2 p_i$$

where,

$p_i$  = the frequency of a RAPD or ISSR phenotype in the sample as a whole.

$H_{pop}$  is the mean within population genetic diversity ( $\sum H_0$ /number of populations).

$H_{pop}/H_{sp}$  is the proportion of the total genetic diversity maintained within populations.

$H_{sp}-H_{pop}/H_{sp}$  is the proportion of the total genetic diversity maintained between populations.

In the present study, Shannon's diversity index analyses were carried out by hand.

### **Allozyme data analysis**

#### ***2.11.7 General statistics***

Allozyme phenotype frequencies were entered into the programme BIOSYS-1 (version 1.7; Swofford and Selander, 1981). This programme calculated allele frequencies and estimated various measures of allozyme diversity including:- the mean number of alleles per locus ( $A$ ), the mean percentage of polymorphic loci ( $P$ ), the mean observed heterozygosity and the mean unbiased expected heterozygosity ( $H_e$ ).

Conformity to Hardy-Weinberg equilibrium was tested across loci in all populations using the probability test option (Fisher's method) in the programme GENEPOP (version 3; Raymond and Rousset, 1995).

#### ***2.11.8 F-statistics***

Wright's  $F$ -statistics allow one to partition heterozygote deficit into within and between population components.  $F$ -statistics are useful for assessing levels of structuring in samples of natural populations.  $F_{it}$  measures the global deficit of heterozygotes, giving an

indication of total inbreeding. *Fis* measures the heterozygote deficit within populations, giving an indication of inbreeding within populations. *Fst* measures the heterozygote deficit among populations, giving an indication of subdivision among populations.

In the present study, *F*-statistics were calculated using the programme FSTAT (version 1.2; Goudet, 1995). FSTAT produced Weir and Cockerham's (1984) unbiased estimators of Wright's *F*-statistics (*F*,  $\theta$  and *f*) calculated separately for each locus, for all populations. *F* is analogous to *Fit*,  $\theta$  is analogous to *Fst* and *f* is analogous to *Fis*. Locus specific means and standard deviations were obtained by jackknifing over populations. Overall means and standard deviations were estimated by jackknifing over loci. Ninety-five per cent confidence intervals were estimated by bootstrapping over loci.

### 2.11.9 Nei's unbiased genetic distance

If  $x_i$  and  $y_i$  are the frequencies of the  $i^{\text{th}}$  allele at a locus in samples from populations X and Y respectively, then Nei's standard genetic distance (*D*) between the populations can be estimated as follows:

$$D = -\ln [J_{xy}/J_x J_y]$$

where,

$J_x$ ,  $J_y$  and  $J_{xy}$  are the averages of  $\sum x_i^2$ ,  $\sum y_i^2$  and  $\sum x_i y_i$  ( $\sum$  stands for summation over all alleles at a locus) over the  $r$  loci studied, respectively (Nei, 1978).

Nei's (1978) unbiased genetic distance statistic (*D'*) allows for small sample sizes and can be calculated as follows:

$$D' = -\ln [G'_{xy}/G'_x G'_y]$$

where,

$G'_x$  and  $G'_y$  are the averages of  $(2n_x J_x - 1)/(2n_x - 1)$  and  $(2n_y J_y - 1)/(2n_y - 1)$  over the  $r$  loci studied, respectively;  $G'_{xy} = J_{xy}$ , and  $n_x$  and  $n_y$  are the numbers of individuals sampled from populations X and Y, respectively.

In the present study, Nei's unbiased genetic distance was calculated between each population pair (using the allele frequencies from every locus) with BIOSYS-1. This distance matrix was used as the input for PHYLIP and the data were clustered by the UPGMA and neighbour-joining methods (see above).

# **Chapter 3**

## **The Development of Taxon-specific Molecular Markers**

## **The Development of Taxon-specific Molecular Markers**

### **3.1 Introduction**

When studying hybridisation, molecular markers that clearly and reliably differentiate individuals of one taxon from individuals of a second taxon, are highly informative. Such molecular markers can be used to unambiguously identify cases of hybridisation and introgression. For example, Perron *et al.* (1995) obtained seven RAPD markers that were specific to either black spruce (*Picea mariana*) or red spruce (*Picea rubens*) and used these markers to detect hybrid and introgressant individuals in allopatric and sympatric populations of the two species (Perron and Bousquet, 1997).

Molecular markers may also reveal additional significant details about hybridisation and introgression (Bachmann, 1994). For example, Rieseberg (Rieseberg *et al.*, 1993; Rieseberg *et al.*, 1995) used molecular markers to study the genomic processes that accompany recombinational speciation. Rieseberg used RAPDs to produce linkage maps for *Helianthus annuus*, *H. petiolaris* and their homoploid hybrid species, *H. anomalus*. Rieseberg found that *H. anomalus* had undergone extensive genomic rearrangements relative to its parental species. He also established that both *H. annuus* and *H. petiolaris* had contributed roughly equal proportions of their genome to *H. anomalus*.

Arnold (Arnold *et al.*, 1991; Arnold *et al.*, 1992) used molecular markers to investigate hybridisation between *Iris fulva*, *I. hexagona* and *I. brevicaulis* in Louisiana. RAPD and cpDNA PCR-RFLP markers revealed that pollen dispersal, rather than seed dispersal, was the most important method of gene-flow between the *Iris* species.

In a similar way, it was intended that molecular markers could be used to examine hybridisation between *Senecio aethnensis* and *S. chrysanthemifolius* on Mt. Etna, Sicily, (see chapter four) and their supposed hybrid species, *S. squalidus* (see chapter five).

### 3.1.1 What is a taxon-specific molecular marker?

The definition of a "taxon-specific marker" varies between hybridisation studies. For example, Perron *et al.* (1995) defined species-specific markers as markers that were "*monomorphic or nearly monomorphic in one species...while absent in the other*". However, Neuffer *et al.* (1999) defined species-specific markers as markers that "*occurred in 75% of the individuals in one... species but in none of the other... species*".

In this study, a "taxon-specific marker" is defined as 'a marker unique to one taxon but not necessarily found in all individuals of that taxon' (Howard *et al.*, 1997). The various classes of taxon-specific markers used are described as follows (table 3.1):

A diagnostic taxon-specific molecular marker is found in *all* individuals of one taxon and no individuals of a second taxon (Howard *et al.*, 1997). In contrast, a taxon-specific (I) marker is present at high frequency in one taxon ( $\geq 0.70$ ) and absent in another taxon, while a taxon-specific (II) marker is present at low frequency ( $\geq 0.40$ ) in one taxon and absent in a second taxon. Taxon-specific "private bands" are markers unique to one taxon but present only at very low ( $\geq 0.10$ ,  $<0.40$ ) frequencies (Wolfe *et al.*, 1998).

Many workers are pragmatic and supplement taxon-specific markers with what might be called "frequency-difference" markers, i.e. markers present at high frequency in one taxon and low frequency in the second taxon. Use of such "frequency-difference" markers, in studies of inter-specific hybridisation, has been made by Shoemaker *et al.* (1994), Howard *et al.* (1997) and Wolfe *et al.* (1998). The various classes of "frequency-difference" markers used in the present study are detailed in table 3.1.

**Table 3.1** Categories of useful marker bands as used in the present study (based on definitions of Howard *et al.* (1997) and others).

Category	Overall frequency of marker	
	in taxon x	in taxon y
<b>taxon-specific</b> (for taxon x)		
diagnostic taxon-specific marker	1.00	0.00
taxon-specific (I) marker	≥ 0.70	0.00
taxon-specific (II) marker	≥ 0.40	0.00
private band	≥ 0.10	0.00
<b>frequency-difference</b> (for taxon x)		
frequency-difference (I) marker	≥ 0.70	≤ 0.05
frequency-difference (II) marker	≥ 0.70	≤ 0.10
frequency-difference (III) marker	≥ 0.70	≤ 0.15
frequency-difference (IV) marker	≥ 0.50	≤ 0.10
frequency-difference (V) marker	≥ 0.60	≤ 0.20

### 3.1.2 Choice of molecular marker methods

The "perfect" technique for generating molecular markers would produce many reliable, reproducible, independent, codominant markers at an appropriate taxonomic level for the study to be undertaken. The procedure would be quick, easy to perform, cheap and require no development work. Unfortunately, the perfect technique for generating molecular markers does not exist and one must therefore compromise (section 1.4).

In the present study, the choice of molecular marker methods was a pragmatic decision. RAPDs and ISSRs were favoured for two main reasons. Firstly, RAPDs and ISSRs have many practical advantages; for example, they are easy and quick to use and do not require development work. Secondly, a RAPD and ISSR analysis samples a large proportion of the genome and has the potential to produce many polymorphic markers (Harris, 1999). Furthermore, RAPDs and ISSRs have been successfully used as genetic markers in a number of hybridisation studies (e.g. Arnold (1993), Shoemaker *et al.* (1994), Perron *et al.* (1995) and Wolfe *et al.* (1998) and 1.4.3.3). In addition to using RAPDs and ISSRs as genetic markers, isozymes and cpDNA RFLPs have been employed to a more limited extent. Previous studies indicated that these markers had the potential to distinguish between *S. chrysanthemifolius* and *S. aethnensis* (Abbott *et al.*, 1995; Abbott *et al.*, 2000).

However, the work had also revealed that levels of isozyme and cpDNA RFLP variation in *S. chrysanthemifolius* and *S. aethnensis* were low.

### **3.1.3 Aim:**

The aim of the present study was to develop reliable RAPD, ISSR, isozyme and cpDNA RFLP molecular markers that can distinguish between *S. chrysanthemifolius* and *S. aethnensis*.

## 3.2 Materials and Methods

### 3.2.1 RAPD and ISSR survey of variation within and between *S. aethnensis* and *S. chrysanthemifolius*

RAPD and ISSR analysis was carried out using the procedures outlined in section 2.8.

#### 3.2.1.1 Preliminary survey

Four individuals of 'pure' *S. aethnensis* (from two populations - S13 and S14, table 2.1) and four individuals of 'pure' *S. chrysanthemifolius* (from three populations - C0, C1 and C9, table 2.1) were selected for a preliminary screen of RAPD and ISSR primers. A total of 305 primers were tested (table 3.2). Half way through the screening process, the number of individuals per taxon was raised from four to six. The examination of larger numbers of individuals per taxon will increase the likelihood that any apparent taxon-specific differences observed are 'real' (i.e. will be maintained in a larger survey). However, numbers had to be kept to a minimum to reduce expense and time and, consequently, six individuals per taxon was taken as a sensible compromise.

**Table 3.2** Primers used in the preliminary RAPD/ISSR survey.

Primer type	Set(s)	Number of primers
RAPD	Operon sets A, B, C, F, G, H	120 (6 x 20)
RAPD primer-pairs	Paired combinations of Operon RAPD primers: B/B, F/H, G/A, F/A, F/G, B/F, B/G, G/H	96
ISSR	University of British Columbia set # 9	89

#### Detailed survey

Primers that differentiated individuals of *S. aethnensis* from individuals of *S. chrysanthemifolius* in the preliminary survey of RAPD and ISSR variation were chosen for a more detailed survey. Forty-five individuals of 'pure' *S. aethnensis* and 45 individuals of 'pure' *S. chrysanthemifolius* were screened with 65 primers (table 3.4) that had been shown to generate markers that distinguished the two taxa in the preliminary survey. 'Pure' *S. chrysanthemifolius* came from three geographically separated populations. Two populations (C0 and C1) were from the southern slopes of Mt. Etna and one population (C9) was from the northern slopes of Mt. Etna (table 3.3). 'Pure' *S. aethnensis* came from four populations, two from the southern summit of Mt. Etna, two from the northern summit

of Mt. Etna. *S. aethnensis* is endemic to the high altitude zone on Mt. Etna so it was not possible to select *S. aethnensis* populations that were as geographically separated as those of *S. chrysanthemifolius*.

**Table 3.3** Populations of 'pure' *Senecio chrysanthemifolius* and *S. aethnensis* used in a detailed survey of RAPD and ISSR variation within these two taxa.

Population	Location (altitude)	No. of individuals
<i>S. chrysanthemifolius</i>		
CO <sup>†</sup>	Catania, 11 km south of Mt. Etna (150 m)	16
C1	Pedara, southern base of Mt. Etna (600 m)	14
C9	Randazzo, northern base of Mt. Etna (750 m)	15
<i>S. aethnensis</i>		
VB <sup>‡</sup>	southern summit of Mt. Etna (2600 m)	17
BB	southern summit of Mt. Etna (2525 m)	17
TC	northern summit of Mt. Etna (2550 m)	7
VO/UJ	northern summit of Mt. Etna (2800 m)	2+2

<sup>†</sup> Population C0 was 11 km from C1 and 41 km from C9. C1 was 30 km from C9.

<sup>‡</sup> Population VB was 1.35 km from BB, 6.05 km from TC and 4.55 km from VO/UJ. BB was 6.35 km from TC and 5.05 km from VO/UJ. TC was 1.75 km from VO/UJ.

Banding patterns were examined carefully. Those bands that differentiated between the two taxa, and were easy to score, were categorised as markers for *S. chrysanthemifolius* and *S. aethnensis* according to the scheme set out in table 3.1.

To ascertain if RAPD and ISSR markers were reproducible, eight RAPD primers and one ISSR primer (for all 90 individuals) were repeated. The RAPD and ISSR profiles obtained in the repeats were compared with the original profiles.

**Table 3.4** RAPD/ISSR primers/RAPD primer-pairs used in the full survey of *Senecio chrysanthemifolius* and *S. aethnensis* material.

<b>Primer</b>	<b>Sequence 5'-3'</b>	<b>Primer</b>	<b>Sequence 5'-3'</b>
<u>RAPDS</u>		<u>RAPD PRIMER</u>	
		<u>-PAIRS</u>	
A1	CAGGCCCTTC	A7/G2	GAAACGGGTG
A2	TGCCGAGCTG		GGCACTGAGG
A3	AGTCAGCCAC	A9/G10	GGGTAACGCC
A9	GGGTAACGCC		AGGGCCGTCT
A11	CAATCGCCGT	B2/B10	TGATCCCTGG
A13	CAGCACCCAC		CTGCTGGGAC
A14	TCTGTGCTGG	B9/G19	TGGGGGACTC
A16	AGCCAGCGAA		GTCAGGGCAA
A17	GACCGCTTGT	F2/G9	GAGGATCCCT
A19	CAAACGTCCG		CTGACGTCAC
A20	GTTGCGATCC	F3/G10	CCTGATCACC
B4	GGACTGGAGT		AGGGCCGTCT
B6	TGCTCTGCCC		
B12	CCTTGACGCA	<u>ISSRs</u>	
B15	GGAGGGTGTT		
B19	ACCCCCGAAG	8	AGAGAGAGAGAGAGAGAGC
C2	GTGAGGCGTC	17	CACACACACACACACAA
C5	GATGACCGCC	18	CACACACACACACACAG
C8	TGGACCGGTG	25	ACACACACACACACACT
C11	AAAGCTGCGG	26	ACACACACACACACACC
C16	CACACTCCAG	27	ACACACACACACACACG
C20	ACTTCGCCAC	34	AGAGAGAGAGAGAGAGAYT
F7	CCGATATCCC	36	AGAGAGAGAGAGAGAGYA
F9	CCAAGCTTCC	40	GAGAGAGAGAGAGAGAYT
F12	ACGGTACCAG	42	GAGAGAGAGAGAGAGAYG
G1	CTACGGAGGA	47	CACACACACACACARC
G3	GAGCCCTCCA	55	ACACACACACACACACYT
G5	CTGAGACGGA	57	ACACACACACACACACYG
G6	GTGCCTAACC	60	TGTGTGTGTGTGTGTGRA
G8	TCACGTCCAC	64	ATGATGATGATGATGATG
G10	AGGGCCGTCT	81	GGGTGGGGTGGGGTG
G11	TGCCCGTCGT	86	VDVCTCTCTCTCTCTCT
G12	CAGCTCACGA	88	BDBCACACACACACA
G14	GGATGAGACC	91	HVHTGTGTGTGTGTGTG
G15	ACTGGGACTC		
H4	GGAAGTCGCC		
H7	CTGCATCGTG		
H13	GACGCCACAC		
H18	GAATCGGCCA		
H19	CTGACCAGCC		

R = (A, G), Y = (C, T), B = (C, G, T), D = (A, G, T), H = (A, C, T), V = (A, C, G)

### 3.2.2 Survey of isozyme variation within and between *S. chrysanthemifolius* and *S. aethnensis*

The isozyme analysis was carried out using the procedures outlined in section 2.10. A total of 91 individuals of 'pure' *S. aethnensis*, from four populations (table 3.5) were surveyed for variation at the *Acp-2* and *Pgi-2* loci.

**Table 3.5** Populations of 'pure' *Senecio aethnensis* used in a survey of allozyme variation at the *Acp* and *Pgi* loci.

Population	Location (altitude)	No. of individuals
BB	southern summit of Mt. Etna (2525m)	30
VB	southern summit of Mt. Etna (2600m)	31
TJ	southern summit of Mt. Etna (2590m)	25
TC	northern summit of Mt. Etna (2550m)	5

Population TJ was 0.65 km from VB, 0.85 km from BB, and 6.4 km from TC. See table 3.3 for distances between other populations.

The results were compared to values obtained previously by Abbott *et al.* (2000), for three populations of 'pure' *S. chrysanthemifolius* and three populations of *S. aethnensis* (table 3.6).

**Table 3.6** Populations of *Senecio aethnensis* and 'pure' *S. chrysanthemifolius* used in a previous survey of allozyme variation (Abbott *et al.*, 2000).

Population	Location (altitude)	No. of individuals
<i>S. chrysanthemifolius</i>		
	5 km W of Nicolosi, southern base of Mt. Etna (800 m)	}
	Bronte, southern base of Mt. Etna (760 m)	} 95
	4 km SE of Linguaglossa, northern base of Mt. Etna (750 m)	}
<i>S. aethnensis</i>		
	2 km E Rifúgio Sapienza, southern slopes of Mt. Etna (1890m)	}
	3 km E Rif. Sapienza, southern slopes of Mt. Etna (1650 m)	} 62
	Rif. Citelli, northern slopes of Mt. Etna (1750 m)	}

### 3.2.3 Survey of chloroplast DNA RFLP variation within and between *S. chrysanthemifolius* and *S. aethnensis*.

The cpDNA analysis was carried out using the procedures outlined in section 2.7, using the enzyme *Cla*I and *Lactuca sativa* L. probe C6. Thirty nine 'pure' *S. chrysanthemifolius* individuals and 43 'pure' *S. aethnensis* individuals were selected for cpDNA typing (table 3.7).

**Table 3.7** Populations of 'pure' *Senecio aethnensis* used in a survey of cpDNA variation.

Population	Location (altitude)	No. of individuals
<i>S. chrysanthemifolius</i>		
C0	Catania, 11 km south of Mt. Etna (150 m)	15
C1	Pedara, southern base of Mt. Etna (600 m)	12
C9	Randazzo, northern base of Mt. Etna (750 m)	12
		(total 39)
<i>S. aethnensis</i>		
BB	southern summit of Mt. Etna (2525 m)	17
VB	southern summit of Mt. Etna (2600 m)	17
TC	northern summit of Mt. Etna (2550 m)	7
VO/UJ	northern summit of Mt. Etna (2800 m)	1+1
		(total 43)

See table 3.3 for distances between populations.

### 3.3 Results

#### 3.3.1 Results of the survey of RAPD and ISSR variation within and between *S. chrysanthemifolius* and *S. aethnensis*

##### 3.3.1.1 The preliminary survey

Sixty-five of the 305 primers investigated in the preliminary survey of RAPD and ISSR variation produced banding patterns that differentiated *S. aethnensis* individuals from *S. chrysanthemifolius* individuals (table 3.8). These comprised 40 RAPD primers, 19 ISSR primers and six RAPD primer-pairs. RAPD primers were the most useful for generating markers that distinguished between the two taxa - 33% of RAPD primers revealed potential taxon-specific markers compared to 21% of the ISSR primers and only 6% of the RAPD primer-pairs.

**Table 3.8** Types of banding pattern produced by primers in the preliminary survey.

Primer type	No. of primers generating no potential taxon-specific bands			No. of primers with potential for generating taxon-specific bands	% of primers with potential for generating taxon-specific bands
	Poor resolution <sup>1</sup>	Little variation <sup>2</sup>	Variation <sup>3</sup>		
<u>RAPD</u> (n = 120)	11	21	48	40	33%
<u>ISSR</u> (n = 89)	43	9	18	19	21%
<u>RAPD primer-pairs</u> (n = 96)	8	16	66	6	6%
All primers (n = 305)	62 (20%)	46 (15%)	132 (43%)	65	21%

<sup>1</sup> Primers produced poorly resolved banding patterns i.e. no amplification, poor amplification, smearing of products or very fuzzy and unclear bands.

<sup>2</sup> Primers produced bands that were monomorphic across virtually all individuals from both taxa.

<sup>3</sup> Primers produced some bands that were polymorphic across individuals of both taxa but the pattern was not taxon-specific.

Primers that did not appear to generate taxon-specific bands fell into three categories: 15% (of the total number) of primers produced banding patterns that varied little across all individuals, 43% produced banding patterns that showed variation between individuals but with no clear differences between taxa, and 20% produced unclear banding patterns due to no or poor DNA amplification, smearing etc. (table 3.8). The proportion of the different primer types in the three classes varied; for example, a high proportion of the ISSR primers were poorly resolved, compared with RAPD primers and RAPD primer-pairs (table 3.8).

### ***3.3.1.2 The detailed survey***

Results of the detailed survey of RAPD and ISSR variation within and between *S. aethnensis* and *S. chrysanthemifolius* using the 65 primers selected after the preliminary survey are displayed in tables 3.9, 3.10 and 3.11.

Some bands that had appeared to distinguish between the two taxa in the preliminary survey were found to be of little use when more material was examined in the full survey. Nevertheless, all but one primer (ISSR 55) produced at least one band that differentiated the two taxa in the full survey.

Table 3.9 lists the 112 bands that appeared to separate the two taxa in the full survey but were not retained for future use. Eighty-one of these bands were too difficult to score accurately. Reasons for this varied, the most common being faintness of bands or the overly close proximity of these bands to others. Nine bands distinguished the two taxa and were easily scored but did not meet the frequency-difference criteria that had been set previously (see note to table 3.9; although such criteria are arbitrary a line must be drawn at some point). Eight "private bands" were rejected because it was felt that they would be of little use in future studies due to their low frequencies in the taxon they were evident in. Also included in the table are a small number (15) of bands that were excluded from future studies *a posteriori*. Nine of these bands proved too difficult to score in later studies - due to faintness, closeness to other bands etc. - and six were unreliable.

**Table 3.9** Number of bands that distinguished *Senecio chrysanthemifolius* from *S. aethnensis* but were rejected for further use. 'C' identifies potential *S. chrysanthemifolius* marker bands; 'A' identifies potential *S. aethnensis* marker bands (e.g. 1C means one potential *S. chrysanthemifolius* marker, 3A means three potential *S. aethnensis* markers).

Primer	Difficult to score <sup>1</sup>	Distinguish taxa and can be scored but not used in further studies <sup>2</sup>		Distinguish taxa in full survey but rejected <i>a posteriori</i> <sup>5</sup>	
		Insufficient level of differentiation <sup>3</sup>	Private band <sup>4</sup>	Difficult to score hybrid individuals <sup>6</sup>	Unreliable <sup>7</sup>
A1			1 C (36/0), 1 C (11/0)		
A2	2 C, 1 A				
A3	1 C, 2 A				
A9	1 A			1 A (66/20)	
A11		1 C (89/33)			
A13	1 C				
A16	1 C				
A17	1 C				
A19	1 C, 3 A				
A20	1 A				
B4	1 C, 2 A				
B6	1 C				
B12	1 A				
B15	1 C, 1 A				
B19	1 A				
C2	1 C				
C8	3 C				
C11	1 C, 1 A		1 C (25/0)		
C16	1 A				1 C
F7	1 A				2 C
F9	1 A			1 C (80/7), 1 C (96/0)	
F12	2 A				
G1	3 A				
G3	1 C				
G5	1 C				
G6	1 A				
G8	1 A				
G10	1 A		1 A (14/0)		
G12	1 A				
G14	1 C, 3 A				
H4	1 A				
H7	1 A	1 C (89/40)			
H13	1 C				
H18		1 A (93/25)			
A7/G2	2 A	1 A (49/2), 1 A (31/2)		1 A (53/9)	
A9/G10			1 A (14/0)	1 C (76/9), 1 A (52/0)	1 C
B2/B10				1 A (96/7)	1 C
B9/G19	1 C	1 A (45/11), 1 A (31/7)			
F2/G9		1 C (95/47)			

Table 3.9 continued.

Primer	Difficult to score <sup>1</sup>	Distinguish taxa and can be scored but not used in further studies <sup>2</sup>		Distinguish taxa in full survey but rejected <i>a posteriori</i> <sup>5</sup>	
		Insufficient level of differentiation <sup>3</sup>	Private band <sup>4</sup>	Difficult to score hybrid individuals <sup>6</sup>	Unreliable <sup>7</sup>
F3/G10	1 C, 1 A		1 A (13/0), 1 A (20/0)		
8					1 A
17	1 C, 2 A				
25	2 C	1 A (53/18)			
26	1 C				
27	1 A				
34	1 C, 1 A				
36	1 C, 2 A				
42	1 A		1 A (30/0)		
47	2 A				
60	4 A				
64	1 C, 2 A				
81	1 C			1 C (91/2)	
86	1 C				
88	1 A				
91	1 C			1 A (93/2)	

## Notes

<sup>1</sup> These bands distinguished between the two taxa but were considered too difficult to score accurately. Reasons for this included: faintness of bands, bands being too close to other bands, bands that varied too greatly in intensity between individuals.

<sup>2</sup> These bands distinguished between the two taxa and could be scored accurately but were not used following the full survey: <sup>3</sup> These bands did not meet the minimum criteria, set out in section 3.1.1 of a frequency of  $\geq 0.50$  in one taxon and  $\leq 0.10$  in the other taxon, or  $\geq 0.60$  in one taxon and  $\leq 0.20$  in the other taxon. Numbers in parentheses represent the frequency of the band in the two taxa e.g. 1 C (89/33) would indicate that the potential *S. chrysanthemifolius* band was present at a frequency of 0.89 in *S. chrysanthemifolius* and at a frequency of 0.33 in *S. aethnensis*. <sup>4</sup> These bands were species-specific private bands - present at a frequency of  $< 0.40$  ( $> 0.10$ ) in one taxon and absent in the other taxon. Numbers in parentheses represent the frequency of the band in the two taxa e.g. 1 A (14/0) would indicate that the band was present at a frequency of 0.14 in *S. aethnensis* and 0.00 in *S. chrysanthemifolius*.

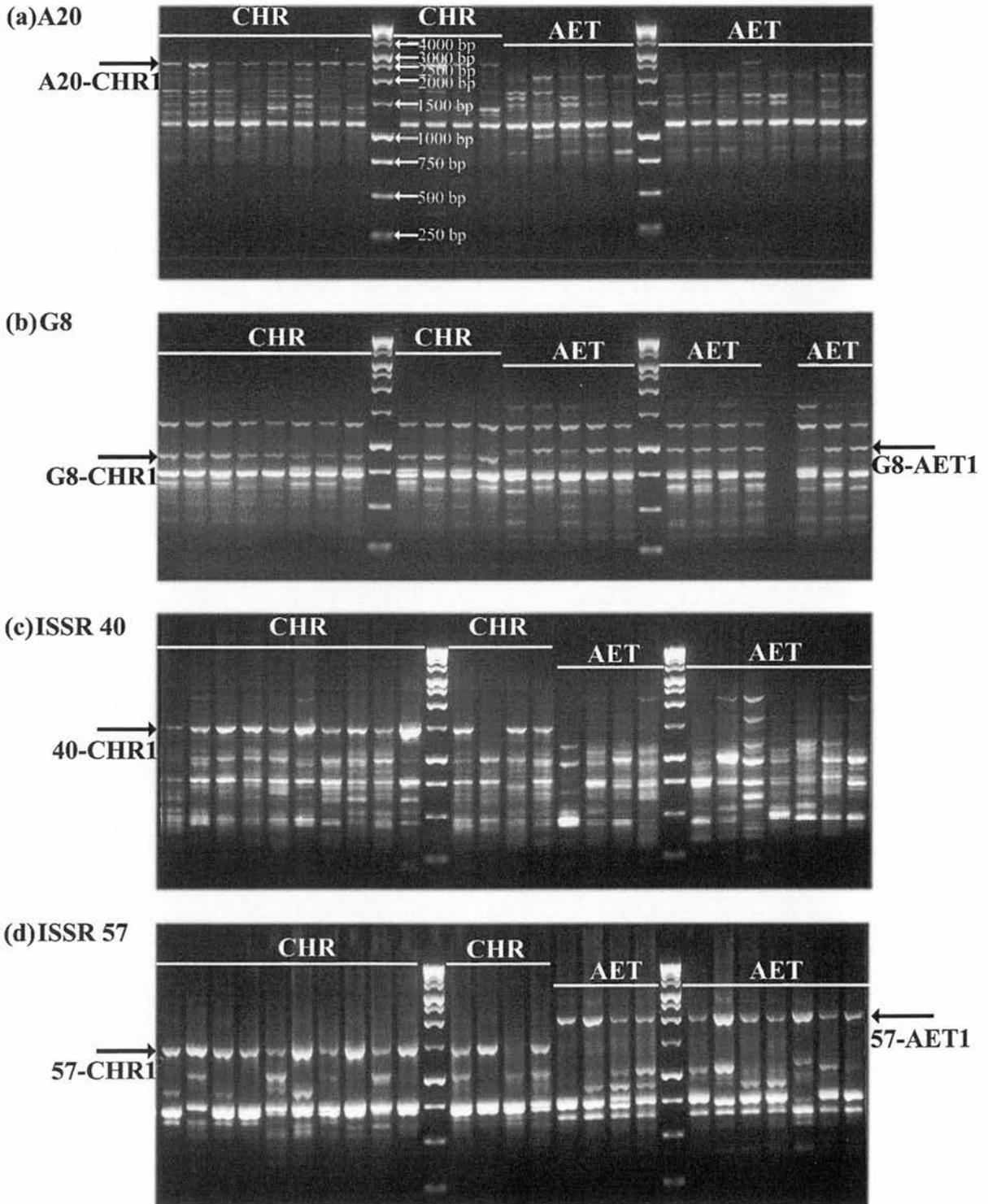
<sup>5</sup> These bands were used to look at Sicilian hybrid material (chapter 4) but were rejected at that stage because:

<sup>6</sup> They could not be scored accurately in hybrid individuals, for similar reasons as those given in <sup>1</sup>; or <sup>7</sup> they appeared to be unreliable (standards did not match, see section 4.2.2).

Fifty-five bands, generated by 35 primers, were found to satisfactorily distinguish between the two taxa in the full survey. They were comparatively easy to score (i.e. clear, bright, reproducible and not too close to other bands), and also proved to be useful and reliable in later studies. Twenty-nine of these bands were *S. chrysanthemifolius* markers and 26 were *S. aethnensis* markers. Examples of these bands are shown in plate 3.1.

Table 3.10 partitions these 55 bands into the categories outlined in section 3.1.1. Five bands were diagnostic taxon-specific markers (monomorphic in one taxon and absent in the other taxon), while six bands were taxon-specific (I) markers (present in one taxon at very

**Plate 3.1** Examples of *Senecio chrysanthemifolius* and *S. aethnensis* marker bands (highlighted by arrows) amplified by (a) RAPD primer A20, (b) RAPD primer G8, (c) ISSR primer 40, and (d) ISSR primer 57 (see table 3.11 for further details). *Senecio chrysanthemifolius* individuals are denoted CHR, *S. aethnensis* individuals are denoted AET. Band sizes in the 1 kb ladder are indicated in (a).



high frequency and absent in the other taxon). Two additional bands were taxon-specific (II) markers (present in one taxon at a lower frequency and absent in the other taxon). Most bands (41) were "frequency-difference" markers, 31 of which were category I, II or III markers (i.e. present at very high frequency in one taxon and at very low frequency in the other taxon).

**Table 3.10** *Senecio chrysanthemifolius* and *S. aethnensis* markers retained for future use, partitioned according to the categories outlined in section 3.1.1. 'CHR' represents *S. chrysanthemifolius* marker bands; 'AET' represents *S. aethnensis* marker bands.

Category of marker	<i>S. chrysanthemifolius</i> markers	No.	<i>S. aethnensis</i> markers	No.
<b>diagnostic taxon-specific</b>	B12-CHR-1,	1	H19-AET-1, F2/G9-AET-1, 26-AET-1, 57-AET-1	4
<b>taxon-specific (I)</b>	A11-CHR-1, A13-CHR-1, A14-CHR-1, G10-CHR-1, G12-CHR-1, 86-CHR-1	6		
<b>taxon-specific (II)</b>			B15-AET-1, C11-AET-1	2
<b>(private band)</b>			A11-AET-2 <sup>1</sup>	1
<b>frequency-difference (I)</b>	A9-CHR-1, B15-CHR1, C2-CHR-1, C5-CHR-1, G1-CHR-1, G8-CHR-1, H7-CHR-1, H19-CHR-1, 26-CHR-1, 40-CHR-1, 57-CHR-1	11	A11-AET-1, F12-AET-1, G6-AET-1, G11-AET-1, G15-AET-1, H13-AET-1	6
<b>frequency-difference (II)</b>	A14-CHR-1, A20-CHR-1, G5-CHR-1, F12-CHR-1, 18-CHR-1, 25-CHR-1	6	A14-AET-1, C2-AET-1, C20-AET-1, F2/G9-AET-1, 18-AET-1, 88-AET-1 G8-AET-1, G12-AET-1	6
<b>frequency-difference (III)</b>				2
<b>frequency-difference (IV)</b>	B2/B10-CHR-1	1	B6-AET-1, C20-AET-2, G5-AET-1	3
<b>frequency-difference (V)</b>	B6-CHR-1, G14-CHR-1, G14-CHR-2, H4-CHR-1	4	G15-AET-2, 26-AET-2	2
<b>all</b>		29		26

<sup>1</sup> Two other bands from primer A11 were scored, so this "private band" was retained.

Table 3.11 presents the frequencies of the 55 markers in the two taxa, split into their constituent populations. It is clear that the overall frequency of a band in a taxon may mask variation between populations within the taxa. For example, the overall frequency of band H7-CHR-1 in *S. chrysanthemifolius* was 0.82 but in population C9 it was present at a frequency of only 0.53. Other bands exhibiting similar discrepancies (i.e. between populations within one taxon) are G15-AET-2, G14-CHR-1, G5-CHR-1 and 18-CHR-1. Several of these discrepancies can, in part, be explained by the increased likelihood of sampling error in the small population of *S. aethnensis* (population TC consisted of seven individuals compared to a minimum of 14 in most other populations).

Eight RAPD primers and one ISSR primer were repeated (for all 90 individuals). One thousand and fifty-five comparisons were made between bands in the repeats and bands in the originals, across 12 markers. The error rate (i.e. the percentage of discrepancies between repeats and originals) was less than 2.1%.

**Table 3.11** Frequencies of taxon-specific RAPD/ISSR markers in 'pure' populations of *Senecio chrysanthemifolius* and *S. aethnensis*. 'CHR' identifies *S. chrysanthemifolius* markers; 'AET' identifies *S. aethnensis* markers.

Primer	Band <sup>†</sup>	<i>S. chrysanthemifolius</i>				<i>S. aethnensis</i>				(VO/UJ) <sup>*</sup> n = 4
		C0 n = 16	C1 n = 14	C9 n = 15	Overall n = 45	VB n = 17	BB n = 17	TC n = 7	Overall n = 45	
A9	CHR-1 1050	0.94 <sup>‡</sup>	0.93	0.80	<b>0.89</b>	0.06	0.00	0.00	<b>0.02</b>	0
A11	CHR-1 800	0.81	0.79	0.87	<b>0.82</b>	0.00	0.00	0.00	<b>0.00</b>	0
	AET-1 650	0.00	0.07	0.00	<b>0.02</b>	1.00	1.00	1.00	<b>1.00</b>	4
	AET-2 550	0.00	0.00	0.00	<b>0.00</b>	0.35	0.41	0.14	<b>0.36</b>	2
A13	CHR-1 700	0.88	0.71	0.87	<b>0.82</b>	0.00	0.00	0.00	<b>0.00</b>	0
A14	CHR-1 1650	1.00	0.85 <sup>13</sup>	0.93	<b>0.93<sup>44</sup></b>	0.00	0.00	0.00	<b>0.00</b>	0
	CHR-2 600	0.94	0.69 <sup>13</sup>	0.80	<b>0.82<sup>44</sup></b>	0.00	0.12	0.14	<b>0.09</b>	1
	AET-1 1450	0.00	0.06 <sup>13</sup>	0.20	<b>0.09<sup>44</sup></b>	0.88	0.76	1.00	<b>0.87</b>	4
A20	CHR-1 2500	0.94	1.00	1.00	<b>0.98</b>	0.00	0.24	0.00	<b>0.09</b>	0
B6	CHR-1 520	1.00 <sup>15</sup>	0.69 <sup>13</sup>	0.60	<b>0.77<sup>43</sup></b>	0.06	0.29	0.29	<b>0.18</b>	0
	AET-1 800	0.00 <sup>15</sup>	0.08 <sup>13</sup>	0.00	<b>0.02<sup>43</sup></b>	0.53	0.65	0.57	<b>0.60</b>	3
B12	CHR-1 500	1.00	1.00	1.00	<b>1.00</b>	0.00	0.00	0.00	<b>0.00</b>	0
B15	CHR-1 1400	0.69	1.00	0.93	<b>0.87</b>	0.00	0.00	0.00	<b>0.02</b>	1
	AET-1 450	0.00	0.00	0.00	<b>0.00</b>	0.41	0.47	0.86	<b>0.53</b>	3
C2	CHR-1 700	0.94	0.93	0.93	<b>0.93</b>	0.12	0.00	0.00	<b>0.04</b>	0
	AET-1 1550	0.00	0.14	0.07	<b>0.07</b>	0.82	0.76	0.86	<b>0.80</b>	3
C5	CHR-1 570	0.94	1.00	1.00	<b>0.98</b>	0.06	0.00	0.00	<b>0.04</b>	1
C11	AET-1 950	0.00 <sup>15</sup>	0.00	0.00	<b>0.00<sup>44</sup></b>	0.47	0.24	0.57	<b>0.40</b>	2
C20	AET-1 1450	0.00	0.14	0.07	<b>0.07</b>	1.00	1.00	1.00	<b>1.00</b>	4
	AET-2 1300	0.00 <sup>15</sup>	0.00	0.13	<b>0.05<sup>44</sup></b>	0.41	0.59	0.57	<b>0.51</b>	2
F12	CHR-1 250	1.00	1.00	1.00	<b>1.00</b>	0.18	0.06	0.00	<b>0.09</b>	0
	AET-1 350	0.06	0.00	0.00	<b>0.02</b>	0.82	0.65	0.71	<b>0.73</b>	3
G1	CHR-1 1400	1.00	1.00	1.00	<b>1.00</b>	0.00	0.06	0.14	<b>0.04</b>	0
G5	CHR-1 450	0.75	1.00	0.93	<b>0.89</b>	0.06	0.06	0.29	<b>0.09</b>	0
	AET-1 350	0.00	0.00	0.07	<b>0.02</b>	0.76	0.59	0.14	<b>0.58</b>	2
G6	AET-1 1500	0.06	0.00	0.00	<b>0.02</b>	1.00	1.00	1.00	<b>1.00</b>	4
G8	CHR-1 900	0.88	1.00	0.93	<b>0.93</b>	0.06	0.00	0.00	<b>0.02</b>	0
	AET-1 950	0.25	0.00	0.13	<b>0.13</b>	1.00	1.00	1.00	<b>1.00</b>	4

Table 3.11 continued.

Primer	Band <sup>†</sup>	<i>S. chrysanthemifolius</i>				<i>S. aethnensis</i>				(VO/ UJ) <sup>*</sup> n = 4
		C0	C1	C9	Overall	VB	BB	TC	Overall	
		n = 16	n = 14	n = 15	n = 45	n = 17	n = 17	n = 7	n = 45	
G10	CHR-1 600	0.88	0.79	0.80	<b>0.82</b>	0.00	0.00	0.00	<b>0.00</b>	0
G11	AET-1 320	0.00	0.07	0.00	<b>0.02</b>	0.94	0.82	1.00	<b>0.89</b>	3
G12	CHR-1 1450	1.00 <sup>‡</sup>	0.93	0.87	<b>0.93</b>	0.00	0.00	0.00	<b>0.00</b>	0
	AET-1 300	0.19	0.14	0.07	<b>0.13</b>	1.00	1.00	1.00	<b>1.00</b>	4
G14	CHR-1 520	0.50	0.69 <sup>13</sup>	0.67	<b>0.61<sup>44</sup></b>	0.06	0.35	0.14	<b>0.18</b>	0
	CHR-2 470	0.94	0.93	1.00	<b>0.96</b>	0.12	0.29	0.29	<b>0.20</b>	0
G15	AET-1 1200	0.06	0.00	0.07 <sup>14</sup>	<b>0.05<sup>44</sup></b>	1.00	1.00 <sup>16</sup>	0.57	<b>0.93<sup>44</sup></b>	4
	AET-2 850	0.00	0.21	0.36 <sup>14</sup>	<b>0.18<sup>44</sup></b>	0.65	0.56 <sup>16</sup>	0.43	<b>0.59<sup>44</sup></b>	3
H4	CHR-1 700	0.94	0.93	0.93	<b>0.93</b>	0.31 <sup>16</sup>	0.18	0.00	<b>0.18<sup>44</sup></b>	0
H7	CHR-1 600	1.00	0.93	0.53	<b>0.82</b>	0.00	0.00	0.00	<b>0.02</b>	1
H13	AET-1 1150	0.06	0.00	0.00	<b>0.02</b>	1.00	1.00	1.00	<b>1.00</b>	4
H19	CHR-1 650	1.00	1.00 <sup>13</sup>	1.00	<b>1.00<sup>44</sup></b>	0.06 <sup>16</sup>	0.00	0.00 <sup>6</sup>	<b>0.02<sup>42</sup></b>	0 <sup>3</sup>
	AET-1 370	0.00	0.00	0.00	<b>0.00</b>	1.00	1.00	1.00	<b>1.00</b>	4
B2/B10	CHR-1 520	0.73 <sup>15</sup>	0.64	0.60	<b>0.66<sup>44</sup></b>	0.12	0.12	0.00	<b>0.09</b>	0
F2/G9	AET-1 900	0.07 <sup>15</sup>	0.14	0.00	<b>0.07<sup>44</sup></b>	0.88	0.94	1.00	<b>0.93</b>	4
	AET-2 520	0.00 <sup>15</sup>	0.00	0.00	<b>0.00<sup>44</sup></b>	1.00	1.00	1.00	<b>1.00</b>	4
18	CHR-1 1300	0.94	0.93	0.67	<b>0.84</b>	0.00	0.00 <sup>16</sup>	0.43	<b>0.07<sup>44</sup></b>	0
	AET-1 2100	0.00	0.07	0.13	<b>0.07</b>	0.88	0.94	1.00	<b>0.89</b>	2
25	CHR-1 1530	0.75	1.00	1.00	<b>0.91</b>	0.18	0.00	0.00	<b>0.07</b>	0
26	CHR-1 2300	0.67 <sup>15</sup>	0.86	0.80	<b>0.77<sup>44</sup></b>	0.00	0.06	0.14	<b>0.04</b>	0
	AET-1 2500	0.00 <sup>15</sup>	0.00	0.00	<b>0.00<sup>44</sup></b>	1.00	1.00	1.00	<b>1.00</b>	4
	AET-2 2100	0.13 <sup>15</sup>	0.36	0.07	<b>0.18<sup>44</sup></b>	1.00	1.00	1.00	<b>1.00</b>	4
40	CHR-1 1450	0.94	0.71	0.93	<b>0.87</b>	0.00	0.06	0.00	<b>0.02</b>	0
57	CHR-1 1450	1.00 <sup>15</sup>	1.00	0.87	<b>0.93<sup>44</sup></b>	0.00	0.00	0.00	<b>0.02</b>	1
	AET-1 2050	0.00 <sup>15</sup>	0.00	0.00	<b>0.00<sup>44</sup></b>	1.00	1.00	1.00	<b>1.00</b>	4
86	CHR-1 1650	0.88	0.86	0.79 <sup>14</sup>	<b>0.84<sup>44</sup></b>	0.00	0.00	0.00	<b>0.00</b>	0
88	AET-1 950	0.06	0.07	0.07	<b>0.07</b>	0.71	0.82	0.71	<b>0.78</b>	4

<sup>†</sup> The figure below each band indicates its approximate size.

<sup>‡</sup> Numbers in superscript after a frequency show the number of individuals in the population scored for that band. No superscript number indicates that all individuals in the population were scored, i.e. the number indicated at the top of that column.

<sup>\*</sup> The number of individuals with the band - a frequency was not calculated for populations VO and UJ due to the very small sample sizes (two individuals per population) which would result in serious bias.

### 3.3.2 Results of the survey of isozyme variation within and between *S. chrysanthemifolius* and *S. aethnensis*

Abbott *et al.* (2000) surveyed isozyme variation in three populations of 'pure' *S. chrysanthemifolius* and three populations of *S. aethnensis* (table 3.6). Variation was assayed for six enzymes, and eight loci in total were scored. Three of the eight loci (*G3pd-1*, *Pgi-1* and *Pgm-1*) were monomorphic in all six populations, and results from the five polymorphic loci are summarised in table 3.12.

**Table 3.12** Allele frequencies at five polymorphic loci in *Senecio chrysanthemifolius* and *S. aethnensis* after pooling data over populations (data from Abbott *et al.*, 2000).

Locus/allele	Taxon	
	<i>S. chrysanthemifolius</i> <sup>†</sup>	<i>S. aethnensis</i> <sup>‡</sup>
<i>Acp-2</i>		
	N	(95)
<i>a</i>		----
<i>b</i>		1.000
		(62)
		0.645
		0.355
<i>Aat-3</i>		
	N	(94)
<i>e</i>		0.925
<i>f</i>		0.075
		(62)
		0.911
		0.089
<i>β-Est-3</i>		
	N	(87)
<i>e</i>		0.632
<i>f</i>		0.368
		(61)
		0.107
		0.893
<i>Pgi-2</i>		
	N	(88)
<i>b</i>		----
<i>c</i>		1.000
		(59)
		0.023
		0.906
		0.119

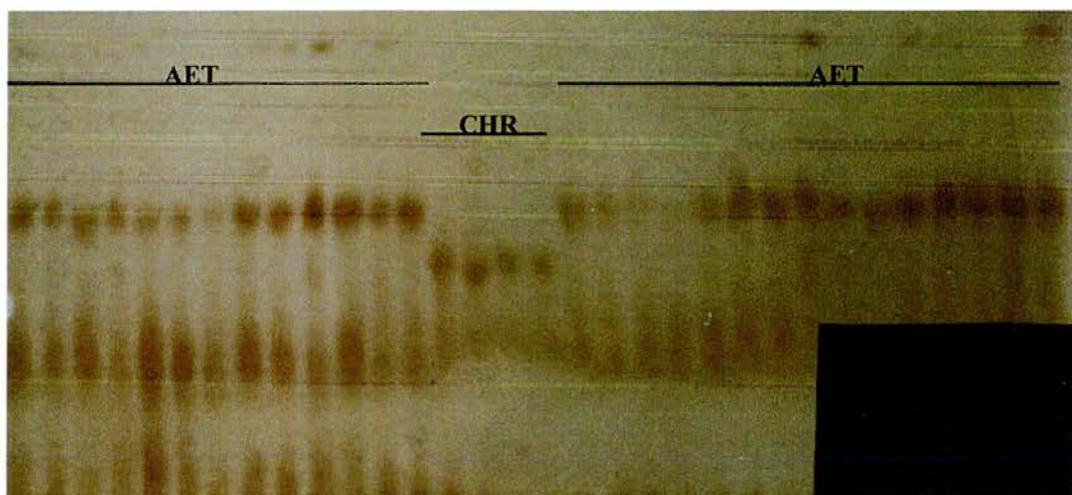
Data from the five polymorphic loci were examined to ascertain whether any alleles could be employed as taxon-specific or frequency-difference markers for *S. chrysanthemifolius* or *S. aethnensis*. Notable differences in allele frequencies, between *S. chrysanthemifolius* and *S. aethnensis*, were observed at three loci - *Pgm-2*, *Pgi-2* and *Acp-2*. At the *Pgm-2* locus, *S. chrysanthemifolius* possessed allele *Pgm-2 a* that was not present in *S. aethnensis*. At the *Acp-2* locus, *S. aethnensis* possessed allele *Acp-2 a* that was not present in *S. chrysanthemifolius*, whilst *Acp-2 b* was present at high frequency in *S. chrysanthemifolius* and low frequency in *S. aethnensis*. At the *Pgi-2* locus, *S. aethnensis* possessed allele *Pgi-2 b* that was not present in *S. chrysanthemifolius*.

Although *Pgm-2 a* was specific to *S. chrysanthemifolius* it could not be employed as a species marker, due to its very low frequency in *S. chrysanthemifolius*. *Pgi-2 b* and *Acp-2 b*, likewise, did not meet the criteria for markers outlined in section 3.1.1. In contrast, *Acp-2 a* did meet the criteria and could be considered a taxon-specific marker for *S. aethnensis*.

However, observations in the field (see section 4.3.1) suggested that the individuals of *S. aethnensis* examined in this isozyme analysis were *S. aethnensis*-like hybrids rather than 'pure' *S. aethnensis*; indeed, it had been acknowledged that this was a possibility (Abbott *et al.*, 1995). It was suggested that if 'pure' *S. aethnensis* populations were to be surveyed, differences in allele frequencies at the *Acp* and *Pgi* loci, between *S. chrysanthemifolius* and *S. aethnensis*, might become more pronounced. In particular, there was speculation that 'pure' *S. aethnensis* might be monomorphic for *Acp-2 a*. Therefore, the decision was taken to survey 'pure' populations of *S. aethnensis* for variation at the *Acp-2* (plate 3.2) and *Pgi-2* loci.

Allele frequencies at the *Acp-2* and *Pgi-2* loci in populations of 'pure' *S. aethnensis* are displayed in table 3.13. All individuals, in all populations, were monomorphic for *Acp-2 a*. The frequency of *Pgi-2 b* varied across populations; ranging from 0.1 to 0.28.

**Plate 3.2** Isozyme gel showing variation at the *Acp-2* locus. *Senecio aethnensis* individuals (denoted AET) are homozygous for the *a* (fast) allele, whereas *S. chrysanthemifolius* individuals (denoted CHR) are homozygous for the *b* (slow) allele.



**Table 3.13** Allele frequencies at *Acp-2* and *Pgi-2* in populations of 'pure' *Senecio aethnensis*.

Locus/allele		population			TC
		BB	VB	TJ	
<i>Acp-2</i>					
	N	(31)	(30)	(25)	(5)
<i>a</i>		1.00	1.00	1.00	1.00
<i>b</i>		----	----	----	----
<i>Pgi-2</i>					
	N	(27)	(30)	(25)	(5)
<i>b</i>		0.241	0.167	0.280	0.100
<i>c</i>		0.759	0.833	0.720	0.900

The data for *Acp-2* and *Pgi-2* were pooled over all populations of 'pure' *S. aethnensis*, for comparison (table 3.14) with the populations of 'pure' *S. chrysanthemifolius* surveyed by Abbott *et al.* (2000). *Acp-2 a* was categorised as a diagnostic taxon-specific marker for *S. aethnensis* and *Acp-2 b* as a diagnostic taxon-specific marker for *S. chrysanthemifolius*. *Pgi-2 b* was classed as a private band (section 3.1.1) for *S. aethnensis*.

**Table 3.14** Allele frequencies at *Acp-2* and *Pgi-2* in populations of 'pure' *Senecio aethnensis* (data from present study) and *S. chrysanthemifolius* (data from Abbott *et al.*, 2000), after pooling data over populations.

Locus/allele		taxon	
		<i>S. chrysanthemifolius</i>	<i>S. aethnensis</i>
<i>Acp-2</i>			
	N	(95)	(91)
<i>a</i>		----	1.000
<i>b</i>		1.000	----
<i>Pgi-2</i>			
	N	(88)	(87)
<i>b</i>		----	0.218
<i>c</i>		1.000	0.782

### 3.3.3 Results of the survey of chloroplast DNA RFLP variation within and between *S. chrysanthemifolius* and *S. aethnensis*.

Abbott *et al.* (1995) surveyed cpDNA RFLP variation in various *Senecio* species, using eighteen restriction enzymes and a library of *Lactuca sativa* L. cpDNA probes. Five clear polymorphisms were resolved, allowing three cpDNA haplotypes - A, B and C - to be

distinguished. Haplotype A differed from B due to a single restriction site mutation and C differed from A and B with regard to three restriction site mutations and a length mutation. To detect the single restriction site mutation, which differentiated haplotypes A and B, the enzyme *Cl*I and *Lactuca sativa* L. probe C6 were employed. Individuals of haplotype A possessed 3.1 kb fragment (the primitive state), whereas B haplotype individuals possessed a 3.3 kb fragment (the derived state - loss of a restriction site having led to the replacement of the 3.1 kb fragment and, a presumed but undetected, 0.2 kb fragment with a single fragment).

Abbott *et al.* (1995) tentatively concluded that haplotype A was primitive and ancestral to haplotypes B and C (since it was present in the outgroup, *S. jacobaea*). Eleven individuals of *S. chrysanthemifolius* and five individuals of *S. aethnensis* were included in the survey. All 11 *S. chrysanthemifolius* possessed the B haplotype, while *S. aethnensis* was polymorphic for A and B (four B haplotype individuals, one A haplotype individual).

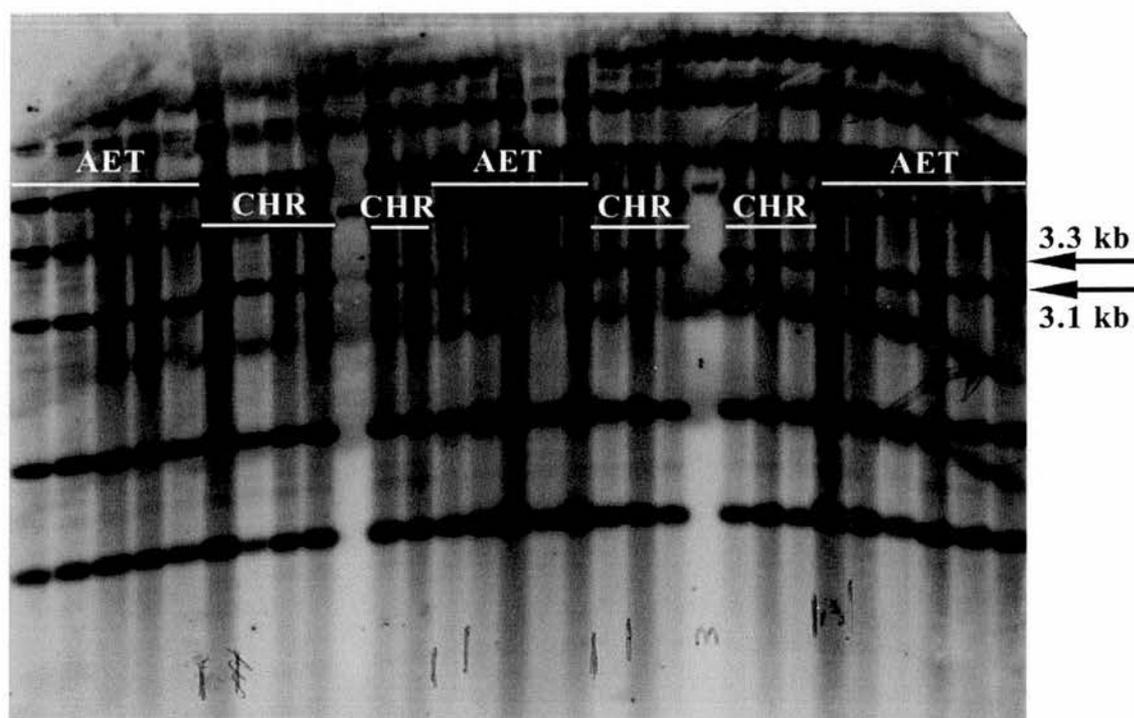
It was believed that the *S. aethnensis* individuals examined in the cpDNA analysis were *S. aethnensis*-like hybrids rather than 'pure' *S. aethnensis* (see section 4.3.1) and this led Abbott *et al.* (1995) to speculate that the A haplotype might be more common in populations of 'pure' *S. aethnensis*. If the A haplotype were to be present at high frequency in 'pure' *S. aethnensis*, then the A and B haplotypes could be useful markers for *S. aethnensis* and *S. chrysanthemifolius*, respectively. Consequently, it was decided that 'pure' populations of *S. aethnensis* would be assessed for their cpDNA haplotype (plate 3.3). In addition, a number of 'pure' *S. chrysanthemifolius* individuals would be examined, as the original survey was based upon so few accessions.

Thirty nine 'pure' *S. chrysanthemifolius* individuals and 43 'pure' *S. aethnensis* individuals were typed for their cpDNA (table 3.15). *Senecio chrysanthemifolius* was monomorphic for the B haplotype. In contrast, *S. aethnensis* individuals possessed either the B or the A haplotype. Pooling data across *S. aethnensis* populations gave an overall frequency of 0.72 for the A haplotype and 0.28 for the B haplotype. These figures mask considerable variation between *S. aethnensis* populations: population VO/UJ was fixed for the A haplotype, whereas less than 50% of the individuals in population TC had the A haplotype. However, the sample sizes of populations VO/UJ and TC were small and the results from these populations should therefore be interpreted with caution.

**Table 3.15** Chloroplast DNA haplotype frequencies of 'pure' *Senecio chrysanthemifolius* and *S. aethnensis* populations.

Taxon/population (N)	Chloroplast DNA haplotype	
	Haplotype A	Haplotype B
<i>S. chrysanthemifolius</i>		
C0 (15)	0.00	1.00
C1 (12)	0.00	1.00
C9 (12)	0.00	1.00
<b>all populations (39)</b>	<b>0.00</b>	<b>1.00</b>
<i>S. aethnensis</i>		
BB (17)	0.94	0.06
VB (17)	0.65	0.35
TC (7)	0.43	0.57
UJ/VO (2)	1.00	0.00
<b>all populations (43)</b>	<b>0.72</b>	<b>0.28</b>

**Plate 3.3** Luminograph of total genomic DNA digested with restriction enzyme *Cla*I and probed with *Lactuca sativa* cloned cpDNA probe pLsC6. The site mutation that differentiates cpDNA haplotype A from haplotype B (possession of 3.1 kb fragment and 3.3 kb fragment, respectively) is highlighted. Individuals of *Senecio chrysanthemifolius* are denoted CHR, individuals of *S. aethnensis* are denoted AET.



## 3.4 Discussion

### 3.4.1 The effectiveness of the survey of RAPD and ISSR variation within and between *S. chrysanthemifolius* and *S. aethnensis*

#### 3.4.1.1 The effectiveness of the preliminary survey

The preliminary survey was a rather crude but effective procedure for rejecting primers thought to have little potential for producing taxon-specific markers, thereby minimising resource wastage. The approach reduced the number of primers employed in the analysis from 305 to a more manageable 65 and was also effective in terms of the eventual yield of useful markers (35 of the 65 primers selected for the detailed survey produced useful markers, 3.3.1.2). However, the preliminary survey was not infallible and it is possible that a number of primers did produce potential taxon-specific bands but were mistakenly rejected due to sampling effects and experimental error.

RAPD primers produced more potential taxon-specific bands per primer than did ISSR primers or RAPD primer-pairs. Many ISSR primers generated poorly resolved banding patterns, for example smeared products or inadequately amplified products. Better resolution of ISSR bands might have been obtained by optimising reaction conditions for each ISSR primer individually, but this would have required more time. It is also possible that some of the ISSR primers had degraded somewhat during storage. The employment of RAPD primer-pairs was decidedly unsuccessful in generating potential taxon-specific markers, although why this was the case is not known.

The approach adopted in the preliminary survey involved screening a large number of primers over just a few individuals (six) per taxon. In contrast, Dawson *et al.* (1996) and Perron *et al.* (1995) utilised a modified form of bulked segregant analysis (Michelmore *et al.*, 1991) to identify taxon-specific markers. In this, the DNA of several individuals from one taxon is pooled, so that a comparison can be made with the pooled DNA of a second taxon. The advantage of this approach is that only two samples are run with each primer. Moreover, each sample may contain a larger number of individuals per taxon than is practical to screen individually. However, a band which is present in a pooled sample is not guaranteed to be present in all individuals within the pool (a band may be visible if present

in only some of the individuals), and conversely the absence of a band from a DNA pool does not guarantee that it is absent from all individuals of that pool, due to PCR competition effects (Michelmore *et al.*, 1991). Thus, it was felt that the non-pooled DNA method used in the present study was more suitable for indicating which primers were likely to yield taxon-specific markers.

#### **3.4.1.2 The effectiveness of the detailed survey**

The detailed survey of RAPD and ISSR variation resulted in the development of 14 taxon-specific markers; seven that were specific to *S. chrysanthemifolius* and seven that were specific to *S. aethnensis* (table 3.10). In addition, a large number (41) of frequency-difference markers were developed and most of these exhibited extreme frequency differences between the two taxa. Use of these taxon-specific and frequency-difference markers should be most informative when examining hybridisation between these taxa; both for the identification of hybrids between them and for the characterisation of their putative hybrid species, *Senecio squalidus*.

The disadvantages of RAPDs and ISSRs as molecular markers are discussed elsewhere (see section 1.4.3.2) and include questions of their reliability and reproducibility and the fact that in heterozygotes their presence is dominant to their absence. However, the RAPD and ISSR markers produced were believed to be reliable and reproducible. Furthermore, dominance need not be a problem if RAPD and ISSR markers are scored as phenotypes rather than genotypes (Haig *et al.*, 1994).

The number of RAPD and ISSR molecular markers developed in this study, that distinguished between the two taxa (55 taxon-specific/frequency-difference markers from an initial screen of 305 primers), was comparable with similar studies. For example, Perron *et al.* (1995) screened 161 RAPD primers and obtained seven taxon-specific markers for *Picea mariana* and *Picea rubens* (closely related spruce). In other studies of closely related species considerably fewer markers have been generated. For example, Howard *et al.* (1997) screened 700 RAPD primers and obtained only two taxon-specific markers and six frequency-difference markers for *Quercus gambelii* and *Quercus grisea* (closely related oaks).

*Development of more RAPD and ISSR molecular markers to distinguish S. chrysanthemifolius and S. aethnensis*

It would have been possible to increase the number of taxon-specific RAPD/ISSR markers for *S. chrysanthemifolius* and *S. aethnensis* in two ways. Firstly, one could have attempted to exploit some of the markers rejected in this study i.e. bands that did meet the frequency-difference criteria set, private bands and even bands considered too difficult to score, for example by optimising PCR reaction conditions for each primer individually. Secondly, one could have surveyed more primers; RAPD primers clearly offered the greatest potential for finding new markers. Many RAPD other primer sets were available but the limit was set by resources (time and money).

**3.4.2 The survey of isozyme variation between *S. chrysanthemifolius* and *S. aethnensis***

Previous work by Abbott *et al.* (2000) had indicated that alleles at two enzyme loci, *Acp-2* and *Pgi-2*, might be suitable candidates for taxon-specific markers. A thorough survey of 'pure' *S. aethnensis* populations from high altitude sites was undertaken to complement the thorough survey of 'pure' *S. chrysanthemifolius* previously undertaken by Abbott *et al.* (2000). At the *Acp-2* locus, alternate alleles were fixed in *S. chrysanthemifolius* and *S. aethnensis*; thus *Acp-2 a* could be used as a diagnostic marker for *S. aethnensis* and *Acp-2 b* as a diagnostic marker for *S. chrysanthemifolius*. The polymorphism at *Pgi-2* did not prove to be so useful. Although *Pgi-2 b* was specific to *S. aethnensis*, it fell into the category of 'private band'. Isozyme markers are particularly useful because they are codominant, nuclear markers and therefore complement RAPD and ISSR markers (which tend to be dominant, nuclear markers).

**3.4.3 The survey of chloroplast DNA RFLP variation between *S. chrysanthemifolius* and *S. aethnensis***

Previous work by Abbott *et al.* (1995) had suggested that 'pure' *S. chrysanthemifolius* and *S. aethnensis* might possess different cpDNA haplotypes. A survey of cpDNA variation in 'pure' populations of *S. chrysanthemifolius* and *S. aethnensis* revealed that the A haplotype was a taxon-specific marker for *S. aethnensis* but the B haplotype, although characteristic of *S. chrysanthemifolius*, did not meet the criteria for markers outlined in section 3.1.1, since it was present at too high a frequency (> 0.20) in *S. aethnensis*. It was hoped that a difference in cpDNA haplotype between the two species would give an indication of the

level of seed-mediated gene-flow (as opposed to pollen-mediated gene-flow) between the two species.

#### **3.4.4 The taxonomic status of *S. chrysanthemifolius* and *S. aethnensis***

The taxonomic status of *S. chrysanthemifolius* and *S. aethnensis* has been a matter of debate. Some botanists have awarded both taxa species status (Alexander, 1979), while others have considered them to be sub-species (Poli Marchese, 1991). This thesis supports the treatment of *S. chrysanthemifolius* and *S. aethnensis* as closely related but *bona fide* species. The two taxa are morphologically and ecologically divergent (Crisp, 1972; table 4.1) and also genetically distinct. The level of genetic differentiation between *S. chrysanthemifolius* and *S. aethnensis* (in terms of the number of RAPD and ISSR markers that distinguish the two taxa) compares very favourably with the level of genetic differentiation between other closely related but well-established pairs of species (again in terms of number of RAPD and ISSR markers that distinguish them; see 3.4.1.2).

Although, it would be desirable to use the RAPD and ISSR data set to obtain an objective measure of genetic distance between the two species, this would not be a valid course of action because the markers had been pre-selected to show species-specific differences which introduces considerable bias into measurement of genetic distances (Harris, 1999). A more accurate way of assessing genetic distance between the two species would be to choose a number of RAPD and ISSR primers at random and to score the RAPD and ISSR bands generated without pre-selection.

#### **3.4.5 Assumptions made and improvements suggested**

Three major assumptions were made in the development of molecular markers to distinguish between *S. chrysanthemifolius* and *S. aethnensis*:

*Assumption one:* - individuals examined were genetically 'pure' representatives of the taxa under consideration.

The first, and perhaps most important, assumption is that individuals used in the development of markers are 'pure' representatives of the taxa compared; i.e. they are not hybrids or introgressants. The presence of hybrids or introgressants can cause serious bias

in the interpretation of data (Perron *et al.*, 1995). To ensure that individuals are 'pure', one should ideally sample only from populations outside the limits of inter-taxon gene-flow.

In this study, the assumption of 'pure' individuals is not necessarily valid. The sampled individuals came from populations whose locations were conceivably within range of gene-flow from the other taxon. Indeed, it is likely that some of the supposedly 'pure' individuals from each taxon contained genetic material introgressed from the other taxon. As a consequence, it is possible that some frequency-difference markers might yet be found to be taxon-specific, if one was able to sample from populations that have never been subject to inter-specific gene-flow. Alternatively, it is feasible that both taxa possess the same band - as in the case of frequency-difference markers - due to sharing of ancestral polymorphisms rather than recent inter-taxa introgression (Shoemaker *et al.*, 1994).

An improvement to this study would have been to have used material from populations away from the area of potential inter-specific gene flow. Such populations of *S. chrysanthemifolius* do exist (e.g. on the Aeolian Islands, on Ustica and in Calabria; Poli Marchese, 1991) but sadly these were not collected from, due to time constraints. However, populations of *S. aethnensis* outside the range of potential inter-specific gene-flow probably do not exist, since *S. aethnensis* is endemic to the summit of Mt. Etna and within range of gene-flow from *S. chrysanthemifolius* populations growing on the mountain at lower altitudes.

As a further safeguard against mistaking introgressant for 'pure' material, Perron *et al.* (1995) recommended *a priori* selection of individuals using morphological traits that clearly separate the two taxa. *A priori* selection was not carried out in the present study. Nevertheless, the *S. chrysanthemifolius* and *S. aethnensis* plants utilised did not deviate noticeably from the expected morphology of 'pure' representatives of each species.

*Assumption two: - individuals examined were representative of each taxon as a whole.*

The second assumption is that the sampled individuals are representative of each taxon as a whole i.e. provide sufficient coverage of the geographical and environmental range of each taxon. This is to ensure that the markers identified are *taxon-specific*, and not merely *population, region or habitat specific*.

The populations of *S. chrysanthemifolius* chosen for study were reasonably widely spaced (up to 40 km apart) and marker frequencies were fairly constant across them. This suggests that, although *S. chrysanthemifolius* was not sampled from its entire distribution, the assumption is broadly satisfied. In the case of *S. aethnensis* the assumption has been met, due to the restricted distribution of the taxon.

*Assumption three:- enough individuals per population and per taxon were sampled.*

The final assumption (common to most population genetic studies) is that large enough numbers of individuals are sampled per population and per taxon to ensure that the marker frequencies obtained are a fair representation of the real situation and not an effect of sampling bias due to small sample sizes. It is also assumed that the number of individuals sampled is sufficient to identify markers present in a population at low frequency. This is important because it can effect the classification of bands (e.g. whether a band is labelled as taxon-specific marker or a frequency-difference marker, see section 3.1.1).

It is possible that the RAPD and ISSR analysis did not satisfy the assumption of adequate sample sizes; although the number of individuals sampled (45 individuals per taxon and population sizes that mostly range between 14-17) was certainly not unusual. For instance, Perron *et al.* (1995) used approximately 22 individuals of each of two species to develop taxon-specific RAPD molecular markers, while Howard *et al.* (1997) used 41 individuals of one species and 20 of another species for the same purpose. Other studies have relied on *very* small sample sizes. For example, the RAPD taxon-specific markers used by Padgett *et al.* (1998) to examine a postulated hybrid species of *Nuphar* were based on less than six accessions of each parental species. Some authors do not even mention sample size (e.g. Arnold, 1993). Needless to say, the knowledge that other studies have been based upon equally inadequate sizes is no cause for complacency and, ideally, many more individuals per population and per taxon should have been sampled to improve confidence in the detection of taxon-specific markers. It is difficult to establish what the 'correct' sample size, per taxon and per population, would have been for this RAPD and ISSR analysis. For example, Lynch and Milligan (1994) suggested that, for RAPD analysis, a minimum of 100 individuals per population was needed. However, this figure was based upon the assumption that one wished to calculate RAPD allele frequencies, whereas the present study used RAPD phenotype frequencies instead.

### 3.4.6 The potential of other molecular marker systems to produce taxon-specific bands for *S. chrysanthemifolius* and *S. aethnensis*

The present study relied upon RAPDs, ISSRs, isozymes and cpDNA RFLPs. Other molecular marker systems that had the potential to produce taxon-specific markers for *S. chrysanthemifolius* and *S. aethnensis* included AFLPs, microsatellites, PCR-RFLP and rDNA (see below). The AFLP technique can produce many markers, very rapidly. Nevertheless, AFLPs were rejected because their similarity to RAPDs (Harris, 1999) meant they could provide little in the way of additional information. Moreover AFLPs are expensive and require development work. Microsatellites require considerable development work and may not be ideal markers for hybridisation studies (Jarne and Lagoda, 1996). PCR-RFLP of cpDNA could be employed to analyse cpDNA variation between *S. chrysanthemifolius* and *S. aethnensis* in greater detail.

#### *Ribosomal DNA*

A RFLP survey of ribosomal DNA in *S. chrysanthemifolius* and *S. aethnensis* (Abbott, personal communication) produced interesting preliminary results. Three rDNA length variants (a, b and c) were detected in the Sicilian *Senecio* material. Most *S. aethnensis* individuals sampled (five of eight) possessed the a and b length variants, whilst the remainder possessed just the b length variant. Of 14 *S. chrysanthemifolius* individuals, 12 possessed the a, b and c length variants and two possessed the b and c variants only. Thus, the c length variant had potential to be a taxon-specific marker for *S. chrysanthemifolius*. Two restriction site polymorphisms were also detected in Sicilian material - an *EcoRV* cutting site and a *BstEII* cutting site. Five of the seven *S. aethnensis* individuals sampled had an *EcoRV* cutting site and a *BstEII* cutting site, the remaining two individuals had just the *BstEII* site. Two of the 14 *S. chrysanthemifolius* individuals sampled had the *EcoRV* cutting site, the remaining 12 had neither of the cutting sites. Thus, the *BstEII* cutting site had the potential to be a taxon-specific marker for *S. aethnensis*. Future work might investigate the potential of rDNA in greater detail, using larger sample sizes.

### 3.4.7 Summary

Molecular markers that can clearly and reliably differentiate between individuals of *S. aethnensis* and *S. chrysanthemifolius* were developed using RAPD/ISSR primers, isozymes and cpDNA RFLPs. Only a fraction of these markers were taxon-specific in the strict sense but many showed extreme frequency differences between the two taxa. It was concluded that both the taxon-specific and frequency-difference markers should be very informative for studying hybridisation between *S. chrysanthemifolius* and *S. aethnensis*, on Mt. Etna, and their putative hybrid species - *S. squalidus*. As both taxon-specific and frequency-difference markers were treated in the same way (both qualitatively and quantitatively) in future chapters it seems sensible to refer to both as taxon-specific markers.

# **Chapter 4**

**Analysis of the Hybrid Zone Between  
*Senecio chrysanthemifolius* and *Senecio aethnensis*  
on Mt. Etna, Sicily, Using Molecular Markers**

## **Analysis of the Hybrid Zone Between *Senecio chrysanthemifolius* and *Senecio aethnensis* on Mt. Etna, Sicily, Using Molecular Markers.**

### **4.1 Introduction**

#### **4.1.1 Early work on the postulated hybrid zone between *S. chrysanthemifolius* and *S. aethnensis* on Mt. Etna, Sicily**

In the late 1960's, Crisp (1972) visited Sicily to investigate the origins of British *S. squalidus*. In the region around Mt. Etna, he found that *Senecio* plants similar to British *S. squalidus* could be grouped into three categories: a montane taxon (above ~1000m), a lowland taxon (below ~1000m) and a taxon of intermediate altitude (~ 1000m ±200m). Italian botanists (Pojero, 1902 and Poli, 1959; in Crisp, 1972) recognised these taxa as separate species: with *S. aethnensis* Jan. as the montane species (endemic to Mt. Etna), *S. chrysanthemifolius* Poir. as the lowland species (endemic to Sicily, surrounding islands and Calabria in Southern Italy) and *S. incisus* Presl. (= *S. glaber* Ucria; endemic to Mt. Etna) as the intermediate species.

From an examination of these *Senecio* taxa on Mt. Etna, Crisp noted that many morphological characters differentiated *S. chrysanthemifolius* from *S. aethnensis*, most conspicuously leaf dissection, glaucosity and ligule size (plate 4.1). *S. aethnensis* possessed entire, glaucous leaves and large ligules, while *S. chrysanthemifolius* had highly dissected, non-glaucous leaves and small ligules. Crisp stated that individuals of *S. incisus* "formed a series of intermediates with respect of glaucosity, leaf dissection and ligule size" and "always marked the upper limits of local *S. chrysanthemifolius*, and the lower limits of local *S. aethnensis*". These field observations suggested that "*S. incisus* was, in fact, a series of hybrid swarms between *S. aethnensis* and *S. chrysanthemifolius*, and that probably a degree of gene-flow takes place between the two species".

From seeds collected in the Mt. Etna area, progeny of *S. chrysanthemifolius*, *S. aethnensis* and *S. incisus* were raised for comparison under glass (Crisp, 1972). Several morphological characters that distinguished *S. chrysanthemifolius* and *S. aethnensis* in the field, including leaf dissection, glaucosity and ligule size (table 4.1), appeared to be under genetic control

**Plate 4.1** (a) *Senecio chrysanthemifolius*, (b) *S. aethnensis* and (c) artificial F<sub>1</sub> hybrids between *S. chrysanthemifolius* and *S. aethnensis* (all greenhouse grown).



in that they were inherited from one generation to the next. These characters could, therefore, be used as morphological markers.

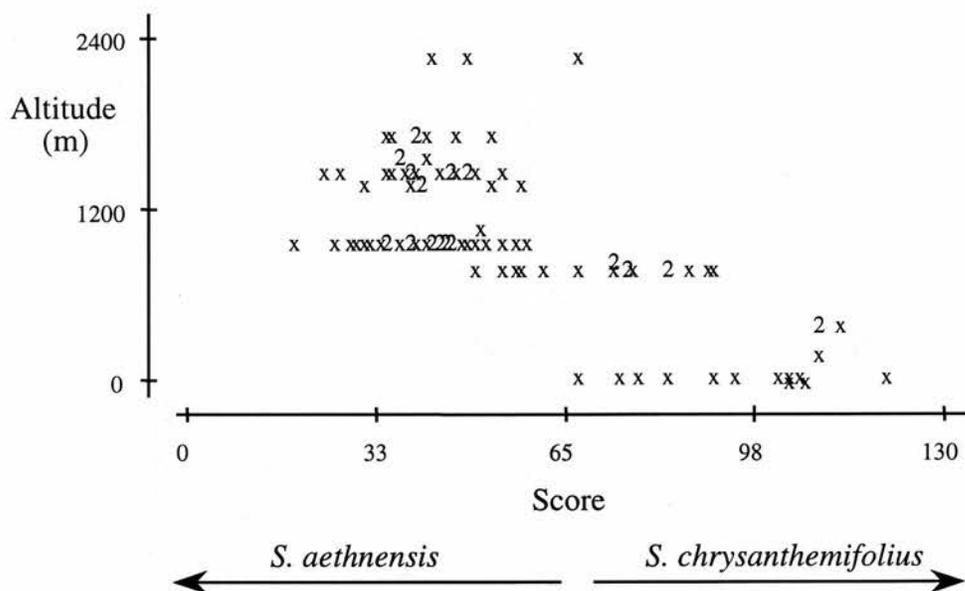
Morphological marker scores of progeny from the *S. chrysanthemifolius*/*S. aethnensis*/*S. incisus* complex were related to the altitude of parental material on Mt. Etna and, for each individual, these scores were combined to form a composite score (Crisp, 1972). The distribution of scores with altitude showed the degree to which intergradation of character differences between *S. chrysanthemifolius* and *S. aethnensis* had taken place (fig. 4.1).

Crisp also crossed *S. chrysanthemifolius* with *S. aethnensis* to produce artificial F<sub>1</sub> hybrids, confirming the inter-fertility of the two species. Artificial F<sub>1</sub> hybrids were morphologically intermediate between the parental species and resembled the putative natural hybrids (plate 4.1). An investigation of the potential fertility of both the artificial F<sub>1</sub>s and putative natural hybrids showed that neither type had substantially reduced fertility relative to *S. chrysanthemifolius* and *S. aethnensis*.

Crisp concluded that there was indeed a hybrid zone between *S. chrysanthemifolius* and *S. aethnensis* on Mt. Etna, that there was reciprocal gene-flow between *S. chrysanthemifolius* and *S. aethnensis* and that *S. incisus* populations should be thought of as hybrid populations. He suggested that "the main control of introgression between *S. aethnensis* and *S. chrysanthemifolius* in the field is almost certainly altitudinal separation of environmentally adapted genotypes coupled with selection of hybrid progeny phenotypes".

**Fig. 4.1** Composite morphological marker scores of progeny from the *Senecio chrysanthemifolius*/*S. aethnensis*/*S. incisus* complex in Sicily, related to altitude on Mt. Etna of the parental material (from Crisp, 1972).

The most 'pure' *S. chrysanthemifolius* individuals would score 130, the most 'pure' *S. aethnensis* individuals would score 0.



**Table 4.1** Morphological characters that differentiate *Senecio chrysanthemifolius* from *S. aethnensis* and which appear to be under genetic control (from Crisp, 1972).

Character	Typical <i>S. chrysanthemifolius</i>	Typical <i>S. aethnensis</i>
leaf dissection	highly dissected	entire
leaf shape in outline	rhomboid	long-spathulate to oblong
hairiness of leaves and stems	slightly hairy	glabrous
glaucosity of leaves	non-glaucous	glaucous
relative leaf auricle size	absent	very large
mean achene length	small (e.g. < 2.6 mm)	large (e.g. > 3.9 mm)
hairiness of achenes	dense, in grooves	glabrous
achene colour	dark brown beneath hair	light brown
inner involucre bract tips	dense black tips	colourless
inner involucre bract length	small (e.g. 5.0 mm)	large (e.g. 10.0 mm)
receptacle breadth	small (e.g. 3.5 mm)	large (e.g. 8.5 mm)
ligule length	small (e.g. < 5.5 mm)	large (e.g. < 15.5 mm)
ligule breadth	small (e.g. < 2.2 mm)	large (e.g. < 5.2 mm)

#### 4.1.2 The use of molecular markers

Morphological methods were the mainstay of hybridisation studies until the advent of isozyme markers in the 1970's and DNA markers in the 1980's. It is believed that molecular markers have many advantages over morphological markers for the study of hybridisation. For example, compared to morphological markers, molecular markers are more likely to be numerous, under simple genetic control, independent, heritable and selectively neutral and less likely to be prone to convergence (Sytsma, 1990; Rieseberg and Wendel, 1993). As a result, many examples of plant hybridisation based on morphological data were re-examined using molecular markers. Sometimes molecular markers confirmed these hybridisation events and their consequences, sometimes not. For instance, based upon morphological studies, Heiser (1951) proposed that *Helianthus annuus* ssp. *texanus* and the weedy race of *H. bolanderi* were hybrid taxa. Rieseberg used molecular markers (cpDNA RFLPs, rDNA RFLPs and isozymes) to confirm the hybrid status of the former (Rieseberg *et al.*, 1990a) and disprove the hybrid status of the latter (Rieseberg and Soltis, 1988).

Crisp's conclusions about hybridisation between *S. chrysanthemifolius* and *S. aethnensis*, like those of Heiser, were based on the tools available at the time. Therefore, as part of an investigation into the origins of British *S. squalidus*, Abbott (Abbott *et al.*, 1995; Abbott

and Milne, 1995; Abbott *et al.*, 2000) re-examined hybridisation between the *Senecio* species on Mt. Etna, using molecular markers. Isozyme and cpDNA RFLP analyses were carried out on the progeny of plants collected (as seed) from various *Senecio* populations on Mt. Etna in 1988. Populations of *S. chrysanthemifolius* and *S. aethnensis* shared the same alleles at most polymorphic isozyme loci (see section 3.2.2), but one isozyme locus, *Acp-2*, was most informative (table 4.2). Populations of *S. chrysanthemifolius* were fixed for the *Acp-2 b* allele, whereas *S. aethnensis* and the putative hybrid populations were polymorphic for *Acp-2 a* and *b*. The *Acp-2 a* allele increased in frequency with altitude along two transects, one on the southern slopes of Mt. Etna and one on the northern slopes, suggesting that high altitude populations of *S. aethnensis*, which were not sampled, might be monomorphic for *Acp-2 a*. A similar pattern was revealed by cpDNA RFLP analysis (table 4.3). Chloroplast DNA haplotype B was present in all samples of *S. chrysanthemifolius*, whereas *S. aethnensis* and the putative hybrids possessed either haplotype B or A. Abbott *et al.* (1995) speculated that haplotype A might be diagnostic for *S. aethnensis*, with high altitude populations of *S. aethnensis* being monomorphic for this haplotype.

From the distribution of *Acp-2* alleles and chloroplast DNA haplotypes in the Mt. Etna material, it was suggested that gene-flow was polarised from *S. chrysanthemifolius* into *S. aethnensis* (Abbott *et al.*, 1995). However, the isozyme and cpDNA studies were limited. First, by the fact that high altitude *S. aethnensis* populations were not sampled. Second, by the small number of individuals surveyed for cpDNA, and third, by the few molecular markers examined. Thus, it was felt that the postulated hybrid zone should be examined once more, using a molecular marker system that would generate a large number of taxon-specific markers (compared to isozymes and cpDNA RFLP) and with the inclusion of high altitude *S. aethnensis* populations.

**Table 4.2** Distribution of *Acp-2 a* and *b* among *Senecio chrysanthemifolius*, *S. aethnensis* and putative hybrid (*S. chrysanthemifolius* x *S. aethnensis*) populations along two transects on Mt. Etna, Sicily (Abbott *et al.*, 2000).

Population	Altitude (m)	N <sup>1</sup>	<i>Acp-2</i> allele frequency	
			<i>a</i>	<i>b</i>
south transect				
<i>S. chrysanthemifolius</i>	760	46	0.000	1.000
<i>S. chrysanthemifolius</i>	800	24	0.000	1.000
putative hybrid	1195	32	0.016	0.984
putative hybrid	1500	28	0.232	0.768
<i>S. aethnensis</i>	1650	24	0.542	0.458
<i>S. aethnensis</i>	1890	13	0.615	0.385
north transect				
<i>S. chrysanthemifolius</i>	750	25	0.000	1.000
putative hybrid	1400	16	0.531	0.469
<i>S. aethnensis</i>	1750	25	0.760	0.240

<sup>1</sup>N = number of accessions.**Table 4.3** Chloroplast haplotypes of *Senecio chrysanthemifolius*, *S. aethnensis* and putative hybrid (*S. chrysanthemifolius* x *S. aethnensis*) individuals on Mt. Etna, Sicily (Abbott *et al.*, 1995).

Taxon	cpDNA haplotype (no. of individuals)	
	A	B
<i>S. chrysanthemifolius</i>	0	11
hybrid	2	5
<i>S. aethnensis</i>	1	4

#### 4.1.3 Aim:

The aim of the present study was to examine the postulated hybrid zone on Mt. Etna, using molecular markers, and to ascertain whether there was good molecular evidence for its existence. The shortcomings of previous studies would be overcome by employing large numbers of taxon-specific *S. chrysanthemifolius* and *S. aethnensis* molecular markers and ensuring that 'pure' populations of *S. aethnensis*, from high altitude, were sampled.

## 4.2 Materials and methods

### 4.2.1 Choice of transect and populations

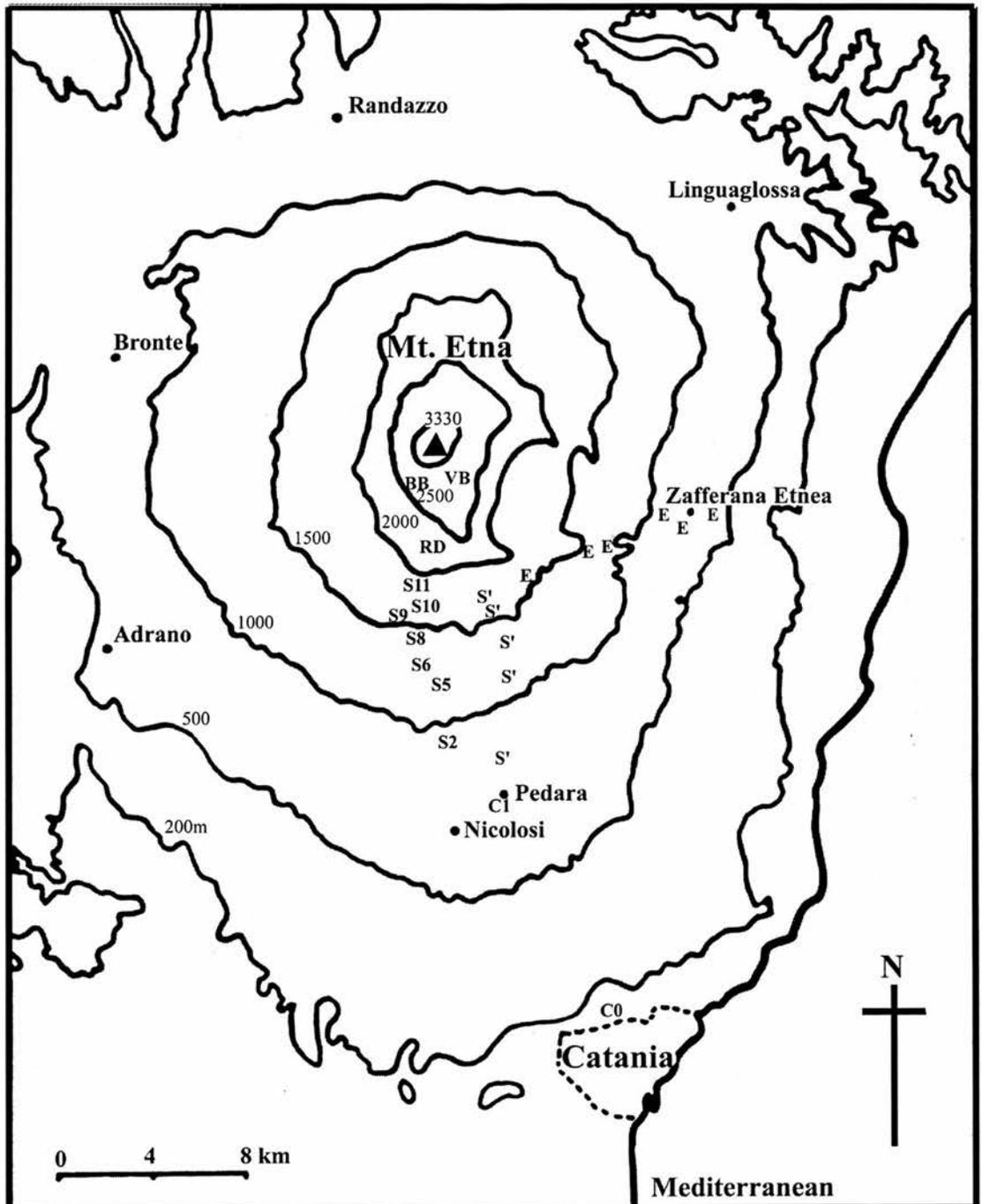
**Table 4.4** Populations along transect S, through the postulated hybrid zone between *Senecio chrysanthemifolius* and *S. aethnensis* on Mt. Etna, used in the survey of molecular marker variation.

Popn.	Altitude (m)	Distance (km) <sup>1</sup>	Field identification	Habitat
C0	150 m	0	'pure' <i>S. chrysanthemifolius</i>	urban waste-ground
C1	600 m	11.00	'pure' <i>S. chrysanthemifolius</i>	urban waste-ground
S2	950 m	17.50	<i>S. chrysanthemifolius</i>	waste-ground by roadside
S5	1300 m	21.00	<i>S. chrysanthemifolius</i> -like hybrids	car-park verge
S6	1400 m	21.75	<i>S. chrysanthemifolius</i> -like hybrids	lava flow by road
S8	1500 m	22.60	<i>S. chrysanthemifolius</i> -like hybrids	roadside verge
S9	1600 m	23.15	mixed hybrids	roadside verge
S10	1650 m	23.80	mixed hybrids	beneath pine trees
S11	1800 m	24.68	<i>S. aethnensis</i> -like hybrids	roadside verge
RD	2175 m	26.18	<i>S. aethnensis</i> -like hybrids	lava substrate in skiing area
BB	2525 m	27.93	'pure' <i>S. aethnensis</i>	fine lava substrate
VB	2600 m	29.28	'pure' <i>S. aethnensis</i>	coarse lava substrate

<sup>1</sup> Cumulative distance from population C0.

The transect S, running up the southern incline of Mt. Etna (fig. 4.2), was chosen because it contained a more continuous set of populations, and more individuals per population, than other potential transects such as S' or E (fig. 4.2). The transect ran from the outskirts of Catania (~150 m) to the upper limit of vegetation on the southern summit of Mt. Etna (approximately 2600 m). Populations in transect S were discrete and a majority comprised less than 50 individuals. Populations along the transect that were included in the survey of molecular marker variation are detailed in table 4.4. Populations were chosen to ensure that the region of highest hybridity, estimated from morphological observation in the field, was thoroughly sampled and to provide a reasonable distribution of populations, in terms of both altitude and horizontal distance between populations. A classification, based upon morphology, was made of each population in the field. The two populations at lowest altitude, C0 and C1, were considered to represent 'pure' *S. chrysanthemifolius*. S2 was identified as a *S. chrysanthemifolius* population whereas S5, S6 and S8 were identified as '*S. chrysanthemifolius*-like' hybrid populations. Populations S9 (less than ten individuals) and S10 contained a mixture of '*S. chrysanthemifolius*-like' and '*S. aethnensis*-like' hybrid

**Fig 4.2** Locations of populations along transect S, through the postulated hybrid zone between *Senecio chrysanthemifolius* and *S. aethnensis*, on Mt. Etna, Sicily (see also tables 2.1 and 4.4). The locations of populations in transects S' and E are also shown (see text for details).



individuals. S11 and RD were recorded as '*S. aethnensis*-like' hybrid populations and populations BB and VB were thought to represent 'pure' *S. aethnensis*.

#### **4.2.2 Survey of RAPD and ISSR variation along transect S, through the postulated hybrid zone between *S. chrysanthemifolius* and *S. aethnensis*, on Mt. Etna**

RAPD and ISSR analyses were carried out using procedures outlined in section 2.8. The development of RAPD and ISSR markers that distinguished *S. chrysanthemifolius* from *S. aethnensis* is discussed in chapter three. During this process, four populations from transect S were surveyed: two populations believed to represent 'pure' *S. chrysanthemifolius* (C0 (16 individuals) and C1 (14 individuals) and two populations believed to represent 'pure' *S. aethnensis* (BB (17 individuals) and VB (17 individuals). Five more populations along transect S were then examined - S2 (14 individuals), S5 (14 individuals), S6 (15 individuals), S8 (15 individuals) and S11 (14 individuals). These nine populations (C0, C1, S2, S5, S6, S8, S11, BB and VB) were surveyed with all 55 RAPD and ISSR markers (using 35 primers; table 3.11).

A preliminary examination of the data suggested that three other populations along the transect - S9 (only six individuals available), S10 (only eight individuals available) and RD (14 individuals) - should also be included in the analysis, along with a number of F<sub>1</sub> plants produced from crosses between *S. chrysanthemifolius* and *S. aethnensis* (i.e. ten F<sub>1</sub>s produced by crossing individuals of C1 and C0 with individuals of VB). Due to resource constraints, populations S9, S10, S11 and the F<sub>1</sub>s were not surveyed for all 55 RAPD and ISSR markers. Instead, a subset of 26 markers (table 4.5) was selected, allowing 13 *S. chrysanthemifolius* markers and 13 *S. aethnensis* markers (from a total of 19 RAPD and ISSR primers) to be surveyed. The criteria for choosing these 26 markers was the ease with which they could be scored. Repeats of individual DNA samples were carried out, as necessary, to ensure that the data set of 26 markers surveyed across all 13 populations (12 transect populations plus the F<sub>1</sub>s) was complete.

**Table 4.5** Markers used to survey all 12 populations along transect S and F<sub>1</sub> individuals (see also table 3.11).

Primer	Band	Primer	Band	Primer	Band	Primer	Band
A9	CHR-1	C2	CHR-1	G5	CHR-1	G11	AET-1
A11	CHR-1	C2	AET-1	G5	AET-1	G12	CHR-1
A11	AET-1	C5	CHR-1	G6	AET-1	G12	AET-1
A20	CHR-1	C20	AET-1	G8	CHR-1	H7	CHR-1
B15	CHR-1	F12	CHR-1	G8	AET-1	H13	AET-1
B15	AET-1	F12	AET-1	G10	CHR-1	H19	AET-1
18	AET-1	40	CHR-1				

Certain precautions were taken to ensure that RAPD and ISSR profiles were reliable and reproducible. Six individuals of pure *S. chrysanthemifolius* and six of pure *S. aethnensis* were run as standards on each gel. These standards were previously scored individuals from populations C0, C1, C9, VB, BB, TC, UJ and VO. The inclusion of standards assisted in the accurate scoring of RAPD and ISSR fragments across gels and acted as a check for reproducibility (by comparing the banding pattern obtained for each standard in a gel with that obtained for the same individual in earlier gel runs). As an additional check for reproducibility, amplifications using six RAPD primers and two ISSR primers (for 72 individuals) were repeated. RAPD and ISSR profiles obtained in repeat runs were compared with original profiles. Homology tests, using the restriction digest method (see section 2.9), were also carried out for the sub-sample of 26 markers.

#### 4.2.3 Survey of isozyme variation along transect S

A survey of isozyme variation across transect S was carried out using the procedures outlined in section 2.10. Isozyme variation was surveyed in all 12 populations along the transect on Mt. Etna described in table 4.4. Twenty individuals per population were examined, except for populations S9 and S10 where only 12 and eight individuals, respectively, were available. Variation was assayed for six enzymes and six loci were scored: *Acp-2*, *Aat-3*, *Aco-1*, *Idh-1*, *Pgi-2* and *Pgm-1*.

#### 4.2.4 Survey of chloroplast DNA RFLP variation along transect S

A survey of chloroplast DNA RFLP variation was carried out along transect S using the procedures outlined in section 2.7. Four populations in the transect had already been surveyed for cpDNA haplotype (chapter three). These included two populations, C0 (15 individuals) and C1 (12 individuals), believed to represent 'pure' *S. chrysanthemifolius*, and

two populations, BB (17 individuals) and VB (17 individuals), believed to represent 'pure' *S. aethnensis*. The additional populations examined were - S2 (15 individuals), S5 (15 individuals), S6 (15 individuals), S8 (15 individuals) S9 (six individuals), S10 (four individuals), S11 (15 individuals) and RD (14 individuals).

#### **4.2.5 Data analysis**

##### ***4.2.5.1 Survey of RAPD and ISSR variation***

###### *Single locus variation*

The frequency of each RAPD and ISSR marker band in each population was plotted against (i) population altitude (estimated from a map) and (ii) the cumulative distance of each population from the start of the transect (i.e. from the lowest altitude site, C0). The direct distance between sequential populations was estimated from a map; and, as the transect only approximated to a straight line, this method led to an overestimation of total distance.

###### *Multilocus variation*

A binary data matrix was constructed for the 26 RAPD and ISSR markers scored in all populations (12 transect populations plus the F<sub>1</sub>s). Markers were recorded as present (1) or absent (0) in each individual and versions of this data matrix were used in the following analyses:

###### *(i) Hybrid index analysis*

A hybrid index (Grant, 1981; Howard *et al.*, 1997) was constructed for each individual in the transect and for each F<sub>1</sub> individual. A *S. chrysanthemifolius* marker was assigned a score of +1 and a *S. aethnensis* marker a score of -1. The hybrid index of an individual represented the sum of its scores over all 26 markers. As there were equal numbers of markers for each parent taxon, namely 13 for *S. chrysanthemifolius* and 13 for *S. aethnensis*, the hybrid index of an individual could range from +13 to -13. Frequency distributions of hybrid indices were constructed for each population and for the group of F<sub>1</sub> individuals. The mean and range of hybrid indices within each population, and the mean

and range of the number of *S. chrysanthemifolius* and *S. aethnensis* markers within each population were calculated.

*(ii) Principal coordinate analysis (PCO)*

Principal coordinate analysis was carried out on two matrices of binary data; one matrix included all individuals sampled from the transect, while the other included all of these individuals plus the ten artificially produced F<sub>1</sub> individuals. Three coefficients were used in the principal coordinate analysis: Nei and Li's similarity coefficient, Jaccard's similarity coefficient and a Euclidean distance coefficient (see sections 2.11.2 and 2.11.3).

*(iii) Cluster analysis*

Distance matrices derived from Nei and Li, and Jaccard's similarity coefficients were used in UPGMA and neighbour-joining cluster analyses (see section 2.11.5). Cluster analyses were carried out on two binary data matrices as in (ii) above.

*Population level analyses*

*(i) Analysis of molecular variance (AMOVA)*

An AMOVA (see section 2.11.4) was carried out to estimate variance components within and between the 12 populations along the transect. The AMOVA also calculated the fractional proportion of the total variance partitioned between each population pair, and these pair-wise  $\Phi_{ST}$  values were used as a measure of inter-population genetic distance. Phenograms were constructed from the pair-wise  $\Phi_{ST}$  values using the UPGMA and neighbour-joining clustering methods. A second AMOVA was performed on the same data set as the first but with the addition of the "population" of F<sub>1</sub>s.

*(ii) Shannon's diversity index*

Shannon's diversity index (see section 2.11.6) was used to quantify the level of genetic diversity present in the 12 populations along the transect and to partition total genetic diversity (in the sample as a whole) into its within and between population components.

#### **4.2.5.2 Survey of isozyme variation**

Allele frequencies at the six polymorphic enzyme loci were calculated for each population along the transect. Frequencies of *Acp-2 a* and *Pgi-2 b* were plotted against (i) population altitude and, (ii) the cumulative distance of each population from the start of the transect.

Allozyme variation was estimated for each population in the transect and Hardy-Weinberg equilibrium was tested across loci in all populations (2.11.7). *F*-statistics were also estimated (2.11.8). Genetic distances between populations were calculated according to Nei (1978) (2.11.9) and the distance matrix was subjected to UPGMA and neighbour-joining cluster analyses (2.11.5).

#### **4.2.5.3 Survey of chloroplast DNA RFLP variation**

The frequency of cpDNA haplotype A in each population along the transect was plotted against (i) population altitude and, (ii) the cumulative distance of each population from the start of the transect.

## 4.3 Results

### 4.3.1 Description of the postulated hybrid zone between *S. chrysanthemifolius* and *S. aethnensis* on Mt. Etna, Sicily

#### Personal observations

*Senecio chrysanthemifolius* was found growing on Mt. Etna up to altitudes of approximately 1000 m, while *S. aethnensis* grew on the uppermost slopes, typically at altitudes above 2400 m. The highest population of *S. aethnensis* was found at 2810 m on the northern summit of Mt. Etna. In other populations, individuals that combined morphological features of *S. chrysanthemifolius* and *S. aethnensis* were recorded as putative hybrids, those which bore greater resemblance to *S. chrysanthemifolius* were referred to as '*S. chrysanthemifolius*-like' hybrids, and those which bore greater resemblance to *S. aethnensis* were referred to as '*S. aethnensis*-like' hybrids. Putative hybrids were observed at altitudes ranging from 1150 m to 2300 m and '*S. chrysanthemifolius*-like' hybrids were typically found at lower elevations (up to ~1600 m) than '*S. aethnensis*-like' hybrids (above ~1650 m).

Some samples which I considered to be '*S. aethnensis*-like' hybrids would have been treated by Crisp and Abbott as 'pure' *S. aethnensis* (Crisp, 1972; Abbott *et al.*, 1995; Abbott *et al.*, 2000), and I would have reclassified many of what they considered to be *S. aethnensis* populations as '*S. aethnensis*-like' hybrids. This field-based separation of *S. aethnensis*-like hybrids from pure *S. aethnensis* material was based on morphological observations of several populations of *S. aethnensis* at very high altitude (2550 m and above). *Senecio aethnensis* at high altitude is most likely to be pure, since it is furthest from *S. chrysanthemifolius* and potential gene-flow from that species should be much reduced. I believe that my classification of material on Mt. Etna was accurate and did not suffer from the deficiencies of those of Abbott and Crisp, who did not have the benefit of observing the majority of high altitude *S. aethnensis* populations upon which my conclusions were based. Moreover, it was acknowledged that populations of '*S. aethnensis*' used previously in isozyme and cpDNA surveys had possibly been subject to extensive gene-flow from *S. chrysanthemifolius* (Abbott *et al.*, 1995; Abbott *et al.*, 2000).

**Plate 4.2** (a) Site of S2 population (950 m) on transect S, Mt. Etna (arrow highlights *Senecio* individuals), (b) Site of S8 population (1500 m) on transect S, Mt. Etna.

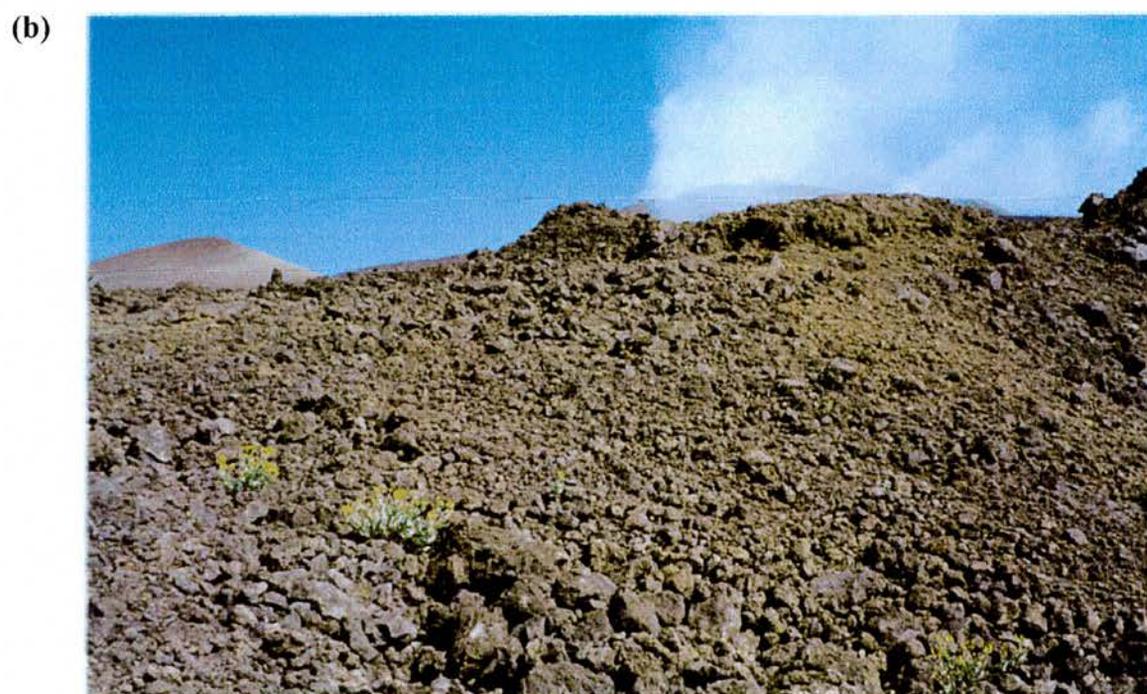
(a)



(b)



**Plate 4.3** (a) Hybrid individuals growing by roadside at ~1400 m on transect S, Mt. Etna, (b) Site of VB population of 'pure' *Senecio aethnensis* (2600 m) on transect S, Mt. Etna.



The distribution of putative hybrid *Senecio* material in the Mt. Etna area was not uniform. Populations in three transects on the southern slopes of Mt. Etna (transects S, S' and E, fig. 4.2) displayed a reasonably continuous transition from *S. chrysanthemifolius*, through 'S. *chrysanthemifolius*-like' hybrids and 'S. *aethnensis*-like' hybrids, to pure *S. aethnensis*. However, on the northern slopes of Mt. Etna, *S. chrysanthemifolius* populations were recorded up to 1100m, followed by a gap in which no *Senecio* material was observed until 'S. *aethnensis*-like' hybrids and *S. aethnensis* were recorded at 1660 m and above. In 1988 Abbott (Abbott *et al.*, 2000) recorded a hybrid swarm at 1400 m on this northern transect, but this was not found in 1996.

*S. chrysanthemifolius* occupied mainly man-made and, potentially, ruderal habitats such as urban waste-ground and roadside verges. Putative hybrid sites tended to resemble those of *S. chrysanthemifolius*, such as road-side verges (plates 4.2 and 4.3). The habitat of pure *S. aethnensis* was rather different, being entirely natural and generally consisting of coarse and fine lava flows (plate 4.3). Indeed, *S. aethnensis* is one of the few species capable of surviving in the high volcanic desert on Mt. Etna.

Field observation suggested that *Senecio* species on Mt. Etna differ in flowering time: with *S. chrysanthemifolius* populations at low elevation flowering earliest and populations located at successively higher altitudes flowering progressively later in the year. High altitude *S. aethnensis* did not flower until late July or early August, approximately two months after the peak flowering time of populations of *S. chrysanthemifolius* at low altitude.

It must be noted, that this description of the postulated hybrid zone is based upon information that was both temporally and spatially limited. Only two weeks (one in mid-June 1996 and one in early August 1997) were spent in the Mt. Etna area. Ideally, one would have studied the Sicilian *Senecio* populations over at least two full field seasons. Secondly, observations were made in a restricted geographic area. Due to time constraints it was not possible to observe and sample material from areas of Mt. Etna that were not easily accessible by car. This was especially the case for the western and north-western mid to upper slopes of Mt. Etna.

### **4.3.2 RAPD and ISSR variation along transect S, through the postulated hybrid zone between *S. chrysanthemifolius* and *S. aethnensis*, on Mt. Etna**

Examples of RAPD gels produced for populations along transect S, and the  $F_{1S}$ , are presented in plates 4.4 - 4.6.

Six RAPD primers and two ISSR primers were repeated on extracts from 72 individuals. In total, 912 comparisons were made between presence/absence of bands in repeats and originals, across 13 markers. The error rate, i.e. the percentage of discrepancies between repeats and originals, was less than 4.0%. Homology between bands in putative hybrid populations and the equivalent bands in *S. chrysanthemifolius* or *S. aethnensis* was confirmed in 22 of the 26 cases (plate 6.3). Homology tests for the remaining four bands were unsuccessful (for example, not enough band product was amplified to restrict, see section 2.9).

#### **4.3.2.1 Single locus variation**

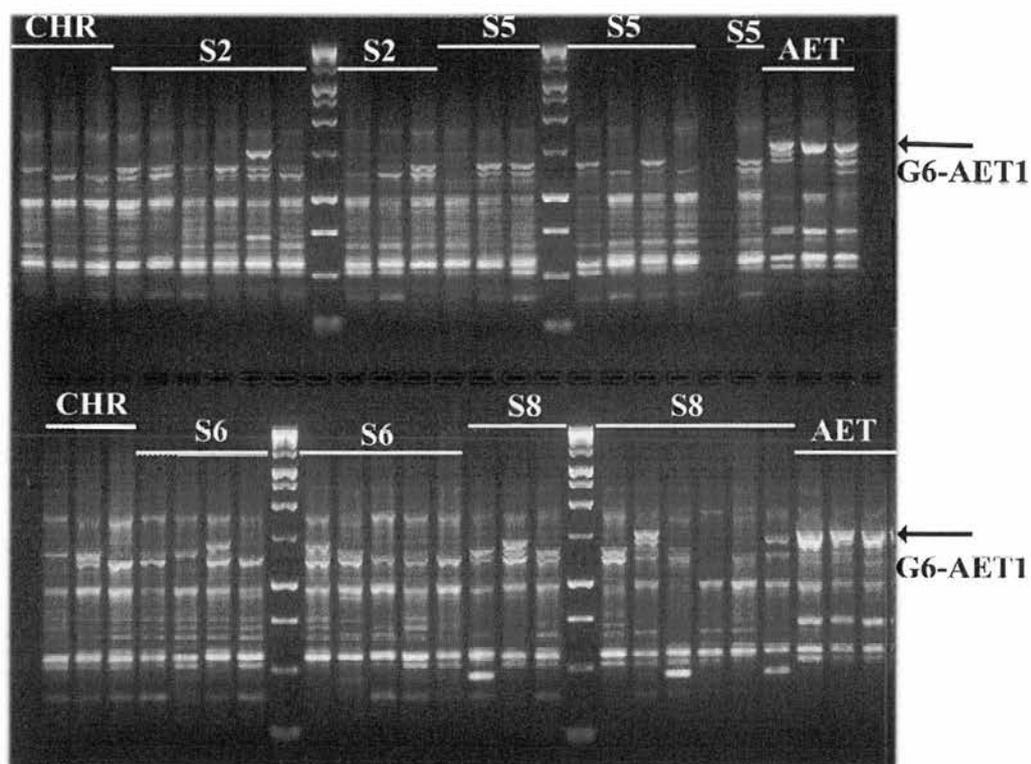
The frequency of each marker in each population was calculated and plotted against: (i) population altitude (altitude plots) and (ii) cumulative distance of each population from the start (C0) of the transect (distance plots). Plots for 15 *S. chrysanthemifolius* markers and 18 *S. aethnensis* markers are shown in figs. 4.3-4.8. Plots of the remaining 22 markers examined, over distance, are in appendix C.

Changes in marker frequency towards the start of the transect (i.e. less than ~ 1500 m altitude or less than ~ 20 km distance from C0) were more abrupt over altitude than over distance (fig. 4.3 versus fig. 4.4, fig. 4.5 versus fig. 4.6, fig. 4.7 versus fig. 4.8). In contrast, changes in marker frequency towards the end of the transect (i.e. more than ~ 1500 m altitude or more than ~ 20 km distance from C0) were more abrupt over distance than for altitude. However, these differences were simply due to the changing gradient along the transect; with a shallow incline at the start of the transect replaced by a steep incline at the end of it. Therefore, the choice of altitude or distance for the  $x$  axis did not affect the interpretation of data and the overall patterns of change in marker frequencies along the transect remained the same.

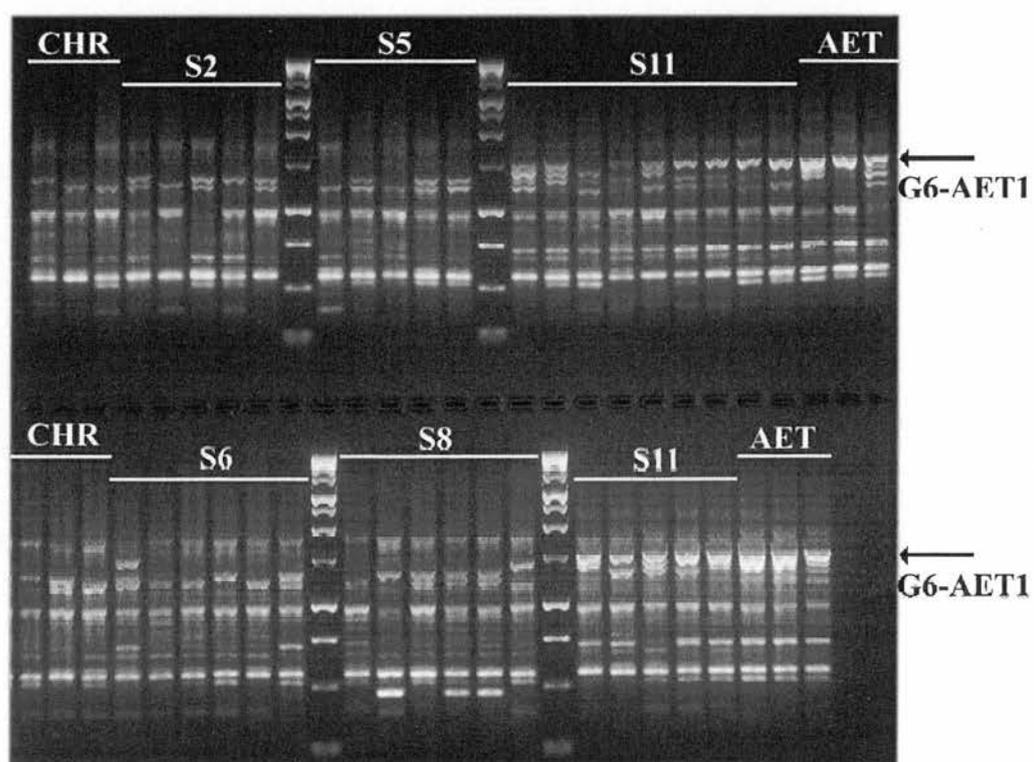
Some graphs were based upon data from nine populations, while others were for 12 populations (see section 4.2.2). However, similar overall patterns were observed

**Plate 4.4 (a) and (b)** Presence/absence of the G6-AET1 RAPD marker band in five populations along transect S, Mt. Etna. AET indicates *Senecio aethnensis* standards, CHR indicates *S. chrysanthemifolius* standards.

(a) G6

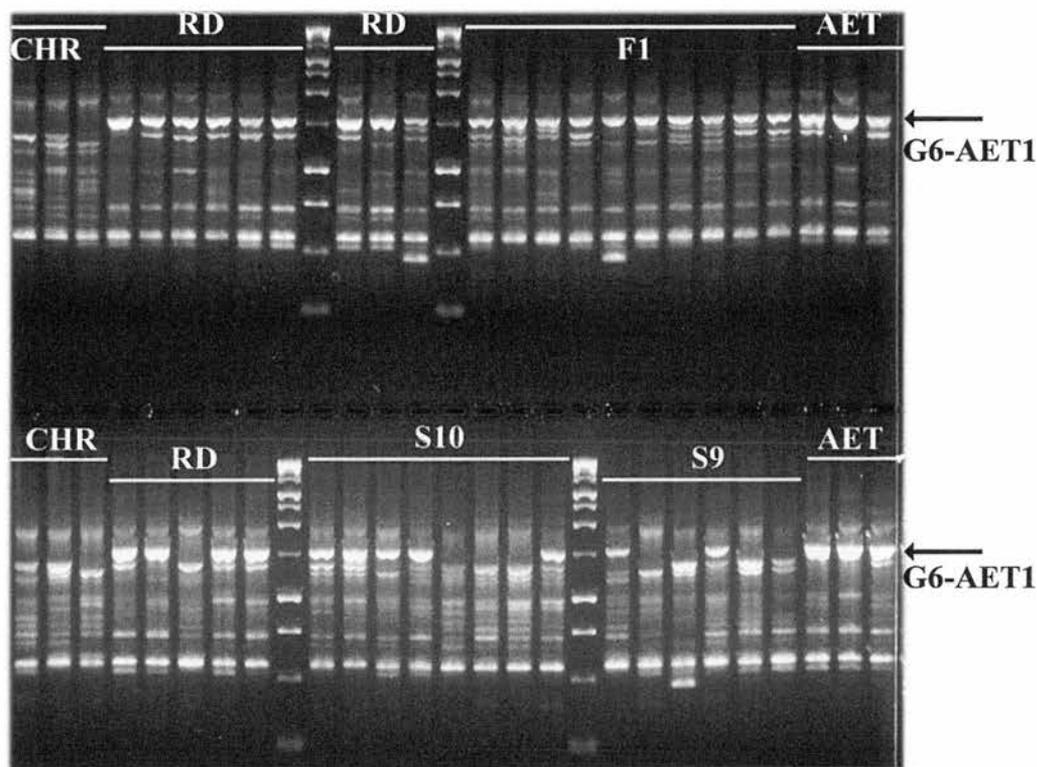


(b) G6

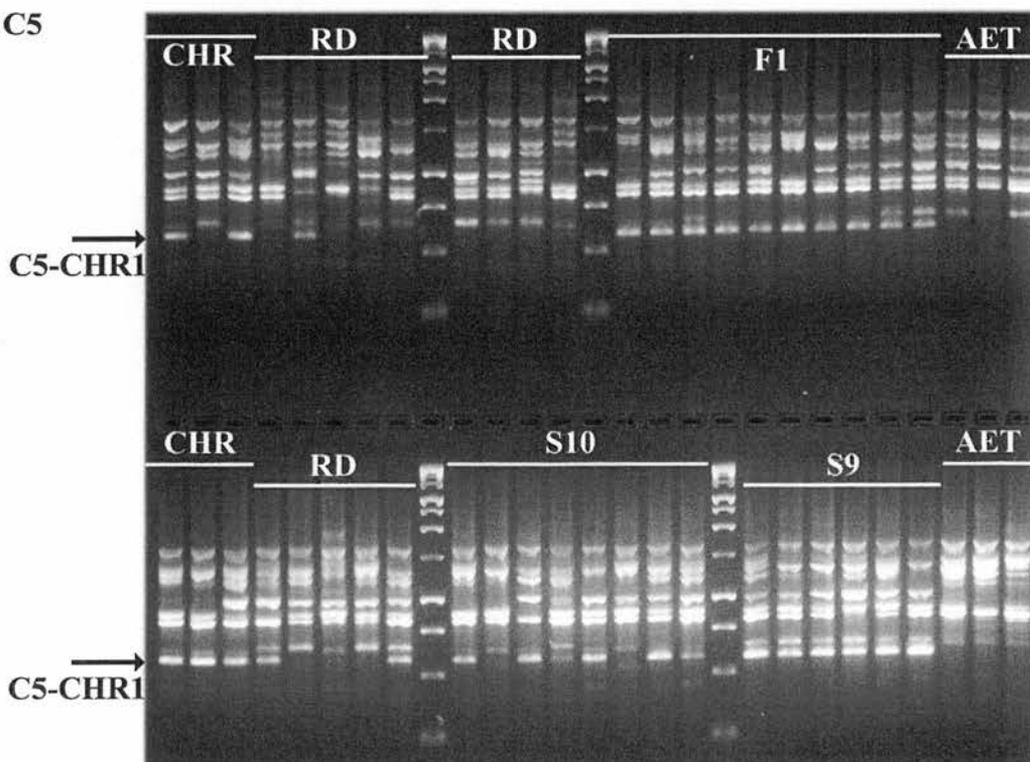


**Plate 4.5 (a)** Presence/absence of the G6-AET1 RAPD marker band in three populations along transect S, Mt. Etna; and in the group of artificial F<sub>1</sub> hybrids between *Senecio chrysanthemifolius* and *S. aethnensis*. **(b)** Presence/absence of the C5-CHR1 RAPD marker band in three populations along transect S, Mt. Etna; and in the group of artificial F<sub>1</sub> hybrids between *S. chrysanthemifolius* and *S. aethnensis*. AET indicates *S. aethnensis* standards, CHR indicates *S. chrysanthemifolius* standards.

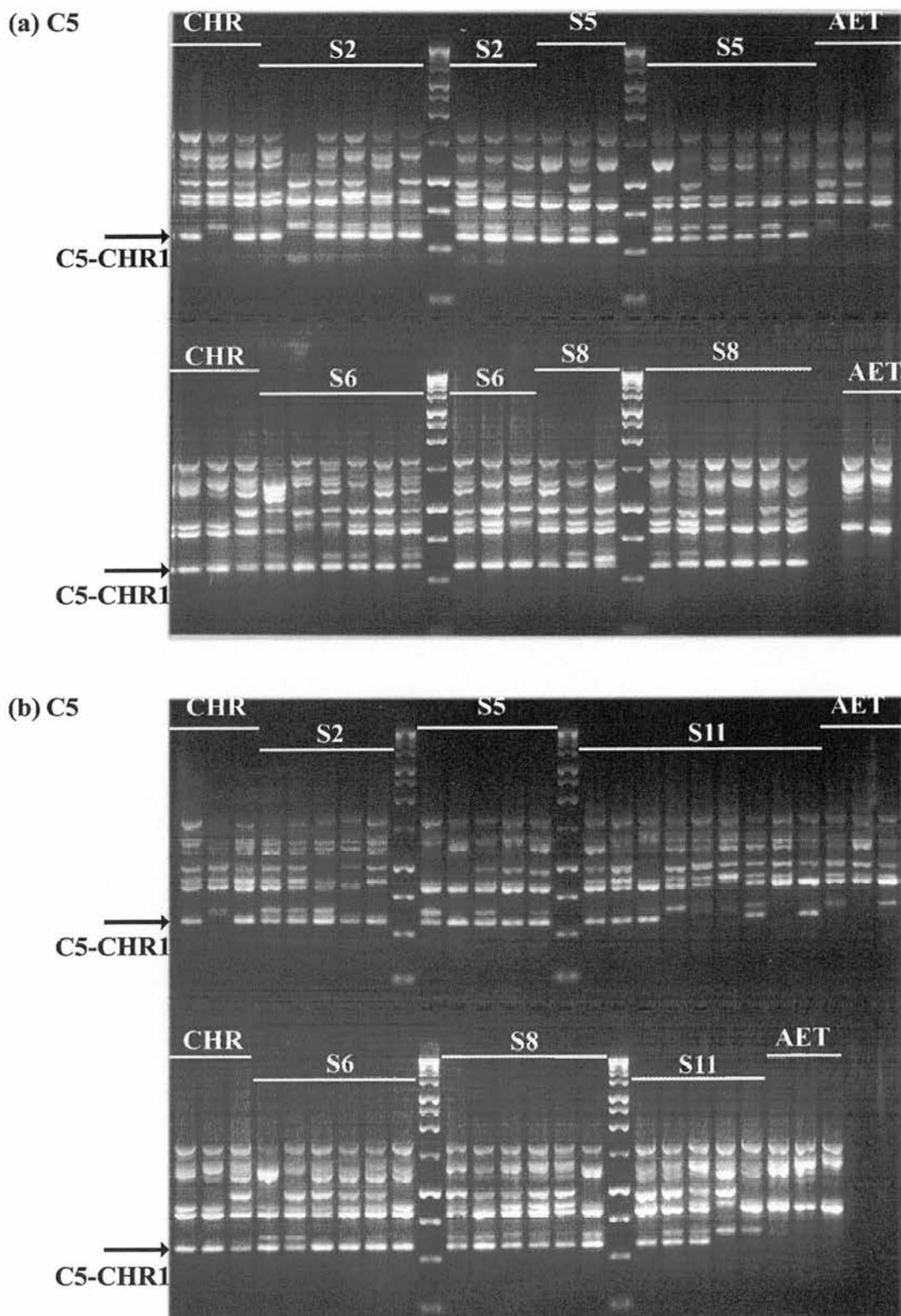
(a) G6

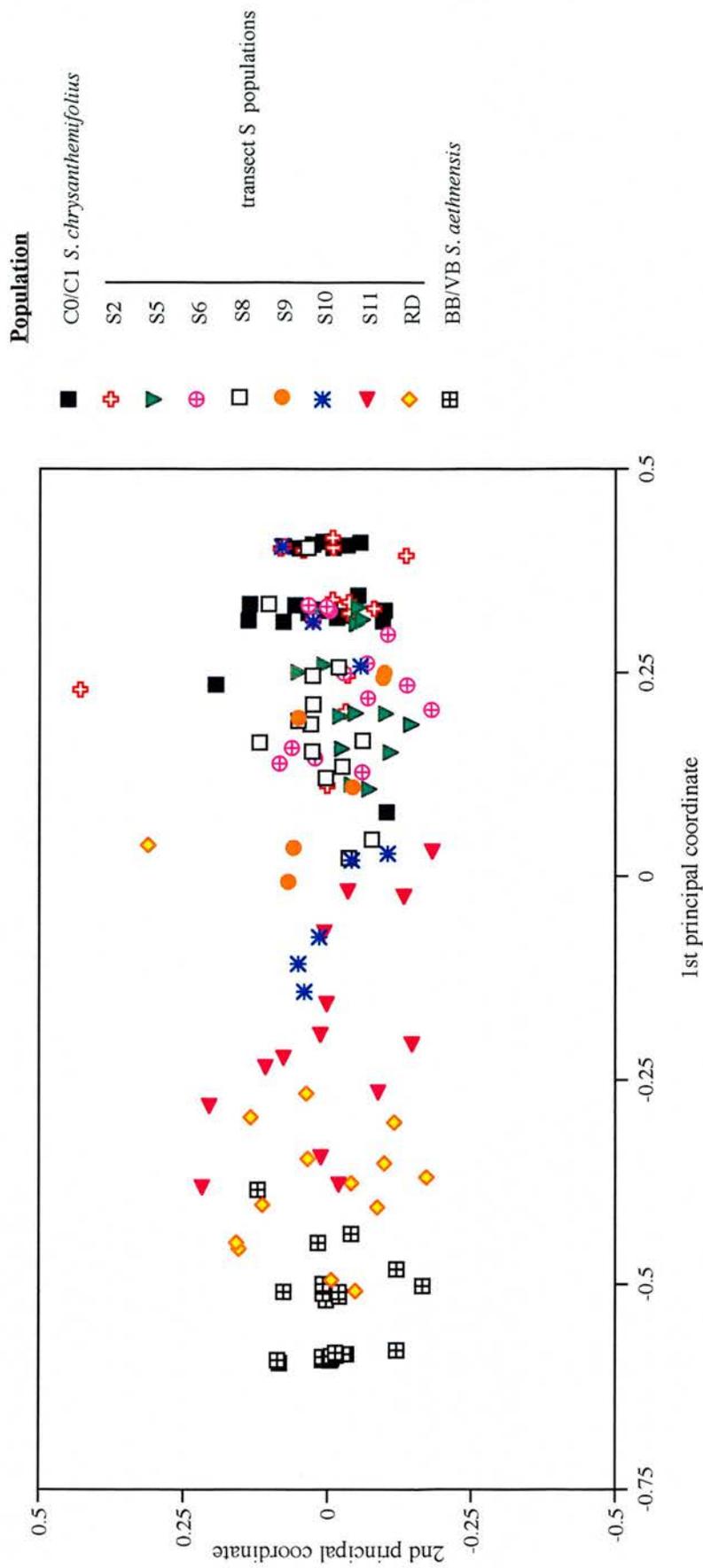


(b) C5



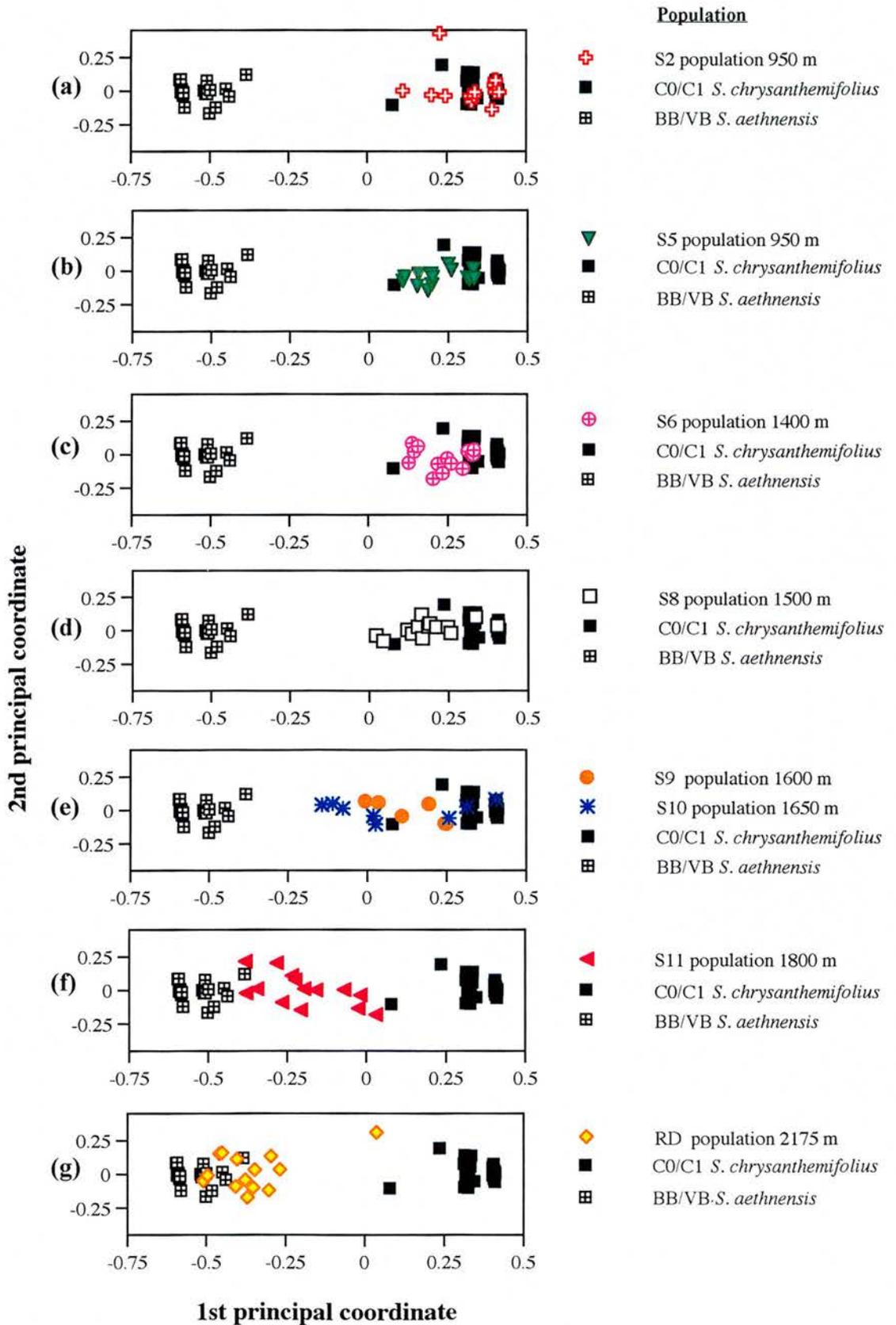
**Plate 4.6 (a) and (b)** Presence/absence of the C5-CHR1 RAPD marker band in five populations along transect S, Mt. Etna. AET indicates *Senecio aethnensis* standards, CHR indicates *S. chrysanthemifolius* standards.

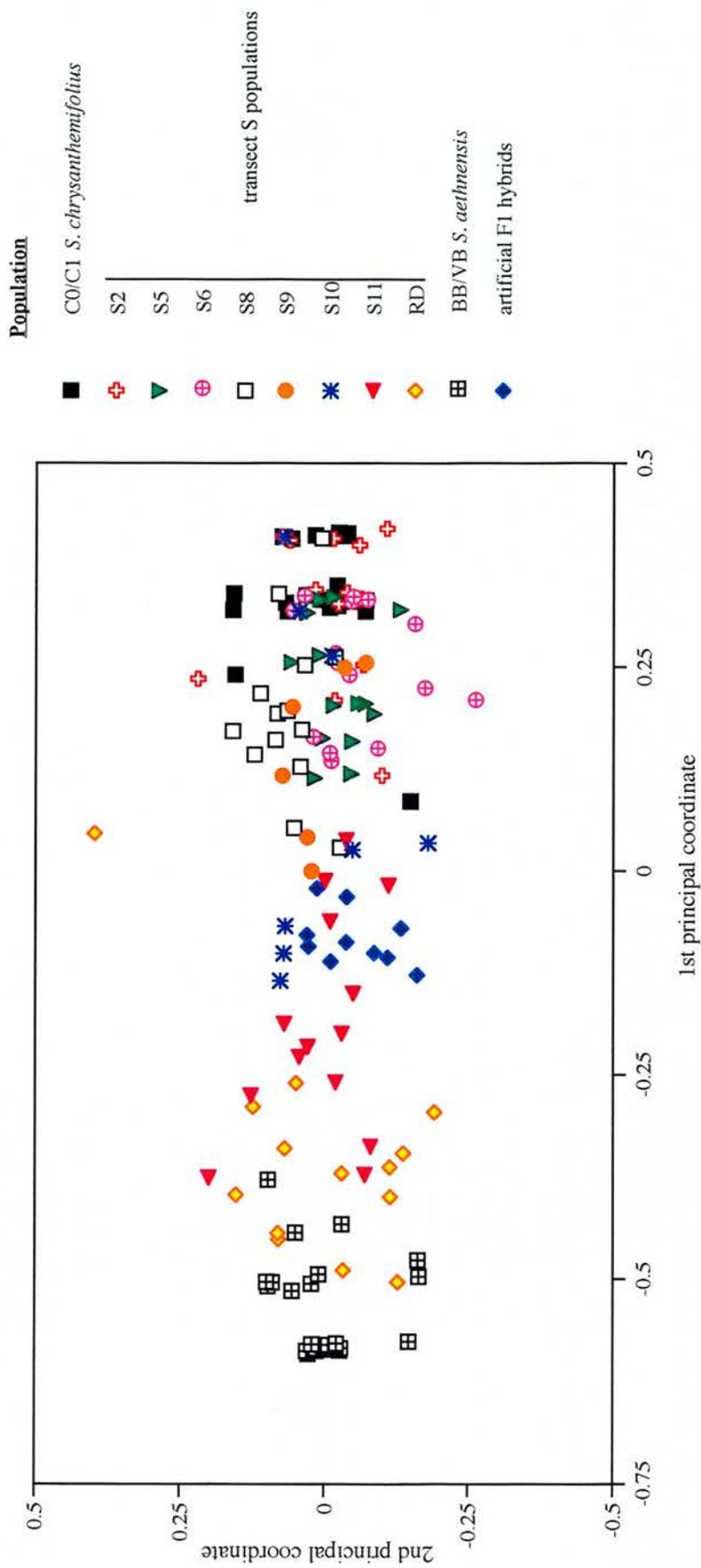




**Fig. 4.11** The first two coordinates of a principal coordinate analysis of RAPD and ISSR data from individuals in populations along transect S, through the postulated hybrid zone between *S. chrysanthemifolius* and *S. aethnensis*, on Mt. Etna.

**Fig. 4.12** Edited version of fig. 4.11. Each edited sub-section was produced by plotting only the principal coordinates of individuals from the 'pure' populations of *Senecio chrysanthemifolius* and *S. aethnensis*, plus the principal coordinates of individuals from one (or two) other population(s) from transect S (rather than plotting the principal coordinates of all sampled individuals in the transect, as in fig. 4.11).





**Fig. 4.13** The first two coordinates of a principal coordinate analysis of RAPD and ISSR data from individuals in populations along transect S, through the postulated hybrid zone between *S. chrysanthemifolius* and *S. aethnensis*, on Mt. Etna; plus the group of artificial F1 individuals.

irrespective of these differences (compare, for example, fig. 4.3 a and b with fig. 4.3 c and d, and fig. 4.5 a and b with fig. 4.5 c and d).

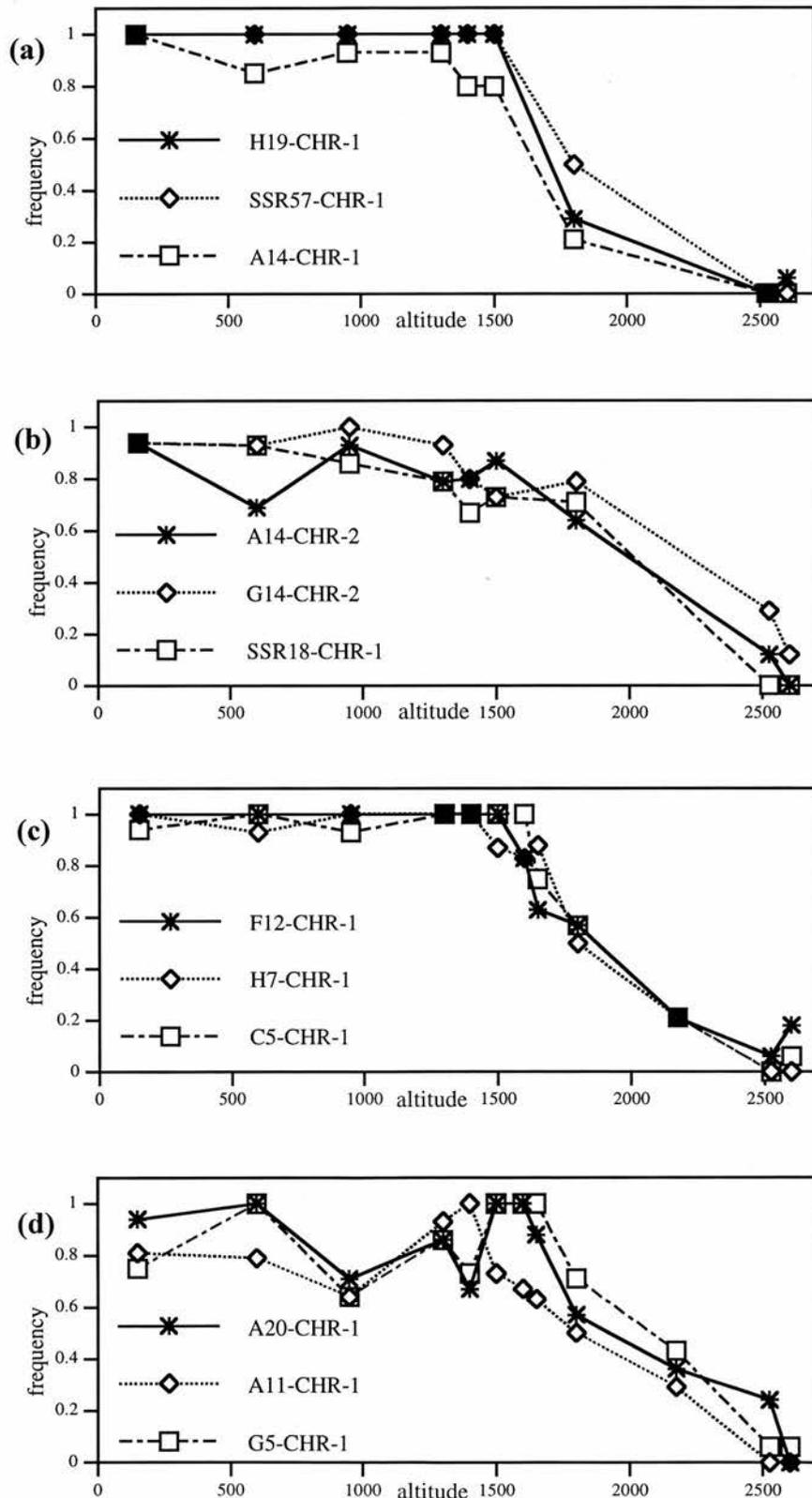
Figs. 4.3 and 4.4 are representative of a common pattern of change recorded for *S. chrysanthemifolius*, with most markers showing clinal variation such that transitions in frequencies took the form of a simple gradient without striking reversals.

*Senecio chrysanthemifolius* markers were present at high frequency from the start of the transect (C0 at 0 km/150 m) until approximately 1500 m or 23 km from C0, at which point their frequency began to decrease rapidly. Marker frequencies tended to reach a minimum at 2500 m or above (28 km from C0 and beyond). Visual comparison suggested that *S. chrysanthemifolius* marker clines were broadly coincident, that is the clines were located at roughly the same place along the transect (Futuyma, 1998). However, within this overall trend, a degree of variation was observed. For example, some *S. chrysanthemifolius* markers demonstrated modest reversals in frequency between 150 m and 1500 m (0 km and 23 km from C0), whilst others did not (fig. 4.3 d versus fig. 4.3 c).

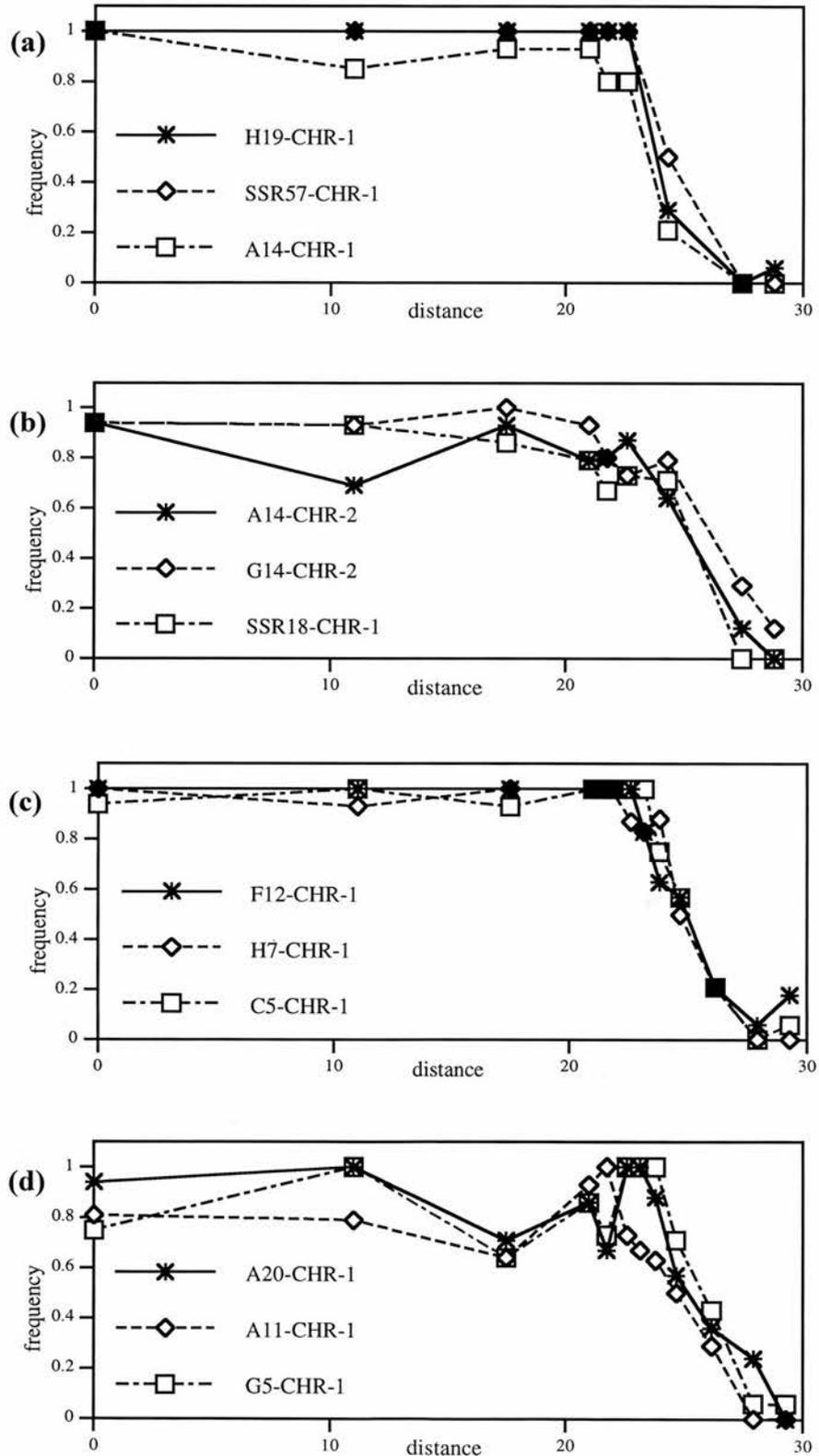
Figs. 4.5 and 4.6 represent the majority of *S. aethnensis* marker plots, with most markers showing smooth clinal variation. Markers occurred at low frequency at the start of the transect until approximately 1500 m or 23 km from C0, at which point their frequency began to increase rapidly. Frequencies of *S. aethnensis* markers tended to reach a maximum at altitudes of 1800 m or above (25 km from C0 and beyond). Visual inspection of the graphs indicated that marker clines were broadly coincident, although again a degree of variation was seen. For example, some *S. aethnensis* markers demonstrated a fairly modest increase, and subsequent decrease, in frequency between 1000 m and 1500 m, whilst others did not (fig. 4.5 b versus fig. 4.5 a). Fig. 4.7 d shows that *S. aethnensis* markers that were present at a comparatively low frequency in 'pure' *S. aethnensis*, nevertheless, exhibited a similar pattern of variation to markers present in 'pure' *S. aethnensis* at high frequency. Overall, it was evident that major decreases in the frequencies of many *S. chrysanthemifolius* markers coincided with major increases in frequencies of many *S. aethnensis* markers.

Whilst a majority of markers showed clinal variation that approximated to the patterns described above, this was not the case for all markers. For example, the *S. chrysanthemifolius* and *S. aethnensis* markers in fig. 4.7 and 4.8 a, b and c decreased or

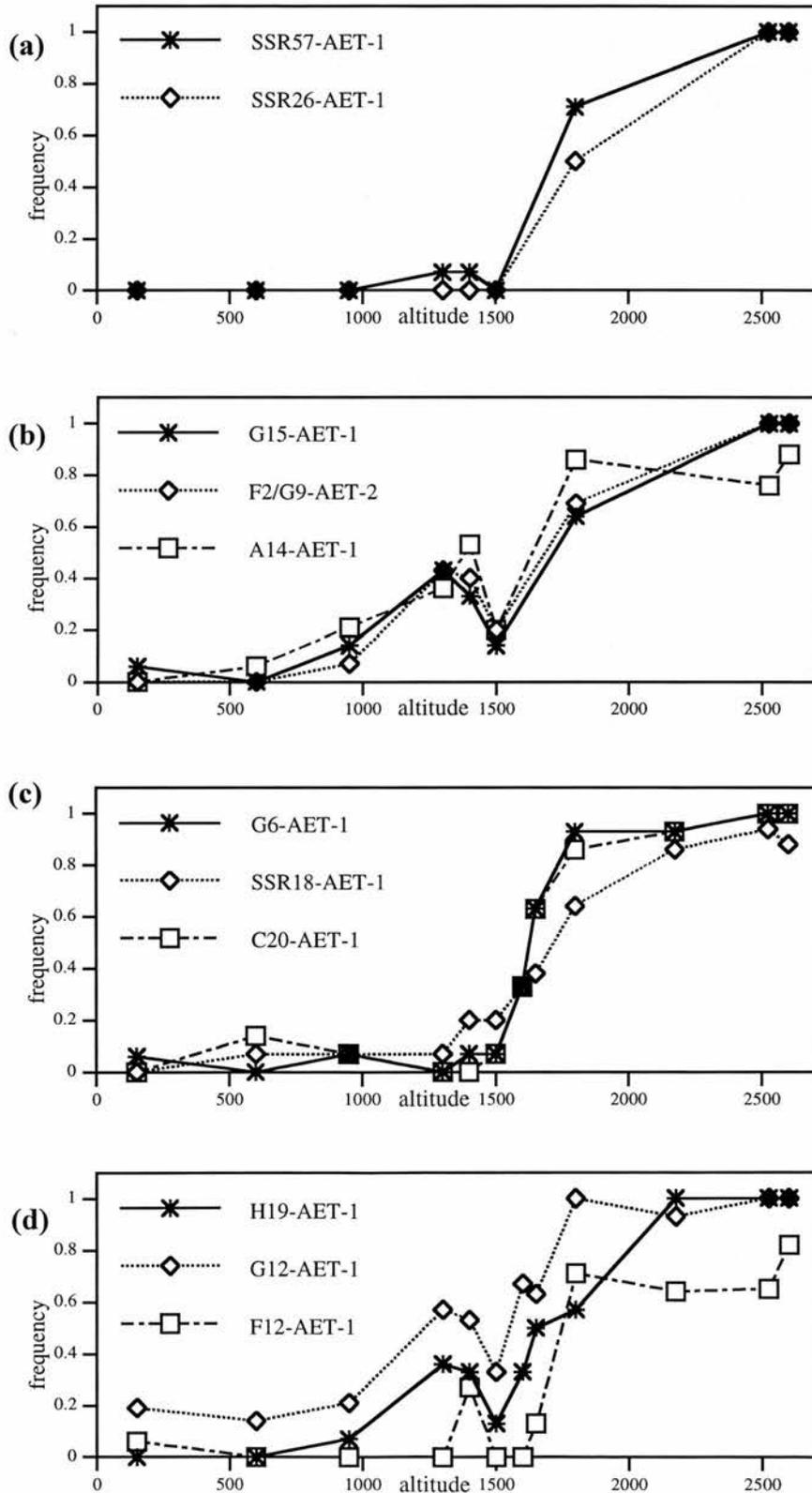
**Fig. 4.3** Frequency of selected *Senecio chrysanthemifolius* RAPD and ISSR (denoted SSR) markers in populations along transect S, through the postulated hybrid zone between *S. chrysanthemifolius* and *S. aethnensis* on Mt. Etna, plotted against population altitude (m).



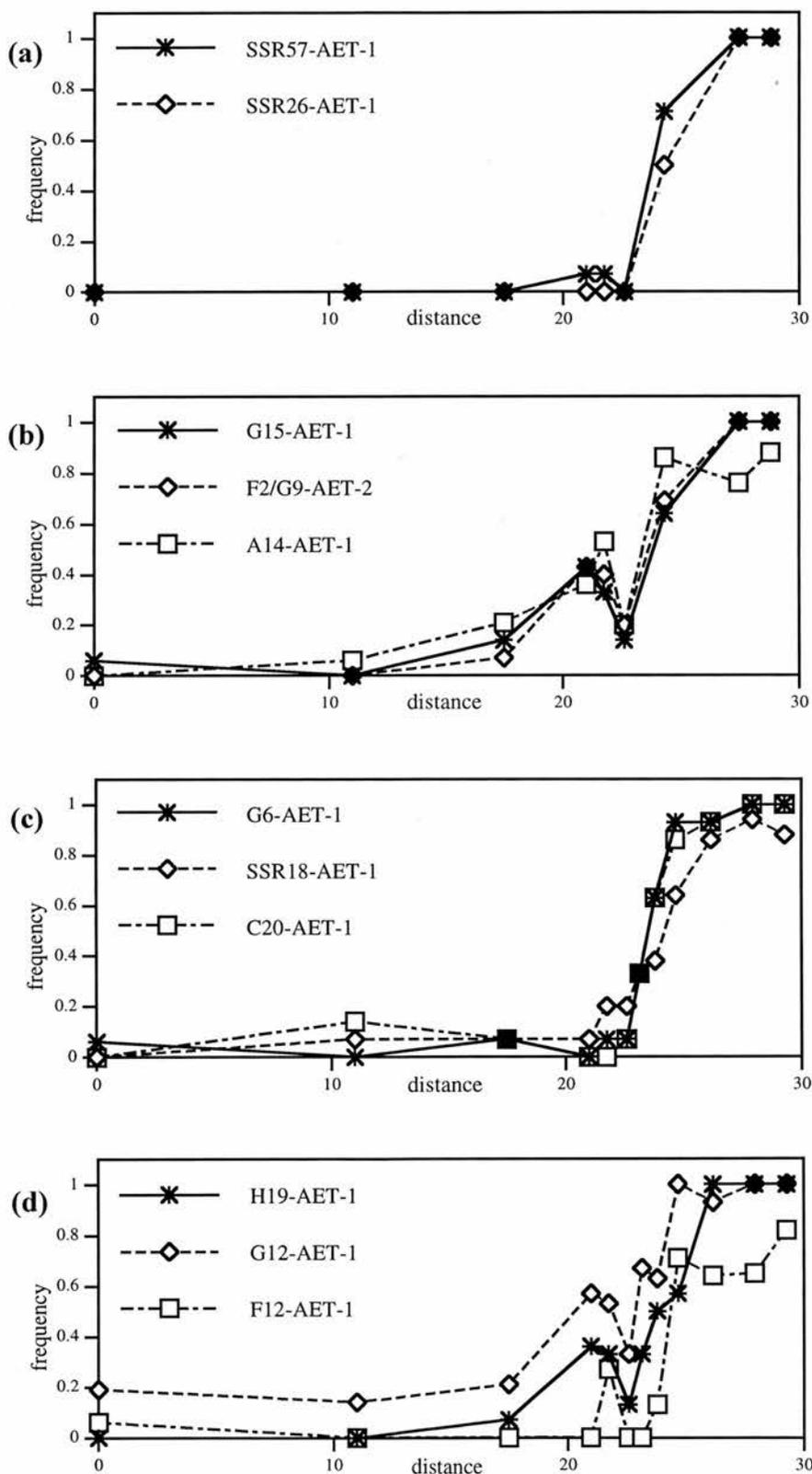
**Fig. 4.4** Frequency of selected *Senecio chrysanthemifolius* RAPD and ISSR (denoted SSR) markers in populations along transect S, through the postulated hybrid zone between *S. chrysanthemifolius* and *S. aethnensis* on Mt. Etna, plotted against cumulative distance of each population from the start of the transect (km).



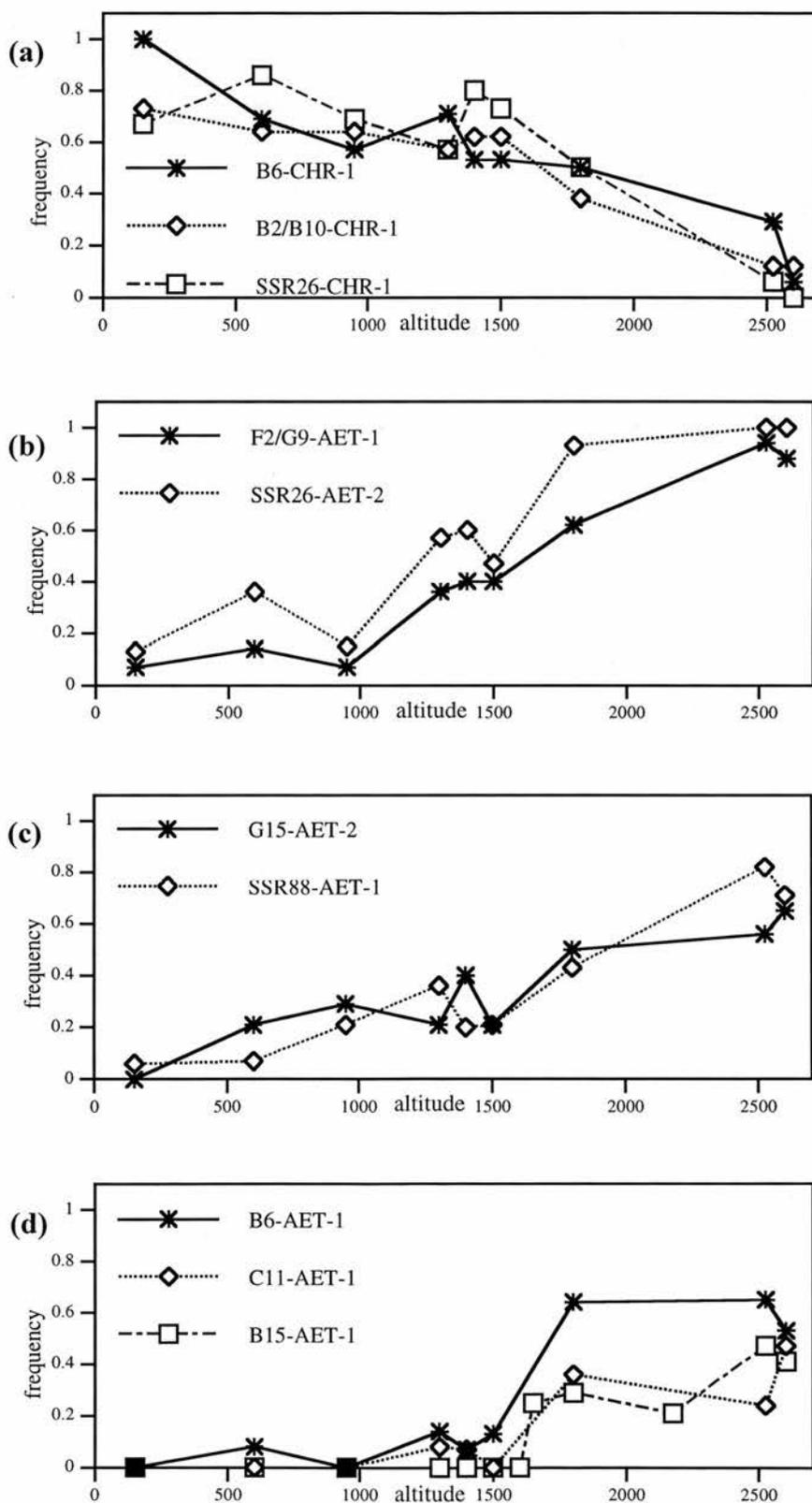
**Fig. 4.5** Frequency of selected *Senecio aethnensis* RAPD and ISSR (denoted SSR) markers in populations along transect S, through the postulated hybrid zone between *S. chrysanthemifolius* and *S. aethnensis* on Mt. Etna, plotted against population altitude (m).



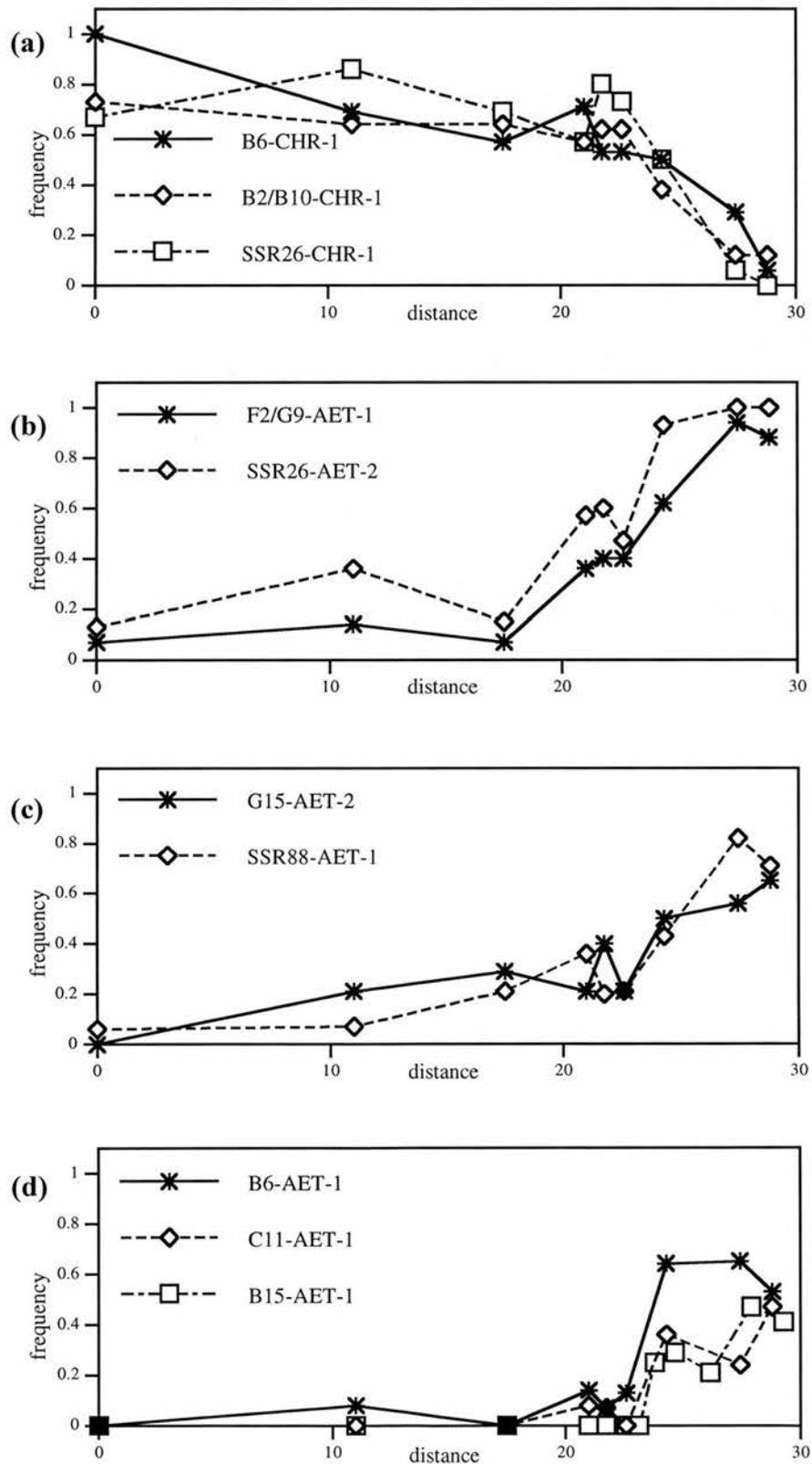
**Fig. 4.6** Frequency of selected *Senecio aethnensis* RAPD and ISSR (denoted SSR) markers in populations along transect S, through the postulated hybrid zone between *S. chrysanthemifolius* and *S. aethnensis* on Mt. Etna, plotted against cumulative distance of each population from the start of the transect (km).



**Fig. 4.7** Frequency of selected *Senecio chrysanthemifolius* and *S. aethnensis* RAPD and ISSR (denoted SSR) markers in populations along transect S, through the postulated hybrid zone between *S. chrysanthemifolius* and *S. aethnensis* on Mt. Etna, plotted against population altitude (m).



**Fig. 4.8** Frequency of selected *Senecio chrysanthemifolius* and *S. aethnensis* RAPD and ISSR (denoted SSR) markers in populations along transect S, through the postulated hybrid zone between *S. chrysanthemifolius* and *S. aethnensis* on Mt. Etna, plotted against cumulative distance of each population from the start of the transect (km).



increased in frequency over a greater distance and over a broader altitudinal range than was general.

### **Multilocus variation**

The binary data matrix constructed for the 26 RAPD and ISSR markers scored in all populations is presented in appendix B.

#### **4.3.2.2 Hybrid index analysis**

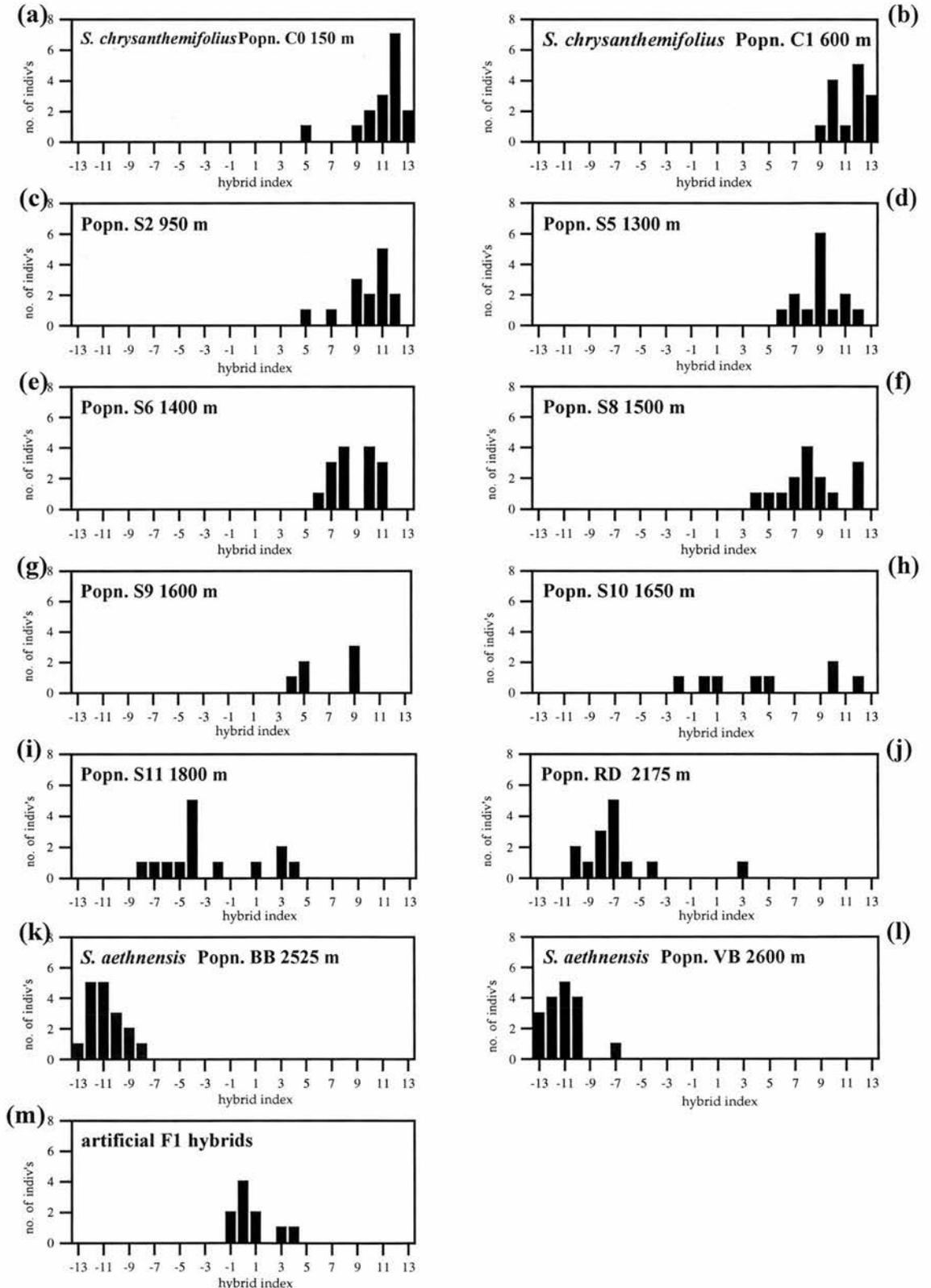
'Pure' populations of *S. chrysanthemifolius* (C0 and C1) and *S. aethnensis* (BB and VB) were clearly separated on the basis of their hybrid indices (fig. 4.9 a and b, versus k and l). This was expected, as the 26 RAPD and ISSR markers, upon which the hybrid indices were based, were chosen for their ability to differentiate individuals of *S. chrysanthemifolius* and *S. aethnensis*. The hybrid indices of 'pure' *S. aethnensis* individuals (populations BB and VB) ranged from -13 to -7, with a mean of -11 (fig. 4.10 a), while those of 'pure' *S. chrysanthemifolius* individuals (populations C0 and C1) ranged from +13 to +9 (excluding an outlier of +5), with a mean of +11.2. Few of the *S. chrysanthemifolius* and *S. aethnensis* markers employed were present at a frequency of 1.00 in one species and 0 in the other species (see chapter 3). Therefore, not all individuals of 'pure' *S. chrysanthemifolius* were expected to score +13 and, similarly, all individuals of 'pure' *S. aethnensis* were not expected to score -13. Moreover, there was evidence (e.g. the outlier visible in fig. 4.9 a) that not all individuals in what were designated as 'pure' populations were in fact 'pure' (see section 3.4.5).

Hybrid indices of all individuals in populations S2, S5, S6, S8 and S9 were positive, indicating that *S. chrysanthemifolius* markers were more prevalent than those of *S. aethnensis*. Between population S2 (950 m) and population S9 (1600 m) there was a gradual shift in the frequency distributions, away from those of 'pure' *S. chrysanthemifolius* (populations C0 and C1) and towards the centre of the hybrid index scale (fig. 4.9 c, d, e, f, g). This was accompanied by a decrease in mean hybrid index, which fell from +9.9 in population S2 to +6.8 in population S9 (fig. 4.10 a).

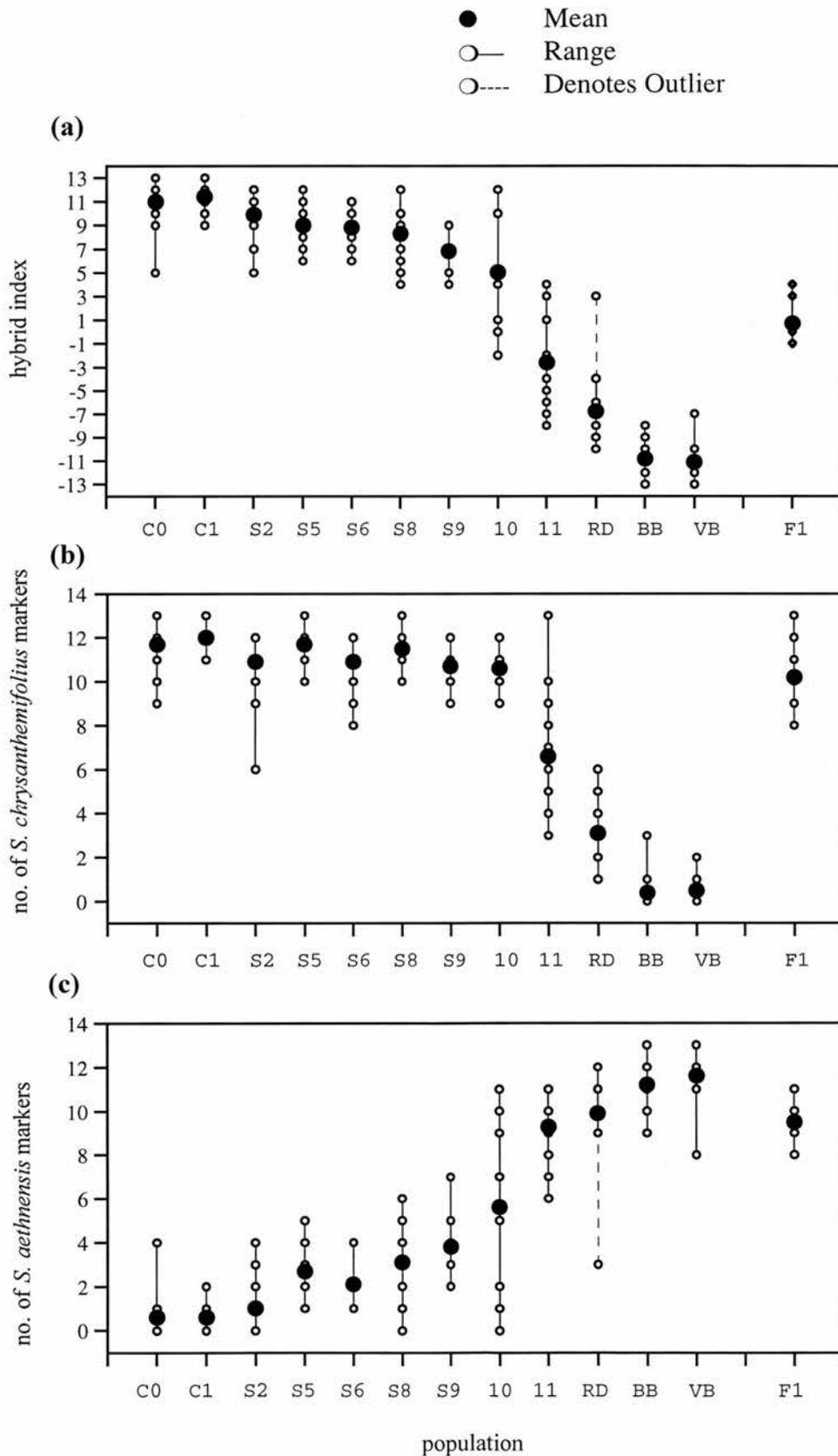
The ranges of hybrid indices in populations S10 and S11 were particularly wide (-2 to +12 and -8 to +4 respectively; fig. 4.10 a). Both positive and negative hybrid indices were present in S10 (fig. 4.9 h) and S11 (fig. 4.9 i). However, whereas population S10 had a bias

**Fig. 4.9 (a) - (l)** Frequency distributions of hybrid indices for the 12 populations along transect S, through the postulated hybrid zone between *Senecio chrysanthemifolius* and *S. aethnensis*, on Mt. Etna; and for the group of artificial F<sub>1</sub> individuals (**m**).

The most 'pure' *S. chrysanthemifolius* individuals would score +13, the most 'pure' *S. aethnensis* individuals would score -13.



**Fig. 4.10** (a) Mean and range of hybrid indices within populations along transect S, through the postulated hybrid zone between *Senecio chrysanthemifolius* and *S. aethnensis*, on Mt. Etna; and within the group of artificial F<sub>1</sub> individuals. (b) Mean and range of the number of *S. chrysanthemifolius* markers within each population along transect S, and within the group of artificial F<sub>1</sub> individuals. (c) Mean and range of the number of *S. aethnensis* markers within each population along transect S, and within the group of artificial F<sub>1</sub> individuals.



towards positive hybrid indices, population S11 had a bias towards negative hybrid indices (i.e. most individuals contained more *S. aethnensis* markers than *S. chrysanthemifolius* markers). This was reflected in a reduction in mean hybrid indices from +5 in S10 to -2.6 in S11 (fig. 4.10 a).

With one exception, all individuals in population RD had negative hybrid indices, and the population had a frequency distribution (fig. 4.9 j) and mean hybrid index (-6.8; fig. 4.10 a) that were tending towards those of 'pure' *S. aethnensis*.

It is to be expected that F<sub>1</sub> individuals would have hybrid index scores close to zero (i.e. that the number of parental markers in F<sub>1</sub> individuals would be roughly equivalent). The hybrid indices of the F<sub>1</sub> individuals in the present study ranged from -1 to +4 with a mean of +0.7 (fig. 4.10 a); thus the result matched the expectation.

The mean and range of the number of *S. chrysanthemifolius* and *S. aethnensis* markers in each population implied that the overall decline in mean hybrid index between populations C0 and S10 was due more to an increase in the number of *S. aethnensis* markers than to a drop in number of *S. chrysanthemifolius* markers (fig. 4.10 b and c). In contrast, the decline in mean hybrid index between populations S11 and VB was due to a decrease in the number of *S. chrysanthemifolius* markers, combined with an increase in the number of *S. aethnensis* markers (fig. 4.10 b and c). Population S10 exhibited the widest range in number of *S. aethnensis* markers, while population S11 showed the widest range in number of *S. chrysanthemifolius* markers (fig. 4.10 b and c).

#### 4.3.2.3 Principal coordinate analysis

Hybrid indices were an effective means of visualising multilocus RAPD and ISSR data from the postulated hybrid zone. However, calculation of hybrid indices leads to a loss of information. For instance, two individuals with the same hybrid index could possess very different combinations of markers (e.g. an individual with a hybrid index of +3 might have 13 *S. chrysanthemifolius* markers and 10 *S. aethnensis* markers or three *S. chrysanthemifolius* markers and no *S. aethnensis* markers). Principal coordinate analysis, based upon Nei and Li's, or Jaccard's, coefficient, provides greater resolution of the data because the similarity between individuals is based upon the shared presence of RAPD and ISSR markers. PCO can, therefore, be considered as a valuable counterpart to hybrid index analysis.

The PCO plots based upon all three coefficients were very similar and therefore only the results derived from Nei and Li's coefficient are displayed (figs. 4.11 and 4.13). The latter were chosen because more of the total variation was accounted for by the first two principal coordinates, relative to plots based on the other coefficients.

The first principal coordinate of the PCO plot that included all individuals on the transect represented 91.04% of the total variation, the second principal coordinate explained 4.83%, while the third principal coordinate represented 4.59% of the total variation (fig. 4.11). The first principal coordinate axis produced a convincing separation of 'pure' populations of *S. chrysanthemifolius* (C0 and C1, fig. 4.11) and *S. aethnensis* (BB and VB). This was expected, as the 26 RAPD and ISSR markers, upon which the PCO was based, were chosen for their ability to differentiate individuals of *S. chrysanthemifolius* and *S. aethnensis*. *Senecio chrysanthemifolius* individuals formed a closely aggregated group, except for one obvious outlier. As this outlier was positioned towards the centre of the first principal coordinate axis, it was clear that it was not representative of 'pure' *S. chrysanthemifolius*. *Senecio aethnensis* individuals formed a more loosely aggregated group.

A majority of individuals from population S2 clustered very closely with individuals of 'pure' *S. chrysanthemifolius* (fig. 4.12 a), while most individuals from populations S5, S6 and S8 formed closely aggregated groups near to, but distinct from 'pure' *S. chrysanthemifolius* (fig. 4.12 b-d). Individuals from population S9 were less aggregated and two S9 individuals, together with two individuals from population S8, tended towards intermediacy between *S. chrysanthemifolius* and *S. aethnensis*.

Individuals from populations S10 and S11 (which comprised a range of parental, hybrid and backcrossed phenotypes) were widely scattered across the PCO plot (fig. 4.12 e-f) and a number occupied positions intermediate with respect to 'pure' *S. chrysanthemifolius* and *S. aethnensis*. However, some S10 individuals were closer to 'pure' *S. chrysanthemifolius*, whilst some S11 individuals were closer to 'pure' *S. aethnensis*.

Individuals from population RD (bar one obvious outlier) clustered towards or with 'pure' *S. aethnensis* individuals.

The first principal coordinate of the PCO plot that included all individuals from the transect plus the artificial F<sub>1</sub>s represented 91.51% of the total variation, the second principal coordinate explained 5.00% and the third principal coordinate represented 4.86% of the total variation (fig. 4.13). F<sub>1</sub> individuals in this plot were positioned between the two parental species (fig. 4.13), and close to certain individuals from populations S9, S10 and S11.

#### 4.3.2.4 Cluster analysis

Cluster analysis was carried out on two data sets, one of which contained all individuals in transect S, and a second that contained all individuals in the transect plus the F<sub>1</sub>s. Neighbour-joining and UPGMA phenograms were produced for both data sets but only the former are displayed (fig. 4.14 and fig. 4.15). Although most of the information conveyed by the phenograms is very similar to that derived from the PCO plots, the phenograms do serve to emphasise a number of features; for example, the intermediate nature of individuals from populations S10 and S11, which cluster between populations of 'pure' *S. aethnensis* (VB, BB) and 'pure' *S. chrysanthemifolius* populations (C0, C1; fig. 4.14), and on either side of the F<sub>1</sub> individuals (fig. 4.15).

#### 4.3.2.5 Analysis of molecular variance

The AMOVA (table 4.6) revealed that 56% of the total variation was present between populations, compared to 44% present within populations (significant at  $p < 0.001$ ).

**Table 4.6** Analysis of molecular variance (AMOVA) for 12 populations along transect S, through the postulated hybrid zone between *S. chrysanthemifolius* and *S. aethnensis*, on Mt. Etna.

Source of variation	d.f. <sup>1</sup>	SSD <sup>2</sup>	MSD <sup>3</sup>	Variance component	% of the total variance	P-value <sup>4</sup>
Between populations	11	562.31	51.12	3.56	56.44	< 0.001
Within populations	152	417.25	2.75	2.75 ( $\Phi_{ST} = 0.564$ )	43.56	< 0.001

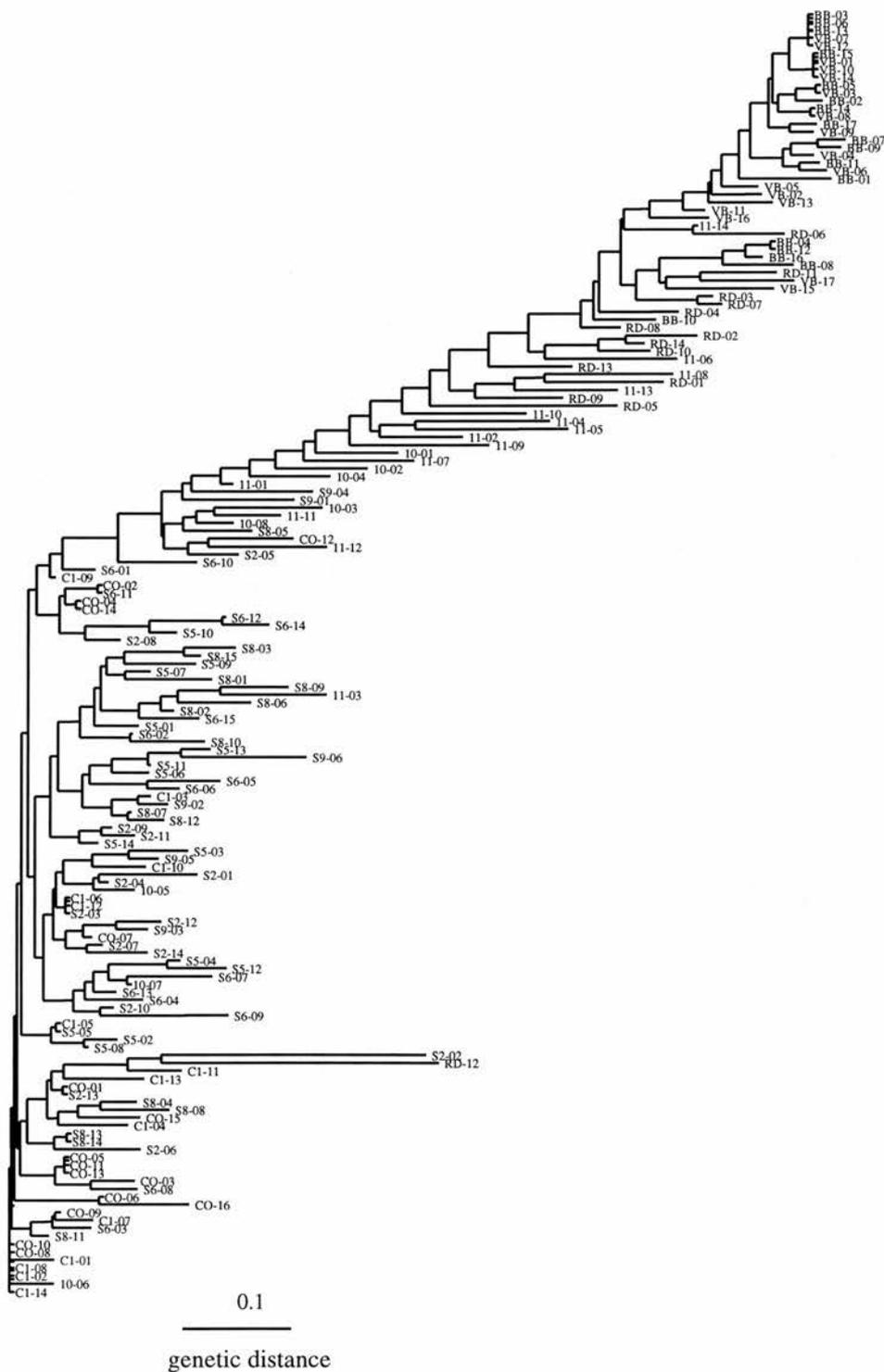
<sup>1</sup> Degrees of freedom.

<sup>2</sup> Sum of squared deviations.

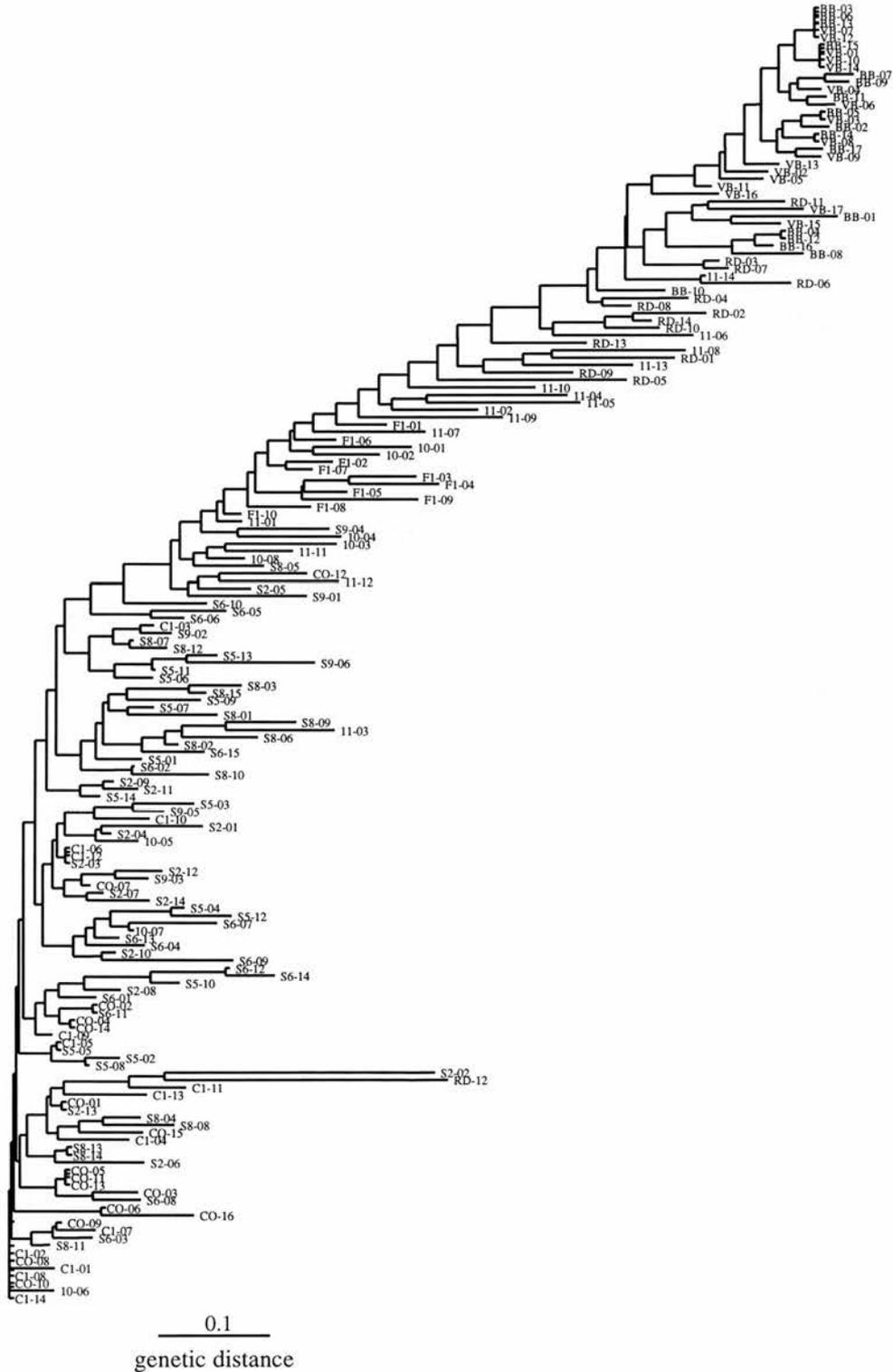
<sup>3</sup> Mean squared deviations.

<sup>4</sup> Significance of the variance components (the probability of obtaining a more extreme component estimate by chance alone).

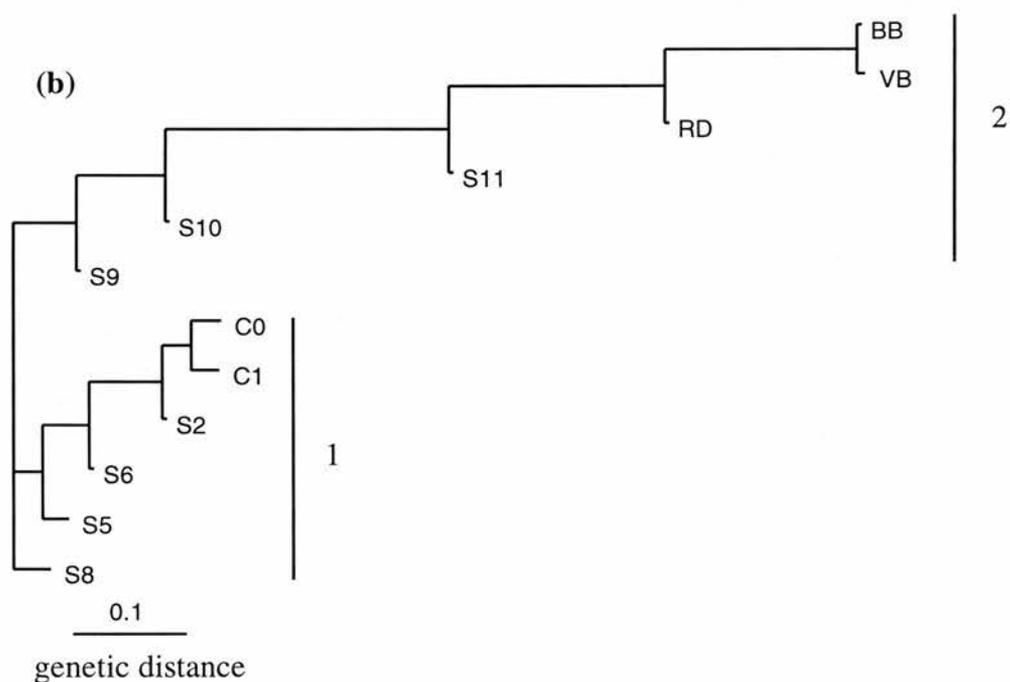
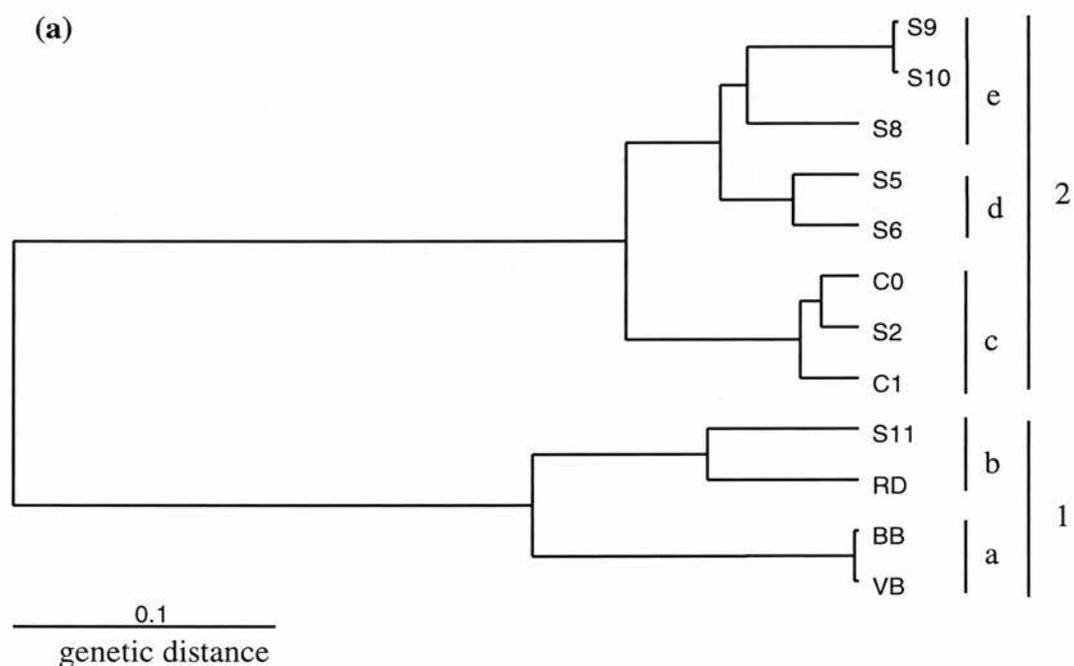
**Fig. 4.14** Neighbour-joining dendrogram of all sampled individuals in populations along transect S, through the postulated hybrid zone between *Senecio chrysanthemifolius* and *S. aethnensis*, on Mt. Etna. The clustering was based upon genetic distances, derived from Nei and Li's similarity coefficient, from RAPD and ISSR data.



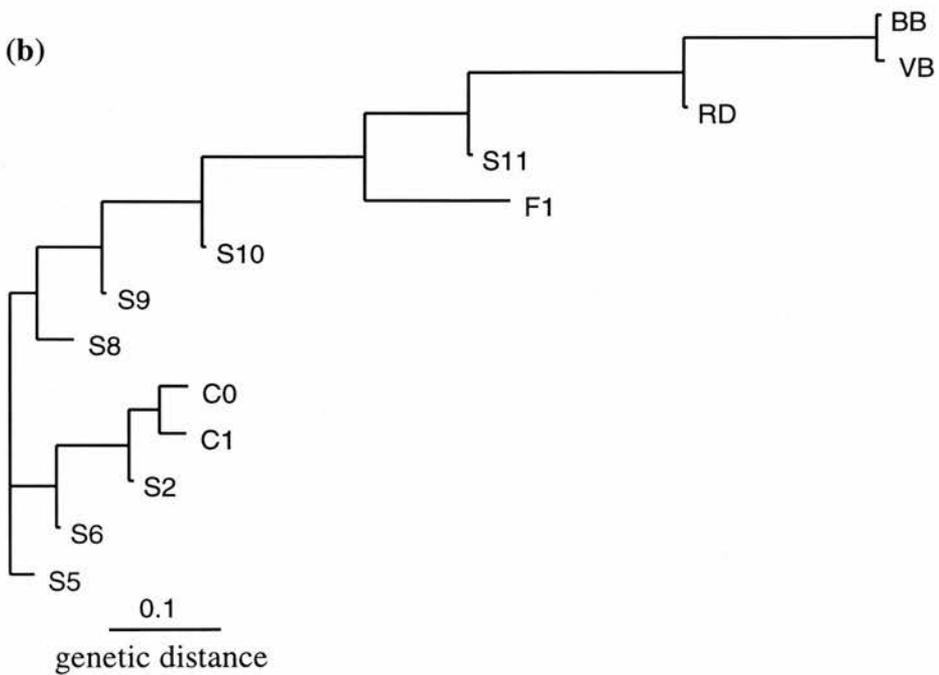
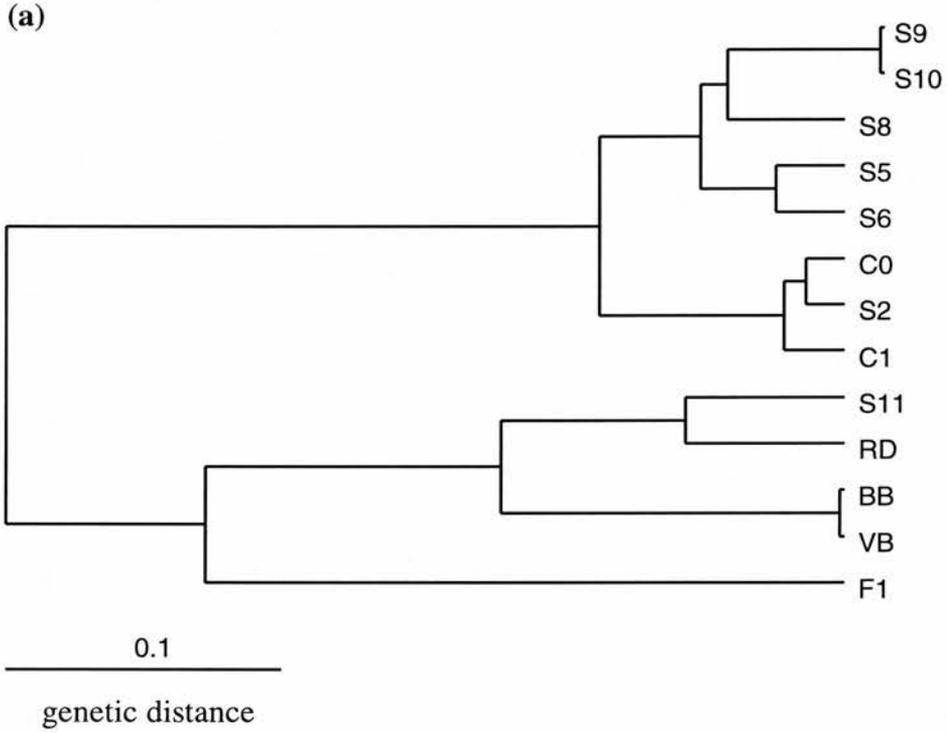
**Fig. 4.15** Neighbour-joining dendrogram of all sampled individuals in populations along transect S, through the postulated hybrid zone between *Senecio chrysanthemifolius* and *S. aethnensis*, on Mt. Etna; plus the 10 artificial F<sub>1</sub> individuals. The clustering was based upon genetic distances, derived from Nei and Li's similarity coefficient, from RAPD and ISSR data.



**Fig. 4.16** (a) UPGMA, and (b) neighbour-joining dendrograms of populations along transect S, through the postulated hybrid zone between *Senecio chrysanthemifolius* and *S. aethnensis*, on Mt. Etna. The clustering was based upon pair-wise  $\Phi_{ST}$  values from an AMOVA of RAPD and ISSR data.



**Fig. 4.17** (a) UPGMA, and (b) neighbour-joining dendrograms of populations along transect S, through the postulated hybrid zone between *Senecio chrysanthemifolius* and *S. aethnensis*, on Mt. Etna; and the group of artificial F<sub>1</sub> individuals. The clustering was based upon pair-wise  $\Phi_{ST}$  values from an AMOVA of RAPD and ISSR data.



Most pair-wise  $\Phi_{ST}$  values (used as a measure of inter-population genetic distance) obtained in this study were significantly larger than a random  $\Phi_{ST}$  value (table 4.7), indicating that the majority of populations were significantly genetically different from each other. Populations VB and C1 were the most distinct (pair-wise  $\Phi_{ST}$  of 0.8716) and populations S10 and S9 were most similar (pair-wise  $\Phi_{ST}$  of -0.026).

**Table 4.7** Inter-population genetic distances (pair-wise  $\Phi_{ST}$  values) between 12 populations along transect S, through the postulated hybrid zone between *S. chrysanthemifolius* and *S. aethnensis*, on Mt. Etna.

	C0	C1	S2	S5	S6	S8	S9	S10	S11	RD	BB	VB
C0		0.0513	0.0288†	0.2047	0.1094	0.1837	0.1997	0.3399	0.5552	0.7113	0.8536	0.8605
C1			0.0379	0.1708	0.1397	0.1932	0.1527	0.3385	0.5516	0.7169	0.8647	0.8716
S2				0.1241	0.0337‡	0.1432	0.1095	0.2329	0.4763	0.6440	0.8105	0.8176
S5					0.0504	0.0795	0.0228†	0.1802	0.4308	0.6279	0.8033	0.8094
S6						0.1119	0.0558	0.1835	0.4158	0.5977	0.7781	0.7846
S8							0.0246†	0.1464	0.3943	0.5962	0.7724	0.7796
S9								-0.026†	0.2215	0.4715	0.7299	0.7427
S10									0.1667	0.4207	0.6857	0.6981
S11										0.1157	0.3517	0.3557
RD											0.1254	0.1673
BB												0.0031†
VB												

All inter-population genetic distances were significantly larger than a random  $\Phi_{ST}$  value at  $P < 0.001$  except: ‡ significant at  $P < 0.05$  and † not significant.

UPGMA clustering of pair-wise  $\Phi_{ST}$  values produced two main clusters (fig. 4.16 a). Cluster one comprised two sub-clusters: sub-cluster (a) contained both populations of 'pure' *S. aethnensis* (VB and BB) and sub-cluster (b) contained populations RD and S11. Cluster two comprised three sub-clusters: (c), (d) and (e). Sub-cluster (e) consisted of populations S8, S9 and S10 and was linked as a sister to sub-cluster (d), which consisted of populations S5 and S6. Sub-clusters (d) and (e) were connected to sub-cluster (c), which contained population S2 and both populations of 'pure' *S. chrysanthemifolius* (C0 and C1).

Neighbour-joining clustering of the same  $\Phi_{ST}$  values produced a somewhat different phenogram (fig. 4.16 b). Two main clusters branched away from the principal axis of the phenogram. In cluster one, population S5 was closest to the main axis, followed by populations S6 and S2, and the 'pure' *S. chrysanthemifolius* populations (C0, C1) were at the apex. In cluster two, population S9 was closest to the main axis, followed by populations S10, S11 and RD, and the 'pure' *S. aethnensis* populations (VB, BB) were at the apex. Population S8 branched off the main axis separately from clusters one and two.

An AMOVA was also carried out on the 12 populations along the transect plus the group of  $F_1$  individuals. The variance components etc. were very similar to those of the AMOVA described above (data not shown). UPGMA and neighbour-joining clustering of pair-wise  $\Phi_{ST}$  values produced phenograms that were comparable to those described above. However, these additional phenograms are of interest because they show the position of the group of  $F_1$  individuals relative to the 12 populations along the transect. In the UPGMA phenogram (fig. 4.17 a), the  $F_1$  group is positioned in the same cluster as populations S11, RD and the two 'pure' *S. aethnensis* populations (BB and VB). In the neighbour-joining phenogram the  $F_1$  group is situated between population S10 and population S11 (fig. 4.17 b).

#### 4.3.2.6 Shannon's diversity index

The level of genetic diversity ( $H_0$ ) within each of the 12 populations along the transect and the  $F_1$  individuals was quantified using Shannon's diversity index (table 4.8). Population S11 maintained the highest level of genetic diversity (compared to other transect populations), closely followed by populations RD, S9 and S10. Populations S8, S6, S2 and S5 possessed intermediate levels of genetic diversity, while populations of 'pure' *S. chrysanthemifolius* (C0, C1) and *S. aethnensis* (BB, VB) had the lowest diversity values.

**Table 4.8** Estimates of genetic diversity ( $H_0$ ) within populations along transect S, through the postulated hybrid zone between *S. chrysanthemifolius* and *S. aethnensis*, on Mt. Etna.

Population	Genetic diversity ( $H_0$ ) <sup>†</sup>	$H_0/26$ <sup>‡</sup>
1	3.363.	0.129
C1 <i>S. chrysanthemifolius</i>	3.274	0.126
S2	5.665	0.218
S5	4.899	0.188
S6	6.317	0.243
S8	6.369	0.245
S9	8.169	0.314
S10	8.063	0.310
S11	9.654	0.371
RD	8.418	0.324
BB <i>S. aethnensis</i>	3.209	0.123
VB <i>S. aethnensis</i>	3.316	0.128
$F_1$ *	5.434	0.209

<sup>†</sup> Mean of values from 26 RAPD and ISSR markers.

<sup>‡</sup> Genetic diversity per marker scored (i.e.  $H_0$  divided by the total number of bands scored).

\* The genetic diversity of the group of  $F_1$  individuals was calculated but the  $F_1$  group was not included in the calculations of the  $H_{pop}/H_{sp}$  etc..

Partitioning genetic diversity within and between populations (table 4.9) revealed that approximately equal amounts of the total variation were present between and within populations (49% and 51% respectively). These values were very similar to those obtained in the AMOVA analysis (table 4.6).

**Table 4.9** Partitioning genetic diversity within and between the 12 populations along the transect through the postulated hybrid zone, between *S. chrysanthemifolius* and *S. aethnensis*, on Mt. Etna.

$H_{pop}^1$	$H_{pop}/26^1$	$H_{sp}^2$	$H_{sp}/26^2$	$H_{pop}/H_{sp}^3$	$(H_{sp}-H_{pop})/H_{sp}^4$
5.893	0.227	12.027	0.463	0.490	0.510

<sup>1</sup> Mean within population genetic diversity ( $H_{pop}$ ) and mean within population genetic diversity per marker scored ( $H_{pop}/26$ ).

<sup>2</sup> Total genetic diversity ( $H_{sp}$ ) and total genetic diversity per marker scored ( $H_{sp}/26$ ).

<sup>3</sup> Proportion of the total genetic diversity maintained within populations.

<sup>4</sup> Proportion of the total genetic diversity maintained between populations.

### 4.3.3 Isozyme variation along transect S, through the postulated hybrid zone between *S. chrysanthemifolius* and *S. aethnensis*, on Mt. Etna

Allele frequencies recorded at the six loci surveyed in 12 populations along the transect are shown in table 4.10. All six loci were polymorphic and the total number of alleles detected was 15 (range 2-3 per locus). In previous studies of Sicilian *Senecio* material, *Pgi-1* and *Pgm-1* were scored but were found to be monomorphic (Abbott, personal communication). They were therefore ignored in the present study.

Two loci, *Acp-2* and *Pgi-2*, were known to be useful for distinguishing *S. chrysanthemifolius* from *S. aethnensis* (see section 3.4.2) and so the pattern of variation at these loci was of particular interest.

*Acp-2 b* and *Acp-2 a* are diagnostic taxon-specific markers for *S. chrysanthemifolius* and *S. aethnensis*, respectively. The frequency of *Acp-2 a* increased with increasing altitude and distance from the start of the transect (C0), forming a cline that approximated to a sigmoidal curve (fig. 4.18). The frequency of this allele remained very low from the start of the transect, at an altitude of 150 m, to a distance of 17.5 km and an altitude of 1000 m. Between 17.5 km/1000 m and 21 km/1300 m the frequency of this allele increased slightly, to 0.15, before falling to 0.05 at 22.6 km/1500 m. From 22.6 km/1500 m its frequency rose sharply, then levelled off to a value of 1.00 at 28 km/2500 m. The frequency of *Acp-2 b*

correspondingly decreased with increasing altitude and distance from the start of the transect (table 4.10).

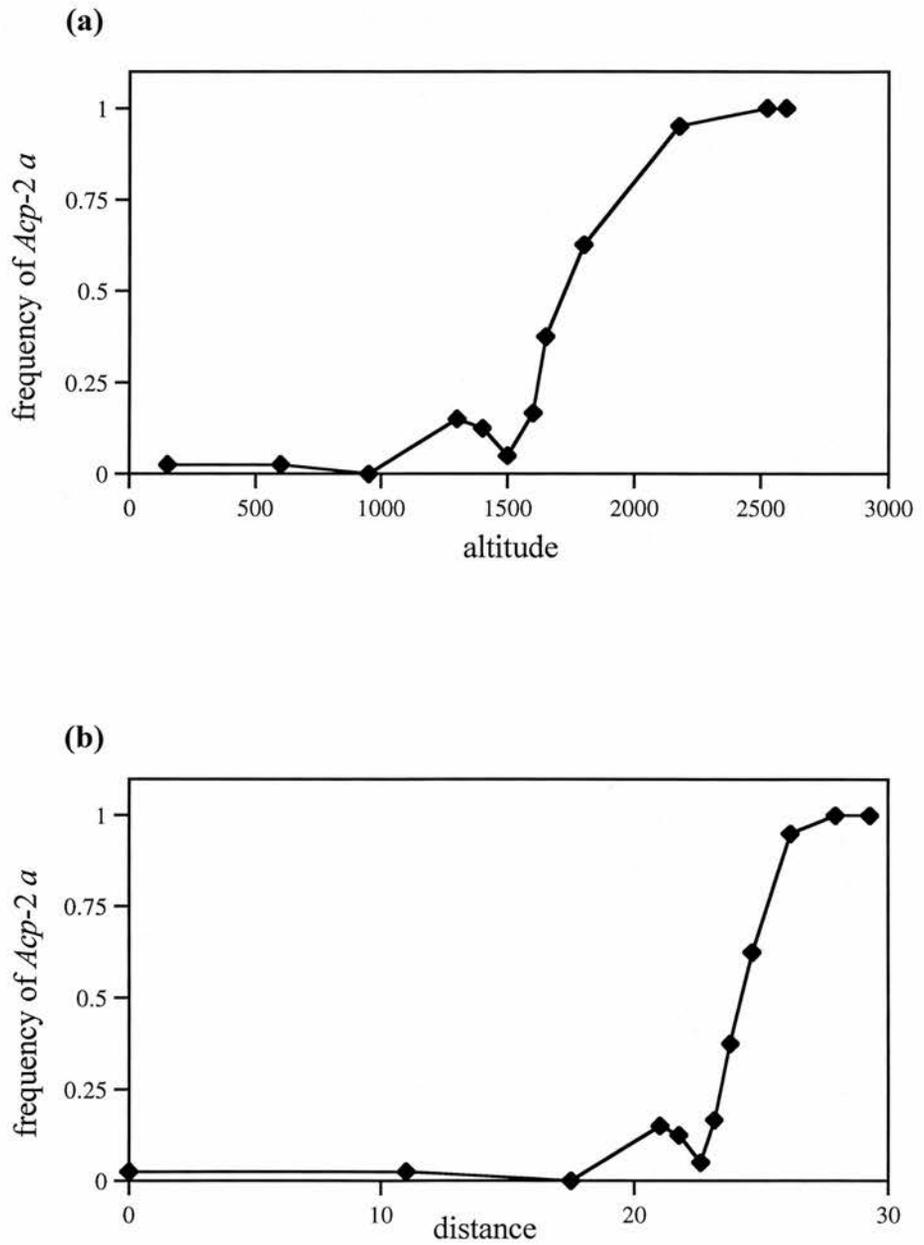
*Pgi-2 b* and *Idh-1 c* were established to be 'private bands' specific to *S. aethnensis*; i.e. they were absent from populations of 'pure' *S. chrysanthemifolius* and present, at low frequency, in populations of 'pure' *S. aethnensis*. Unlike *Acp-2 a*, the frequencies of *Pgi-2 b* and *Idh-1 c* did not increase smoothly with increasing altitude or distance from the start of the transect (C0). *Pgi-2 b* and *Idh-1 c* were first recorded in population S5, (21 km/1300 m) and then increased and decreased in frequency along the remainder of the transect (fig. 4.19; table 4.10).

Alleles at the *Aat-3*, *Aco-1* and *Pgm-2* loci did not distinguish 'pure' *S. chrysanthemifolius* from 'pure' *S. aethnensis*, and could not, therefore, be used to study changes in allele frequency along the transect.

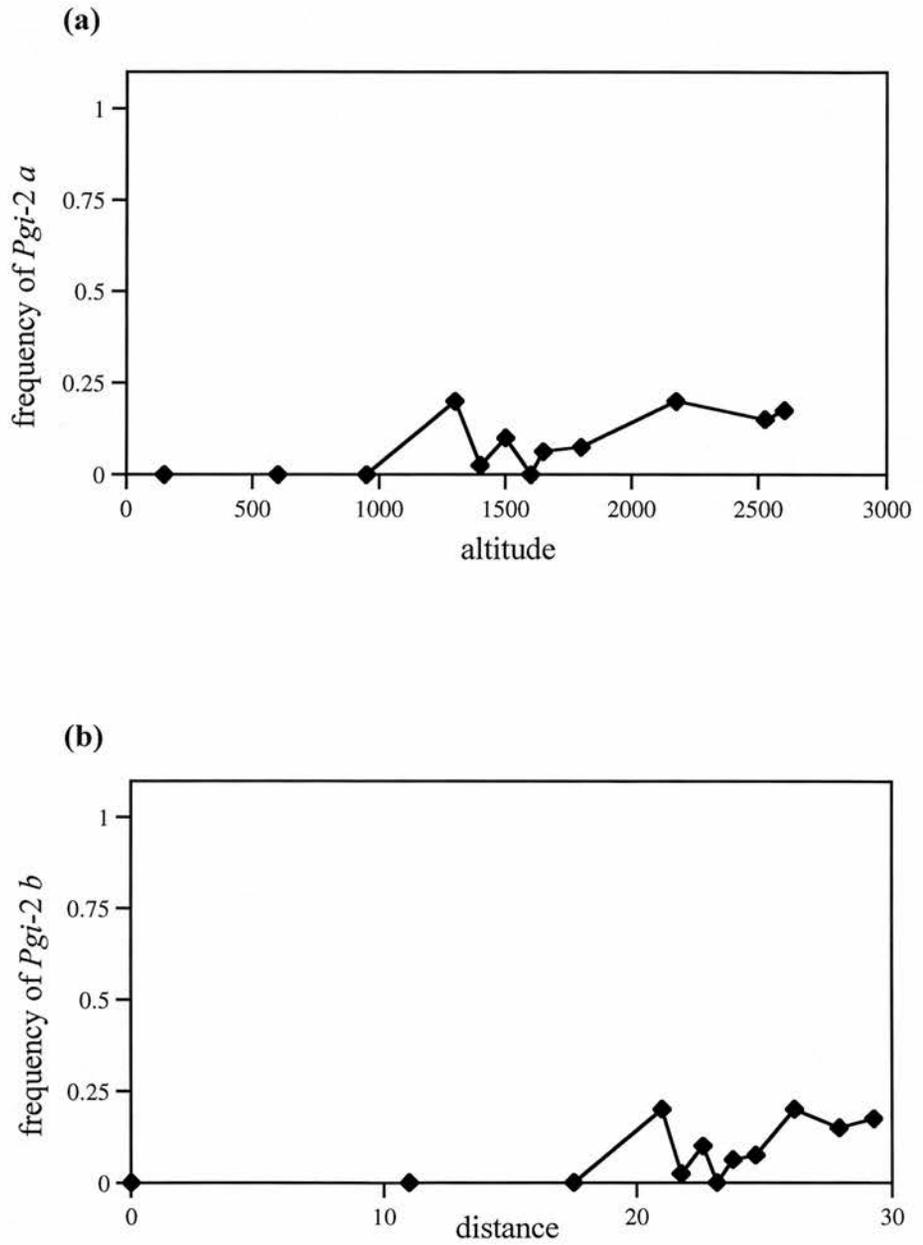
**Table 4.10** Allele frequencies at six polymorphic isozyme loci in populations along transect S, through the postulated hybrid zone between *Senecio chrysanthemifolius* and *S. aethnensis*, on Mt. Etna.

Locus/ allele	CO n = 20	C1 n = 20	S2 n = 20	S5 n = 20	S6 n = 20	S8 n = 20	S9 n = 12	S10 n = 8	S11 n = 20	RD n = 20	BB n = 20	VB n = 20
<b>Aat-3</b>												
<i>d</i>	0.975	0.950	0.900	0.775	0.800	0.925	0.792	0.938	0.900	0.850	0.850	0.975
<i>e</i>	0.025	0.050	0.100	0.225	0.200	0.075	0.208	0.063	0.100	0.150	0.150	0.025
<b>Acp-2</b>												
<i>a</i>	0.025	0.025	0.000	0.150	0.125	0.050	0.167	0.375	0.625	0.950	1.000	1.000
<i>b</i>	0.975	0.975	1.000	0.850	0.875	0.950	0.833	0.625	0.375	0.50	0.000	0.000
<b>Aco-1</b>												
<i>a</i>	0.975	1.000	1.000	1.000	0.925	0.950	0.958	0.938	0.900	0.775	0.925	0.925
<i>b</i>	0.025	0.000	0.000	0.000	0.075	0.050	0.042	0.000	0.100	0.175	0.075	0.075
<i>c</i>	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.063	0.000	0.050	0.000	0.000
<b>Idh-1</b>												
<i>a</i>	0.050	0.000	0.000	0.000	0.000	0.050	0.000	0.000	0.000	0.000	0.000	0.000
<i>b</i>	0.950	1.000	1.000	0.950	1.000	0.950	0.958	0.938	0.900	0.700	0.875	0.975
<i>c</i>	0.000	0.000	0.000	0.050	0.000	0.000	0.042	0.063	0.100	0.300	0.125	0.025
<b>Pgi-2</b>												
<i>b</i>	0.000	0.000	0.000	0.200	0.025	0.100	0.000	0.063	0.075	0.200	0.150	0.175
<i>c</i>	1.000	1.000	1.000	0.800	0.975	0.900	1.000	0.938	0.925	0.800	0.850	0.825
<b>Pgm-2</b>												
<i>a</i>	0.000	0.000	0.025	0.025	0.025	0.000	0.042	0.188	0.000	0.000	0.000	0.000
<i>b</i>	0.950	0.975	0.850	0.925	0.975	1.000	0.917	0.813	1.000	0.950	0.875	0.975
<i>c</i>	0.050	0.025	0.125	0.050	0.000	0.000	0.042	0.000	0.000	0.050	0.125	0.025

**Fig. 4.18** Frequency of the *Acp-2 a* allozyme allele in populations along transect S, through the postulated hybrid zone between *Senecio chrysanthemifolius* and *S. aethnensis*, on Mt. Etna; plotted against (a) population altitude (m), and (b) cumulative distance of each population from the start of the transect (km).



**Fig. 4.19** Frequency of the *Pgi-2 b* allozyme allele in populations along transect S, through the postulated hybrid zone between *Senecio chrysanthemifolius* and *S. acthensis*, on Mt. Etna; plotted against (a) population altitude (m), and (b) cumulative distance of each population from the start of the transect (km).



Levels of allozyme diversity varied across the populations in the transect (table 4.11). Lower levels of diversity were present in populations of 'pure' *S. chrysanthemifolius* (C0 and C1) compared to populations of 'pure' *S. aethnensis* (VB and BB), as shown by mean values of  $A$  (mean number of alleles per locus),  $P$  (percentage of polymorphic loci) and  $H_e$  (mean unbiased expected heterozygosity). Population RD had the highest level of diversity, followed by populations S10, S5 and S11. Populations S6, S8 and S9 showed levels of diversity that were similar to those of the 'pure' *S. aethnensis* populations, while population S2 exhibited a low level of genetic diversity compared to all populations, except those of 'pure' *S. chrysanthemifolius*. The coefficient of inbreeding ( $f$ ) was low in all populations and there was no significant deviation from Hardy-Weinberg equilibrium, indicating that random mating was taking place at all points in the transect.

**Table 4.11** Estimates of allozyme variability and Weir and Cockerham's (1984) coefficient of inbreeding ( $f = Fis$ ) for the 12 populations along transect S, through the postulated hybrid zone between *S. chrysanthemifolius* and *S. aethnensis*, on Mt. Etna. Hardy-Weinberg equilibrium across loci within samples (Fisher's method) was tested.

Population	Sample size <sup>1</sup>	$A$ <sup>2†</sup>	$P$ <sup>3</sup>	$H_e$ <sup>4†</sup>	$f$ <sup>5‡</sup>
C0 <i>S. chrysanthemifolius</i>	20	1.8 (0.2)	83.3	0.057 (0.015)	0.280
C1 <i>S. chrysanthemifolius</i>	20	1.5 (0.2)	50.0	0.033 (0.016)	-0.013
	mean	1.65	66.5	0.045	
S2	20	1.5 (0.3)	33.3	0.075 (0.049)	-0.108
S5	20	2.0 (0.3)	83.3	0.198 (0.057)	0.077
S6	20	1.8 (0.2)	83.3	0.132 (0.051)	0.122
S8	20	1.8 (0.2)	83.3	0.103 (0.025)	0.114
S9	12	2.0 (0.3)	83.3	0.161 (0.054)	0.051
S10	8	2.0 (0.0)	100	0.221 (0.065)	0.355
S11	20	1.8 (0.2)	83.3	0.196 (0.064)	-0.020
RD	20	2.2 (0.2)	100	0.265 (0.058)	-0.168
BB <i>S. aethnensis</i>	20	1.8 (0.2)	83.3	0.186 (0.041)	0.151
VB <i>S. aethnensis</i>	20	1.8 (0.2)	83.3	0.098 (0.044)	-0.108
	mean	1.8	83.3	0.142	

<sup>†</sup> Standard errors in parentheses.

<sup>1</sup> All loci had the sample size indicated.

<sup>2</sup> Mean number of alleles per locus.

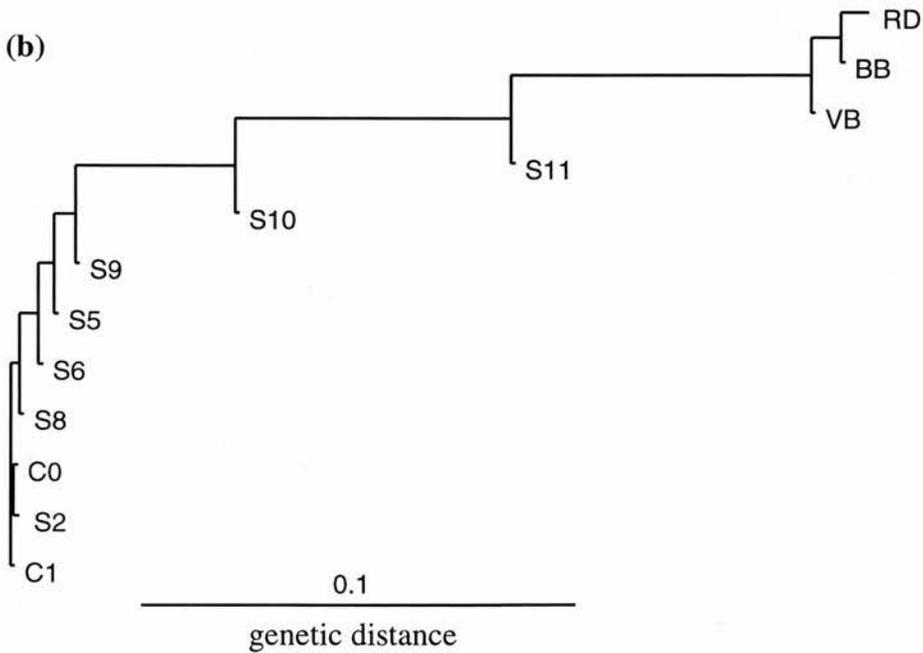
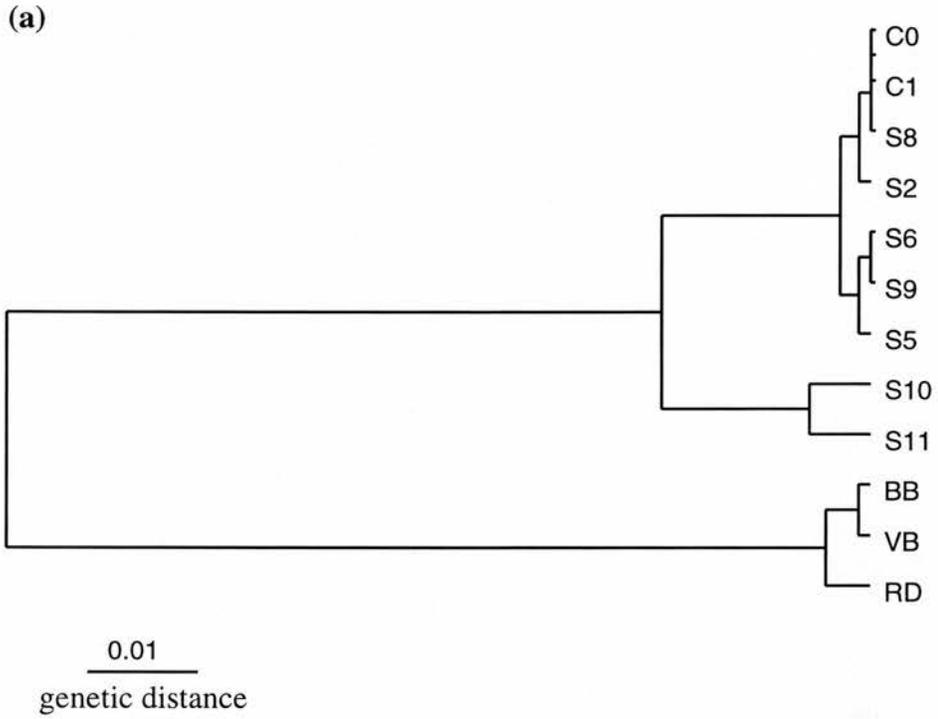
<sup>3</sup> Percentage of polymorphic loci (this number is artificially inflated, because two monomorphic loci were not scored, see section 4.3.3).

<sup>4</sup> Mean expected heterozygosity (unbiased estimate).

<sup>5</sup> Weir and Cockerham's (1984) coefficient of inbreeding,  $f = Fis$ .

<sup>‡</sup> Hardy-Weinberg equilibrium was accepted for all populations at  $P < 0.01$ .

**Fig. 4.20** (a) UPGMA, and (b) neighbour-joining dendrograms of populations along transect S, through the postulated hybrid zone between *Senecio chrysanthemifolius* and *S. aethnensis*, on Mt. Etna. The clustering was based upon Nei's genetic distances between populations, derived from allele frequencies at 6 polymorphic isozyme loci.



The mean level of subdivision (table 4.12) between the transect populations ( $\theta = F_{st}$ ) over all loci was significantly different from zero (as shown by bootstrap estimates of the 95% confidence intervals). This indicated that there was significant population differentiation within transect S on Mt. Etna. The mean coefficient of inbreeding ( $f = F_{is}$ ) across transect populations over all loci (table 4.12) was not significantly different from zero (as shown by the bootstrap estimates of the 95% confidence intervals).

The UPGMA and neighbour-joining trees produced from Nei's genetic distances between populations, based on allele frequencies at the six polymorphic loci, are shown in fig. 4.20. Overall, the genetic distances between populations were low.

The UPGMA tree separated populations into two main clusters (fig. 4.20 a). One cluster consisted of both populations of 'pure' *S. aethnensis* (VB and BB) and population RD. Within the second cluster, two sub-clusters were apparent. One sub-cluster comprised populations S10 and S11. The other sub-cluster contained populations S9, S6 and S5 in one group and the two populations of 'pure' *S. chrysanthemifolius* (C0 and C1), S8 and S2 in another group.

The neighbour-joining tree showed a rather different structure (fig. 4.20 b). The two populations of 'pure' *S. chrysanthemifolius* (C0 and C1) clustered with population S2. Populations S8, S6, S5 and S9 clustered progressively further away from C0, C1 and S2. From a node close to population S9, a long branch extended to populations S10 and S11. From population S11, the branch continued to a tight cluster containing population RD and both populations of 'pure' *S. aethnensis* (VB and BB).

**Table 4.12** Weir and Cockerham's (1984) estimators of Wright's  $F$ -statistics calculated separately for each locus for the 12 populations from the transect through the postulated hybrid zone, between *S. chrysanthemifolius* and *S. aethnensis*, on Mt. Etna. Locus specific means and standard deviations were obtained by jackknifing over populations. Overall means and standard deviations were estimated by jackknifing over loci. 95% Confidence intervals were estimated by bootstrapping over loci.

Locus	$F^{1\ddagger}$	$\theta^{2\ddagger}$	$f^{3\ddagger}$
<i>Aat-3</i>	0.101 (0.055)	0.020 (0.015)	0.082 (0.058)
<i>Aco-1</i>	-0.058 (0.017)	0.046 (0.035)	-0.106 (0.046)
<i>Acp-2</i>	0.740 (0.114)	0.719 (0.116)	0.068 (0.120)
<i>Idh-1</i>	0.111 (0.177)	0.122 (0.086)	0.011 (0.256)
<i>Pgi-2</i>	0.084 (0.065)	0.056 (0.023)	0.029 (0.065)
<i>Pgm-2</i>	0.029 (0.088)	0.030 (0.016)	-0.001 (0.082)
Mean	0.417 (0.256)	0.392 (0.257)	0.038 (0.025)
95% Confidence interval	0.029-0.590	0.033-0.570	-0.019-0.072

<sup>†</sup> Standard deviations in parentheses.

<sup>1</sup>  $F = F_{it}$  (within total).

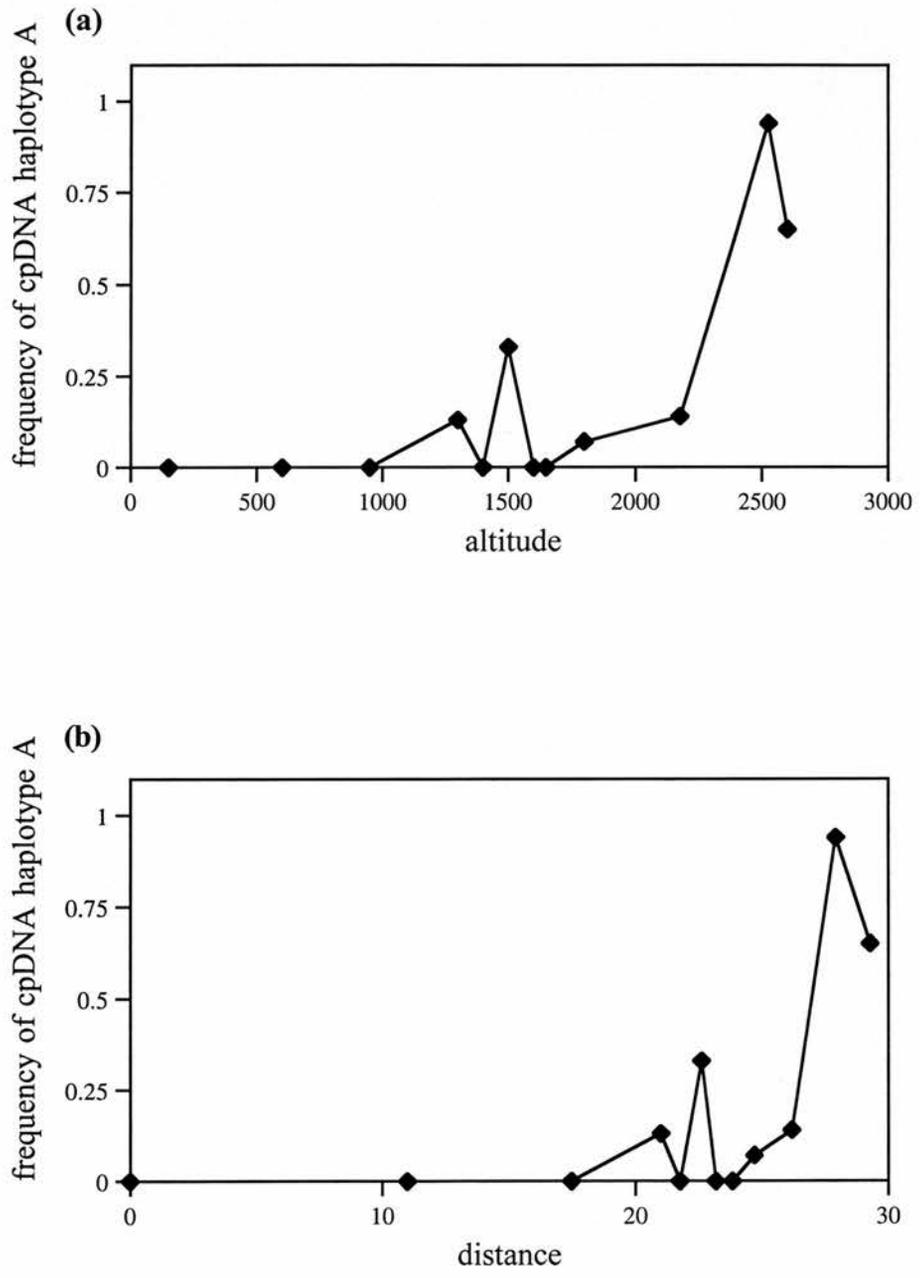
<sup>2</sup>  $\theta = F_{st}$  (among populations).

<sup>3</sup>  $f = F_{is}$  (within populations).

#### 4.3.4 Chloroplast DNA RFLP variation along transect S, through the postulated hybrid zone between *S. chrysanthemifolius* and *S. aethnensis*, on Mt. Etna

Haplotype A was absent from populations along the transect until population S5 at 1300 m (21 km from the start of the transect) (fig. 4.21). Between population S5 and population RD (at 2175 m or 26.2 km from the start of the transect) the frequency of haplotype A remained low, ranging from 0 to 0.33. After population RD, the frequency of haplotype A increased considerably to 0.94 in population BB and 0.65 in population VB (at 2600 m or 29.3 km from the start of the transect). The frequency of haplotype B correspondingly decreased with increasing altitude and distance from the start of the transect.

**Fig. 4.21** Frequency of cpDNA haplotype A in populations along transect S, through the postulated hybrid zone between *Senecio chrysanthemifolius* and *S. aethnensis*, on Mt. Etna; plotted against (a) population altitude (m), and (b) cumulative distance of each population from the start of the transect (km).



## 4.4 Discussion

### 4.4.1 Evidence for a hybrid zone between *S. chrysanthemifolius* and *S. aethnensis* on Mt. Etna

"Hybrid zones occur when genetically distinct groups of individuals meet and mate, resulting in at least some offspring of mixed ancestry" (Harrison, 1993).

Molecular evidence supports the hypothesis that there is a hybrid zone between *S. chrysanthemifolius* and *S. aethnensis* on Mt. Etna. RAPD, ISSR and isozyme loci exhibit the clinal variation that is characteristic of hybrid zones (Barton and Hewitt, 1985). In addition RAPD, ISSR and isozyme markers enabled individuals that are hybrids, that is of mixed ancestry, to be identified (from their additive profile of parental nuclear genetic markers).

Crisp (1972) and Abbott (Abbott *et al.*, 2000) made the implicit assumption that the postulated hybrid zone on Mt. Etna was a zone of secondary contact (i.e. that *S. chrysanthemifolius* and *S. aethnensis* were formerly allopatric). However, Crisp admitted "*another possibility exists: that the two taxa [S. chrysanthemifolius and S. aethnensis] are in the process of speciation from a single taxon*". The question must therefore be asked: is the hybrid zone on Mt. Etna primary or secondary in origin?

### 4.4.2 Primary or secondary intergradation

Distinguishing between primary and secondary intergradation is extremely difficult (see section 1.2.1.1). Regrettably, there is no historical information concerning the hybrid zone between *S. chrysanthemifolius* and *S. aethnensis*. This means that conclusions must be based upon circumstantial evidence drawn from phylogenetics, ecology and the type of clinal variation (no biogeographic information was available).

If the hybrid zone on Mt. Etna was primary in origin one would expect the molecular marker clines to be scattered (1.2.1.1). Therefore, the large number of coincident clines observed in the present study argues strongly for secondary contact (Futuyma, 1998). Phylogenetic data are helpful. Although *S. chrysanthemifolius* and *S. aethnensis* appear to be closely related, they are nevertheless morphologically and genetically quite distinct (chapter 3; table 4.1). Abbott *et al.* (2000) argued "*such nuclear and cytoplasmic*

*differentiation between these two species indicates that they are likely to have originated in isolation and recently come into contact*". Unfortunately, the phylogenetic approach of Harrison (1990; see section 1.2.1.1) could not be used because only one transect through the hybrid zone was examined with molecular markers. Ecological information also suggests that the hybrid zone on Mt. Etna is due to secondary rather than primary intergradation. Firstly, the hybrid zone appears to be associated with disturbance. Crisp (1972) suggested that the natural deciduous woodland barrier separating *S. chrysanthemifolius* and *S. aethnensis* may have been removed by man, through increased grazing of livestock. A recent review of vegetation change on Mt. Etna confirms that loss of woodland from the mountain has been considerable, and attributes this to destructive lava flows, and tree removal for timber and agricultural land gain (Poli Marchese, 1999). The effect of road-building on Mt. Etna may also have been significant, producing corridors of dispersal for *S. chrysanthemifolius*, *S. aethnensis* and hybrids (personal observation). Secondly, within the distribution of *S. chrysanthemifolius* (across Sicily and into southern Calabria) there are a number of mountainous regions. If primary intergradation was occurring on Mt. Etna one might expect to see some degree of altitudinal differentiation of *S. chrysanthemifolius* on these other mountains. However, there are no records of altitudinal differentiation of *S. chrysanthemifolius* anywhere apart from Mt. Etna, where *S. aethnensis* is endemic (Poli Marchese, 1991).

In summary, it is impossible to establish with absolute certainty the origin of the hybrid zone on Mt. Etna. Circumstantial evidence does, however, indicate the hybrid zone was formed by secondary contact between *S. chrysanthemifolius* and *S. aethnensis*. This inability to state with confidence whether a hybrid zone is primary or secondary is not unusual. Indeed, most authors have had to resort to speculation (e.g. Millar, 1983; Perron and Bousquet, 1997) when making decisions of this kind.

#### **4.4.3 Ancestral polymorphism**

Individuals from certain populations, notably S10 (altitude 1650 m) and S11 (altitude 1800 m), on Mt. Etna had additive profiles of parental nuclear genetic (RAPD and ISSR) markers, relative to 'pure' *S. chrysanthemifolius* and *S. aethnensis*. The genetic additivity was interpreted as evidence of hybridity but would also be expected if the populations were ancestral to *S. chrysanthemifolius* and *S. aethnensis*, with the apparent parental species

derived by phylogenetic sorting (Dawson *et al.*, 1996). However, this suggestion of plesiomorphy can be rejected on several grounds. If the putative hybrids were ancestral to *S. chrysanthemifolius* and *S. aethnensis*, they would be expected to possess unique markers and possibly an ancestral organelle haplotype (Allan *et al.*, 1997; Rieseberg *et al.*, 1990b). However, the putative hybrids exhibited no unique RAPD or ISSR markers and shared the cpDNA haplotypes of *S. chrysanthemifolius* and *S. aethnensis*. Moreover, the principal coordinate analysis of RAPD and ISSR data revealed that individuals from populations S10 and S11 had a wide range of multilocus phenotypes (see below) and were, therefore, not likely to be in linkage equilibrium. Dawson *et al.* (1996) argued that pronounced linkage disequilibrium was not to be anticipated in ancestral populations.

#### 4.4.4 Single locus variation

The RAPD and ISSR markers, and also the *Acp* and *Pgi* allozyme markers, examined in the current study occurred at high frequency in one parental species and low frequency in the other parental species. Therefore, assuming that these markers are selectively neutral, and also the existence of a non-mosaic hybrid zone (see below), markers should have decreased or increased in frequency, across the transect, in a gradual fashion (fig. 15.20, Futuyma, 1998). However, a majority of markers did not exhibit this pattern, instead they showed a sharp increase or decrease in frequency at some point along the transect. This implies that selection is acting on many of the RAPD, ISSR and allozyme loci (Futuyma, 1998). How can this paradox, of 'neutral' markers under selection, be resolved? It is possible that some of the RAPD, ISSR and isozyme loci were under direct selection (i.e. not neutral) but it is likely that most of the RAPD, ISSR and allozyme loci were tightly linked to loci subject to selection (Futuyma, 1998). The assumption that 'neutral' markers are linked to loci under selection has been made in many hybrid zone studies (e.g. Szymura and Barton, 1991).

Most RAPD and ISSR markers, and also the *Acp-2 a* and *b* allozyme alleles, showed smooth clinal variation in parts of the hybrid zone where changes in frequency occurred. This indicates that the hybrid zone was not mosaic-like in structure. If a hybrid zone is a mosaic (see section 1.2.2.4) one would expect to observe "a patchy distribution of different genotypes" (Arnold, 1997), i.e. many markers would have exhibited major reversals in frequency along the transect. For example, Shoemaker *et al.* (1996) concluded that the fire ant hybrid zone they were investigating was a mosaic because, although the overall pattern

was, at first glance, clinal there were "*significant reversals in the transitions of... genetic markers...*" along the transects studied rather than "*persistent changeovers*". Howard *et al.* (1997) investigated the hybrid zone between two oak species along an altitudinal transect. They concluded that the hybrid zone was a mosaic because there was not a gradual transition from one oak species to the other oak species as they moved through the transect, instead the distribution of the two oak taxa and their hybrids was "*decidedly patchy*".

Clinal variation of markers can be used to provide information on the role of selection and dispersal in the hybrid zone. If the hybrid zone is maintained by selection alone, one would expect many clines to be scattered, as selection would not act on each locus at the same place along the transect. In fact, many of the *S. chrysanthemifolius* RAPD and ISSR marker clines, and also the *Acp-2 a* cline, were broadly coincident, as were many of the *S. aethnensis* marker clines and the *Acp-2 b* cline. Furthermore, the major decrease in the frequency of many *S. chrysanthemifolius* markers coincided with the major increase in frequency of many *S. aethnensis* markers. Coincidence of clines implies that the hybrid zone is maintained by a balance between selection and dispersal (Barton and Hewitt, 1985). Assuming this is so, one can ask whether the clines indicate the form of selection acting (in conjunction with dispersal): i.e. intrinsic (environment-independent) selection against hybrids or environment-dependent selection (different genes being favoured in different places). Barton and Hewitt (1985) argue that if the form of selection is environment-dependent, clines should be more scattered than if intrinsic selection is acting against hybrids. They also suggest that if selection is environment-dependent, clines from different transects across the same hybrid zone will be less similar to each other (than if environment-independent selection is acting against hybrids) because environmental conditions will vary from place to place through the hybrid zone. However, it is not easy to use these criteria to judge a hybrid zone and in the present study only one transect through the hybrid zone was investigated. Thus, although the general coincidence of many marker clines suggests that dispersal and selection are both important in the hybrid zone on Mt. Etna, the clines alone cannot be used to predict the type of selection that is at work.

Regardless of the type of selection, the strength of selection can be calculated from a selection-dispersal cline (Barton and Gale, 1993). According to single-locus models, the width of a cline is proportional to the ratio between dispersal and the square root of

selection. Therefore, if one knows the width of the cline and dispersal rates of the organisms involved, the strength of selection can be calculated. In rough terms, a narrow cline, relative to dispersal rates, indicates strong selection but a wide cline, relative to dispersal rates, indicates weaker selection.

Unfortunately, as Barton and Gale admit, it is not normally practicable to compute the strength of selection because of the difficulties in obtaining accurate figures for dispersal rates. Barton and Hewitt (1985) believed that dispersal rates for taxa in hybrid zones were often grossly underestimated. Furthermore, most of the hybrid zones from which Barton and Hewitt drew this conclusion were animal hybrid zones. To estimate the dispersal rate of an animal (e.g. by the capture-mark-recapture method) is fairly difficult. To obtain an accurate figure for the dispersal rate of the *Senecio* species in this hybrid zone would be exceedingly problematic as one would have to estimate the movement of seed by wind and of pollen by insects, both of which are enormously variable processes and liable to be affected by a large number of variables e.g. local topology and weather conditions. For this reason, it was felt that any attempt to calculate the strength of selection would be of little value, indeed little better than a guess. However, one might speculate that the strength of selection is fairly high in the hybrid zone on Mt. Etna, because cline width is likely to be narrow relative to the potential dispersal rate of *Senecio* species.

#### **4.4.5 Hybrid indices and principal coordinate analysis**

Hybrid index analysis and principal coordinate analysis suggested that there was a progression from 'pure' *S. chrysanthemifolius*, such as populations C0 and C1, at low altitudes to 'pure' *S. aethnensis*, such as populations VB and BB, at high altitudes, via a number of introgressed and hybrid populations (S2, S5, S6, S8, S9, S10, S11) along transect S on Mt. Etna.

Hybrid indices and principal coordinate analysis indicated that populations S2, S5, S6, S8 and S9 (altitude 950 m to 1600 m) were genetically similar to, although distinct from, populations of 'pure' *S. chrysanthemifolius* (populations C0 and C1). With increasing altitude, there was a tendency for populations S2, S5, S6, S8 and S9 to have fewer individuals that were genetically indistinguishable from 'pure' *S. chrysanthemifolius* individuals and to possess more individuals that were verging towards intermediacy between the two parent species. In other words, populations S2, S5, S6, S8 and S9 were *S.*

*chrysanthemifolius*-like populations but subjected to an increasing degree of introgression from *S. aethnensis*. Thus, population S2 (altitude 950 m) comprised mainly 'pure' *S. chrysanthemifolius* individuals with a small number of introgressed *S. chrysanthemifolius* individuals, while population S9 (altitude 1600 m) contained no 'pure' *S. chrysanthemifolius* individuals, only introgressed *S. chrysanthemifolius* individuals.

The analyses indicated that populations S10 and S11 were genetically diverse populations and included individuals that were genetically intermediate between *S. chrysanthemifolius* and *S. aethnensis*. Thus, populations S10 (altitude 1650 m) and S11 (1800 m) were mixed hybrid populations; that is, populations containing a variety of parental, introgressed and intermediate types. Population S10 comprised 'pure' *S. chrysanthemifolius*, introgressed *S. chrysanthemifolius* and individuals intermediate between the two parental species, while population S11 comprised introgressed *S. aethnensis*, introgressed *S. chrysanthemifolius*, and intermediate individuals.

Analysis indicated that population RD, from the upper altitudinal range of the transect (2175 m), was genetically similar to populations of 'pure' *S. aethnensis*. Thus, population RD was a *S. aethnensis*-like population that had been subject to a degree of introgression from *S. chrysanthemifolius*, and contained both introgressed *S. aethnensis* and 'pure' *S. aethnensis* individuals.

Finally, analysis indicated that F<sub>1</sub> individuals were intermediate between the parental species and that their genetic composition was similar to certain hybrid individuals from populations S9, S10 and S11.

Other similar studies have interpreted hybrid index and PCO results in a comparable manner to the present study. Dawson *et al.* (1996) assessed a hybrid zone between two tropical tree species *Gliricidia sepium* and *G. maculata*. Principal coordinate analysis of RAPD data produced two well-separated groups that defined the species. However, some individuals, from three populations, occupied intermediate positions in the PCO and were interpreted as hybrids. Dawson *et al.* (1996) combined the RAPD data with two PCR-RFLP markers to construct a 'nuclear' hybrid index that ranged from one for 'pure' *G. sepium* individuals to zero for 'pure' *G. maculata* individuals. The three hybrid populations (as identified by the principal coordinate analysis) had mean 'nuclear' index values of 0.61,

0.65 and 0.35. Dawson *et al.* (1996) suggested that populations with values of 0.61 and 0.65 were intermediate hybrid populations, whereas the population with the value of 0.35 was formed by back-crossing to *G. maculata* following inter-specific hybridisation. Howard *et al.* (1997) employed a hybrid index to estimate the amount of introgression between two oak species in North America. Isolated populations of *Quercus grisea* had character index scores -6 to -2 and isolated populations of *Quercus gambelii* +5 to +8. One population, in the area of sympatry between the species, contained individuals with character index scores ranging from -6 to +8, with a majority of scores between -2 and +3. Howard *et al.* (1997) concluded that this population was a mixed hybrid population including parental and intermediate types

Clustering procedures, such as UPGMA and neighbour-joining, assume hierarchical rather than reticulate patterns of evolution (Quicke, 1993). Therefore, one might predict that phenograms would be unsuitable for hybrid zone data and would misrepresent relationships among hybrids and parents. However, the relationships depicted in the neighbour-joining tree constructed for individuals in transect S (and the tree constructed for individuals in transect S plus the F<sub>1</sub>s) were similar to those depicted by principal coordinate analysis. Nevertheless, ordination techniques, such as principal coordinate analysis, remain the best way of dealing with hybrid zone data because they are non-hierarchical and make no *a priori* assumptions about the existence or otherwise of groupings among the individuals (Quicke, 1993).

#### **4.4.6 Population level analyses**

The results of the AMOVA and Shannon's diversity index, for the 12 populations along the transect, showed that the total genetic variation was partitioned fairly evenly into within and between population components. In an out-crossing plant species, such as *S. chrysanthemifolius* or *S. aethnensis*, one would expect most variation to be maintained within populations (Hamrick and Godt, 1990). However, in the present study, the analyses were carried out on populations of *S. chrysanthemifolius*, *S. aethnensis* and their hybrids and, consequently, a greater amount of genetic variation was partitioned between populations than would have been expected if looking at either parental species alone (Gillies *et al.*, 1997).

The level of genetic diversity, based upon RAPD and ISSR variation, within each population along the transect was quantified using Shannon's diversity index. Not surprisingly, hybrid populations maintained higher levels of genetic diversity than 'pure' populations of *S. chrysanthemifolius* or *S. aethnensis*, because hybrid individuals combined markers from both parental species. The population with the highest level of genetic diversity was population S11 - which was also the most genetically diverse population according to the PCO. The level of genetic diversity in each population based on Shannon's index was compared to the level of genetic diversity in each population according to the various measures based on allozyme loci (expected heterozygosity, etc.). There was broad agreement between RAPD and ISSR, and allozyme results; both showed that the parental species had lower levels of genetic diversity than most of the hybrid populations. However there were also differences between data sets. For example, according to allozyme data, population RD, rather than population S11, had the highest level of diversity and 'pure' *S. aethnensis* was more diverse than 'pure' *S. chrysanthemifolius*. Disagreement between RAPD and ISSR, and allozyme results can be attributed to the difference in the number of loci sampled (26 RAPD and ISSR loci versus 6 isozyme loci) and the manner in which the loci were selected (the 26 RAPD and ISSR loci were chosen because they were present at high frequency in one parental species and low frequency in the other parental species but only 2 of the 6 isozyme loci were chosen for this reason, the remaining 4 loci were picked at random).

Most populations along transect S were significantly genetically different from each other based upon pair-wise  $\Phi_{ST}$  values from the AMOVA. This result concurred with that from *F*-statistics analysis of isozyme data, which indicated that there was significant population differentiation within the transect. A neighbour-joining phenogram constructed from the pair-wise  $\Phi_{ST}$  values proved to be most informative. It showed that, at the population level, the conclusions drawn from hybrid indices and principal coordinate analysis were broadly supported, i.e. relationships between the various populations were the same. Thus, population RD was a *S. aethnensis*-like hybrid population, populations S2, S5, S6 and S8 were *S. chrysanthemifolius*-like hybrid populations and populations S10 and S11 were genetically diverse populations. The neighbour-joining phenogram constructed from allozyme data (based upon Nei's genetic distances between populations) was similar to the phenogram based upon pair-wise  $\Phi_{ST}$  values, i.e. RAPD and ISSR data. Disagreement

between the two phenograms can be attributed to a difference in the number of loci sampled and the way in which loci were selected for analysis (see above).

#### **4.4.7 Maintenance of the hybrid zone**

The hybrid zone, between *S. chrysanthemifolius* and *S. aethnensis*, on Mt. Etna, has been documented for over 30 years (from the late 1960's onwards; Crisp, 1972) and is likely to be of relatively ancient origin (see chapter 5). Due to its longevity, the hybrid zone may be thought of as stable. Various models have been proposed that try to account for the existence of stable hybrid zones, including: the bounded hybrid superiority model, mosaic hybrid zone model, environment-dependent selection and dispersal model, and the tension zone model (see section 1.2.2). It is of interest to see which one of these models might apply to the hybrid zone on Mt. Etna.

The bounded hybrid superiority model and mosaic hybrid zone model can be rejected fairly decisively as descriptors for the maintenance of the hybrid zone on Mt. Etna. The single locus variation recorded was incompatible with both of these models. The smooth clinal variation observed rules out the possibility of a mosaic hybrid zone (see above). That many of the clines were coincident is inconsistent with the bounded hybrid superiority model, which proposes that hybrid zones are maintained by selection alone (see above). The bounded hybrid superiority model can be further discarded because the model assumes that the effect of dispersal is negligible. This cannot be a sensible assumption in this instance; wind-distributed *Senecio* seed may be carried far from their origin and insects carrying pollen may also travel considerable distances. Moreover, if the effect of dispersal is negligible, marker clines should be broad in comparison to the potential dispersal distance of the organisms (Barton and Hewitt, 1985) and this is clearly not the case with regards to the hybrid zone on Mt. Etna.

Coincidence of clines suggests that a balance between selection and dispersal maintains the hybrid zone on Mt. Etna (see above). This finding leads to a choice between two hybrid zone models. Thus, the hybrid zone on Mt. Etna may be a tension zone, maintained by a balance between dispersal and intrinsic selection against hybrids between *S. chrysanthemifolius* and *S. aethnensis*. Alternatively, the hybrid zone on Mt. Etna may be maintained by a balance between dispersal and environment-dependent selection.

Unfortunately, it is difficult to differentiate between a tension zone and a hybrid zone maintained by environment-dependent selection and dispersal (Barton and Gale, 1993). With this proviso in mind, is it possible to determine which of these two models is more appropriate to the situation on Mt. Etna?

Barton and Hewitt (1985) claim to be able to distinguish between the two selection-dispersal models by careful examination of patterns of clinal variation along a transect and by comparing clines from different transects across the same hybrid zone. However, as previously mentioned, Barton and Hewitt's criteria are not easy to use and only one transect was studied in this hybrid zone. Indeed, many would argue that it is extremely difficult to distinguish between the two selection-dispersal models by looking at patterns of clinal variation alone (Barton and Gale, 1993; Arnold, 1997). Therefore, we need to employ other sources of information to discriminate between the two models.

If the tension zone model is applicable to the hybrid zone on Mt. Etna, there should be strong evidence of intrinsic (environment-independent) selection against the hybrid offspring of *S. chrysanthemifolius* and *S. aethnensis*. Consequently, all classes of hybrids between *S. chrysanthemifolius* and *S. aethnensis* should be uniformly unfit (i.e. exhibit reduced viability and/or fertility) relative to 'pure' *S. chrysanthemifolius* and *S. aethnensis*, across all environments. In the greenhouse, F<sub>1</sub> hybrids between *S. chrysanthemifolius* and *S. aethnensis* have been germinated, cultivated, and successfully crossed amongst themselves and to their parental taxa. F<sub>2</sub> and backcross hybrids have also been germinated and grown, and F<sub>2</sub>s, have been successfully crossed amongst themselves (Forbes, personal communication; personal observation). There is no indication that survivorship (as measured in terms of percentage germination and survival of seedlings to maturity) is lower among F<sub>1</sub>, F<sub>2</sub> or backcross individuals than among 'pure' *S. chrysanthemifolius* or *S. aethnensis* individuals (Forbes, personal communication; personal observation). There is, however, a suggestion that some F<sub>1</sub> hybrids may exhibit lower fertility (reduced seed set) than 'pure' representatives of their parental taxa (Forbes, personal communication). Information about the fitness of later generation hybrids is even more sparse, although later generation hybrids, grown in the greenhouse from seed collected in natural populations, do not appear to be less viable than 'pure' *S. chrysanthemifolius* or *S. aethnensis* (personal observation). Crisp (1972) investigated the potential fertility of hybrids between *S.*

*chrysanthemifolius* and *S. aethnensis*, by looking for signs of meiotic disturbance in pollen mother cells and staining pollen to estimate levels of pollen sterility. Based upon his sample of around 30 natural hybrids (the progeny of seed from Sicily) and ten F<sub>1</sub> plants (the product of just a single cross between *S. chrysanthemifolius* and *S. aethnensis*), Crisp concluded that there was no evidence to suggest that hybrids had substantially reduced potential fertility relative to their parents. Thus, while the possibility of intrinsic selection against some classes of hybrids between *S. chrysanthemifolius* and *S. aethnensis* cannot be ruled out, there is no suggestion that hybrids are uniformly unfit across all environments. Consequently, the hybrid zone on Mt. Etna does not appear to fulfil a main requirement of the tension zone model. However, this inference is highly speculative (as the evidence upon which it is based is anecdotal) and much more detailed analysis of hybrid fitness is required before a firm conclusion is drawn on the point.

To properly assess whether hybrids are uniformly unfit across all environments, a careful, quantitative study of comparative fitness under a range of controlled conditions is required. Burke *et al.* (1998b) examined components of fitness in *Iris hexagona*, *I. fulva* and their reciprocal F<sub>1</sub> hybrids under greenhouse conditions. Seed germination, vegetative growth, clonal reproduction and sexual reproduction were measured and compared between the cross types. Hybrids outperformed one or both parental species for all four fitness components and the composite fitness of the F<sub>1</sub>s was greater than either parent. Burke *et al.* (1998b) cautioned that their results did not necessarily reflect lifetime fitness of Irises; nevertheless, they concluded that their findings "*clearly contradict the assumption of environment-independent hybrid inferiority*" that characterises the tension zone model. Comprehensive studies such as this are, unfortunately, all too rare. For example, the hybrid zone between two species of *Bombina* is held to be classic example of a tension zone (Barton and Hewitt, 1985), but Arnold (1997) argued that environment-independent inferiority of *Bombina* hybrids had yet to be adequately demonstrated.

If the environment-dependent selection/dispersal model is applicable to the hybrid zone on Mt. Etna, the hybrid zone should occur along an environmental gradient or gradients. Conditions on Mt. Etna are extremely divergent. For example, at the base of Mt. Etna, the climate is Mediterranean, habitats are often urban or agricultural, soils are well-developed and many plant species co-exist (Poli Marchese, 1991). In contrast, the summit of Mt. Etna experiences an alpine climate (sporting a cap of snow throughout the winter months) and

habitats are volcanic (frequent lava flows, sulphurous, etc.) and open (Poli Marchese, 1991). Thus, a large number of environmental gradients can be hypothesised as coinciding with the altitudinal hybrid zone, including: temperature, snow cover, wind speed, fog, frost, soil moisture, substrate-type, edaphic sulphur content, UV-light exposure, disturbance and inter-specific competition. There is a good basis for the existence of some of these environmental gradients. For example, a decrease in temperature and increase in snow cover, frost-risk, wind-speed and UV light exposure, with ascending altitude are well-documented phenomena (Fitter and Hay, 1981). Other environmental gradients are more speculative. For example, one can suggest that plants from higher elevations on Mt. Etna will be more disturbed than those at lower elevations due to lava flows and the action of freeze-thaw. Unfortunately, detailed altitudinal-environmental data for the Mt. Etna area (based upon regular sampling points along an altitudinal transect) are not available (Poli Marchese, 1999). If such data were to be obtained, a multiple linear regression analysis could be performed to see whether any of the hypothesised environmental gradients are closely correlated with the hybrid zone, i.e. which environmental variables explain the most variation in hybrid index scores (Moore, 1977; Gillies *et al.*, 1999). More work with respect to environmental gradients on Mt. Etna is clearly required.

Moreover, it is not enough to show that there are environmental gradients that coincide with the hybrid zone. One must demonstrate a causal relationship between the hybrid zone and the environmental gradient(s), i.e. provide evidence of selection in response to the differing environmental conditions on Mt. Etna. It is easy to imagine that genes that increase fitness, and are therefore selected, in one parental habitat on Mt. Etna might gradually decrease fitness and be progressively selected against as one moves through the hybrid zone and into the other parental habitat. For example, an adaptation that improved survival under snow would be favoured in plants at higher altitudes on Mt. Etna, but the adaptation might become costly to maintain or even positively disadvantageous at lower altitudes.

A number of different approaches can be used to assess genotype-environment associations. Firstly, 'greenhouse studies' can be employed. For example, based on field observation, a number of biotic and abiotic factors have been suggested as major determinants of the distribution of Louisiana *Iris* species and their hybrids (Arnold and Bennett, 1993). Greenhouse trials were used to examine the effects of shade tolerance,

salinity tolerance and inter-specific competitive ability on the relative fitness of *Iris fulva*, *I. hexagona* and two classes of natural hybrids (Bennett and Grace, 1990; Arnold and Bennett, 1993). The authors found that all three variables had measurable effects on the performance of the parental taxa and their hybrids. However, greenhouse studies are, by their nature, limited. Although, they may provide an indication of the comparative importance of various environmental variables, one cannot extrapolate the findings into the natural habitats of the parents and hybrids. Secondly, the fitness of hybrid and parental genotypes in their native habitats can be estimated using cohort analysis of natural populations (Emms and Arnold, 1997). For example, Bennett (1989; in Arnold and Bennett, 1993) used a three-year demographic study of natural populations of *Iris hexagona*, *I. fulva* and their hybrids to estimate relative fitness. However, this method has several disadvantages. Practically, it can be hard to locate suitable plots to work with and correctly identify the various genotypes (Emms and Arnold, 1997). Moreover, natural cohort analysis reveals nothing about the relative fitness of hybrids and parents other than in their native habitats.

Perhaps the best method for assessing genotype-environment associations is to use a reciprocal transplant experiment. In such an experiment, the fitness of parents and/or hybrids in their native habitats can be compared to their fitness in the alternate habitats. If the environment-dependent selection/dispersal model is appropriate for a hybrid zone, each genotype should do best in its native habitat. Reciprocal transplant experiments have been carried out in a number of hybrid zones. For example, Emms and Arnold (1997) conducted reciprocal transplant experiments with rhizomes of *Iris fulva*, *I. hexagona* and F<sub>1</sub> and F<sub>2</sub> hybrids. Their results indicated that early generation hybrids between *I. hexagona* and *I. fulva* can survive and reproduce clonally at least as well as their parents, in both hybrid and parental habitats. Young (1996) investigated the hybrid zone between *Iris douglasiana* and *I. innominata*. Reciprocal transplant experiments were performed with individuals of *I. douglasiana* and *I. innominata* (they neglected to use hybrids) at two life history stages - seed and adult. Adult plants survived much better in their own species' habitat, suggesting that habitat association plays an important role in structuring the hybrid zone. Wang *et al.* (1997) conducted reciprocal transplant experiments with two sub-species of big sagebrush (*Artemisia tridentata*) and their hybrids. Hybrids were most fit within the hybrid common garden, while in the parental common gardens the native parental taxon exhibited greater

fitness than hybrids or the alien parental taxon. For reciprocal transplant experiments to be as accurate as possible they need to be long-term and geographically widespread (important selective events may vary temporally and spatially) and many fitness components must be measured to gain a reasonable estimate of lifetime reproductive success (Emms and Arnold, 1997). Unfortunately, cohort analysis, greenhouse studies and reciprocal transplant experiments have not been carried out with *S. chrysanthemifolius*, *S. aethnensis* and their hybrids. Thus, there is, as yet, no proof of genotype by environment interactions in the hybrid zone on Mt. Etna.

### *Conclusion*

I would suggest that the environment-dependent selection/dispersal model is more applicable to the hybrid zone on Mt. Etna than the tension zone model. However, this conclusion is based upon very limited information; for example, (i) hybrids between *S. chrysanthemifolius* and *S. aethnensis* do not seem to be uniformly unfit across all environments, and (ii) the apparent coincidence of several environmental gradients with the hybrid zone on Mt. Etna. It is clear that a great deal more work is required before this conclusion can be considered more than mere speculation. Therefore, a priority for future work must be to determine if there is good evidence for genotype by environment interactions in the hybrid zone on Mt. Etna. Greenhouse studies could provide useful data. For example, such a study could be used to examine the effect of temperature or soil moisture upon seedling/adult survival, with relative ease. Other environmental factors would be more difficult to imitate artificially, e.g. disturbance and snow cover. However, as mentioned previously, greenhouse studies have their disadvantages and, ideally, reciprocal transplant experiments should also be carried out. Reciprocal transplant experiments would be costly and difficult to set up, unless considerable assistance was available in Sicily. The experiments would also need to be long-term to allow measurement of life-time fitness. Furthermore, since the Mt. Etna region is a protected area, experimental work of this kind might not be welcomed.

Of course, further work might reveal that the hybrid zone between *S. chrysanthemifolius* and *S. aethnensis* on Mt. Etna is actually maintained by a balance between dispersal and both environment-dependent and environment-independent selection. If this were the case

it would imply that neither the tension zone model nor the environment-dependent selection model are wholly appropriate as descriptors for the hybrid zone on Mt. Etna. Indeed, it may be unwise to assume that any hybrid zone will conform exactly to a traditional hybrid zone model (mosaic, tension zone, bounded hybrid superiority, environment-dependent selection/dispersal) because the models may be too simplistic to represent reality. Therefore, we can look for hybrid zone models that are better able to reflect the complexity of hybrid zones. One such example is Arnold's (1997) "evolutionary novelty" model, which proposes that both exogenous and endogenous selection may act in a hybrid zone and does not specify the role of dispersal (see section 1.2.2.5). However, Arnold's model has been criticised for being vague and un-testable (Ritchie, 1998; Eckenwalder, 1998). Alternatively, we can simply employ the traditional hybrid zone models in a less dogmatic fashion. For example, we can classify a hybrid zone as a tension zone if it is primarily maintained by a balance between dispersal and intrinsic selection against hybrids, while at the same time recognising that environment-dependent selection may also be taking place within the zone. This seems to be the most sensible approach. It reminds us that models are theoretical constructs designed to increase our understanding of hybrid zones and perhaps we should not expect them to be directly applicable to real situations. Nevertheless, it acknowledges that hybrid zone models may provide a useful basis for stimulating further investigation, by promoting the formulation of testable hypotheses.

#### **4.4.8 The future of the hybrid zone**

The hybrid zone on Mt. Etna has been in existence for possibly many years and, all things being equal, might be expected to persist for a considerable length of time. Selection can maintain steep clines at some loci, even while introgression causes the clines in neutral alleles to disappear (Futuyma, 1998).

A hybrid zone primarily structured by environment-independent selection and dispersal, i.e. a tension zone, may move, for example, into a region of low population density (see section 1.2.2.3), whereas a hybrid zone predominantly structured by environment-dependent selection and dispersal is expected to remain in the same position, if the environmental conditions remain constant. I believe that observations made by Crisp, Abbott and myself suggest that the centre of the hybrid zone has moved to a higher altitude on Mt. Etna during the past 30 years. Crisp stated that the region of highest hybridity was at about 800 m, a

noticeably lower elevation than that identified by Abbott ( $1500 \pm 400$  m) and myself (~1800 m). Careful examination of our records indicates that this difference is real and not the result of alternative interpretations of the same situation. I suggest that this finding does not contradict my hypothesis that the hybrid zone on Mt. Etna is predominantly maintained by a balance between environment-dependent selection and dispersal. Instead, I would argue that the shift in the hybrid zone is the result of changing environmental conditions, that is global warming causing an upward movement of the climatic zones on Mt. Etna.

If this supposition is correct it suggests that the future of *S. aethnensis* may be less than secure. Were the hybrid zone to continue to increase in elevation, *S. aethnensis* might eventually be introgressed into extinction. This is because *S. aethnensis* is endemic to Mt. Etna and there is only a finite amount of space at the very top of the mountain for *S. aethnensis* to migrate into.

#### **4.4.9 Improvements to method of analysis and future work**

The DNA extracts used in the present study were not from individuals growing along the transect on Mt. Etna but from greenhouse grown progeny of seed material collected from the transect. This was unavoidable because collecting tissue samples from individuals in the transect proved impractical. It was difficult to collect large numbers of samples using silica gel and, in any case, silica gel dried material did not produce good quality DNA. A consequence of the way in which samples were collected was that instead of surveying genetic variation along the transect directly, I sampled potential variation in the next generation of individuals. Moreover, the environmental selection pressures placed on the seedlings grown in the greenhouse were likely to have been different from those experienced had they been left to germinate on Mt. Etna. This may have caused bias in the genetic variation recorded for each population along the transect and this issue could be addressed in a future study.

Improvements could be made to the sampling strategy of the molecular analysis. For example, more populations could be chosen for analysis. This would produce a more detailed and, therefore, accurate description of the hybrid zone. In the present study, there was under-representation of populations at the start (below 1200 m) of the transect and towards the end (between 1800 m and 2500 m) of the transect. However, the main transitional region of the hybrid zone was well represented by populations - indeed between

1300 m and 1800 m all available populations were surveyed along transect S. Increasing the number of individuals per population would have decreased the likelihood of erroneous results due to bias caused by small sample sizes. More individuals were available from certain populations along the transect, but this was not the case for all populations (e.g. in populations S9 and S10 viable seed was obtained from a very small number of plants).

Instead of increasing numbers of individuals and populations examined in the transect, it would probably be more useful to analyse a second transect through the hybrid zone on Mt. Etna. If this second transect showed patterns of genetic variation similar to those observed in the first transect, this would provide strong support for the conclusions reached. In other words, a second transect would be a means of replicating the investigation, albeit imperfectly. A suitable second transect might be that referred to as S' (fig. 4.2) but it is rather close to the original transect. No suitable transect was available on the north side of Mt. Etna. Several studies have examined more than one transect through a hybrid zone. For example, Shoemaker *et al.* (1996) observed somewhat different distributional patterns of genetic and morphological markers between two differently orientated transects across a fire ant hybrid zone. From a comparison of the two transects, Shoemaker *et al.* concluded that the fire ant hybrid zone was better represented by the mosaic model than by the tension zone model, and differences between the transects could be attributed to contrasting histories of colonisation and interaction of fire ants in the two areas. Szymura and Barton (1991) compared two transects through a *Bombina* (toad) hybrid zone. They found that there was a striking similarity between the patterns of morphological and electrophoretic variation across the transects. This similarity helped confirm Szymura and Barton's belief that there is a substantial barrier to gene flow between the two *Bombina* species, produced by strong intrinsic selection against hybrids.

# **Chapter 5**

## **Molecular Analysis of the Hybrid Origin of British *Senecio squalidus***

## **Molecular Analysis of the Hybrid Origin of British *Senecio squalidus***

### **5.1 Introduction**

#### **5.1.1 Early history - introduction and spread of British *S. squalidus***

*Senecio squalidus* was introduced to Britain in the seventeenth century, when a population was established in the Oxford Botanic Garden (founded 1621). Herbarium sheets indicate that the species was present in the Botanic Garden by 1690, at the latest, and that the material was derived from Mt. Etna, Sicily. Unfortunately, other early records of the Botanic Garden were lost in a fire and, consequently, the exact date of the introduction and the precise source location of the material remain unknown (Crisp, 1972). *Senecio squalidus* was described as a species by Linnaeus in 1753, from material cultivated in Uppsala, and classed as a southern European annual (Crisp, 1972). It is not known, for certain, how Linnaeus obtained the seed of *S. squalidus*. Some have suggested that seed was supplied to Linnaeus from the Oxford Botanic Garden (Walker, 1833; in Kent, 1956) but this has been disputed (Smith, 1828; in Kent, 1956).

After a century or so of cultivation, *S. squalidus* escaped from the confines of the Oxford Botanic Garden. The first report of this species outside the Garden dates from 1792 (Crisp, 1972). *Senecio squalidus* spread extensively throughout the city of Oxford and by 1800 was "very plentiful on almost every wall in Oxford" (Smith, 1828; in Kent 1956). The species soon caught the attention of botanists throughout Britain; as demonstrated by the large number of early nineteenth century specimens of *S. squalidus* in major herbaria. The earliest record of the species outside Oxford dates from 1801, when it was noted in Worcester (Lees, 1868; in Kent, 1956). In the 1820's *S. squalidus* was observed in Taunton in Somerset, Bideford in Devon and Allesley in Warwickshire (Kent, 1956). These, and other, early sightings can be attributed to the distribution of the species by botanists, who deliberately introduced it to waste-ground or cultivated it in gardens, from whence it escaped.

The subsequent dispersal of *S. squalidus* across Britain is thought to have been aided by the development of the railway system. The species reached the railway network in the late

1800's and spread rapidly along it (Kent, 1960). Railway tracks provided ideal migration corridors and the light, plumed seeds of *S. squalidus* were easily carried in the currents of air produced by passing trains (Druce, 1927). *Senecio squalidus* was also distributed by the movement of industrial ballast (Kent, 1960). In the early 1900's the species was recorded in counties of southern England, Wales and the Midlands (Kent, 1960) and, by 1940, it was well established across much of non-rural England and Wales (Kent, 1960).

### 5.1.2 Current distribution

The post 1940 distribution of *S. squalidus* has been chronicled by Kent (1955, 1957, 1963, 1964a, 1964b, 1964c, 1966), Crisp (1972) and Abbott (1992). The species is currently found as far north as Inverness, west into Ireland and south into northern France and the Channel Islands (Crisp, 1972; Abbott, 1992). It continues to spread, often along the verges and central reservations of major roads (Abbott *et al.*, 2000; Cook, 1998). Today it is one of the most common plant species of the urban environment; ubiquitous on waste-ground and walls of towns and cities.

### 5.1.3 The hybrid origin of *S. squalidus*

Crisp (1972) was intrigued by the history of *S. squalidus* and wanted to know more: Had the species been introduced to Oxford more than once? Had changes taken place in British *S. squalidus* relative to Sicilian *S. squalidus*? What caused British *S. squalidus* to escape from the Oxford Botanic Garden after over a century of quiescent cultivation there? To answer questions such as these, Crisp decided to compare British *S. squalidus* with "Sicilian *S. squalidus*". He examined herbarium specimens of Sicilian *S. squalidus* and then travelled to Mt. Etna, Sicily, to observe live Sicilian material. However, Crisp did not find "Sicilian *S. squalidus*" in Sicily. Instead, he found three *Senecio* taxa similar to the British taxon: *S. aethnensis*, *S. chrysanthemifolius* and *S. incisus*. He deduced that *S. aethnensis* and *S. chrysanthemifolius* were undergoing hybridisation and that *S. incisus* was an array of hybrid swarms between these two species (see chapter three).

Crisp compared *S. aethnensis*, *S. chrysanthemifolius* and hybrid (*S. incisus*) material with British *S. squalidus*, and discovered that "*British S. squalidus, although polymorphic, falls in every character observed within the morphological range of the Sicilian taxa. Its range is, however, more narrow and occupies a more or less intermediate position between the*

*extremes of the Sicilian taxa*" (Crisp, 1972). This morphological data led Crisp to conclude that "*British S. squalidus is of hybrid origin*". He argued that British *S. squalidus* was probably derived from a limited sample of plants taken from the hybrid zone, between *S. chrysanthemifolius* and *S. aethnensis*, on Mt. Etna.

The introduction of a small number of individuals might explain the narrow morphological range observed in British *S. squalidus*, relative to its putative parental species. However, genetic drift and/or selection in the cultivated population in Oxford could also have contributed to this change. Crisp drew attention to the occasional extremes of morphology encountered in British *S. squalidus* (e.g. individuals with more or less entire leaves). Although these extreme types could have resulted from mutation, they normally fell within the morphological range of the Sicilian taxa. Crisp, therefore, concluded that such extreme types were attributable to the hybrid ancestry of British *S. squalidus* and resulted from "*residual variation and the rare recombination of ancestral characters*".

Abbott (Abbott and Milne, 1995; Abbott *et al.*, 1995 and Abbott *et al.*, 2000) re-examined Crisp's hypothesis, that British *S. squalidus* was derived from the Sicilian hybrid zone between *S. chrysanthemifolius* and *S. aethnensis*. The identification of hybrid species using morphological characters can be difficult (see section 4.1.2) and, for this reason, Abbott wished to establish the hybrid origin of British *S. squalidus* using molecular markers.

British *S. squalidus* and Sicilian *Senecio* material - *S. chrysanthemifolius*, *S. aethnensis* and their hybrid - were surveyed for chloroplast DNA, ribosomal DNA and isozyme variation. All British *S. squalidus* individuals were found to possess the cpDNA haplotype B, as did all *S. chrysanthemifolius* individuals and some *S. aethnensis* and hybrid individuals (Abbott *et al.*, 1995). The isozyme data were used to calculate Nei's genetic distance between each population pair. Genetic distances between the British *S. squalidus* and Sicilian *Senecio* populations were low, indicating that they were closely related (Abbott *et al.*, 2000). However, the pattern of variation at one isozyme locus was particularly informative. British *S. squalidus*, *S. aethnensis* and Sicilian hybrid material were polymorphic for the *Acp-2 a* and *b* alleles, whereas *S. chrysanthemifolius* was monomorphic for the *b* allele. This suggested that *S. aethnensis* was a parent of British *S. squalidus*. Further analysis showed that *S. aethnensis* was monomorphic for a rDNA

restriction site mutation (*Bst*EII) that was absent from both British *S. squalidus* and *S. chrysanthemifolius*, it was therefore argued that *S. chrysanthemifolius* must also be a parent of the British taxon (Abbott, unpublished).

The molecular data (Abbott and Milne, 1995; Abbott *et al.*, 1995; Abbott *et al.*, 2000 and Abbott, unpublished) - although consistent with Crisp's hypothesis of a Sicilian hybrid origin of British *S. squalidus* - was not entirely satisfactory. Firstly, the cpDNA and rDNA surveys were based upon small sample sizes. Secondly, it was suspected that the *S. aethnensis* populations examined were not 'pure' *S. aethnensis* but instead were introgressed (see section 4.3.1). Thirdly, only very few taxon-specific molecular markers were available to distinguish *S. chrysanthemifolius* from *S. aethnensis*. Consequently, it was felt that additional molecular work was required to verify the hybrid status of British *S. squalidus*.

The isozyme results also raised a second line of inquiry. In a phenogram (produced by cluster analysis using Nei's genetic distances, as above) some populations of British *S. squalidus* clustered closely with *S. chrysanthemifolius* and certain Sicilian hybrid populations (Abbott *et al.*, 2000). This suggested that British *S. squalidus* might have inherited more of its genome from *S. chrysanthemifolius* than from *S. aethnensis*; it was hoped that additional molecular investigation might reveal if this was indeed the case.

#### **5.1.4 Aim:**

In summary, Crisp (1972) and Abbott (Abbott and Milne, 1995; Abbott *et al.*, 1995 and Abbott *et al.*, 2000) have postulated that British *S. squalidus* is of homoploid hybrid origin; derived from material from the hybrid zone, between *S. chrysanthemifolius* and *S. aethnensis*, on Mt. Etna, Sicily. However, the use of morphological characters for identifying hybrid species has been criticised and the molecular results were not conclusive.

Therefore, the aim of the present study was to use molecular markers to ascertain if there was good molecular evidence to confirm the homoploid hybrid origin of British *S. squalidus*. If the hybrid origin was confirmed, an analysis would be made of what proportion of the genome of British *S. squalidus* could be attributed to each of the two parental taxa.

## 5.2 Materials and methods

### 5.2.1 Survey of RAPD and ISSR variation in British *S. squalidus* populations

Six geographically widespread populations - Oxford, Portland Bill, Cardiff, Norwich, Hull and Kirkcaldy - of British *S. squalidus* were chosen for analysis (table 2.2) and DNA was obtained from 12 individuals per population. Individuals were surveyed for the 26 RAPD and ISSR markers used to examine variation within the 12 populations along transect S on Mt. Etna (table 4.5). RAPD and ISSR analyses were carried out using the procedures outlined in section 2.8. Repeats of individual DNA samples were carried out, as necessary, to ensure that the data set was complete.

Precautions were taken to ensure that the RAPD and ISSR results were reliable and reproducible. Standards were run on all RAPD and ISSR gels, as detailed in section 4.2.2. A number of RAPD and ISSR reactions were repeated as an additional check for reproducibility. Homology tests, using the restriction digest method (2.9), were carried out for all the *S. chrysanthemifolius* and *S. aethnensis* markers believed to be present in British *S. squalidus*.

#### 5.2.1.1 Data analysis

The frequency of each *S. chrysanthemifolius* and *S. aethnensis* marker in each population was calculated and pooled over the six populations.

A binary data matrix was constructed for the 26 RAPD and ISSR markers and this matrix was used in the following analyses:

#### Multilocus variation

##### (i) Hybrid index analysis

A hybrid index was constructed for each individual in the six British *S. squalidus* populations, as described in section 4.2.5.1. Frequency distributions of hybrid indices were made for each population. The mean and range of hybrid indices and the mean and range of the number of *S. chrysanthemifolius* and *S. aethnensis* markers within each population were calculated. The results were compared with those of *Senecio* material from the transect on Mt. Etna (chapter four).

*(ii) Principal coordinate analysis (PCO)*

Principal coordinate analysis (see section 2.11.3) was carried out on the binary data matrix of the British *S. squalidus* populations. Two coefficients were used in the PCO: Nei and Li's similarity coefficient and Jaccard's similarity coefficient (see section 2.11.2). In addition, two further PCOs were carried out: the first involved all British *S. squalidus* individuals plus *S. chrysanthemifolius*, *S. aethnensis* and artificial F<sub>1</sub> individuals. The second involved all British *S. squalidus* individuals, all Sicilian hybrid individuals and individuals from one population of *S. chrysanthemifolius* (C1) and one population of *S. aethnensis* (BB).

Population level analyses

*(i) Analysis of molecular variance (AMOVA)*

An AMOVA (2.11.4) was carried out to estimate variance components within and between the six British *S. squalidus* populations. The AMOVA also calculated the fractional proportion of the total variance partitioned between each population pair and these pair-wise  $\Phi_{ST}$  values were used as a measure of inter-population genetic distance. Phenograms were constructed from the pair-wise  $\Phi_{ST}$  values using the UPGMA and neighbour-joining clustering methods (2.11.5).

*(ii) Shannon's diversity index*

Shannon's diversity index (see section 2.11.6) was used to quantify the level of genetic diversity present in the six populations of British *S. squalidus* and to partition the total genetic diversity (in the sample as a whole) into within and between population components. Estimates of genetic diversity for the six populations of British *S. squalidus* were compared to estimates of genetic diversity for the 12 populations along the transect, between *S. chrysanthemifolius* and *S. aethnensis*, on Mt. Etna.

### **5.2.2 Survey of chloroplast and isozyme variation**

Sufficient individuals of British *S. squalidus* had already been surveyed (Abbott *et al.*, 1995; Abbott and Milne, 1995 and Abbott *et al.*, 2000) for chloroplast and isozyme variation. These data were compared to the chloroplast DNA RFLP and isozyme data for *S. chrysanthemifolius*, *S. aethnensis* and their hybrid obtained in the present study (chapter four).

## 5.3 Results

### 5.3.1 Survey of RAPD and ISSR variation in British *S. squalidus* populations

Examples of RAPD and ISSR gels of the British *S. squalidus* populations are shown in plates 5.1 - 5.2.

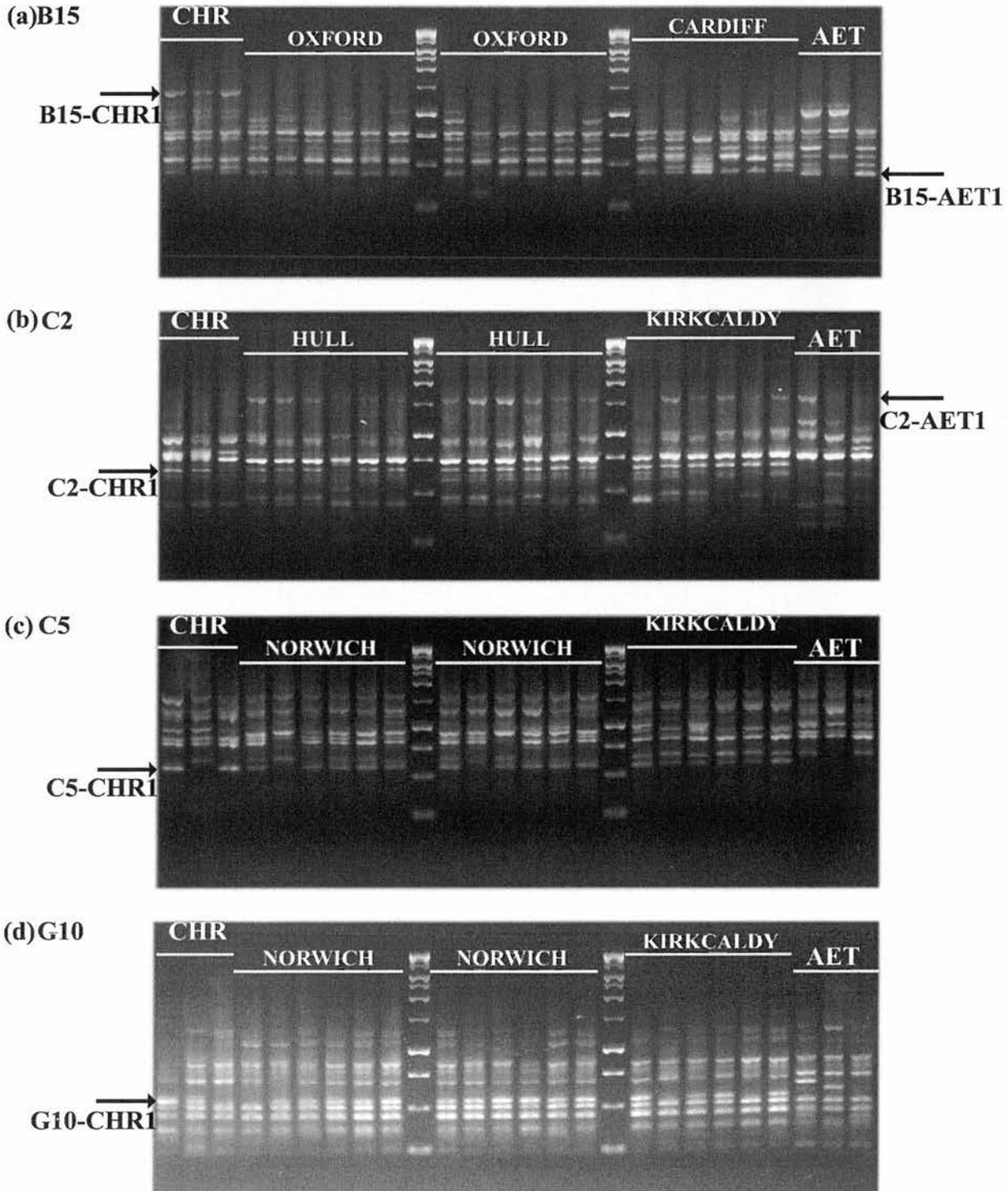
Four RAPD primers were repeated (for 72 individuals). Two hundred and seventy three comparisons were made between the presence/absence of bands in repeats and originals, across four markers. The error rate (i.e. the percentage of discrepancies between repeats and originals) was less than 4.4%. Homology between bands in British *S. squalidus* and the equivalent bands in *S. chrysanthemifolius* or *S. aethnensis* was confirmed in 17 of the 21 cases (plate 6.3). Homology tests for the remaining four bands were unsuccessful (for example, not enough band product was amplified to restrict, see section 2.9).

The frequency of each *S. chrysanthemifolius* and *S. aethnensis* marker in each population, and after data were pooled over the six populations, are shown in table 5.1. Some markers were present at a similar frequency across all populations, whereas other markers showed considerable variation between populations. For example, A11-AET-1 was present at low frequency in all populations bar Kirkcaldy, and H7-CHR-1 was present at high frequency in all populations bar Oxford. The mean frequency of all *S. chrysanthemifolius* markers (0.62) was greater than the mean frequency of all *S. aethnensis* markers (0.41). Overall, therefore, *S. squalidus* was highly polymorphic for marker band presence/absence, both within and between populations.

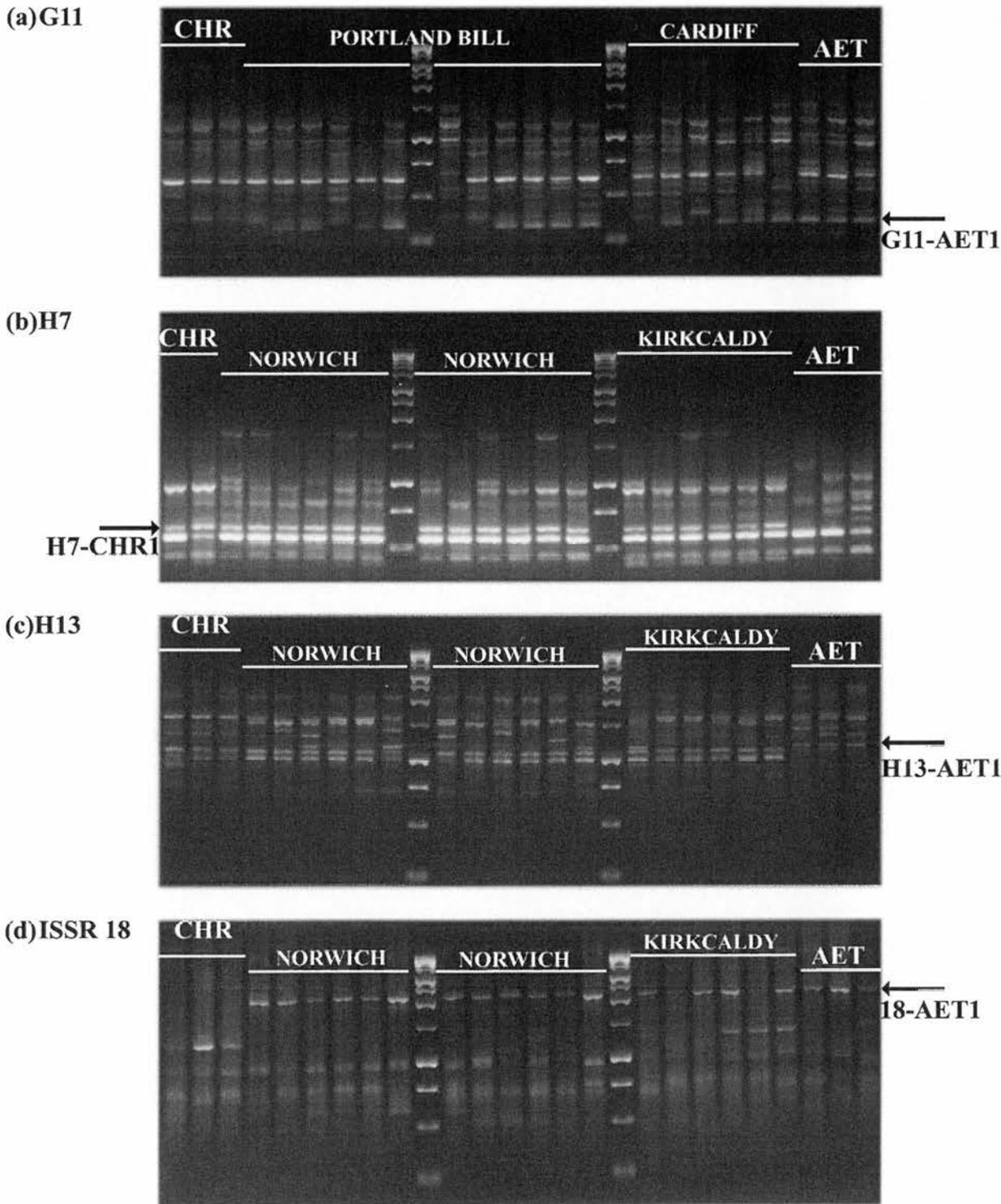
Furthermore, five of the 26 markers were absent from the British *S. squalidus* individuals surveyed; three markers specific to *S. aethnensis* (B15-AET-1, G5-AET-1 and H19-AET-1) and two specific to *S. chrysanthemifolius* (B15-CHR-1 and G12-CHR-1). However, one of the absent *S. aethnensis* markers, B15-AET-1, was only present in 'pure' *S. aethnensis* at a fairly low frequency of 0.53 (table 3.11). Another *S. aethnensis* marker (F12-AET-1) was detected in British *S. squalidus* at a very low frequency of 0.01 (i.e. in one individual of 72 surveyed).

The binary data matrix constructed for the 26 RAPD and ISSR markers scored in all six populations is presented in appendix B. It shows that individuals of British *S. squalidus*

**Plate 5.1** Examples of *Senecio chrysanthemifolius* and *S. aethnensis* marker bands amplified by RAPD primers in populations of British *S. squalidus*. Photograph (a) primer B15, (b) primer C2, (c) primer C5, and (d) primer G10. AET indicates *S. aethnensis* standards, CHR indicates *S. chrysanthemifolius* standards.



**Plate 5.2** Examples of *Senecio chrysanthemifolius* and *S. aethnensis* marker bands amplified by RAPD and ISSR primers in populations of British *S. squalidus*. Photograph (a) primer G11, (b) primer H7, (c) primer H13, and (d) primer 18. AET indicates *S. aethnensis* standards, CHR indicates *S. chrysanthemifolius* standards.



combine species-specific *S. chrysanthemifolius* markers and species-specific *S. aethnensis* markers.

British *S. squalidus* individuals were also examined to see if they had any unique RAPD and ISSR markers; that is markers present in British *S. squalidus*, but absent from the putative parental species, *S. chrysanthemifolius* and *S. aethnensis* (and the Sicilian hybrids). The number of bands in *S. squalidus* with the potential to be unique markers was considered to be six, out of a total of more than 100. However, it was difficult to judge whether bands in British *S. squalidus* were also present in the putative parental taxa because only a limited number of *S. chrysanthemifolius* and *S. aethnensis* individuals were run as standards on the British *S. squalidus* gels. Therefore, more work would be required to confirm whether the six bands really were unique markers and whether any other potential unique markers had been overlooked.

**Table 5.1** Frequencies of taxon-specific *Senecio chrysanthemifolius* and *S. aethnensis* RAPD and ISSR markers in six populations of British *S. squalidus*.

Population		All populations n = 72	six Portland Bill n = 12	Oxford n = 12	Cardiff n = 12	Norwich n = 12	Hull n = 12	Kirkcaldy n = 12
Primer	Band							
A9	CHR-1	0.63	0.50	0.58	0.67	0.58	0.67	0.75
A11	CHR-1	0.51	0.58	0.33	0.33	0.75	0.42	0.67
	AET-1	0.29	0.08	0.00	0.25	0.25	0.25	0.92
A20	CHR-1	0.51	0.58	0.25	0.08	0.67	0.75	0.75
B15	CHR-1	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	AET-1	0.00	0.00	0.00	0.00	0.00	0.00	0.00
C2	CHR-1	0.96	1.00	1.00	1.00	0.75	1.00	1.00
	AET-1	0.86	0.83	0.92	0.83	0.83	1.00	0.75
C5	CHR-1	0.65	0.75	0.33	0.25	0.83	0.83	0.92
C20	AET-1	0.63	0.58	0.92	0.25	0.42	0.83	0.75
F12	CHR-1	0.86	0.92	0.42	1.00	0.83	1.00	1.00
	AET-1	0.01	0.00	0.08	0.00	0.00	0.00	0.00
G5	CHR-1	0.75	0.75	1.00	0.92	0.92	0.58	0.33
	AET-1	0.00	0.00	0.00	0.00	0.00	0.00	0.00
G6	AET-1	0.52	0.75	0.67	0.42	0.25	0.83	0.17
G8	CHR-1	0.87	0.83	0.83	0.83	1.00	0.92	0.83
	AET-1	0.60	0.67	0.42	0.67	0.83	0.58	0.42
G10	CHR-1	0.63	0.92	0.67	0.50	0.92	0.17	0.58
G11	AET-1	0.24	0.67	0.17	0.50	0.00	0.00	0.08
G12	CHR-1	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	AET-1	0.72	0.92	0.83	0.83	0.67	1.00	0.08
H7	CHR-1	0.88	1.00	0.33	1.00	1.00	1.00	0.92
H13	AET-1	0.52	0.75	0.67	0.33	0.25	0.92	0.17
H19	AET-1	0.00	0.00	0.00	0.00	0.00	0.00	0.00
18	AET-1	0.88	0.92	0.75	0.83	1.00	1.00	0.75
40	CHR-1	0.76	0.58	0.92	0.58	0.92	0.75	0.83

### **Multilocus variation**

#### **5.3.2 Hybrid index analysis**

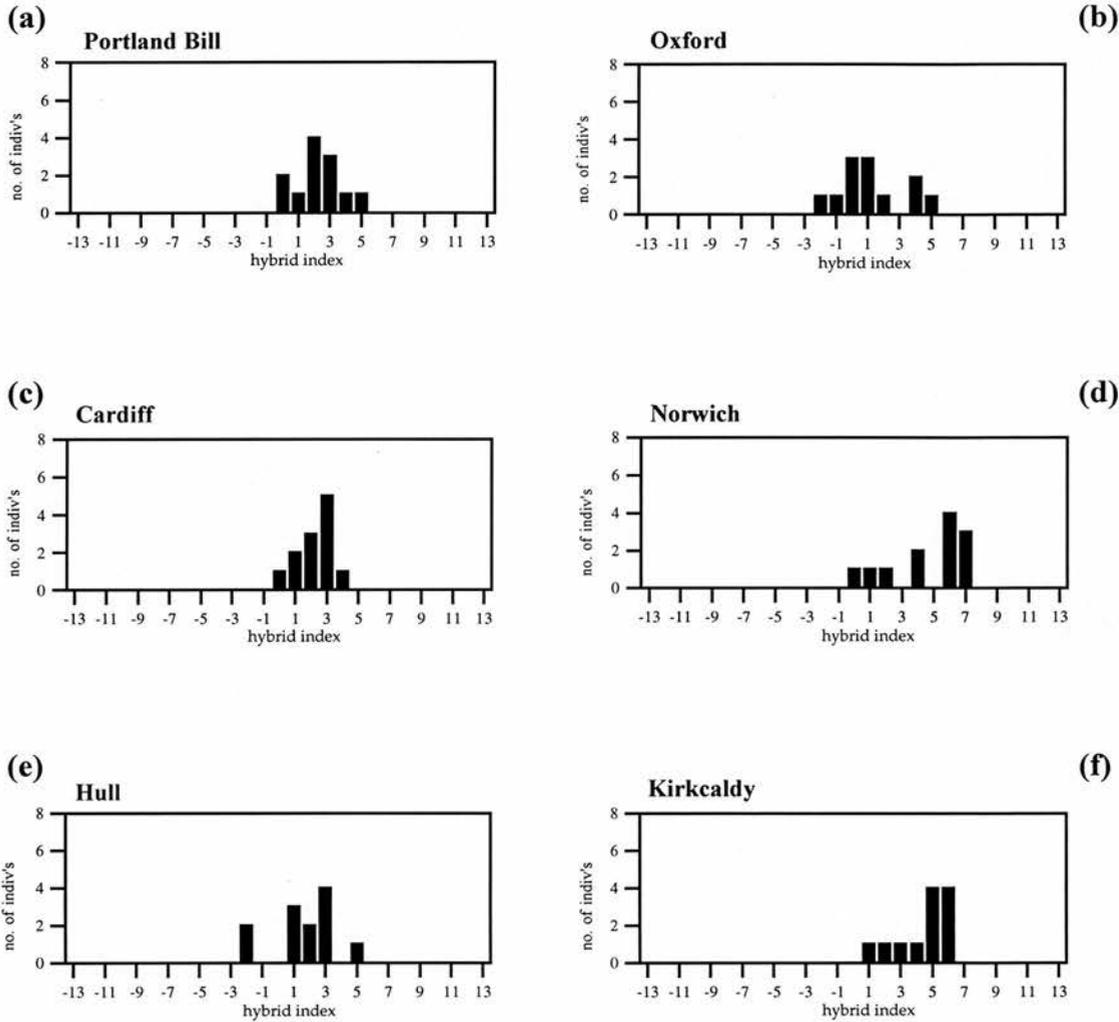
Hybrid indices of individuals within populations of British *S. squalidus* ranged from -2 to +7, with an overall mean of +2.76. A positive hybrid index indicated that *S. chrysanthemifolius* markers occurred in greater numbers than did *S. aethnensis* markers, a negative hybrid index indicated the opposite, while a hybrid index of zero indicated that numbers of *S. chrysanthemifolius* and *S. aethnensis* markers were evenly balanced. Hybrid indices of individuals from Oxford and Hull ranged from -2 to +5 (fig. 5.1 (b), (e)), with a mean of +1.3 for the Oxford population and +1.7 for the Hull population (fig. 5.2). Individuals from Cardiff and Portland Bill had hybrid indices ranging from 0 to +4 and 0 to +5, respectively, with population means of +2.3. Hybrid indices of individuals from Kirkcaldy and Norwich ranged from 0 to +6 and 0 to +7, with population means of +4.5 and +4.7, respectively. In all six populations, the mean number of *S. chrysanthemifolius* markers (from 6.7 in the Oxford population to 9.2 in the Norwich population) was greater than the mean number of *S. aethnensis* markers (from 4.1 in the Kirkcaldy population to 6.4 in the Hull population; fig. 5.2 (b), (c)).

The six British *S. squalidus* populations had hybrid index frequency distributions roughly intermediate between those of 'pure' *S. chrysanthemifolius* and *S. aethnensis* populations but with a bias towards *S. chrysanthemifolius* (fig. 5.1 cf. fig. 4.9). Likewise, the mean hybrid indices of the six British *S. squalidus* populations were roughly intermediate between those of 'pure' *S. chrysanthemifolius* (+11.2) and *S. aethnensis* (-11) but with a bias towards *S. chrysanthemifolius* (fig. 5.2 cf. fig. 4.10). The bias towards *S. chrysanthemifolius* was especially pronounced in the populations from Norwich and Kirkcaldy.

#### **5.3.3 Principal coordinate analysis**

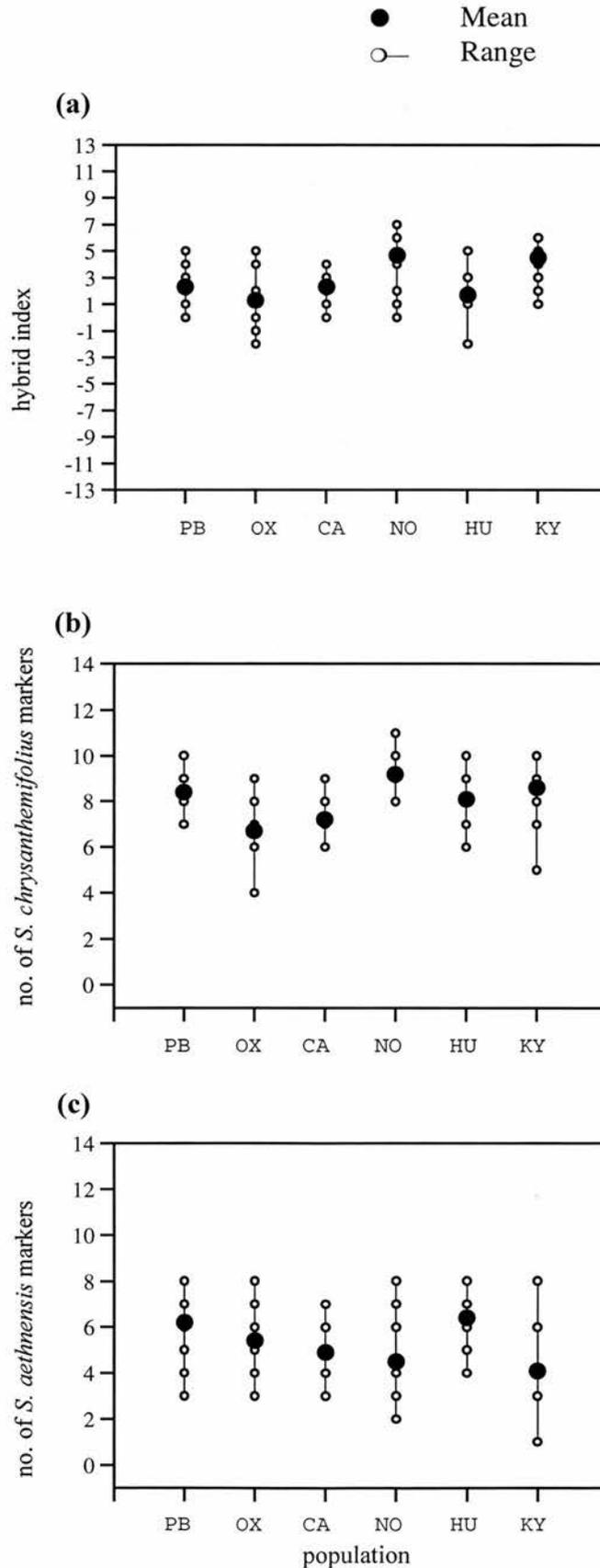
The PCO results based upon Jaccard's, and Nei and Li's, coefficients were very similar and therefore only those for the latter are displayed. These were chosen because more of the total variation was accounted for by the first two principal coordinates in these plots.

**Fig. 5.1** Frequency distributions of hybrid indices for six British *Senecio squalidus* populations.



**Fig. 5.2** (a) Mean and range of hybrid indices within six British *Senecio squalidus* populations. (b) Mean and range of the number of *S. chrysanthemifolius* markers within six British *S. squalidus* populations. (c) Mean and range of the number of *S. aethnensis* markers within six British *S. squalidus* populations.

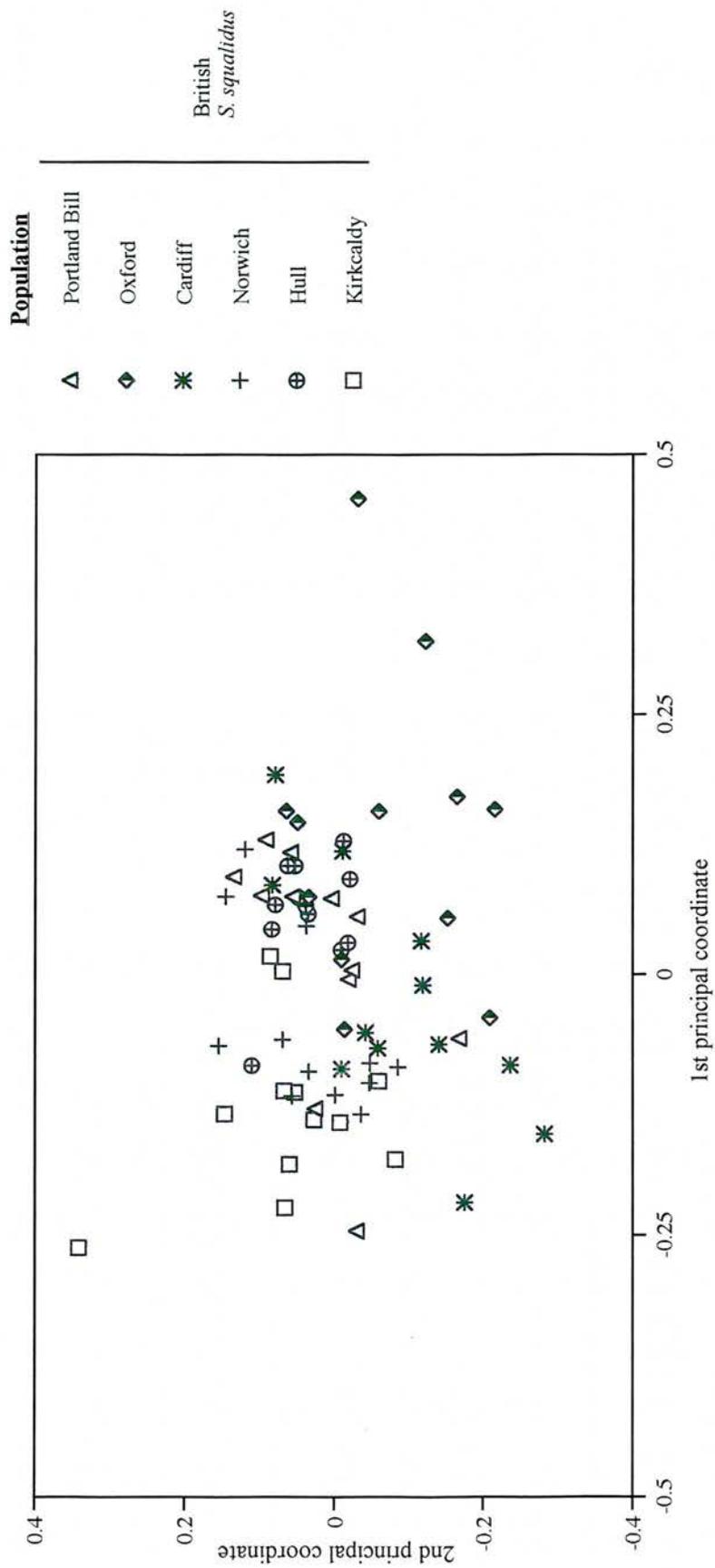
PB - Portland Bill, OX - Oxford, CA - Cardiff, NO - Norwich, HU - Hull and KY - Kirkcaldy.



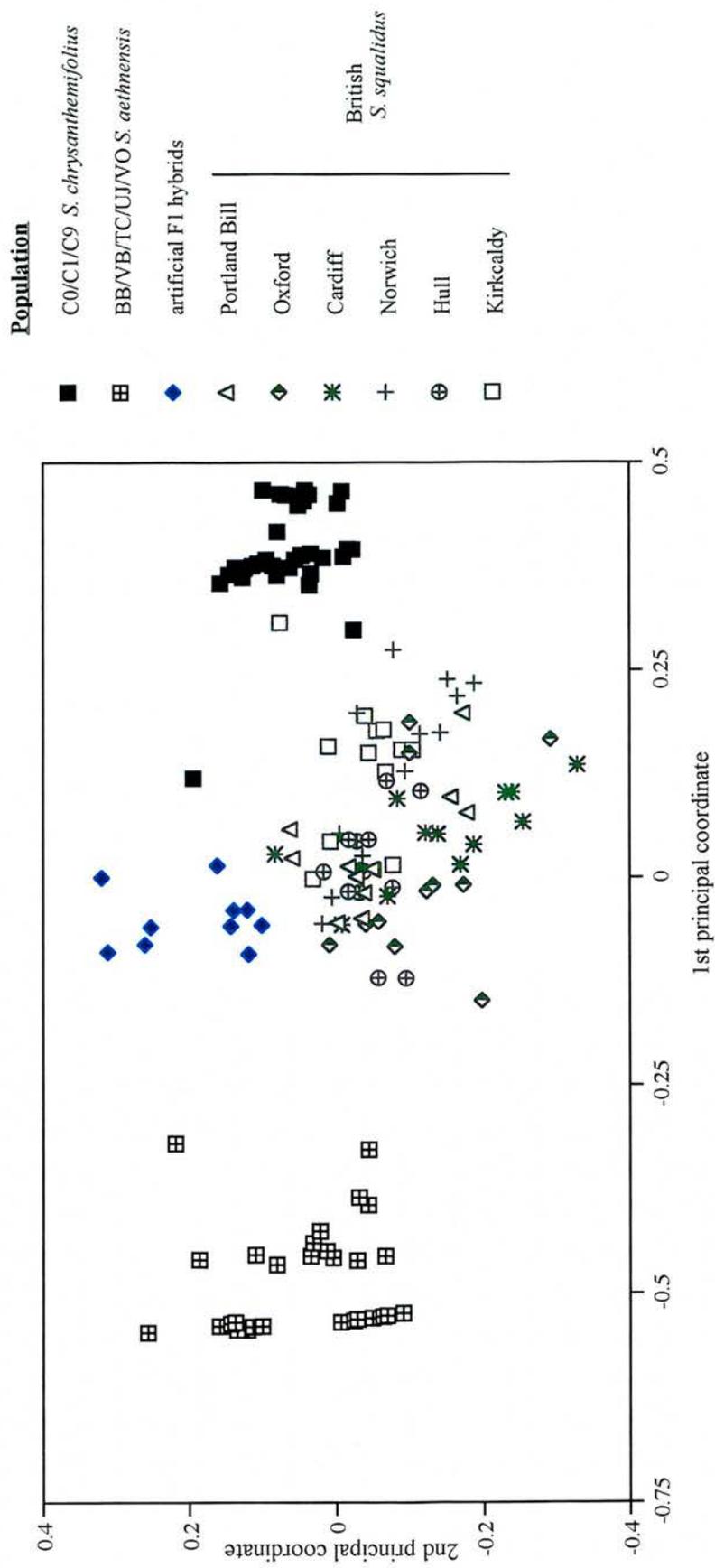
The first principal coordinate, of the PCO plot of British *S. squalidus* individuals, represented 34.28% of the total variation, the second principal coordinate explained 21.85% and the third principal coordinate represented 19.13% of the total variation (fig. 5.3). None of the populations formed distinct groups. Along the first principal coordinate axis, individuals from Oxford were least aggregated and individuals from Hull were most aggregated. The PCO did not resolve any obvious geographic structure in the data, populations that were geographically close to one another did not necessarily cluster together.

The first principal coordinate, of the PCO plot that included British *S. squalidus*, *S. chrysanthemifolius*, *S. aethnensis* and artificial F<sub>1</sub> individuals, represented 86.32% of the total variation, the second principal coordinate explained 8.51% and the third principal coordinate represented 5.20% of the total variation (fig. 5.4). In this plot, British *S. squalidus* formed a group that was more loosely aggregated than the *S. chrysanthemifolius*, *S. aethnensis* and F<sub>1</sub> groups. The first principal coordinate axis produced a clear separation of *S. chrysanthemifolius* and *S. aethnensis* populations, with the artificial F<sub>1</sub>s located centrally. Some British *S. squalidus* individuals were positioned close to F<sub>1</sub> individuals, on the first axis. However, the remaining British *S. squalidus* individuals (notably individuals from Kirkcaldy and Norwich) tended more towards the *S. chrysanthemifolius* cluster, on this axis. On the second principal coordinate axis, all British *S. squalidus* individuals were separated from the group of F<sub>1</sub> individuals.

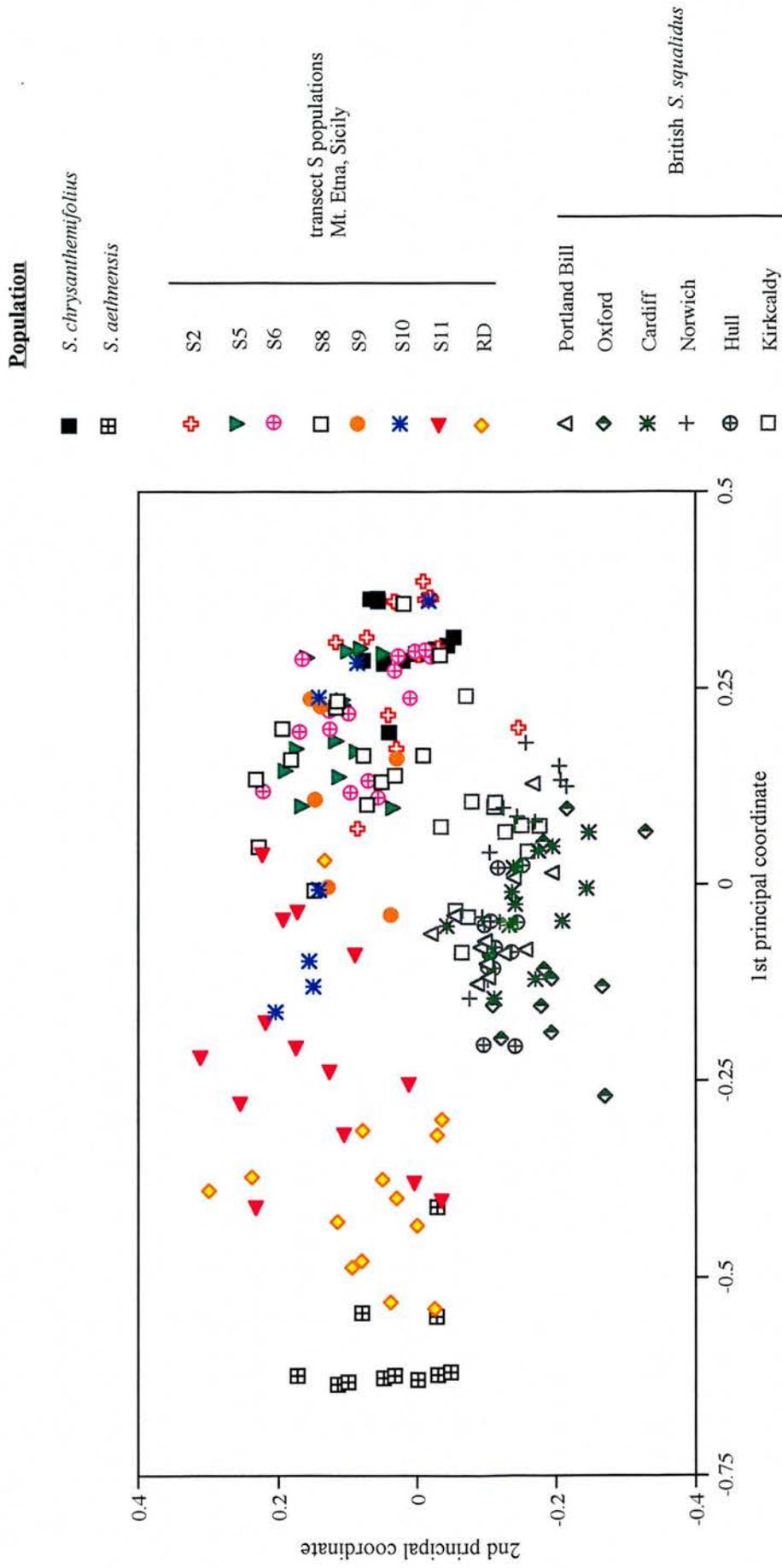
The first principal coordinate, of the PCO plot that included all British *S. squalidus* individuals, all Sicilian hybrid individuals and individuals from one population of *S. chrysanthemifolius* (C1) and one population of *S. aethnensis* (BB), represented 67.41% of the total variation, the second principal coordinate explained 15.63% and the third principal coordinate represented 8.51% of the total variation (fig. 5.5). On the first principal coordinate axis some British *S. squalidus* individuals were located close to Sicilian hybrid individuals from populations S9, S10 and S11, others were located close to Sicilian hybrid individuals from populations S8, S6, S5 and S2 (introgressed *S. chrysanthemifolius*). However, along the second principal coordinate axis, British *S. squalidus* was separated from these Sicilian hybrid populations.



**Fig. 5.3** The first two coordinates of a principal coordinate analysis of RAPD and ISSR data from individuals in six British *S. squalidus* populations.



**Fig. 5.4** The first two coordinates of a principal coordinate analysis of RAPD and ISSR data from individuals in six British *S. squalidus* populations, plus individuals of *S. chrysanthemifolius*, *S. aethnensis* and artificial F1 hybrids between them.



**Fig. 5.5** The first two coordinates of a principal coordinate analysis of RAPD and ISSR data from individuals in six British *S. squalidus* populations; plus individuals from populations along transect S, through the hybrid zone on Mt. Etna, including individuals of *S. chrysanthemifolius* (C1) and *S. aethnensis* (BB).

### 5.3.4 Analysis of molecular variance

The AMOVA (table 5.2) revealed that most of the total variation was present within populations, rather than between populations (79% within, compared to 21% between).

**Table 5.2** Analysis of molecular variance (AMOVA) for six populations of British *Senecio squalidus*.

Source of variation	d.f. <sup>1</sup>	SSD <sup>2</sup>	MSD <sup>3</sup>	Variance component	% of the total variance	P-value <sup>4</sup>
Between populations	5	64.96	12.99	0.82	20.99	< 0.001
Within populations	66	204.75	3.10	3.10 ( $\Phi_{ST} = 0.210$ )	79.01	< 0.001

<sup>1</sup> Degrees of freedom.

<sup>2</sup> Sum of squared deviations.

<sup>3</sup> Mean squared deviations.

<sup>4</sup> Significance of the variance components (the probability of obtaining a more extreme component estimate by chance alone).

All pair-wise  $\Phi_{ST}$  values (used as a measure of inter-population genetic distance) were significantly larger than a random  $\Phi_{ST}$  value (table 5.3), indicating that all populations were significantly genetically different from each other. Populations from Kirkcaldy and Oxford were the most distinct (pair-wise  $\Phi_{ST}$  of 0.3363), while populations from Cardiff and Portland Bill were most similar (pair-wise  $\Phi_{ST}$  of 0.0933).

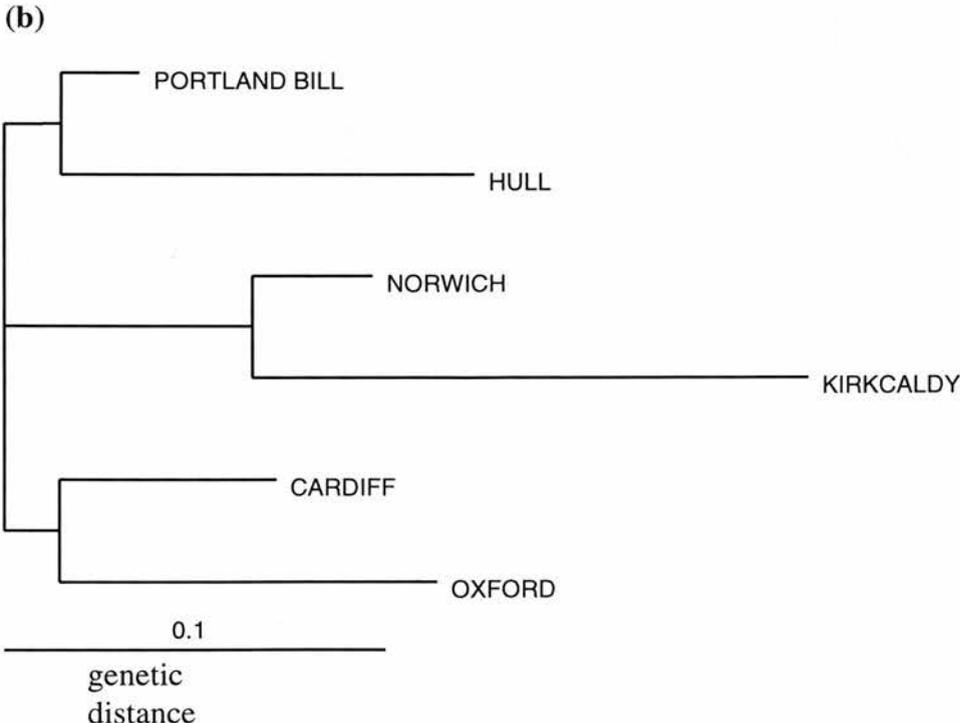
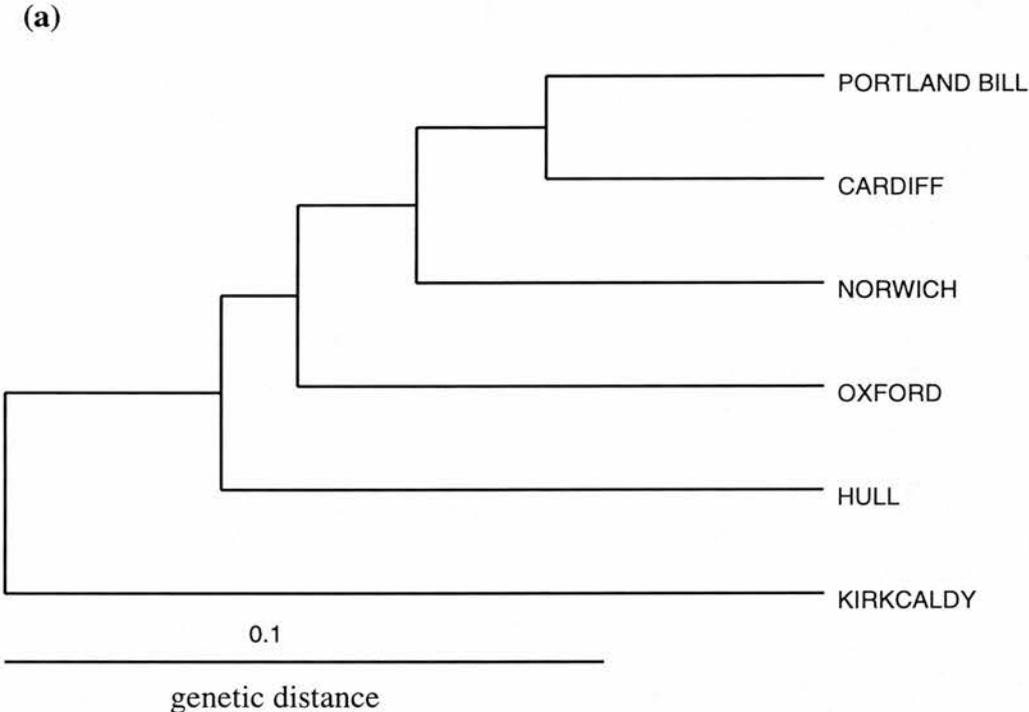
**Table 5.3** Inter-population genetic distances (pair-wise  $\Phi_{ST}$  values) among six populations of British *Senecio squalidus*.

	PB	OX	CA	NO	HU	KY
PB		0.1511	0.0933	0.1211	0.1293	0.2793
OX			0.1563	0.2222	0.2128	0.3363
CA				0.1519	0.2306	0.2764
NO					0.2357	0.1787
HU						0.2986
KY						

All inter-population genetic distances were significantly larger than a random  $\Phi_{ST}$  value at  $P < 0.001$ .

UPGMA and neighbour-joining clustering of pair-wise  $\Phi_{ST}$  values produced phenograms that differed in structure. In the former, the Portland Bill and Cardiff populations clustered together, with the progressive addition of Norwich, Oxford and Hull and, finally, Kirkcaldy (fig. 5.6). In the neighbour-joining tree, three clusters were apparent: one contained the

**Fig. 5.6** (a) UPGMA, and (b) neighbour-joining dendrograms of six British *Senecio squalidus* populations. The clustering was based upon pair-wise  $\Phi_{ST}$  values from an AMOVA of RAPD and ISSR data.



Oxford and Cardiff populations, one the Portland Bill and Hull populations and the other the Norwich and Kirkcaldy populations.

### 5.3.5 Shannon's diversity index

The level of genetic diversity ( $H_0$ ) within each of the six populations of British *S. squalidus* was quantified using Shannon's diversity index (table 5.4). The Cardiff population had the highest level of genetic diversity, followed by the Oxford, Kirkcaldy, Portland Bill and Norwich populations. The Hull population had the lowest level of genetic diversity.

When these results were compared to those of populations along the transect on Mt. Etna, Sicily, (section 4.3.2.6) two facts became apparent. First, the range of genetic diversity was smaller for the British *S. squalidus* populations than for the Sicilian transect populations ( $H_0$  in British *S. squalidus* populations ranged from 4.272 to 6.606;  $H_0$  in Sicilian transect populations ranged from 3.209 to 9.654). Secondly, British *S. squalidus* populations had levels of genetic diversity that were equivalent to the introgressed populations of *S. chrysanthemifolius* (S2, S5, S6 and S8); that is, higher levels of genetic diversity than the 'pure' *S. chrysanthemifolius* and *S. aethnensis* populations and lower levels than the most diverse hybrid populations (such as S11).

**Table 5.4** Estimates of genetic diversity ( $H_0$ ) within six populations of British *Senecio squalidus*.

Population	Genetic diversity ( $H_0$ †)	$H_0/26$ ‡
Portland Bill	5.523	0.212
Oxford	6.567	0.253
Cardiff	6.606	0.254
Norwich	5.103	0.196
Hull	4.272	0.164
Kirkcaldy	5.683	0.219

† Mean of values from 26 RAPD and ISSR markers.

‡ Genetic diversity per marker scored (i.e.  $H_0$  divided by the total number of bands scored).

Partitioning of the genetic diversity within and between populations (table 5.5) revealed that most of the total variation was present within populations, 80%, rather than between populations, 20%. These values were very similar to those obtained in the AMOVA analysis (table 5.2).

**Table 5.5** Partitioning of the genetic diversity within and among the six populations of British *Senecio squalidus*.

$H_{pop}^1$	$H_{pop}/26^1$	$H_{sp}^2$	$H_{sp}/26^2$	$H_{pop}/H_{sp}^3$	$(H_{sp}-H_{pop})/H_{sp}^4$
5.626	0.216	7.037	0.271	0.799	0.201

<sup>1</sup> Mean within population genetic diversity ( $H_{pop}$ ) and mean within population genetic diversity per marker scored ( $H_{pop}/26$ ).

<sup>2</sup> Total genetic diversity ( $H_{sp}$ ) and total genetic diversity per marker scored ( $H_{sp}/26$ ).

<sup>3</sup> Proportion of the total genetic diversity maintained within populations.

<sup>4</sup> Proportion of the total genetic diversity maintained between populations.

## 5.4 Discussion

### 5.4.1 The identification of homoploid hybrid species

The unambiguous documentation of homoploid hybrid species is difficult. The patterns of variation produced by homoploid hybrid speciation may be confused with the patterns produced by convergent evolution, plesiomorphy and symplesiomorphy. In convergent evolution, a species resembles a hybrid taxon because it has independently acquired characters similar to those of the putative parental species (O'Hanlon *et al.*, 1999). Due to plesiomorphy (the retention of ancestral character states), an ancestral population may appear to be a hybrid species, with the apparent parents derived by phylogenetic sorting (O'Hanlon *et al.*, 1999). In symplesiomorphy, a taxon shares characters with related taxa due to the joint retention of characters following speciation of a polymorphic ancestor (Rieseberg, 1997). Thus, a supposed hybrid species may, in reality, be a sister species to one or both of the postulated parents (Allan *et al.*, 1997). Particular confusion may arise if divergent speciation is followed by introgression (i.e. there are two sister species and one of these undergoes introgression with a third species; Allan *et al.*, 1997).

To judge whether a particular taxon is of homoploid hybrid origin, evidence may be gathered from a number of sources:

#### *Morphological evidence*

A homoploid hybrid species has traditionally been expected to display morphological intermediacy between its parental taxa. For example, the homoploid hybrid species *Encelia virginensis* is morphologically intermediate between its parents (Allan *et al.*, 1997). However, the situation is not always so straightforward and a homoploid hybrid species may be a mosaic of parental, extreme and intermediate morphological characters (Rieseberg and Ellstrand, 1993).

Care must be taken when using morphological markers because morphological intermediacy can result from processes other than hybridisation, namely convergence, plesiomorphy and symplesiomorphy.

*Molecular evidence*

The pattern of molecular variation indicative of a homoploid hybrid origin is an additive profile of parental nuclear genetic markers (Gallez and Gottlieb, 1982; Morrell and Rieseberg, 1998; Wolfe *et al.*, 1998). An additive profile of nuclear genetic markers can take several forms, depending upon the effects of segregation and recombination in hybrid populations (Allan *et al.*, 1997). Additivity may be present at the level of the individual; if a marker is codominant there may additivity at a single locus, (i.e. one allele is from parent A and other allele is from parent B) or there may be additivity across loci, for both codominant and dominant, markers (i.e. locus 1 has a marker from parent A, locus 2 has a marker from parent B and so on; Gallez and Gottlieb, 1982). Alternatively, there may be additivity at the level of the population (i.e. individual 1 has a marker from parent A, individual 2 has a marker from parent B and so on) or there may even be additivity between populations (i.e. population 1 contains individuals with markers from parent A, population 2 contains individuals with markers from parent B and so on; Allan *et al.*, 1997).

Complete additivity of parental markers in a hybrid species (all parental markers are represented in the hybrid species) frequently does not occur and, therefore, multiple loci may need to be sampled to detect additivity (Morrell and Rieseberg, 1998). Lack of complete additivity may be caused by parental markers becoming extinct following hybridisation (e.g. due to founder effects, drift in small populations, selection) or because not all parental markers were represented in the hybridising material (Morrell and Rieseberg, 1998).

An additive profile of a few nuclear genetic markers does not, on its own, rule out alternative hypotheses such as convergence, plesiomorphy and symplesiomorphy. Indeed, it is easier to reject a hypothesis of homoploid hybrid origin, using molecular data, than to confirm it (Rieseberg, 1997). However, there are a number of ways in which homoploid hybrid speciation can be distinguished from alternative hypotheses. Firstly, if a putative hybrid has an additive nuclear profile but an ancestral organelle haplotype it is likely to be ancestral to the apparent parents (Rieseberg *et al.*, 1990b). For this reason, many workers suggest that a homoploid hybrid species should possess the organelle genome (usually cpDNA) of at least one of its parents (Wolfe *et al.*, 1998). Secondly, a recent hybrid species should possess few or no unique markers, compared to its parent species (Allan *et al.*,

1997; Morrell and Rieseberg, 1997). Possession of few or no unique markers enables us to reject plesiomorphy and symplesiomorphy (Allan *et al.*, 1997). Thirdly, if a putative homoploid hybrid species possesses multiple linked markers of the likely parental taxa and/or an additive profile of diagnostic parental markers at multiple loci, the likelihood of the observed situation having arisen through convergence or symplesiomorphy is low (Rieseberg, 1997). Finally, gene lineage data may be useful. As species divergence times increase the probability that species share alleles of identical sequence due to hybridisation (rather than an alternative explanation) rises (Rieseberg, 1997)

Molecular markers are a better means of identifying homoploid hybrid species than morphological characters, because genetic additivity is confirmed rather than assumed. It is believed that morphology, unlike molecular data, is prone to considerable convergence and that convergence in molecular data is more easily detected, understood, and dealt with than convergence in morphological data (Sytsma, 1990). Furthermore, it is easier to distinguish symplesiomorphy and plesiomorphy from hybridisation using molecular markers than morphological markers (for example by determining the polarity of an inter-taxon relationship using organelle genomes, see above). Molecular markers also have other advantages; for example, compared to morphological characters, molecular markers are more heritable, numerous and independent (Sytsma, 1990).

#### *Other evidence*

Evidence for a homoploid hybrid origin may also come from secondary chemistry, geographical and ecological data and the production of synthetic hybrids. In the case of secondary chemistry, homoploid hybrid species are expected to exhibit additivity of chemical compounds (Padgett *et al.*, 1998). Hybrid species are sometimes predicted to demonstrate a geographical range that lies within the region of parental sympatry (Padgett *et al.*, 1998) and an ecological distribution intermediate to those of its parents (Allan *et al.*, 1997). It should also be possible to produce synthetic hybrids between the parental species that bear a close resemblance to the hybrid taxon (Allan *et al.*, 1997). However, chemical additivity, ecological intermediacy, etc. do not prove that a putative hybrid exhibits genetic additivity (Rieseberg *et al.*, 1990b). In other words, a species that meets some or all of the above criteria is not necessarily a homoploid hybrid, nor will a genuine homoploid hybrid species inevitably conform to the criteria. For example, *Helianthus anomalous*, *H.*

*deserticola* and *H. petiolaris* are homoploid hybrid species that do not exhibit an additive chemical profile and the three species occupy extreme habitats, rather than habitats intermediate to those of their parents' (Rieseberg, 1991). Therefore, the sources of information described above are, like morphological characters, best used in conjunction with molecular data.

#### *Introgression versus hybrid speciation*

Unambiguous evidence of hybridity does not automatically mean that hybrid speciation has taken place. On the basis of hybridity alone, introgression or hybrid swarm formation might be mistaken for hybrid speciation (Rieseberg, 1997). It is necessary to establish that the proposed hybrid taxon is isolated from parental gene flow and capable of following an independent evolutionary trajectory (Rieseberg, 1997). For example, although the homoploid hybrid origin of *Nuphar x rubrodisca* was confirmed (using morphological and molecular data), the taxon was not considered to be a separate species because proof of reproductive isolation and independent evolution was lacking (Padgett *et al.*, 1998). Evidence of reproductive isolation may come from studying pre-mating reproductive barriers (crosses between hybrids and parents do not occur because of ecological, ethological, etc. separation) or post-mating reproductive barriers (crosses between hybrids and parents are unsuccessful).

#### *Ancient and recent hybrid species*

Following speciation, hybrid and parental lineages will accumulate genetic changes and, gradually, evidence of hybridity may be erased (Morrell and Rieseberg, 1998). Consequently, characteristics of recent hybrid species (discussed above) are not necessarily shared by ancient hybrid species (Wolfe and Elisens, 1994). For example, an ancient hybrid species may, over time, lose many parental alleles and gain new alleles through mutation. Therefore, an ancient hybrid species will not always exhibit a clear additive pattern of parental genetic markers and may possess several unique genetic markers, compared to its parents (Wolfe and Elisens, 1994; Morrell and Rieseberg, 1998). Identifying ancient hybrid species using morphological characters is likely to be even more difficult than when using molecular characters, because the former may diverge more rapidly than the latter (Morrell and Rieseberg, 1998).

Clearly, confirming the origin of putative ancient hybrids is difficult, and this can be illustrated by two examples: in the case of *Gilia achilleifolia*, molecular (isozymes, rDNA, cpDNA) and morphological data could not discriminate between the hypotheses of ancient hybrid origin and divergent evolution (Morrell and Rieseberg, 1998). In the second case, discordant organellar (cpDNA) and nuclear phylogenies (isozymes and rDNA) indicated that hybridisation had played a role in the evolution of *Gossypium bickii*, but the data were unable to distinguish between an ancient hybrid speciation event and introgression (Wendel *et al.*, 1991).

#### 5.4.2 Hybrid origin of British *S. squalidus*

Crisp (1972) and Abbott (Abbott and Milne, 1995; Abbott *et al.*, 1995 and Abbott *et al.*, 2000) hypothesised that British *S. squalidus* was a homoploid hybrid species. The evidence, from the present study and other sources, that supports this hypothesis will now be discussed.

##### *Molecular evidence*

To test the molecular criteria for a homoploid hybrid origin the postulated parental taxa must be extant and differentiated by a number of nuclear genetic markers (Allan *et al.*, 1997). Fortunately, in the case of British *S. squalidus* these conditions were met.

British *S. squalidus* possesses the additive profile of parental nuclear genetic markers that is characteristic of homoploid hybrid speciation. Individuals of the species combine taxon-specific RAPD and ISSR markers of *S. chrysanthemifolius* and *S. aethnensis*. The pattern of additivity observed is across loci, some loci in an individual have *S. chrysanthemifolius* markers, some *S. aethnensis* markers. In addition, British *S. squalidus* has an additive profile with respect to two allozyme markers, *Acp-2 a* and *b*, which are diagnostic markers for *S. chrysanthemifolius* and *S. aethnensis*, respectively. Some populations of British *S. squalidus* are monomorphic for the *b* allele, other populations contain individuals monomorphic for the *a* allele, individuals monomorphic for the *b* allele and heterozygous individuals containing both alleles (Abbott *et al.*, 2000 and Abbott, unpublished).

As mentioned previously (5.4.1), an additive profile of a few nuclear markers is insufficient to rule out alternatives to homoploid hybrid speciation, such as convergence,

symplesiomorphy and plesiomorphy. However, the additive nuclear profile of British *S. squalidus* consists of a relatively large number of RAPD and ISSR markers. Moreover, all individuals of British *S. squalidus* possess cpDNA haplotype B (Abbott *et al.* 2000), which is present in all *S. chrysanthemifolius* individuals and in a low percentage of *S. aethnensis* individuals (chapter 3) and British *S. squalidus* appears to have no, or very few, unique markers.

#### *Morphological evidence*

Crisp based his original proposal for the homoploid hybrid origin of British *S. squalidus* on the morphology of British *S. squalidus* and Sicilian *Senecio* material. A recent analysis has compared the morphology of these taxa again, but this time using a sophisticated multivariate analysis approach, that was not available to Crisp. Although the sample sizes were small, the principal components analysis (PCA) confirmed that most British *S. squalidus* was morphologically intermediate between *S. chrysanthemifolius* and *S. aethnensis* (fig. 2 in Abbott *et al.*, 2000).

#### *Other evidence*

British *S. squalidus* bears a morphological and molecular resemblance to naturally occurring hybrids between *S. chrysanthemifolius* and *S. aethnensis* on Mt. Etna. Many natural hybrids possess a sub-set of the morphological features typical of British *S. squalidus* and a smaller number of individuals bear a very strong likeness to British *S. squalidus* (personal observation). This morphological similarity between Sicilian hybrid material and British *S. squalidus* was illustrated by a PCA (fig. 2 in Abbott *et al.*, 2000). The genetic similarity between Sicilian hybrid material and British *S. squalidus* was demonstrated by a principal coordinate analysis of RAPD and ISSR data in the present study (fig. 5.5). Furthermore, crossing *S. chrysanthemifolius* and *S. aethnensis* produces hybrids that bear a likeness to both British *S. squalidus* and Sicilian hybrid material (personal observation).

#### *A new species?*

Introgression and hybrid swarm formation can be mistaken for hybrid speciation (see above). Indeed, naturally occurring hybrids between *S. chrysanthemifolius* and *S. aethnensis* on Mt. Etna share many of the characteristics of British *S. squalidus*, such as an

additive profile of nuclear genetic markers and morphological intermediacy (chapter 4). Crisp (1972) summed up the difference between these two situations as follows: "*material in Britain... despite being... analogous to S. X incisus* [Sicilian hybrid material], *can be treated as a separate species... because it is both geographically isolated from the parental species, and it has evolved over the course of many generations to a state where it is morphologically distinct from either of them, although still polymorphic and in general intermediate between them*". Molecular evidence agrees with morphology: British *S. squalidus* is genetically distinct from *S. chrysanthemifolius* and *S. aethnensis* (second axis, fig. 5.4), although still polymorphic and in general intermediate between them (first axis fig. 5.4, see below). In other words, British *S. squalidus* is a new species because it is isolated from (see below), and evolving independently of its parental taxa.

In summary, molecular data from the present study, morphological and molecular data from previous work and original herbarium documents (Crisp, 1972) confirm that British *S. squalidus* is of allopatric homoploid hybrid origin and is derived from material sampled from the hybrid zone, between *S. chrysanthemifolius* and *S. aethnensis*, on Mt. Etna, Sicily. However questions remain unanswered, for example, were Sicilian *Senecio* individuals introduced to the Oxford Botanic Garden on one occasion or more than one occasion?

### 5.4.3 Comparison of modes of homoploid hybrid speciation

If a homoploid hybrid derivative is to become a homoploid hybrid species, it must be isolated from parental gene flow. There are two ways in which this can occur. The first method is recombinational speciation, when the hybrid is isolated from its parents by chromosomal or genic incompatibility (Rieseberg, 1997; see section 1.3.6.1). Recombinational speciation is typified by the origin of three *Helianthus* hybrid species, isolated from their parents by chromosomal sterility factors (Rieseberg *et al.*, 1990b; Rieseberg, 1991; 1.3.6.1). The second method is speciation with external barriers to gene exchange, when the hybrid is geographically, ecologically or ethologically, etc. isolated from its parents (Grant, 1981; see section 1.3.6.1). For example, the hybrid species *Penstemon clevelandii* favours a different soil type to its parents (Wolfe and Elisens, 1994; Wolfe *et al.*, 1998).

British *S. squalidus* is geographically isolated from its parents, *S. chrysanthemifolius* and *S. aethnensis*. Moreover, *S. chrysanthemifolius* and *S. aethnensis* form vigorous fertile hybrid progeny when intercrossed and are interfertile with British *S. squalidus* (Forbes, personal communication). It is clear, therefore, that British *S. squalidus* must belong to the category of homoploid hybrid taxa formed by speciation with external barriers.

Other examples of homoploid hybrid species thought to be isolated by external barriers, and rigorously tested with molecular markers, are relatively few in number: *Pinus densata*, *Penstemon clevelandii*, *Encelia virginensis*, and six *Paeonia* species (1.3.6.1). Like British *S. squalidus*, *Pinus densata*, *Penstemon clevelandii* and *E. virginensis* have parents that can be easily intercrossed to produce (fertile?) hybrid offspring and *E. virginensis* is known to be interfertile with its parents. Although natural populations of these three species are frequently allopatric with their parents, they are also ecologically isolated from them. It is likely that *Pinus densata*, *Penstemon clevelandii* and *E. virginensis* originated in sympatry with their parental species. In this scenario, gene flow between parents and hybrids would have been limited by ecological isolation, and the allopatric populations would have arisen later (Abbott *et al.*, 2000). Abbott *et al.* (2000) suggested that the homoploid hybrid *Paeonia* species (Sang *et al.*, 1995) might have emerged in a similar fashion.

If the above assumptions are correct, British *S. squalidus* is rather unusual. Although the initial step in the formation of British *S. squalidus* was the hybridisation of sympatric parents, "*the crucial process of stabilisation resulting in the origin of a new hybrid species occurred at an allopatric site where there was no possibility of hybrid material backcrossing to either parent*" (Abbott *et al.*, 2000). In other words, British *S. squalidus* appears to have originated in allopatry rather than in sympatry (Abbott *et al.*, 2000) and the geographical isolation between *S. squalidus* and its parents made the development of ecological or ethological, etc. barriers unnecessary.

Although the original barrier to gene-flow, between British *S. squalidus* and its parents, was geographical distance, this does not preclude the indirect development of additional barriers to gene-flow. It is thought that *S. squalidus* became adapted to British conditions over a number of years (see section 5.4.6), suggesting that British *S. squalidus* may be unable to survive in parental habitats, and *vice versa*. Thus, British *S. squalidus* and its parents may, in effect, have become ecologically isolated from each other. In 300 years,

British *S. squalidus* has not become genetically or chromosomally incompatible with *S. chrysanthemifolius* and *S. aethnensis*. Nevertheless, such incompatibility may eventually develop, as a by-product of their continued independent evolution. However, unless man intervenes (and introduces British *S. squalidus* to Mt. Etna or *S. chrysanthemifolius* and *S. aethnensis* to Britain), it is highly unlikely that British *S. squalidus* and its parental taxa will come into contact. Consequently, such non-geographic barriers to gene-flow may not be put to the test.

Rieseberg (1997) suggested that "*hybrid genomes are likely to be stabilised quickly with little change thereafter*". This may be true of recombinational species such as *H. anomalus* (upon which Rieseberg's hypothesis was based), where selection may favour one or two hybrid types that exhibit high fertility, among semisterile early generations (Rieseberg *et al.*, 1995). In contrast, hybrid species isolated by external barriers may retain more variability, for longer, because they are not subject to intense selection for one or two optimal (fertile) types. This would appear to be the case with regard to British *S. squalidus*, which exhibits a reasonably high level of polymorphism, in terms of both morphological (Crisp, 1972) and molecular variation (this chapter).

#### **5.4.4 Proportion of the genome of British *S. squalidus* attributable to *S. chrysanthemifolius* and *S. aethnensis***

A phenogram, produced from isozyme data, showed that some populations of British *S. squalidus* clustered closely with *S. chrysanthemifolius* and certain Sicilian hybrid populations (Abbot *et al.*, 2000). This led to the suggestion that British *S. squalidus* might have acquired more of its genome from *S. chrysanthemifolius* than from *S. aethnensis*. Furthermore, the morphology of British *S. squalidus* is generally biased towards *S. chrysanthemifolius*; British *S. squalidus* individuals that resemble *S. aethnensis* appear to be fairly rare, whereas British *S. squalidus* individuals that resemble *S. chrysanthemifolius* are more common (personal observation). In the principal components analysis of morphological characters (fig. 2 in Abbott *et al.*, 2000) one specimen of British *S. squalidus* clustered with *S. chrysanthemifolius*.

The results of the present study concur with the previous findings of Abbott *et al.* (2000). Hybrid indices and principal coordinate analyses of RAPD and ISSR variation indicate that

although some British *S. squalidus* individuals are genetically intermediate between *S. chrysanthemifolius* and *S. aethnensis*, other individuals are genetically more similar to *S. chrysanthemifolius* than *S. aethnensis* (fig. 5.4). Indeed, PCO showed that several British *S. squalidus* individuals bore a genetic similarity to individuals from introgressed *S. chrysanthemifolius* hybrid populations (fig. 5.5). Moreover, British *S. squalidus* possesses cpDNA haplotype B that, although present at low frequency in *S. aethnensis*, is characteristic of *S. chrysanthemifolius*.

Other studies have also attempted to apportion the genome of a homoploid hybrid species between its parental progenitors. Arnold (1993) concluded that the hybrid species *Iris nelsonii* had inherited more of its genome from *I. fulva*, than from its other parental taxa, *I. hexagona* and *I. brevicaulis*. This inference was based upon the high frequency of *I. fulva* diagnostic RAPD markers in *I. nelsonii*, the low frequency of *I. hexagona* and *I. brevicaulis* diagnostic markers in the hybrid species and the fact that all *I. nelsonii* individuals possessed the cpDNA haplotype of *I. fulva*. A more sophisticated approach was taken by Rieseberg *et al.* (1993), who constructed a genetic linkage map for *Helianthus anomalus*, a species derived via hybridisation between *H. annuus* and *H. petiolaris*. Analysis of the parental origin of each RAPD marker in the map indicated that the genomic contributions of *H. annuus* and *H. petiolaris* to *H. anomalus* were roughly equivalent (of the 81 markers in *H. anomalus* of known parental origin, 54 per cent were attributable to *H. annuus* and 46 per cent to *H. petiolaris*).

In summary, RAPD, ISSR, isozyme and morphological data support the hypothesis that British *S. squalidus* has inherited more of its genome from *S. chrysanthemifolius* than from *S. aethnensis*. However, the evidence upon which this conclusion is based is rather sparse compared to that offered by, for instance, Rieseberg *et al.* (1993), above. Therefore, the data presented here should be treated as preliminary and further work needs to be undertaken. This could involve producing a saturated genetic linkage map of British *S. squalidus*, for example using RAPDs or AFLPs, and comparing it with maps of the two parent species. However, *S. squalidus* is genetically polymorphic (as shown by the level of polymorphism for band presence/absence at various RAPD loci) and so it might be difficult to construct a genetic map that is representative of the species as a whole. In contrast, the

genetic composition of *H. anomalus* (the hybrid taxon mapped by Rieseberg *et al.*, 1995) was implicitly assumed to be fairly constant across the species.

As an alternative to molecular markers, genomic *in situ* hybridisation (GISH; Bennett, 1995) could be employed. GISH is often described as "chromosome painting" and uses labelled genomic probes that fluoresce when attached to homologous areas on chromosomes. If successful, GISH can establish the parental origin of chromosomes in a hybrid species and, if sufficiently sensitive, reveal inter-genomic translocations (Thomas *et al.*, 1994). However, GISH is not an easy procedure to carry out; I attempted to apply the technique to the analysis of the genomic composition of British *S. squalidus*, but without success. Using GISH to investigate British *S. squalidus* may be difficult because the two parental species are fairly closely related and *Senecio* chromosomes are small (I. J. Leitch, personal communication). Few, if any, homoploid hybrid species have been successfully examined using GISH but the technique has been very effective in investigating the genomic contribution of parental taxa to polyploid hybrids (e.g. *xFestulpia hubbardii*; Bailey *et al.*, 1993) and polyploid hybrid species (e.g. *Milium montianum*; Bennett *et al.*, 1992).

#### 5.4.5 The genetic structure of British *S. squalidus*

The results of the AMOVA and Shannon's diversity index, for the six British *S. squalidus* populations indicated that most of the total variation was maintained within populations (~ 80%), rather than between populations (~ 20%). This was consistent with the expectation for an out-crossing species (Hamrick and Godt, 1990).

The level of genetic diversity, based upon RAPD and ISSR variation, within British *S. squalidus* was quantified using Shannon's index. It was shown that British *S. squalidus* had a higher level of genetic diversity than both *S. chrysanthemifolius* and *S. aethnensis*. One might expect this to be the case because a hybrid taxon will combine genomic elements from both parents. When compared to Sicilian hybrid populations, British *S. squalidus* had a genetic diversity roughly equivalent to introgressed *S. chrysanthemifolius* populations, such as S5 and S6, but lower than the most diverse hybrid populations, such as S10 and S11 (see chapter 4). That the level of genetic diversity in British *S. squalidus* was similar to that maintained in an introgressed hybrid population, rather than a genetically diverse

hybrid population could be due to a number of reasons. Firstly, the hybrid population from which the material brought to Britain was sampled may not, by chance, have included all markers from both parental species. Secondly, the genetic diversity of British *S. squalidus* may have been lowered by founder effects, the introduction of a limited number of hybrid individuals. Thirdly, the genetic base of British *S. squalidus* may have been further narrowed following its arrival in Britain. Genetic drift and inbreeding may have occurred in the Oxford Botanic Garden population especially if the population was small, as seems likely. Selection may also have caused the loss of certain markers, by removing those linked to non-adaptive genes.

Shannon's index results contrasted with those of the allozyme investigation carried out by Abbott *et al.* (2000). The allozyme study showed that level of genetic diversity (in terms of expected heterozygosity) within British *S. squalidus* was roughly one third of that reported in *S. aethnensis* and roughly half that of Sicilian hybrid material. This difference could be ascribed to variation in the number of loci sampled. The Shannon's index results were based upon 26 RAPD and ISSR loci, whereas the allozyme results were based upon just five polymorphic loci. This might suggest that the Shannon's index results were a more accurate reflection of reality. However, comparison between the RAPD and ISSR results and the allozyme results may be unjustified. The RAPD and ISSR markers were selected because they were present at high frequency in one parental species and low frequency in the other parental species, whereas the five isozyme loci were chosen at random. To make the comparison valid, one would have to use the Shannon's index procedure on a sample of randomly selected RAPD and ISSR loci.

The genetic variability of homoploid hybrid species has been discussed by Rieseberg (1997). Early researchers predicted that hybrid species would combine the alleles of parental taxa and would, therefore, have higher levels of genetic diversity than their parents (and an increased evolutionary potential). Rieseberg found that this expectation was not always borne out by isozyme data from confirmed hybrid species. *Pinus densata* was more variable than its parents, while *Iris nelsonii* and *Stephanomeria diegensis* had levels of genetic variation roughly equivalent to those of their respective parents. In contrast, three *Helianthus* hybrid species were less variable than their parents and this was ascribed to founder effects.

Populations of British *S. squalidus* were significantly genetically different from each other, based upon pair-wise  $\Phi_{ST}$  values from the AMOVA. However, when the pair-wise  $\Phi_{ST}$  values were used to construct UPGMA and neighbour-joining phenograms there was no obvious geographic pattern to the variation. Indeed, the phenograms were poorly resolved and differed noticeably in the placement of populations relative to one another (fig. 5.6). Likewise, principal coordinate analysis of the British *S. squalidus* individuals did not reveal any geographic structure to the distribution of populations (fig. 5.3).

*F*-statistics analysis of allozyme data by Abbott *et al.* (2000) produced similar findings.  $\theta$  was significantly different from zero indicating that there was significant population differentiation, but when a UPGMA tree was produced from Nei's genetic distances between populations no clear geographic pattern was evident (fig. 3 in Abbott *et al.*, 2000). Furthermore, the correlation of genetic distance between populations and geographic distance was not significant.

The occurrence of significant genetic differentiation between British *S. squalidus* populations, for isozyme and RAPD variation, is probably the product of long distance dispersal, founder effects and genetic drift in small populations. Alternatively, the pattern could result from selection acting to produce individuals adapted to divergent local conditions. Abbott *et al.* (2000) suggested that the lack of geographic structuring within British *S. squalidus*, i.e. the absence of isolation by distance is due to "*the recent massive range expansion of S. squalidus in the British Isles*".

#### **5.4.6 The success of British *S. squalidus***

*Senecio squalidus* is an extremely successful invasive species (Gray, 1986). In less than 200 years it has expanded from a single site to colonise most of non-rural Britain. Explanations for this success are discussed below:

##### *Adaptation to local conditions*

*S. squalidus* had been grown in the Oxford Botanic Garden for around 100 years before it began to appear in the wild. Crisp (1972) recognised that the sudden escape of *S. squalidus* from cultivation "*might indicate that a fundamental change had taken place*" in the taxon. He postulated that this change may have been brought about by selection producing *S.*

*squalidus* individuals adapted to the British environment. Unfortunately, we do not know what genetic changes might have taken place in British *S. squalidus* to have led to the evolution of such adapted material (Abbott *et al.*, 2000).

The genetic variation, upon which selection could act, may have been generated by a number of processes. *S. squalidus* maintains a relatively high level of genetic diversity (presumably because of its hybrid origin, see section 5.4.5) and this existing variation would have been released by segregation and recombination. Crisp presumed that mutation might also have played an important role in producing variation. There is little evidence to support this proposal, although some potentially unique markers were found in the British taxon (see section 5.3.1). Crisp dismissed introgression, between *S. squalidus* and another taxon in the Botanic Garden, as a possible source of variation because of the lack of evidence of such (see below).

It would be interesting to test the hypothesis of adaptation to British conditions. One possibility would be to carry out a common garden experiment in Britain and monitor the comparative survival and fertility of *S. chrysanthemifolius*, *S. aethnensis*, Sicilian hybrid and British *S. squalidus* material. However, assessing what genetic changes might have taken place in *S. squalidus*, whilst in the Garden, would be difficult.

#### *Availability of suitable habitat*

The successful, and comparatively rapid, spread of *S. squalidus* in the British Isles has been attributed, as previously mentioned, to the development of the railway system (Crisp, 1972). Railway clinker-ash may resemble the volcanic substrate of the native habitat of its parents (Druce, 1927; Mabey, 1997), both are free-draining, similarly textured and have a high sulphur content. In more recent years, *S. squalidus* has been observed migrating along the road network in Britain. The roadside environment may also bear a resemblance to the native habitat of the parents of *S. squalidus* (e.g. due to sulphur from exhaust pollution, etc.; Cook, 1998). The successful establishment of *S. squalidus* stems from its ability to thrive in habitats that few native plants seem able to exploit - railway tracks and roadside verges, urban walls - and disturbed, open sites where there is little competition from other species - car-parks, urban waste-ground, etc.. Therefore, not only does *S. squalidus* owe its origin to man (in introducing the hybrid Sicilian material to Britain) but its subsequent

spread and establishment has been to a large extent dependent on man (with *S. squalidus* distributed via, and colonising, man-made habitats).

### *Breeding system*

Uni-parental reproduction, by selfing or asexual reproduction, has often been associated with colonising success in plants (Gray, 1986). Uni-parental reproduction is advantageous to colonists because an individual can reproduce when mates are absent or rare, a common situation during the early stages of colonisation (Gray, 1986). Abbott and Forbes (1993) measured the level of seed set in selfed and crossed individuals of British *S. squalidus*. Seed set on selfing was very low, indicating that the self-incompatibility system of British *S. squalidus* is extremely effective; moreover British *S. squalidus* does not reproduce asexually. How, then, can we explain its successful invasion of Britain?

Crisp (1972) suggested that a genetically controlled change might have occurred in the breeding system of *S. squalidus*, during its period of cultivation. Thus self-incompatible British *S. squalidus* may have tended towards self-compatibility (either through mutation of incompatibility alleles or an alteration in the overall expression of the incompatibility system). Crisp thought it probable that "*occasional leakage in the self-incompatibility system allowed the establishment of self-maintaining populations of the taxon, but also allowed a considerable degree of heterozygosity to persist*". This theory was based upon the belief that British *S. squalidus* demonstrated weaker self-incompatibility than *S. chrysanthemifolius* and *S. aethnensis*. However, the crossing experiments from which this conclusion was drawn were deeply flawed (Crisp used inter-specific rather than intra-specific crosses, Crisp, 1972, p248). Furthermore, Forbes (personal communication) noticed no difference in self-incompatibility between British *S. squalidus* and its parents.

Abbott and Forbes (1993) considered the question from a different perspective to Crisp. Instead of proposing a breakdown in self-incompatibility, they suggested the colonising success of *S. squalidus* could be attributed to a life-history characteristic. British *S. squalidus* is a short-lived perennial and can flower in more than one year. This means that a new colony of *S. squalidus* can expand and establish itself (without the need for sexual reproduction) by means of seed immigration, over two or more years. Eventually, a sufficient number of *S* alleles may amass in the population to enable cross-fertilisation.

Abbott and Forbes' theory is very plausible. Nevertheless, they recommended that a future investigation should consider whether *S. squalidus*'s self-incompatibility can breakdown under certain environmental conditions.

#### 5.4.7 *Senecio rupestris* - a confounding factor

In 1963, Walters (1963) drew attention to the morphological similarity between British *S. squalidus* and *S. rupestris* Waldst. and Kit., a closely related species of mountainous areas in central and southern Europe. Crisp pointed out that there was "no evidence to suggest that... [*S. rupestris* has] contributed to the British taxon, or that they own a directly common origin" (Crisp, 1972). However, because of the morphological similarity between British *S. squalidus* and *S. rupestris*, he suggested that both species be included in the same species complex. The species complex would be referred to as the *S. squalidus* aggregate (and would include *S. chrysanthemifolius* and *S. aethnensis*). Alexander (1979) subsumed British *S. squalidus* and *S. rupestris* under *S. squalidus* (in his taxonomic revision of Mediterranean *Senecio*, section *Senecio*). Alexander noted that British *S. squalidus* was atypical of the species, but attributed this to founder effects and genetic drift in British populations.

The morphological similarity between *S. rupestris* and British *S. squalidus* led Abbott (Abbott and Milne, 1995; Abbott *et al.*, unpublished) to speculate on the possible connections between *S. rupestris* and British *S. squalidus*: Could British *S. squalidus* have been derived from *S. rupestris* rather than from Sicilian hybrid *Senecio* material? Could British *S. squalidus* have been derived from *S. rupestris* as well as from Sicilian material; for example through introgression of *S. rupestris* and Sicilian material in the Oxford Botanic Garden? Crisp (above) had pointed out that there was no documentary evidence to support the theory that *S. rupestris* was, in any way, involved in the origin of British *S. squalidus* (no herbarium sheets or records). Nevertheless, an attempt was made to clarify the relationship between British *S. squalidus* and *S. rupestris*; by comparing the morphology, cpDNA and allozyme variation of the two species.

An initial survey of cpDNA revealed that most individuals of *S. rupestris* possessed haplotype A or C (Abbott *et al.*, 1995). In contrast, British *S. squalidus* was monomorphic for haplotype B, which was also present in *S. chrysanthemifolius*, *S. aethnensis* and Sicilian

hybrid material. However, one individual of *S. rupestris*, from central Italy, also possessed haplotype B (Abbott *et al.*, 1995). This prompted analysis of more samples of *S. rupestris* from central Italy and the Balkans. Three more samples from central Italy and two of 28 samples from the Balkans were also found to possess haplotype B (Abbott, personal communication). However, even if haplotype B had not been found in *S. rupestris*, cpDNA on its own could not rule out the possibility that *S. rupestris* was involved in the origin of British *S. squalidus*, as a pollen parent.

By eye, British *S. squalidus* and *S. rupestris* can be correctly distinguished 100% of the time (Forbes, personal communication). A morphometric analysis, based on 32 floral and vegetative characters, was conducted on material of *S. rupestris*, British *S. squalidus*, *S. chrysanthemifolius*, *S. aethnensis* and the Sicilian hybrid (James, 1995), and the data subjected to canonical variate analysis (CVA) (Abbott *et al.*, submitted). Canonical variates one and two clearly separated populations (groups) of *S. rupestris* from populations of the other *Senecio* taxa. Subsequently, CVA was repeated 10 times and on each occasion data of a different *S. squalidus* plant was omitted. The omitted plant was then assigned to a population on the basis of group discriminant functions. None of the *S. squalidus* plants were assigned to *S. rupestris* populations; instead two individuals were assigned to the Sicilian hybrid group, seven were assigned to British *S. squalidus* groups and one was assigned to *S. chrysanthemifolius* (Abbott *et al.*, submitted).

The comparison of allozyme variation in *S. rupestris* and British *S. squalidus* was most informative (Abbott *et al.*, submitted; James 1995). It revealed that the *Acp-2 b* allele, which occurred at high frequency in British *S. squalidus*, was absent from *S. rupestris*. Thus, British *S. squalidus* could not have been derived from *S. rupestris* alone.

It was not possible to state, categorically, that *S. rupestris* had played no part in the lineage of British *S. squalidus*. The majority of allozyme alleles found in British *S. squalidus* occurred in *S. rupestris*, and cpDNA haplotype B, which was fixed in British *S. squalidus*, was also present at low frequency in *S. rupestris*. However, Abbott *et al.* have concluded that, taken together, the results of the morphological and isozyme (*Acp*) analyses, combined with the absence of any known documentary evidence of an introduction of *S. rupestris* to

the British Isles, provide a strong argument against the possibility of *S. rupestris* having contributed to British *S. squalidus*.

Abbott *et al.*'s hypothesis (i.e. that *S. rupestris* has not contributed to British *S. squalidus*) could be tested with RAPDs and ISSRs. One approach, would be to develop RAPD and ISSR markers that are present in *S. rupestris* but not in *S. chrysanthemifolius*, *S. aethnensis* or hybrid Sicilian material. If their hypothesis is correct, none of these *S. rupestris* markers should be found in British *S. squalidus*.

It would also be possible to survey *S. rupestris* with the *S. chrysanthemifolius* and *S. aethnensis* markers used in the present study, to test whether those markers that are present in British *S. squalidus* are absent from *S. rupestris*. However, the *S. chrysanthemifolius* and *S. aethnensis* markers were developed without reference to *S. rupestris*. Therefore, the presence of such markers in both British *S. squalidus* and *S. rupestris* could mean either: (i) *S. rupestris* has contributed to British *S. squalidus* or (ii) alternatively, that *S. rupestris* possesses the same band as *S. chrysanthemifolius* or *S. aethnensis* due to shared ancestry of the species. It would not be possible to distinguish between these two scenarios. Thus, it is clear that the most effective way to test Abbott *et al.*'s hypothesis would be to use the first approach suggested.

Recently, M. Coleman (personal communication) has employed ISSRs to re-examine the relationship between *S. rupestris* and British *S. squalidus*. UPGMA clustering of the ISSR data (46 polymorphic bands, from 10 primers) produced a dendrogram with two main clusters. One cluster contained all individuals of *S. rupestris* (from 10 populations), while the other cluster contained British *S. squalidus*, *S. chrysanthemifolius*, *S. aethnensis* and artificial F<sub>1</sub>s between *S. chrysanthemifolius* and *S. aethnensis*. Moreover, five markers found in *S. rupestris* were absent from British *S. squalidus* and the Sicilian taxa; and two markers found in the Sicilian taxa were present in British *S. squalidus* but absent from *S. rupestris*. Coleman's study provides further evidence that *S. rupestris* was not involved in the origin of British *S. squalidus*. Consequently, I suggest that *S. rupestris* should not be aggregated with British *S. squalidus*, as in Alexander (1979), but should be recognised as a distinct species.

# **Chapter 6**

**A Preliminary Study of the Genomic Composition  
of Hybrid Derivatives of *Senecio squalidus*  
and *Senecio vulgaris* var. *vulgaris***

## **A Preliminary Study of the Genomic Composition of Hybrid Derivatives of *Senecio squalidus* and *Senecio vulgaris* var. *vulgaris***

### **6.1 Introduction**

#### **6.1.1 Hybridisation between *S. squalidus* and *S. vulgaris* var. *vulgaris*, common groundsel**

During its spread through the British Isles, *S. squalidus* has frequently encountered *S. vulgaris* var. *vulgaris*, a native, non-radiate groundsel. The two species hybridise and this has led to the evolution of three new hybrid taxa: *S. vulgaris* var. *hibernicus* (Abbott *et al.*, 1992a), York Radiate Groundsel (Irwin and Abbott, 1992) and *S. cambrensis* (Ashton and Abbott, 1992a).

#### **6.1.2 F<sub>1</sub> hybrids between *S. squalidus* and *S. vulgaris* var. *vulgaris***

##### *Crossing studies*

A number of studies (e.g. Ingram *et al.*, 1980) have successfully crossed diploid *S. squalidus* ( $2n = 20$ ) and tetraploid *S. vulgaris* var. *vulgaris* ( $2n = 40$ ) to produce F<sub>1</sub> hybrids. Successful production of F<sub>1</sub>s is normally dependent on *S. vulgaris* acting as the maternal parent, and the F<sub>1</sub>s are usually triploid ( $2n = 30$ ) (Ingram *et al.*, 1980). Although highly sterile, some triploid F<sub>1</sub>s may yield occasional gametes with a balanced chromosome set, allowing production of F<sub>2</sub> and backcross progeny (to both parents). It is, therefore, theoretically possible that introgression could take place in both directions between *S. squalidus* and *S. vulgaris*. Moreover, fertile hybrid progeny might be created not only at the same ( $2x = 20$ ,  $4x = 40$ ) ploidy levels as the parents but at different ( $5x = 50$ ,  $6x = 60$ ) ploidy levels as well (Abbott and Lowe, 1996).

In addition to the triploid hybrids, two tetraploid F<sub>1</sub> hybrids of *S. squalidus* and *S. vulgaris* have been synthesised (one by Taylor, 1984; one by Lowe, 1996). These were assumed to have been produced by the fusion of an unreduced *S. squalidus* gamete with a normal *S. vulgaris* gamete. Tetraploid F<sub>1</sub>s exhibit reasonable fertility when selfed and are interfertile with *S. vulgaris* (Lowe, 1996).

*F*<sub>1</sub> hybrids in the wild

Although *S. vulgaris* and *S. squalidus* often grow together on waste-ground in the British Isles, hybridisation between the two species in the wild is uncommon. Triploid *F*<sub>1</sub> individuals, known as *S. x baxteri*, are formed regularly but at very low frequency. A survey of four sites containing large mixed populations of *S. vulgaris* and *S. squalidus*, yielded only eight triploid hybrids, compared to a total of roughly 30000 recorded *S. vulgaris* individuals. In addition, approximately 16000 offspring of *S. vulgaris* from the same four sites were examined and only two triploid hybrids were discovered (Marshall and Abbott, 1980). It would appear (from herbarium data, botanical reports and crossing studies) that tetraploid *F*<sub>1</sub>s occur at an even lower frequency than *S. x baxteri* individuals (Lowe, 1996).

**6.1.3 Origin of *S. vulgaris* var. *hibernicus*, radiate groundsel**

*Senecio vulgaris* var. *hibernicus* ( $2n = 40$ ) is a radiate form of var. *vulgaris*, producing capitula with an outer whorl of ray florets in contrast to var. *vulgaris*. *Senecio vulgaris* var. *hibernicus* was first described in the British Isles in 1866 (Syme, 1875; in Abbott and Lowe, 1996), from material growing around Cork, Eire. However, Crisp (1972) later identified a herbarium specimen of radiate groundsel, collected from Oxford in 1832. By the end of the 19<sup>th</sup> century, var. *hibernicus* was well established in Oxford, southern Ireland, Bristol/Cardiff and Cheshire/north-west Wales (Crisp, 1972) and the variant has subsequently extended its range across much of the British Isles (Lowe, 1996).

Two hypotheses were proposed to explain the origin of *S. vulgaris* var. *hibernicus* (Ingram *et al.*, 1980). The first suggested that var. *hibernicus* arose from var. *vulgaris* through a mutation in the single gene causing ray floret presence and absence in *Senecio*. The second hypothesis suggested that var. *hibernicus* was the product of introgression between radiate *S. squalidus* and var. *vulgaris*. Initial evidence favoured the hybridisation hypothesis. The spread of *S. vulgaris* var. *hibernicus* across the British Isles was found to correspond to the spread of *S. squalidus*, following its escape from the Oxford Botanic Gardens in the late 1700s (Crisp, 1972). Moreover, Monaghan and Hull (1976; in Lowe, 1996) claimed that some var. *hibernicus* individuals exhibited morphological features similar to those of *S. squalidus*. However, Stace (1977) criticised both the substance and interpretation of these data, concluding that it was not possible to choose between the two hypotheses. Since then,

stronger evidence in support of the hybridisation theory has emerged. Fertile tetraploid plants that bore a close resemblance to *S. vulgaris* var. *hibernicus* were synthesised by backcrossing an artificial triploid hybrid between *S. squalidus* and var. *vulgaris*, to var. *vulgaris* and then selfing the products (Ingram *et al.*, 1980). More recently, a survey of allozyme variation at the *Aat-3* locus revealed that the *c* allele, which occurs at high frequency in British *S. squalidus*, was present at intermediate frequency in *S. vulgaris* var. *hibernicus* but was very rare among *S. vulgaris* var. *vulgaris* individuals sympatric with var. *hibernicus* individuals, and was absent from monomorphic populations of var. *vulgaris* (Abbott *et al.*, 1992a). Further genetic analysis demonstrated that the *Aat-3* locus and the ray floret locus segregated independently. This implies that genes on more than one chromosome have been introgressed into *S. vulgaris* from *S. squalidus*, and that these genes have tended to remain associated in var. *hibernicus*, possibly due to selection favouring a co-adapted gene-complex or positive assortative mating (Abbott *et al.*, 1992a).

Having produced evidence in support of the hybridisation hypothesis, Abbott *et al.* (1992a) considered the pathway of origin of var. *hibernicus*. Var. *hibernicus* could be derived from a tetraploid F<sub>1</sub>, between *S. vulgaris* and *S. squalidus*, followed by backcrossing to *S. vulgaris* or from a sterile triploid F<sub>1</sub>, *S. x baxteri*, followed by backcrossing to *S. vulgaris* and the resumption of tetraploidy (as suggested by the crossing studies of Ingram *et al.*, 1980). However, no decision was made on which of these two alternatives was the most likely. A herbarium study (Crisp, 1972) and an examination of var. *hibernicus* cpDNA (Abbott and Lowe, 1996) indicate that the taxon may have arisen independently on at least three occasions.

Var. *hibernicus* has been shown to contain less variation for esterases and a range of morphological and life history traits than var. *vulgaris* (Abbott, 1986; Abbott *et al.*, 1992b), which Abbott and Lowe (1996) argued was to be expected due to the recent hybrid origin of var. *hibernicus*. They suggested that, over time, genetic diversity in var. *hibernicus* might be boosted by increased maternal outcrossing of the radiate variant relative to non-radiate var. *vulgaris*. However, higher outcrossing of var. *hibernicus* individuals places them at an automatic disadvantage, relative to var. *vulgaris* individuals, with regard to transmitting their genome to their progeny (the 'cost of outcrossing') (Abbott *et al.*, 1992a). Nevertheless, this difficulty does not appear to have hindered the spread of var. *hibernicus*.

#### 6.1.4 Origin of York Radiate Groundsel

York Radiate Groundsel is a fertile, tetraploid ( $2n = 4x = 40$ ), radiate groundsel. It was discovered in York in 1979 and probably arose in the city post-1950, following establishment of *S. squalidus* in the area (Lowe, 1996). The status of York Radiate Groundsel as a hybrid between *S. squalidus* and *S. vulgaris* was confirmed by Irwin and Abbott (1992) and Lowe (1996). Firstly, York Radiate Groundsel was found to be morphologically intermediate between *S. squalidus* and *S. vulgaris*, although it also possesses a number of novel characters such as four pored pollen and large seed (Irwin and Abbott, 1992; Lowe, 1996). Secondly, York Radiate Groundsel exhibits an additive isozyme profile, combining the  $\beta Est-1$  phenotype characteristic of *S. squalidus*, with the  $\alpha Est-1$  phenotype characteristic of *S. vulgaris* (Irwin and Abbott, 1992). Finally, the results of mitotic chromosome, rDNA and cpDNA analyses were consistent with the hybrid origin of York Radiate Groundsel (Lowe, 1996).

York Radiate Groundsel differs from *S. vulgaris* var. *hibernicus*; in that var. *hibernicus* does not possess the  $\beta Est-1$  phenotype. Together with morphological evidence, this suggests that York Radiate Groundsel has received a greater proportion of its genome from *S. squalidus*, than has var. *hibernicus*.

Attempts to resynthesise York Radiate Groundsel, combined with herbarium and other data, enabled Lowe (1996) to ascertain the most likely pathway of origin for York Radiate Groundsel. He concluded that the taxon was probably derived from a fertile tetraploid  $F_1$  between *S. squalidus* and *S. vulgaris*, followed by episodes of backcrossing to one of the parents, possibly *S. vulgaris*.

A strong post-zygotic barrier exists between *S. squalidus* and York Radiate Groundsel, but the latter is interfertile with *S. vulgaris*. Consequently, the maintenance of its distinctive morphology and isozyme profile over a number of years implies that the taxon must, to some extent, be reproductively isolated from *S. vulgaris* in the field. This reproductive isolation has been attributed to spatial separation, the ability of York Radiate Groundsel to self, and ecological and phenological separation e.g. differences in germination behaviour and flowering time between York Radiate Groundsel and var. *vulgaris* (Irwin and Abbott, 1992; Lowe, 1996). However, the future of York Radiate Groundsel is precarious because

it is endemic to only a few waste-ground sites in York, which might be re-developed in the future (Lowe, personal communication).

### 6.1.5 Origin of *S. cambrensis*, the Welsh groundsel

*Senecio cambrensis* is a new British allohexaploid ( $2n = 6x = 60$ ) species. It was first described by Rosser (1955; in Ashton and Abbott, 1992a), after being discovered in 1948 near Wrexham, north Wales. The species is well established as a ruderal in the Wrexham area and has been reported from Mochdre in north Wales, Shropshire and Leith, Edinburgh (Lowe and Abbott, 1996).

By treating artificial triploid hybrids between *S. squalidus* and *S. vulgaris* with colchicine, allohexaploids that closely resemble *S. cambrensis* can be produced (Ingram and Noltie, 1995). It is believed that *S. cambrensis* evolved in a similar fashion, through the polyploidization of the sterile triploid *S. x baxteri*, and that *S. cambrensis* probably arose sometime after 1910, when *S. squalidus* reached the Wrexham area (Ashton and Abbott, 1992a). *Senecio cambrensis* is intermediate between its parents for many characters but produces larger seed and pollen with four pores instead of three; the species is self-compatible, with high seed set on selfing. A comprehensive description of *S. cambrensis* and its morphology, ecology, etc. can be found in Ingram and Noltie (1995).

There is good isozyme evidence for multiple origins of *S. cambrensis*. Variation at the *Acp-1* and  $\alpha$ *Est-1* loci support separate origins of Welsh and Edinburgh *S. cambrensis* (Ashton and Abbott, 1992a). Furthermore, variation at the *Aat-3* locus suggests that two additional independent origins of the species may have occurred in Wales, one in Mochdre and one in Wrexham. Alternatively, variation at the *Aat-3* and  $\beta$ *Est-3* loci could be explained by segregation from a fixed heterozygote following pairing between homoeologous chromosomes (Ashton and Abbott, 1992a). Additional evidence for separate origins of Welsh and Edinburgh populations was obtained by surveying cpDNA; material from Wales contained an insertion of ~330 bp, but this was absent from Scottish samples (Harris and Ingram, 1992; Lowe and Abbott, 1996). It is interesting to note that this insert was not found in British *S. squalidus* or *S. vulgaris* var. *vulgaris*. However, the insert was detected in two British var. *hibernicus* individuals, suggesting that it may be present in British var. *vulgaris*, the presumed maternal parent of *S. cambrensis*, albeit at very low frequency

(Lowe and Abbott, 1996). Unfortunately, Edinburgh *S. cambrensis* may have become extinct since its origin, due to loss of suitable habitat (waste-ground) in the area where it occurred (Abbott, personal communication).

#### **6.1.6 RAPD variation in the hybrid derivatives of *S. squalidus* and *S. vulgaris* var. *vulgaris***

Lowe (1996) surveyed RAPD variation in var. *hibernicus*, York Radiate Groundsel, *S. squalidus* and var. *vulgaris* and found that York Radiate Groundsel individuals exhibited additive RAPD profiles, that combined bands characteristic of *S. squalidus* and var. *vulgaris*. In contrast, RAPD profiles of var. *hibernicus* individuals resembled those of var. *vulgaris*. Lowe's survey was based on a very small number of individuals, 34 in total. More significantly, approximately 25% of Lowe's RAPD reactions failed (101 amplifications out of a total of 408 amplifications). In comparison, the average failure rate in the present study (chapters 2 to 5) was less than 3%. This suggests that Lowe's survey should be considered preliminary. The amplification problems experienced by Lowe were possibly due, in part, to the method of DNA extraction employed. He used a very rapid procedure, based upon NaOH, that did not remove DNases, RNA and contaminants such as polyphenols. It is likely that the presence of these substances (which are removed by more thorough extraction techniques, such as the one used in the present study) affected the RAPD PCR reactions (Weising *et al.*, 1995).

#### **6.1.7 Aim:**

The aim of the study reported in this chapter was to determine whether *S. chrysanthemifolius* and *S. aethnensis* markers present in British *S. squalidus* (chapter 5) are also present in the presumed products of hybridisation between *S. squalidus* and *S. vulgaris* var. *vulgaris*, namely *S. vulgaris* var. *hibernicus*, *S. cambrensis*, and York Radiate Groundsel. Differential possession of the *S. chrysanthemifolius* and *S. aethnensis* markers in *S. vulgaris* var. *hibernicus*, *S. cambrensis*, and York Radiate Groundsel might provide useful information about the genomic contribution of *S. squalidus* to the three hybrid taxa. However, only a few *S. chrysanthemifolius* and *S. aethnensis* markers were available for the analysis and so it was decided to broaden the study. Consequently, the presence/absence in hybrid derivatives of additional markers characteristic of *S. squalidus*, and markers characteristic of *S. vulgaris* var. *vulgaris* was also determined.

## 6.2 Materials and methods

### 6.2.1 RAPD and ISSR survey of *S. vulgaris* var. *vulgaris*, *S. vulgaris* var. *hibernicus*, *S. cambrensis* and York Radiate Groundsel

RAPD and ISSR analysis was carried out using procedures outlined in section 2.8.

#### 6.2.1.1 Analysis of isolated *S. vulgaris* var. *vulgaris*

Eighteen *S. vulgaris* var. *vulgaris* individuals were chosen from populations believed to be isolated from *S. squalidus* and var. *hibernicus* (referred to henceforth as isolated *S. vulgaris* var. *vulgaris*). These individuals were grown from seed collected from plants at Tomintoul (three individuals), Letham Angus (three individuals), Glenluce (two individuals), Puffin Island (five individuals) and Aberffraw (five individuals) (table 2.2). DNA extracts of these plants were surveyed with the 19 RAPD and ISSR primers (table 4.5) used to examine variation within British *S. squalidus* (chapter 5). Standards were run on RAPD and ISSR gels as detailed in section 4.2.2.

For each primer, it was ascertained whether there were bands present in isolated var. *vulgaris* individuals of similar size to the *S. chrysanthemifolius* and/or *S. aethnensis* marker band(s). If this was so, samples were re-run and/or homology tests (2.9, plate 6.3) conducted to determine if the isolated var. *vulgaris* bands were identical to the *S. chrysanthemifolius* and/or *S. aethnensis* marker bands.

#### 6.2.1.2 Analysis of hybrid derivatives of *S. squalidus* and *S. vulgaris*, using *S. chrysanthemifolius* and *S. aethnensis* markers

*Senecio chrysanthemifolius* and *S. aethnensis* markers that were present at reasonable frequency in *S. squalidus*, and which could not be confused with bands in isolated *S. vulgaris* var. *vulgaris*, were examined for presence/absence in the following material (see table 2.2 for population details):

(i) *Senecio vulgaris* var. *hibernicus* and sympatric var. *vulgaris* (referred to as non-isolated var. *vulgaris*).

A total of 30 *S. vulgaris* var. *vulgaris* individuals and 27 var. *hibernicus* individuals from seven populations polymorphic for capitulum type were examined: Edinburgh, Bristol,

Cork (five individuals of var. *vulgaris* and five of var. *hibernicus* from each population), Cardiff (four individuals of var. *vulgaris*, five of var. *hibernicus*), Birmingham (five individuals of var. *vulgaris*, three of var. *hibernicus*), Brymbo (three individuals of var. *vulgaris*, two of var. *hibernicus*) and St. Helens (three individuals of var. *vulgaris*, two of var. *hibernicus*).

(ii) York Radiate Groundsel, plus var. *vulgaris* from the same population.

Ten individuals of York Radiate Groundsel from one population were examined, along with five individuals of *S. vulgaris* var. *vulgaris*.

(iii) *Senecio cambrensis*.

Thirty four individuals of *S. cambrensis* were examined. Thirteen individuals were from populations in North Wales - Ffrith (one individual), Southsea (three individuals) Wrexham (four individuals), Mochdre (five individuals) - and the remaining 21 individuals were from Leith, Edinburgh.

Standards were run on the RAPD and ISSR gels as detailed in section 4.2.2.

### ***6.2.1.3 Analysis of hybrid derivatives of S. squalidus and S. vulgaris, using additional markers characteristic of S. squalidus and markers characteristic of var. vulgaris***

Results from section 6.2.1.2, suggested that only a limited amount of information could be obtained concerning presence/absence of *S. chrysanthemifolius* and *S. aethnensis* markers in hybrid derivatives of *S. squalidus* and *S. vulgaris* var. *vulgaris*. Therefore, the analysis was broadened to determine the presence/absence in these derivatives of additional markers that characterised *S. squalidus*, and markers that characterised *S. vulgaris* var. *vulgaris*.

To this end, reference gels were produced for the primers used in 6.2.1.2, containing individuals of all relevant *Senecio* taxa. Each reference gel contained the following individuals: six *S. chrysanthemifolius* individuals, six *S. aethnensis* individuals, ten *S. squalidus* individuals (from the six populations described in chapter 5), six isolated *S. vulgaris* var. *vulgaris* individuals, six var. *hibernicus* individuals, six York Radiate Groundsel individuals and ten *S. cambrensis* individuals (six from Wales, four from Edinburgh). All of these individuals had been scored before (above or in chapter 5).

Using these reference gels, it was possible to pick out "var. *vulgaris*" (VU) marker bands and "*S. squalidus*" (SQ) marker bands. "Var. *vulgaris*" marker bands were present in

isolated *S. vulgaris* var. *vulgaris* but absent from *S. squalidus*, while "*S. squalidus*" marker bands were present in *S. squalidus* and either *S. chrysanthemifolius* and/or *S. aethnensis* but absent from isolated var. *vulgaris*. These "*S. squalidus*" bands were in addition to the *S. chrysanthemifolius* and *S. aethnensis* marker bands employed in section 6.2.1.2.

The reference gels aided the scoring of "var. *vulgaris*" and "*S. squalidus*" marker bands across the various *Senecio* taxa on different gels. Therefore, presence/absence of these marker bands could be recorded in non-isolated *S. vulgaris* var. *vulgaris*, var. *hibernicus*, York Radiate Groundsel and *S. cambrensis* by re-examining the gels from 6.2.1.2. Similarly, presence/absence of "var. *vulgaris*" and "*S. squalidus*" marker bands in isolated var. *vulgaris* was recorded by re-examining the gels for isolated var. *vulgaris* described in 6.2.1.1. Finally, presence/absence of "var. *vulgaris*" and "*S. squalidus*" marker bands in *S. squalidus*, *S. chrysanthemifolius* and *S. aethnensis* was recorded by examining the profiles of these taxa in the reference gels. Previously scored *S. squalidus* gels (chapter 5) were checked to ensure that individuals of this taxon on the reference gels were representative of the species as a whole.

## 6.3 Results

See plates 6.1. and 6.2 for example RAPD gels.

### 6.3.1 Screen of isolated *S. vulgaris* var. *vulgaris*

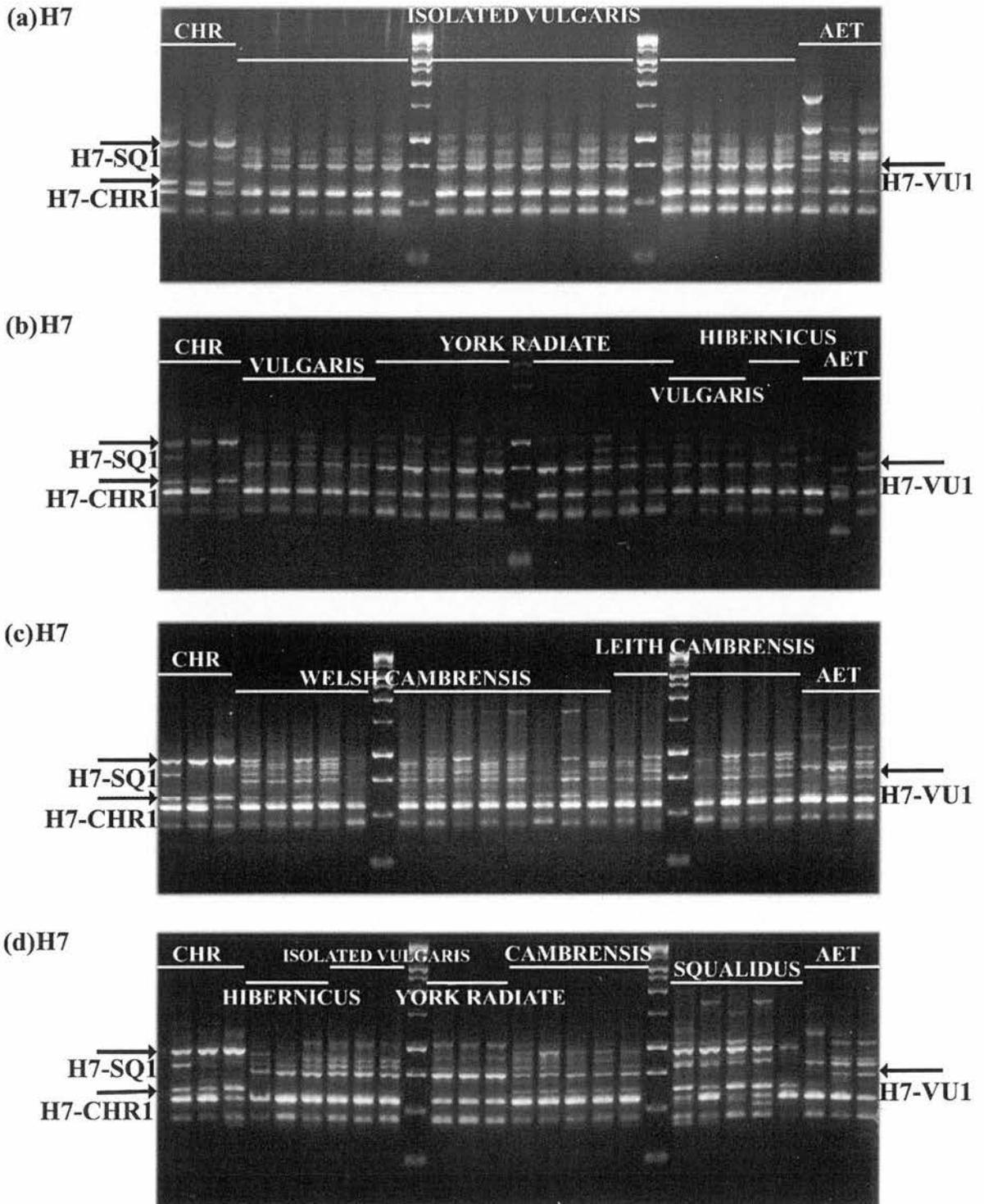
Two of the 26 *S. chrysanthemifolius* and *S. aethnensis* marker bands surveyed - G8 AET-1 and G11-AET-1 - appeared to be present in isolated *S. vulgaris* var. *vulgaris* (table 6.1). In addition, seven of the 26 marker bands were of similar size to bands in isolated var. *vulgaris*. However, running-out relevant samples for longer on gels and/or conducting homology tests on the bands, confirmed that these seven marker bands were not present in isolated var. *vulgaris* (table 6.1). The remaining 17 marker bands were absent from isolated *S. vulgaris* var. *vulgaris* and not easily confused with bands in this material.

### 6.3.2 Presence/absence of *S. chrysanthemifolius* and *S. aethnensis* marker bands in the hybrid derivatives

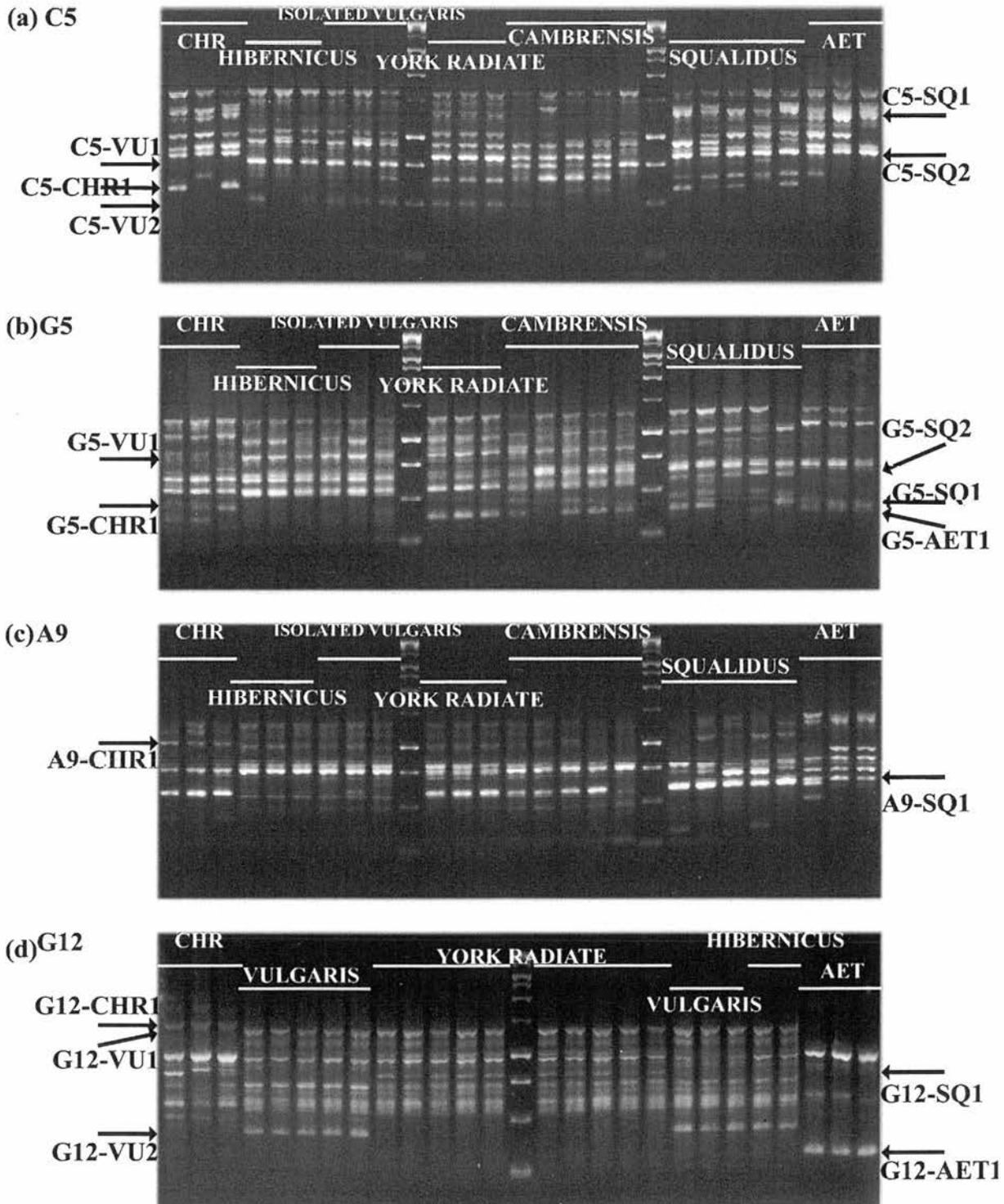
Of the 17 *S. chrysanthemifolius* and *S. aethnensis* marker bands that were absent from isolated *S. vulgaris* var. *vulgaris* and not easily confused with var. *vulgaris* bands, 11 were present in British *S. squalidus* at reasonable frequency (i.e. not including F12 AET-1; table 6.1). Six of these were *S. chrysanthemifolius* markers - A9 CHR-1, C5 CHR-1, F12 CHR-1, G5 CHR-1, H7 CHR-1 and 40 CHR-1 - and five were *S. aethnensis* markers - C2 AET-1, C20 AET-1, G6 AET-1, G12 AET-1 and 18-AET-1 (table 6.2).

None of these 11 *S. chrysanthemifolius* and *S. aethnensis* markers were recorded as present in the populations surveyed of non-isolated *S. vulgaris* var. *vulgaris* and *S. vulgaris* var. *hibernicus* (table 6.2). One *S. aethnensis* marker, 18 AET-1, was definitely present in individuals of York Radiate Groundsel and a further two *S. aethnensis* markers (C2 AET-1, C20 AET-1) may also have been present in this taxon. Three or four, *S. chrysanthemifolius* markers (G5 CHR-1, H7 CHR-1, C5 CHR-1 and possibly A9 CHR-1) were present in *S. cambrensis*, together with one to three *S. aethnensis* markers (G12 AET-1, C2 AET-1, C20 AET-1) (table 6.2). Notable differences between Welsh and Edinburgh *S. cambrensis* were observed for three markers: C5 CHR-1 and G12 AET-1 (present in 20/21 and 21/21 Edinburgh individuals, respectively, and 1/13 Welsh individuals), and C20 AET-1 (present in Welsh individuals only) .

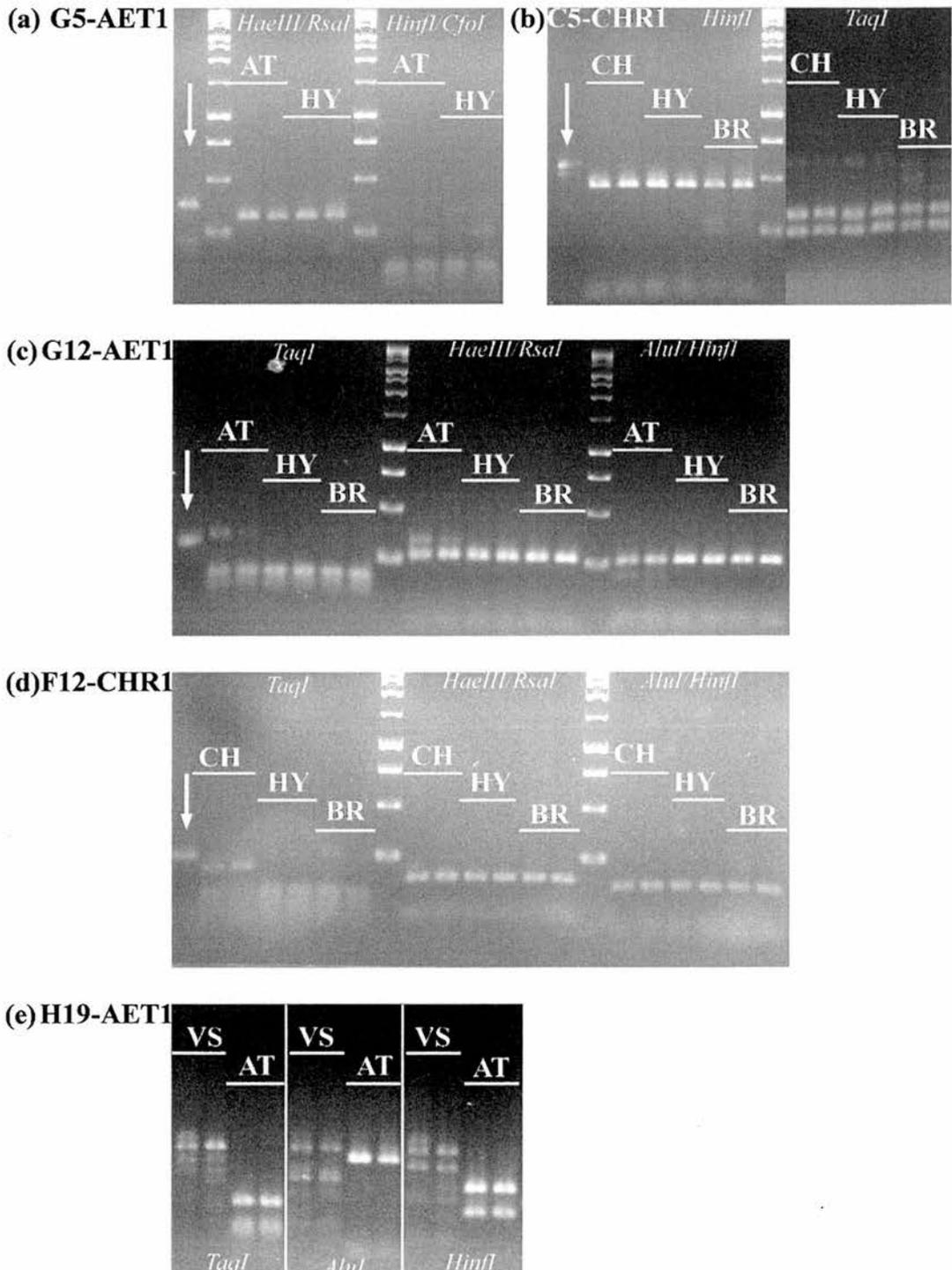
**Plate 6.1** *Senecio chrysanthemifolius*, "var. vulgaris" and "*S. squalidus*" markers amplified by RAPD primer H7 in isolated *S. vulgaris* var. *vulgaris*, var. *hibernicus*, York Radiate Groundsel, *S. cambrensis* and *S. squalidus*. AET indicates *S. aethnensis* standards, CHR indicates *S. chrysanthemifolius* standards.



**Plate 6.2** Examples of *Senecio chrysanthemifolius*, *S. aethnensis*, "var. *vulgaris*" and "*S. squalidus*" markers amplified in isolated *S. vulgaris* var. *vulgaris*, var. *hibernicus*, York Radiate Groundsel, *S. cambrensis* and *S. squalidus*. AET indicates *S. aethnensis* standards, CHR indicates *S. chrysanthemifolius* standards. Photograph (a) RAPD primer C5, (b) RAPD primer G5, (c) RAPD primer A9, (d) RAPD primer G12.



**Plate 6.3** Testing the homology of RAPD bands (see sections 2.9, 4.3.2, 5.3.1 and 6.3.1). Congruent RAPD profiles are observed (a) between *Senecio aethnensis* (AT) individuals and Sicilian hybrid individuals from transect S (HY) for band G5-AET1; (b) between *S. chrysanthemifolius* (CH) individuals and Sicilian hybrid individuals for band C5-CHR1; (c) between *S. aethnensis* individuals, Sicilian hybrid individuals and British *S. squalidus* (BR) individuals for band G12-AET1; and (d) between *S. chrysanthemifolius* individuals, Sicilian hybrid individuals and British *S. squalidus* individuals for band F12-CHR1. Incongruent profiles are observed (e) between *S. aethnensis* individuals and *S. vulgaris* individuals (VS) for band H19-AET1. Arrows indicate uncut bands, restriction enzymes employed are indicated in italics.



**Table 6.1** Presence or absence of 26 *S. chrysanthemifolius* and *S. aethnensis* markers in *S. vulgaris* var. *vulgaris* individuals, from populations isolated from *S. squalidus* in the British Isles.

Primer	Band	Freq. in <i>S. squalidus</i>	Similar var. <i>vulgaris</i> band	Present in isolated var. <i>vulgaris</i>	Comment
A9	CHR-1	0.63	no	no	
A11	CHR-1	0.51	yes	no	Re-ran samples and conducted homology tests to confirm that var. <i>vulgaris</i> band was different from this <i>S. chrysanthemifolius</i> band.
	AET-1	0.29	yes	no	Re-ran samples and conducted homology tests to confirm that var. <i>vulgaris</i> band was different from this <i>S. aethnensis</i> band.
A20	CHR-1	0.51	yes	no	Re-ran samples to confirm that var. <i>vulgaris</i> band was different from this <i>S. chrysanthemifolius</i> band. Homology tests unsuccessful (band did not re-amplify).
B15	CHR-1	0.00	no	no	
	AET-1	0.00	no	no	
C2	CHR-1	0.96	yes	no	Re-ran samples to confirm that var. <i>vulgaris</i> band was different from this <i>S. chrysanthemifolius</i> band. Homology tests unsuccessful (band did not digest).
	AET-1	0.86	no	no	
C5	CHR-1	0.65	no	no	
C20	AET-1	0.63	no	no	
F12	CHR-1	0.86	no	no	
	AET-1	0.01	no	no	
G5	CHR-1	0.75	no	no	
	AET-1	0.00	no	no	
G6	AET-1	0.52	no	no	
G8	CHR-1	0.87	yes	no	Re-ran samples to confirm that var. <i>vulgaris</i> band was different from this <i>S. chrysanthemifolius</i> band. Homology tests not performed.
	AET-1	0.60	yes	yes?	Homology tests did not differentiate between the band in the two Glenluce individuals and this <i>S. aethnensis</i> band.
G10	CHR-1	0.63	yes	no	Re-ran samples to confirm that var. <i>vulgaris</i> band was different from this <i>S. chrysanthemifolius</i> band. Homology tests not performed.
G11	AET-1	0.24	yes	yes?	Homology tests did not differentiate between var. <i>vulgaris</i> and this <i>S. aethnensis</i> band
G12	CHR-1	0.00	no	no	
	AET-1	0.72	no	no	
H7	CHR-1	0.88	no	no	
H13	AET-1	0.52	yes	no	Re-ran samples to confirm that var. <i>vulgaris</i> band was different from this <i>S. aethnensis</i> band. Homology tests not performed.
H19	AET-1	0.00	yes	no	Band in var. <i>vulgaris</i> had a very different intensity to the <i>S. aethnensis</i> band. Homology tests confirmed that var. <i>vulgaris</i> band was different from this <i>S. aethnensis</i> band.
18	AET-1	0.88	no	no	
40	CHR-1	0.76	no	no	

**Table 6.2** Presence or absence of the six *S. chrysanthemifolius* markers and five *S. aethnensis* markers, that occurred at reasonable frequency in *S. squalidus*, in material of non-isolated *S. vulgaris* var. *vulgaris*, *S. vulgaris* var. *hibernicus*, York Radiate Groundsel and *S. cambrensis*.

Marker band	Present in:			
	Non-isolated var. <i>vulgaris</i>	Var. <i>hibernicus</i>	York Radiate Groundsel	<i>S. cambrensis</i>
<b>CHR bands</b>				
A9 CHR-1	no	no	no	yes?? <sup>§</sup>
C5 CHR-1	no	no	no	yes (1/13 Welsh, 20/21 Edinburgh)
F12 CHR-1	no	no	no	? <sup>‡</sup>
G5 CHR-1	no*	no*	no	yes (8/13 Welsh, 10/21 Edinburgh)
H7 CHR-1	no	no	no	yes (13/13 Welsh, 21/21 Edinburgh)
40 CHR-1	no	no	no	no
<b>AET bands</b>				
C2 AET-1	no	no	yes? <sup>†</sup>	yes? <sup>†</sup> (Welsh + Edinburgh)
C20 AET-1	no	no	yes? <sup>†</sup>	yes? <sup>†</sup> (Welsh only)
G6 AET-1	no	no	no	no
G12 AET-1	no	no	no	yes (1/13 Welsh, 21/21 Edinburgh)
18 AET-1	no	no	yes (7/10)	no

\* A very faint band, in a small number (6) of non-isolated var. *vulgaris* and var. *hibernicus* individuals, was close to the *S. chrysanthemifolius* band. These bands seemed different but homology tests would be required to be sure.

<sup>†</sup> Band present but too faint to score accurately.

<sup>‡</sup> Too faint to judge.

<sup>§</sup> Curvature of gel made it difficult to decide if some *S. cambrensis* individuals had this band.

### 6.3.3 Presence/absence of var. *vulgaris* and *S. squalidus* marker bands in the hybrid derivatives

Using the reference gels, 12 "var. *vulgaris*" (VU) marker bands and nine "*S. squalidus*" (SQ) marker bands were identified from the 11 primers. The presence or absence of these 21 bands in isolated *S. vulgaris* var. *vulgaris*, non-isolated var. *vulgaris*, var. *hibernicus*, York Radiate Groundsel, *S. cambrensis*, *S. squalidus*, *S. chrysanthemifolius* and *S. aethnensis* is recorded in table 6.3. Eight "var. *vulgaris*" bands were present in non-isolated var. *vulgaris*, var. *hibernicus*, York Radiate Groundsel and *S. cambrensis*. The remaining four "var. *vulgaris*" bands were present in non-isolated var. *vulgaris*, var. *hibernicus* and *S. cambrensis* but not York Radiate Groundsel. Five "*S. squalidus*" bands were present in

York Radiate Groundsel and *S. cambrensis* but not in non-isolated var. *vulgaris* or var. *hibernicus*. Two "*S. squalidus*" bands were present in *S. cambrensis* only, while one "*S. squalidus*" band was present in York Radiate Groundsel only. The final "*S. squalidus*" band was not present in any of the hybrid derivatives examined nor non-isolated var. *vulgaris*. Only one notable difference between Welsh and Edinburgh *S. cambrensis* individuals was observed - C5 VU-2 was present in 12 of the 13 Welsh individuals but was absent from all Edinburgh individuals.

**Table 6.3** Presence (+) or absence (-) of 21 "var. *vulgaris*" (VU) and "S. *squalidus*" (SQ) marker bands amplified by 11 RAPD and ISSR primers in: isolated *S. vulgaris* var. *vulgaris*, non-isolated var. *vulgaris*, var. *hibernicus*, York Radiate Groundsel, *S. cambrensis* (Ws - Welsh; Ed - Edinburgh), *S. squalidus*, *S. chrysanthemifolius* and *S. aethnensis*. The number of individuals with each band, out of the total number of individuals scored, is recorded below the presence/absence symbols.

Primer	Band size - bp	Isolated var. <i>vulgaris</i>	Non-isolated var. <i>vulgaris</i>	Var. <i>hibernicus</i>	York Radiate Groundsel	<i>S. cambrensis</i>	<i>S. squalidus</i>	<i>S. chrysan.</i>	<i>S. aethnensis</i>
A9*	SQ-1 600	- 0/18	- 0/35	- 0/27	+ 10/10	+ 33/34 13/13 Ws, 20/21 Ed	+ 10/10	+ 6/6	- 0/6
C2*	VU-1 1000	+ 18/18	+ 35/35	+ 27/27	+ 10/10	+ 30/33 12/12 Ws, 18/21 Ed	- 0/10	- 0/6	- 0/6
C5	SQ-1 1400	- 0/18	- 0/35	- 0/27	+ 10/10	+ ? ‡	+ 10/10	+ 6/6	+ 6/6
C5	SQ-2 800	- 0/18	- 0/35	- 0/27	+ 10/10	+ 33/34 13/13 Ws, 20/21 Ed	+ 9/10	+ 6/6	+ 6/6
C5	VU-1 750	+ 18/18	+ 35/35	+ 27/27	- 0/10	+ 34/34 13/13 Ws, 21/21 Ed	- 0/10	- 0/6	- 0/6
C5	VU-2 500	+ 18/18	+ 35/35	+ 25/27	+ 10/10	+ 12/34 12/13 Ws, 0/21 Ed	- 0/10	- 0/6	- 0/6
C20*	SQ-1 900	- 0/17	- 0/35	- 0/27	- 0/10	+ ? ‡	+ 10/10	+ 6/6	- 0/6

Table 6.3 continued.

Primer	Band	Band size	Isolated var. <i>vulgaris</i>	Non-isolated var. <i>vulgaris</i>	Var. <i>hibernicus</i>	York Radiate Groundsel	<i>S. cambrensis</i>	<i>S. squalidus</i>	<i>S. chrysan.</i>	<i>S. aethnensis</i>
C20	VU-1	850	+	+	+	+	+	-	-	-
			17/17	35/35	27/27	10/10	34/34 13/13 Ws, 21/21 Ed	0/10	0/6	0/6
C20	VU-2	450	+	+	+	+	+	-	-	-
			18/18	35/35	27/27	10/10	34/34 13/13 Ws, 21/21 Ed	0/10	0/6	0/6
F12*	VU-1	1000	+	+	+	-	+	-	-	-
			18/18	34/34	27/27	0/10	33/34 13/13 Ws, 20/21 Ed	0/10	0/6	0/6
G5	SQ-1	400	-	-	-	+	+	+	+	+
			0/18	0/35	0/27	10/10	29/34 8/13 Ws, 21/21 Ed	8/10	4/6	6/6
G5	SQ-2	550	-	-	-	+	- ? ‡	+	+	-
			0/18	0/35	0/27	10/10		8/10	6/6	0/6
G5	VU-1	850	+	+	+	+	+	-	-	-
			18/18	35/35	27/27	10/10	34/34 13/13 Ws, 21/21 Ed	0/10	0/6	0/6
G6	SQ-1	1400	-	-	-	-	- ? 0/34† 0/13 Ws, 0/21 Ed	+	+	+
			0/17	0/34	0/27	0/10		10/10	4/6	5/6
G6	VU-1	800	+	+	+	+	+	-	-	-
			17/17	34/34	27/27	10/10	34/34 13/13 Ws, 21/21 Ed	0/10	0/6	0/6

Table 6.3 continued.

Primer	Band size	Isolated var. <i>vulgaris</i>	Non-isolated var. <i>vulgaris</i>	Var. <i>hibernicus</i>	York Radiate Groundsel	<i>S. cambrensis</i>	<i>S. squalidus</i>	<i>S. chrysan.</i>	<i>S. aethnensis</i>
G12*	SQ-1 800	0/18	0/35	0/27	+	34/34 13/13 Ws, 21/21 Ed	+	+	+
G12	VU-1 1400	+	+	+	+	34/34 13/13 Ws, 21/21 Ed	-	-	-
G12	VU-2 400	+	+	+	-	34/34 13/13 Ws, 21/21 Ed	-	-	-
H7	SQ-1 950	0/18	0/35	0/27	-	31/34 12/13 Ws, 19/21 Ed	+	+	+
H7*	VU-1 750	+	+	+	+	34/34 13/13 Ws, 21/21 Ed	-	-	-
18*	VU-1 1100	+	+	+	-	29/34 11/13 Ws, 18/21 Ed	-	-	-

‡ Too difficult to score accurately - usually due a combination of factors such as gel curvature, faintness of bands, band in question close to other bands, etc.. These gels would need to be repeated for accurate scoring.

† There is a very faint trace of what could be this band in two individuals of *S. cambrensis*.

\* These primers were recorded as producing VU (var. *vulgaris*) or SQ (*S. squalidus*) bands in addition to those noted above: A9 (1 x SQ), C2 (2 x SQ), C20 (1 x SQ, 1 x VU), F12 (1 x SQ), G12 (1 x VU), H7 (1 x SQ) and 18 (1 x SQ); plus, 40 (1 x SQ). It was not possible to accurately score these bands for a variety of different reasons (difficult to score across gels, gel curvature, faintness of bands, bands close to other bands, etc.) but it was believed that they could prove to be of future use.

## 6.4 Discussion

### 6.4.1 Screen of isolated *S. vulgaris* var. *vulgaris*

The original aim of this study was to determine if *S. chrysanthemifolius* and *S. aethnensis* molecular markers present in British *S. squalidus* (chapter 5) were also present in products of hybridisation between *S. squalidus* and *S. vulgaris* var. *vulgaris*, namely *S. vulgaris* var. *hibernicus*, *S. cambrensis*, and York Radiate Groundsel. Isolated var. *vulgaris* was surveyed to ensure that the *S. chrysanthemifolius* and *S. aethnensis* marker bands were not shared with *S. vulgaris* and could not be confused with bands in *S. vulgaris* of similar size. *Senecio chrysanthemifolius* and *S. aethnensis* marker bands that fell into either of these categories would obviously be of limited use for surveying the hybrid products of var. *vulgaris* and *S. squalidus*, because it would be unclear from which species they had originated.

It was found that two *S. aethnensis* marker bands appeared to be present in both British *S. squalidus* and isolated var. *vulgaris* (table 6.1). There are a number of possible explanations for this observation:

(i) The bands apparently shared by *S. squalidus* and isolated var. *vulgaris* might not be the same. Restriction-digestion tests indicated that the bands were the same; however, such tests provide only an indication of band homology (1.4.3.2). (ii) The shared presence of bands in *S. squalidus* and isolated var. *vulgaris* might result from introgression of genetic material from *S. squalidus* into var. *vulgaris*, in particular via the stabilised introgressant var. *hibernicus*. However, populations of var. *vulgaris*, understood to be isolated from *S. squalidus* and var. *hibernicus*, had been specifically chosen to minimise this likelihood. Nevertheless, it is possible that some 'isolated' var. *vulgaris* populations were not as 'isolated' as believed. This theory is more reasonable with respect to the G8-AET1 marker than to the G11-AET1 marker, because the former was detected in just one isolated var. *vulgaris* population, while the latter was detected in all five isolated var. *vulgaris* populations. (iii) The shared presence of bands in *S. squalidus* and isolated var. *vulgaris* could, quite feasibly, be due to the joint retention of ancestral polymorphisms. (iv) The shared presence of bands in *S. squalidus* and isolated var. *vulgaris* might be a consequence of introgression from var. *vulgaris* to *S. squalidus*. There is no isozyme evidence of such introgression (Abbott *et al.*, 1992a, Lowe, 1996). Furthermore, var. *hibernicus*, York

Radiate Groundsel and *S. cambrensis* are unlikely to act as 'genetic bridges' between *S. squalidus* and *S. vulgaris* because crosses between *S. squalidus* and these taxa tend to produce sterile offspring (Ingram and Noltie, 1995; Lowe, 1996). Nevertheless the possibility of a low level of introgression from var. *vulgaris* to *S. squalidus* cannot be refuted. Intriguingly, four of the six RAPD bands in *S. squalidus* that were not present in its parental species (*S. chrysanthemifolius* and *S. aethnensis*; see section 5.3.1), were of a similar size to bands present in var. *vulgaris*. However, the homology of these four bands, across the two species, would need to be tested before they were recognised as evidence of introgression from var. *vulgaris* to *S. squalidus*. (v) Finally, the shared presence of bands in *S. squalidus* and isolated var. *vulgaris* could be attributed to *S. squalidus* being one of the diploid progenitors of tetraploid *S. vulgaris*, as suggested by Ashton and Abbott (1992b). If this were the case, one would have anticipated greater difficulty in locating markers that distinguish one species from another. Moreover, several alternative explanations for the origin of *S. vulgaris* exist and the truth remains an enigma, despite numerous studies (Comes *et al.*, 1997)

In addition to the two marker bands discussed above, a number of other *S. chrysanthemifolius* and *S. aethnensis* marker bands, present in British *S. squalidus*, were deemed unsuitable for surveying the hybrid products of *S. squalidus* and var. *vulgaris* because there were bands in isolated var. *vulgaris* of a similar size. This reduced the number of *S. chrysanthemifolius* and *S. aethnensis* marker bands that could be used to survey the hybrid products to only 11.

#### **6.4.2 Presence/absence of *S. chrysanthemifolius*, *S. aethnensis*, "*S. squalidus*" and "*S. vulgaris*" bands in the hybrid derivatives of *S. squalidus* and *S. vulgaris***

The survey of presence/absence of bands in non-isolated *S. vulgaris* var. *vulgaris*, *S. vulgaris* var. *hibernicus*, *S. cambrensis* and the York Radiate Groundsel was initially conducted with the 11 *S. chrysanthemifolius* and *S. aethnensis* marker bands, that were present at a reasonable frequency in *S. squalidus* and could not be confused with bands in isolated var. *vulgaris*. However, with hindsight, it was apparent that these 11 marker bands provided only a limited amount of useful information. A major reason for this was that *S. squalidus* was highly polymorphic for presence/absence of some of the 11 bands (table 5.1), lowering the probability of these bands being passed on to hybrid offspring.

Consequently, reference gels were produced and used to define 12 "var. *vulgaris*" marker bands (present in isolated *S. vulgaris* var. *vulgaris* but absent from *S. squalidus*) and nine "*S. squalidus*" marker bands (present in *S. squalidus* and *S. chrysanthemifolius* and/or *S. aethnensis* but absent from isolated var. *vulgaris* and distinct from the *S. chrysanthemifolius* and *S. aethnensis* marker bands described above). Non-isolated *S. vulgaris* var. *vulgaris*, *S. vulgaris* var. *hibernicus*, *S. cambrensis* and the York Radiate Groundsel were then examined for these additional bands. This decision proved to be sensible, because the combination of the two data sets (11 + 21 bands) provided a more valuable insight into the genomic composition of the products of hybridisation, between *S. squalidus* and var. *vulgaris*, than the original data set of 11 bands.

#### 6.4.3 *S. vulgaris* var. *hibernicus*

All 12 "var. *vulgaris*" bands were present in var. *hibernicus*. Indeed, the RAPD and ISSR profiles of var. *hibernicus* individuals were indistinguishable from the profiles of isolated var. *vulgaris* individuals. These data are not unexpected, as var. *hibernicus* and var. *vulgaris* are variants of the same species and very similar morphologically and in terms of isozyme variation (Lowe, 1996). Var. *hibernicus* was found not to possess any of the 20 RAPD and ISSR bands (11 *S. chrysanthemifolius* and *S. aethnensis* marker bands plus nine additional "*S. squalidus*" marker bands) characteristic of *S. squalidus*. Previous work has emphasised the considerable differences between var. *hibernicus* and *S. squalidus*, both morphologically and in terms of isozyme variation. For example, var. *hibernicus* lacks the  $\beta$ Est-1 loci characteristic of *S. squalidus* (Abbott *et al.*, 1992b), and var. *hibernicus* and *S. squalidus* cluster far apart on a UPGMA dendrogram based upon allozyme phenotype frequencies at eight loci (using Nei's genetic distance; Lowe, 1996).

However, isozyme analysis clearly indicated the hybrid nature of var. *hibernicus*: var. *hibernicus* contains the *Aat-3 c* allele, which is present at high frequency in *S. squalidus* and absent from isolated var. *vulgaris* populations, at intermediate frequency (Abbott *et al.*, 1992a). Moreover, the independent segregation of the ray floret locus and the *Aat-3* locus suggests that genes on more than one chromosome have been introgressed from *S. squalidus* to var. *hibernicus*. Therefore, if a large enough number of taxon-specific RAPD and ISSR markers for *S. squalidus* were employed in a survey of var. *hibernicus*, I would

predict that a difference between the RAPD and ISSR profiles of var. *vulgaris* and var. *hibernicus* would eventually be discovered.

#### **6.4.4 Non-isolated *S. vulgaris* var. *vulgaris***

Non-isolated var. *vulgaris*, i.e. var. *vulgaris* from populations polymorphic for var. *vulgaris* and var. *hibernicus*, did not possess any of the 20 RAPD and ISSR bands characteristic of *S. squalidus*, but did possess all 12 "var. *vulgaris*" bands. In fact, the RAPD and ISSR profiles of non-isolated var. *vulgaris* individuals were identical to the profiles of isolated var. *vulgaris* individuals. The lack of bands characteristic of *S. squalidus* in non-isolated var. *vulgaris* is unsurprising, given that var. *hibernicus*, which is believed to contain *S. squalidus* genes, did not exhibit any of these bands.

The absence of bands characteristic of *S. squalidus* from var. *hibernicus* means that the RAPD and ISSR data, relating to var. *hibernicus* and co-occurring var. *vulgaris*, cannot be used to ascertain if var. *hibernicus* acts as a 'genetic bridge' between *S. squalidus* and var. *vulgaris*, allowing the cryptic introgression of genetic material from *S. squalidus* to var. *vulgaris*.

However, earlier isozyme work suggests that a very low level of cryptic introgression from *S. squalidus* to var. *vulgaris*, via var. *hibernicus* does take place. The *Aat-3 c* allele, which was present at high frequency (0.74) in *S. squalidus* and absent from isolated var. *vulgaris*, occurred at intermediate frequency (0.47) in var. *hibernicus* and at very low frequency (0.03) in var. *vulgaris* sympatric with var. *hibernicus* (Abbott *et al.*, 1992a). This means that in populations polymorphic for var. *hibernicus* and var. *vulgaris*, individuals identified as var. *vulgaris* by the absence of ray florets, are not necessarily 'pure' var. *vulgaris* but may contain a very small amount of genetic material of *S. squalidus* origin.

If, in a future survey, *S. squalidus*-specific RAPD or ISSR bands were found in var. *hibernicus* (see above) then one would predict that some of these bands would also be present at very low frequency in non-isolated var. *vulgaris*. However, it might be necessary to sample a large number of non-isolated var. *vulgaris* individuals to detect such a low level of introgression; for example, in the survey of variation at the *Aat-3* locus, 384 non-isolated var. *vulgaris* plants were examined.

#### 6.4.5 York Radiate Groundsel

York Radiate Groundsel individuals exhibited noticeably different RAPD and ISSR profiles to non-isolated var. *vulgaris* and var. *hibernicus* individuals. The presence, in a typical York Radiate Groundsel individual, of two thirds of the nine "*S. squalidus*" bands plus one to three of the 11 *S. chrysanthemifolius* and *S. aethnensis* bands suggests that the genetic contribution of *S. squalidus* to York Radiate Groundsel was considerably greater than that to var. *hibernicus*. The presence of eight of the 12 "var. *vulgaris*" bands in York Radiate Groundsel indicates that the taxon also owes much of its genome to *S. vulgaris*. However, whereas York Radiate individuals contain eight "var. *vulgaris*" bands, var. *hibernicus* individuals contain all 12. The lack of four "var. *vulgaris*" bands, in conjunction with the presence of bands characteristic of *S. squalidus*, suggests that the genetic relationship between var. *vulgaris* and York Radiate Groundsel is more distant than the relationship between var. *vulgaris* and var. *hibernicus*. The loss of the "var. *vulgaris*" bands (which were present in 100% of var. *vulgaris*) could be the result of genetic drift or selection against linked genes, following the formation of York Radiate Groundsel. Alternatively, absence of the bands might be due to the substitution of null alleles from *S. squalidus* (1.4.3.2). Previously, the danger of using the absence of a RAPD or ISSR band as a marker for a particular taxon has been emphasised (because the possibility of multiple null alleles at a locus cannot easily be tested; 1.4.3.2). However, York Radiate Groundsel is known to have arisen very recently (post 1950) and therefore it may be more valid, than would normally be the case, to assume that the absence of a var. *vulgaris* band is indicative of *S. squalidus*.

In summary, RAPD and ISSR analysis suggests that York Radiate Groundsel is genetically more intermediate between its parental species, *S. squalidus* and *S. vulgaris*, than is var. *hibernicus*. However a reliable assessment of the relative contributions of *S. squalidus* and *S. vulgaris* to the genome of York Radiate Groundsel would require, at the very minimum, many more *S. vulgaris*-specific bands, a similar number of *S. squalidus*-specific bands and possibly a larger sample of York Radiate individuals.

The conclusion reached from the RAPD and ISSR results agrees with morphological data. Multivariate (principal component and canonical variate) analyses of a large number (26-38) of vegetative and floral traits showed that York Radiate Groundsel was

morphologically intermediate between its parents, *S. squalidus* and *S. vulgaris* (Irwin and Abbott, 1992; Lowe, 1996). In contrast, var. *hibernicus* individuals tended to cluster towards var. *vulgaris* individuals, differing only by characters associated with the possession of ray florets. However, the difference between York Radiate Groundsel and var. *hibernicus* was less pronounced with respect to isozyme variation. Both hybrid types clustered with *S. vulgaris* var. *vulgaris*, rather than *S. squalidus*, in a UPGMA dendrogram based upon allozyme phenotype frequencies at eight loci (Lowe, 1996). Moreover, York Radiate Groundsel and var. *hibernicus* each exhibited just one allozyme phenotype or allele characteristic of *S. squalidus*,  $\beta$ Est-1 a and Aat-3 c, respectively (Abbott *et al.*, 1992a; Irwin and Abbott, 1992; Lowe, 1996). This disparity between the isozyme data, and the RAPD, ISSR and morphological results is difficult to explain (perhaps it could be due to the low number of isozyme loci examined?).

#### 6.4.6 *S. cambrensis*

*Senecio cambrensis*, as previously mentioned, is an allohexaploid derived from the polyploidization of the sterile triploid between *S. squalidus* and *S. vulgaris*. RAPD markers are known to display additivity in allopolyploids (Cook *et al.*, 1998). Consequently, one would predict that *S. cambrensis* individuals would combine bands characteristic of *S. squalidus* with those characteristic of *S. vulgaris*. A typical *S. cambrensis* individual possessed between two and seven of the 11 *S. chrysanthemifolius* and *S. aethnensis* marker bands, seven of the nine "*S. squalidus*" bands and all 12 "var. *vulgaris*" bands. Overall, therefore, the results correspond well with the expectation, as *S. cambrensis* individuals, to a large extent, do exhibit additive RAPD and ISSR profiles.

A completely additive profile of bands characteristic of *S. squalidus* and *S. vulgaris* was not expected for several reasons. Firstly, not all bands were necessarily present in the *S. vulgaris* and *S. squalidus* individuals involved in the formation of *S. cambrensis*. This is particularly true of the 11 *S. chrysanthemifolius* and *S. aethnensis* markers scored, because all 11 were polymorphic in *S. squalidus*. Secondly, RAPD and ISSR bands may have been lost, subsequent to the formation of *S. cambrensis*, through genetic drift or selection against genes linked to the RAPD or ISSR bands. Thirdly, the absence of the bands could be attributed to the effect of ploidy level upon competition processes in RAPD PCR reactions (Wolff and Peters-Van Rijn, 1993; Weising *et al.*, 1995). *Senecio cambrensis* ( $2n = 60$ ) has

a larger genome than *S. squalidus* ( $2n = 20$ ), consequently the former will contain more sites of 100% homology, to a particular primer, than the latter. Primers preferentially bind to sites with which they are most homologous, therefore, if many sites of 100% homology are available, primers are unlikely to anneal to sites of lower homology. As a result, bands amplified with a degree of primer to priming-site mismatch in *S. squalidus* will not necessarily be amplified in *S. cambrensis*. This hypothesis explains the absence of a correlation between ploidy level and the number of RAPD fragments generated per primer (Wolff and Peters-Van Rijn, 1993); a similar lack of correlation was noted during a survey of RAPD variation in chrysanthemum species and cultivars at various ploidy levels (Wolff and Peters-Van Rijn, 1993).

Isozyme and cpDNA evidence has shown that Welsh and Edinburgh *S. cambrensis* originated separately (Ashton and Abbott, 1992a; Lowe and Abbott, 1996). Moreover, isozyme data indicate that there may have been two additional independent origins of *S. cambrensis* within Wales, at Wrexham and Mochdre (Ashton and Abbott, 1992a). In the present study, a number of differences were observed between *S. cambrensis* material from Wales and Edinburgh; in particular, Welsh individuals exhibited a "var. *vulgaris*" band and a *S. aethnensis* band that were absent from Edinburgh individuals. Therefore, RAPD and ISSR data do not contradict the theory of separate origins for Welsh and Edinburgh *S. cambrensis*. However, a proper interpretation of the data cannot be made because *S. squalidus* and *S. vulgaris* sympatric with the Welsh and Edinburgh *S. cambrensis* were not investigated. It was not possible to establish if the RAPD and ISSR data support the hypothesis of two separate origins of *S. cambrensis* within Wales due to the very small sample sizes of the Welsh populations and the lack of sympatric *S. squalidus* and *S. vulgaris*, as above. Nevertheless, the results suggest that a more comprehensive survey of RAPD and ISSR variation in *S. cambrensis* and its parents could be most informative about the multiple origins of *S. cambrensis*. For example, Cook *et al.* (1998) employed RAPDs to quantify the number of independent origins of the allotetraploid species *Tragopogon mirus* and *T. miscellus*. The seven populations of *T. mirus* and three populations of *T. miscellus* each had a unique RAPD marker profile and Cook *et al.* (1998) argued that this provided good evidence of multiple origins of the two species.

#### 6.4.7 Improvements to the study and future work

The present work must be regarded as preliminary. Although the study provided interesting information regarding the hybrid products of *S. squalidus* and *S. vulgaris* var. *vulgaris*, it could be improved in several ways.

It would be desirable to examine *S. squalidus* from the same area as the non-isolated var. *vulgaris* and var. *hibernicus* populations, to examine *S. vulgaris* and *S. squalidus* sympatric with Welsh and Edinburgh *S. cambrensis*, and to examine *S. squalidus* from York. However, the study was constrained by a lack of viable seeds from suitable populations and restricted time. The number of populations of each taxon, and the number of individuals sampled from each population, could be increased to reduce sampling error. In particular, it would be beneficial to survey more individuals from the various Welsh *S. cambrensis* populations. It might also be interesting to examine more York Radiate Groundsel individuals, including plants from the other main population of the taxon. However, morphological and allozyme data suggest that these two populations are very similar (Lowe, 1996).

The importance of testing the homology of RAPD and ISSR bands (see section 1.4.3.2 and 2.9) increases as relationships between the taxa under investigation become more distant. Therefore, it would be sensible to conduct homology tests to confirm that the *S. chrysanthemifolius* and *S. aethnensis* bands recorded in York Radiate Groundsel and *S. cambrensis* were indeed one and the same. Furthermore, it would be valuable to check the homology of at least a sub-sample of the *S. squalidus* and *S. vulgaris* bands scored.

A more fundamental criticism of the study concerns its objectives. On reflection, it is clear that the preliminary aim of the study was flawed. Instead of surveying the products of hybridisation between *S. squalidus* and *S. vulgaris* var. *vulgaris* with the pre-developed *S. chrysanthemifolius* and *S. aethnensis* markers, it would have been better to have abandoned these markers at the outset and instead concentrated on developing a set of taxon-specific markers for *S. squalidus* and isolated *S. vulgaris* var. *vulgaris*. Using the strategy outlined in chapter 3, and based on the results of section 6.3.3, it should be possible to discover a large number of RAPD and ISSR taxon-specific markers for the two species, with relative ease. Any future RAPD or ISSR survey would be advised to follow this course of action.

An additional approach to the study of the genomic composition of the hybrid products of *S. squalidus* and *S. vulgaris* would be to employ genomic *in situ* hybridisation (GISH). The advantages and difficulties associated with this technique are discussed elsewhere (5.4.4).

# **Chapter 7**

## **General Conclusions**

## **General Conclusions**

### **7.1 General Conclusions**

The first objective of the research reported in this thesis was to develop molecular markers that distinguish *S. chrysanthemifolius* from *S. aethnensis*. To this end, 'pure' populations of the two taxa, from Mt. Etna in Sicily, were surveyed for RAPD, ISSR, isozyme and cpDNA RFLP variation. RAPD and ISSR analysis proved to be a very effective means of producing taxon-specific markers (chapter 1). In addition, *S. chrysanthemifolius* and *S. aethnensis* differed at a small number of isozyme loci and in chloroplast DNA haplotype. These taxon-specific markers proved to be reliable and reproducible throughout their subsequent use.

The taxonomic status of *S. chrysanthemifolius* and *S. aethnensis* has been a matter of debate. This thesis supports the treatment of *S. chrysanthemifolius* and *S. aethnensis* as closely related but *bona fide* species, as the two taxa are morphologically, ecologically and genetically very distinct (chapter 3).

Based upon morphological evidence, Crisp (1972) proposed that *S. chrysanthemifolius* and *S. aethnensis* form a hybrid zone on Mt. Etna. However, morphological markers are subject to numerous problems (Sytsma, 1990) and, consequently, cannot always be relied upon. Moreover, a later study (Abbott *et al.*, 1995; Abbott *et al.*, 2000, Abbott unpublished) of isozyme, cpDNA and rDNA variation in the postulated hybrid zone was limited by the absence of 'pure' populations of *S. aethnensis*, the low number of molecular markers employed and small sample sizes in the cpDNA and rDNA analyses.

Therefore, the second objective of the present study was to re-examine the putative hybrid zone between *S. chrysanthemifolius* and *S. aethnensis*. The shortcomings of previous work were counteracted by employing a large number of taxon-specific molecular markers and ensuring that 'pure' *S. aethnensis* was sampled. Populations along a linear altitudinal transect, through the putative hybrid zone on Mt. Etna, were surveyed for the taxon-specific RAPD, ISSR, isozyme and cpDNA markers.

The results (chapter 4) confirmed that there is a hybrid zone between *S. chrysanthemifolius* and *S. aethnensis* on Mt. Etna and circumstantial evidence (genetic and ecological) indicates that this zone is secondary in origin. Single locus variation of the RAPD, ISSR and two allozyme markers was most informative. Most of the markers showed fairly smooth clinal variation in the parts of the hybrid zone where change in frequency occurred. Moreover, many of the *S. chrysanthemifolius* marker clines were broadly coincident, as were many marker clines for *S. aethnensis*, and the major decrease in frequency of *S. chrysanthemifolius* markers corresponded with the major increase in the frequency of *S. aethnensis* markers. These facts suggest that the hybrid zone on Mt. Etna is not mosaic-like in structure, nor is it maintained by selection alone (in the absence of dispersal); rather, the persistence of the hybrid zone can be attributed to a balance between selection and dispersal (Barton and Hewitt, 1985).

Hybrid indices and principal co-ordinate analysis of RAPD and ISSR data provided pictorial representations of the genetic composition of individuals, within populations, on the transect through the hybrid zone. With increasing altitude there was a transition from 'pure' *S. chrysanthemifolius* individuals to *S. chrysanthemifolius*-like hybrids, then intermediate hybrids, *S. aethnensis*-like hybrids and finally, near to the summit of Mt. Etna, 'pure' *S. aethnensis* individuals. Clustering of the transect populations, based upon (i)  $\Phi_{ST}$  values from an analysis of molecular variance of RAPD and ISSR data and (ii) Nei's genetic distances from allozyme data, reflected the patterns demonstrated by the hybrid indices and PCO. Moreover, both AMOVA and allozyme results indicated that hybrid populations were genetically more diverse than parental populations.

It is possible to speculate on how the hybrid zone between *S. chrysanthemifolius* and *S. aethnensis* is maintained. The coincidence of clines, as mentioned previously, suggests that the zone is maintained by a balance between dispersal and selection. Choosing between the environment-dependent selection and dispersal model, and the tension zone model (intrinsic selection against hybrids and dispersal) is not an easy task (Barton and Hewitt, 1985). However, circumstantial evidence would appear to favour the former; first, there is no evidence that hybrids between *S. chrysanthemifolius* and *S. aethnensis* are uniformly unfit across all environments and, second, several environmental gradients appear to run parallel with the hybrid zone.

Tradition, and work by Crisp (1972) and Abbott (Abbott *et al.*, 2000), implied that the hybrid zone between *S. chrysanthemifolius* and *S. aethnensis* was the source of British *S. squalidus*. Consequently, the third objective of the present study was to prove or disprove the hypothesis of a Sicilian homoploid hybrid origin of British *S. squalidus*. RAPD and ISSR data (chapter 5) and allozyme and cpDNA data (Abbott *et al.*, 2000; chapter 3) confirm that British *S. squalidus* is indeed the homoploid hybrid derivative of *S. chrysanthemifolius* and *S. aethnensis*. The British taxon combines RAPD, ISSR and allozyme markers characteristic of *S. chrysanthemifolius* and *S. aethnensis*, possesses cpDNA typical of *S. chrysanthemifolius* and has very few unique markers. These results are consistent with morphological and other evidence (Abbott *et al.*, 2000).

British *S. squalidus* remains morphologically and genetically polymorphic, and in general intermediate between its parents. Nevertheless, British *S. squalidus* can be considered a new species because it is isolated from parental gene-flow and evolving independently. Furthermore, British *S. squalidus* may already be ecologically isolated from *S. chrysanthemifolius* and *S. aethnensis* and, eventually, one would predict the indirect evolution of chromosomal or genic incompatibility barriers between *S. squalidus* and its Sicilian progenitors. A comparison with other recorded cases of homoploid hybrid speciation suggests that British *S. squalidus* may be rather unusual; it appears to be the only homoploid hybrid species that has arisen in allopatry.

Principal co-ordinate analysis of RAPD and ISSR data (chapter 5), combined with previous morphological and isozyme analysis (Abbott *et al.*, 2000) indicates that British *S. squalidus* may have inherited more of its genome from *S. chrysanthemifolius* than from *S. aethnensis*. However, further work needs to be undertaken to confirm this theory.

During its spread across Britain, *S. squalidus* has frequently come into contact with native *S. vulgaris* var. *vulgaris*. The final objective of the present study was to use the *S. chrysanthemifolius* and *S. aethnensis* RAPD and ISSR markers present in British *S. squalidus* to examine the products of hybridisation between *S. squalidus* and *S. vulgaris* var. *vulgaris*, namely *S. vulgaris* var. *vulgaris*, *S. cambrensis* and York Radiate Groundsel. However, due to a lack of suitable *S. chrysanthemifolius* and *S. aethnensis* markers, the study was broadened to include additional markers characteristic of *S. squalidus*, plus

markers characteristic of *S. vulgaris* var. *vulgaris*. *Senecio vulgaris* var. *hibernicus* individuals exhibited RAPD and ISSR profiles that were identical to those of var. *vulgaris* individuals. In contrast, York Radiate Groundsel individuals combined markers typical of *S. squalidus* and var. *vulgaris*. These findings agree with previous isozyme and morphological studies (Lowe, 1996), which suggest that var. *hibernicus* contains only a very small amount of *S. squalidus* genetic material, while York Radiate Groundsel contains considerably more *S. squalidus* genetic material. As expected, individuals of the allohexaploid species, *S. cambrensis*, exhibited RAPD and ISSR profiles that were to a large extent additive between the parents.

### **Future Work**

Areas of future work have been highlighted throughout the thesis. One priority is to undertake a more detailed examination of the maintenance of the hybrid zone on Mt. Etna. Environment-dependent selection may be the most important force at work in the hybrid zone and therefore, the zone should be investigated from an ecological perspective, for example using reciprocal transplant experiments and controlled environment studies.

A second priority is to gain greater understanding of the genetic composition of British *S. squalidus*. Two approaches have been suggested; comparative genomic mapping of *S. squalidus* and its parents, and genomic *in situ* hybridisation (GISH).

# References

## **References**

- Aagaard, J.E., Vollmer, S.S., Sorensen, F.C. and Strauss, S.H. 1995 Mitochondrial DNA products among RAPD profiles are frequent and strongly differentiated between races of Douglas-fir. *Molecular Ecology* **4**: 441-447
- Abbott, R.J. 1986 Life history variation associated with the polymorphism for capitulum type and outcrossing rate in *Senecio vulgaris* L. *Heredity* **56**: 381-391
- Abbott, R.J. 1992 Plant invasions, interspecific hybridisation and the evolution of new plant taxa. *Trends in Ecology and Evolution* **7**: 401-405
- Abbott, R.J. 1994 Ecological risks of transgenic plants. *Trends in Ecology and Evolution* **9**: 280-281
- Abbott, R.J., Ashton, P.A. and Forbes, D.G. 1992a Introgressive origin of the radiate groundsel, *Senecio vulgaris* L. var. *hibernicus* Syme: Aat-3 evidence. *Heredity* **68**: 425-435
- Abbott, R.J., Curnow, D. J. and Irwin, J. A. 1995. Molecular systematics of *Senecio squalidus* L. and its close diploid relatives. In: *Advances in Compositae Systematics*. Hind, D. J. N., Jeffrey, C. and Pope, G. V., eds. pp. 223-237. Royal Botanic Gardens, Kew.
- Abbott, R.J. and Forbes, D.G. 1993 Outcrossing rate and self-incompatibility in the colonising species *Senecio squalidus*. *Heredity* **71**: 155-159
- Abbott, R.J., Irwin, J.A., and Ashton, P.A. 1992b Genetic diversity for esterases in the recently evolved stabilised introgressant, *Senecio vulgaris* L. var. *hibernicus* Syme, and its parental taxa *S. vulgaris* L. var. *vulgaris* L. and *S. squalidus* L. *Heredity* **68**: 547-556
- Abbott, R.J., James, J.K., Irwin, J.A. and Comes, H.P. 2000 Hybrid origin of the Oxford Ragwort, *Senecio squalidus* L. *Watsonia* **23**: 123-138
- Abbott, R.J., James, J.K., Irwin, J.A. and Comes, H.P. Submitted. Allopatric origin of a new diploid hybrid species, the Oxford Ragwort, *Senecio squalidus* L.
- Abbott, R.J. and Lowe, A.J. 1996 A review of hybridisation and evolution in British *Senecio*. In: *Compositae: Systematics*. Proceedings of the International Compositae Conference, Kew, 1994. Hind, D.J.N., ed. Vol. 1 pp. 679-689. Royal Botanic Gardens, Kew.
- Abbott, R.J. and Milne, R.I. 1995 Origins and evolutionary effects of invasive weeds. In: *Weeds in a Changing World*. BCPC Symposium Proceedings no. 64. Stirton, C., ed. pp. 53-64. British Crop Protection Council.
- Abernethy, K. 1994 The establishment of a hybrid zone between red and sika deer (genus *Cervus*). *Molecular Ecology* **3**: 551-562
- Adams, R.P. and Demeke, T. 1993 Systematic relationships in *Juniperus* based on random amplified polymorphic DNAs (RAPDs). *Taxon* **42**: 553-571
-

- Alexander, J.C.M. 1979 The Mediterranean species of *Senecio* sections *Senecio* and *Delphinifolius*. *Notes from the Royal Botanic Garden Edinburgh* **37**: 387-428
- Allan, G.J., Clark, C. and Rieseberg, L.H. 1997 Distribution of parental DNA markers in *Encelia virginensis* (Asteraceae: *Heliantheae*), a diploid species of putative hybrid origin. *Plant Systematics and Evolution* **205**: 205-221
- Arens, P., Coops, H., Jansen, J. and Vosman, B. 1998 Molecular genetic analysis of black poplar (*Populus nigra* L.) along Dutch rivers. *Molecular Ecology* **7**: 11-18
- Armstrong, J.S., Gibbs, A.J., Peakall, R. and Weiller, G. 1994 'The RAPDistance Package' <ftp://life.anu.edu.au/pub/software/RAPDistance> or <http://life.anu.edu.au/molecular/software/rapd.html>
- Arnold, M.L. 1993 *Iris nelsonii* (Iridaceae): origin and genetic composition of a homoploid hybrid species. *American Journal of Botany* **80**: 577-583
- Arnold, M.L. 1997 *Natural Hybridisation and Evolution*. Oxford University Press, New York.
- Arnold, M.L. and Bennett, B.D. 1993 Natural hybridisation in Louisiana irises: genetic variation and ecological determinants. In: *Hybrid Zones and the Evolutionary Process*. Harrison, R.G., ed. pp. 115-139. Oxford University Press, Oxford.
- Arnold, M.L., Bennett, B.D. and Zimmer, E.A. 1990 Natural hybridisation between *Iris fulva* and *Iris hexagona*: pattern of ribosomal DNA variation. *Evolution* **44**: 1512-1521
- Arnold, M.L., Buckner, C.M., and Robinson, J.J. 1991 Pollen-mediated introgression and hybrid speciation in Louisiana irises. *Proceedings of the National Academy of Sciences of the USA* **88**: 1398-1402
- Arnold, M.L. and Hodges, S.A. 1995 Are natural hybrids fit or unfit relative to their parents? *Trends in Ecology and Evolution* **10**: 67-71
- Arnold, M.L., Robinson, J.J., Buckner, C.M. and Bennett, B.D. 1992 Pollen dispersal and interspecific gene flow in Louisiana irises. *Heredity* **68**: 399-404
- Ashton, P.A. and Abbott, R.J. 1992a Multiple origins and genetic diversity in the newly arisen allopolyploid species, *Senecio cambrensis* Rosser (Compositae). *Heredity* **68**: 25-32
- Ashton, P.A. and Abbott, R.J. 1992b Isozyme evidence and the origin of *Senecio vulgaris* (Compositae). *Plant Systematics and Evolution* **179**: 167-174
- Avise, J.C. 1994 *Molecular Markers, Natural History and Evolution*. Chapman and Hall, New York.
- Bachmann, K. 1992 Nuclear DNA markers in angiosperm taxonomy. *Acta Botanica Neerlandica* **39**: 369-384
- Bachmann, K. 1994 Molecular markers in plant ecology. *New Phytologist* **126**: 403-418
-

- Bailey, J.P., Bennett, S.T., Bennett, M.D. and Stace, C.A. 1993 Genomic *in situ* hybridisation identifies parental chromosomes in the wild grass hybrid *xFestulpia hubbardii*. *Heredity* **71**: 413-420
- Barton, N.H. and Gale, K.S. 1993 Genetic analysis of hybrid zones. In: *Hybrid Zones and the Evolutionary Process*. Harrison, R.G., ed. pp. 13-45. Oxford University Press, Oxford.
- Barton, N.H. and Hewitt, G.M. 1985 Analysis of hybrid zones. *Annual Review of Ecology and Systematics* **16**: 113-148
- Beismann, H., Barker, J.H.A., Karp, A and Speck, T. 1997 AFLP analysis sheds light on distribution of two *Salix* species and their hybrid along a natural gradient. *Molecular Ecology* **6**: 989-993
- Bennett, B.D. and Grace, J.B. 1990 Shade tolerance and its effects on the segregation of two species of Louisiana iris and their hybrids. *American Journal of Botany* **77**: 100-107
- Bennett, M.D. 1995 The development and use of genomic *in situ* hybridisation (GISH) as a new tool in plant biosystematics. In: *Kew Chromosome Conference IV*. Brandham, P.E. and Bennett, M.D., eds. pp. 167-183. Royal Botanic Gardens, Kew.
- Bennett, S.T., Kenton, A.Y. and Bennett, M.D. 1992 Genomic *in situ* hybridisation reveals the allopolyploid nature of *Milium montianum* (Gramineae). *Chromosoma* **101**: 420-424
- Black, W.C. 1993 PCR with arbitrary primers: approach with care. *Insect Molecular Biology* **2**: 1-6
- Brookes, M. 1999 Live and let live. *New Scientist* **3<sup>rd</sup> July**: 32-36
- Bucci, G. Anzidei, M., Madaghiele, A. and Vendramin, G.G. 1998 Detection of haplotypic variation and natural hybridisation in *halepensis*-complex pine species using chloroplast simple sequence repeat (SSR) markers. *Molecular Ecology* **7**: 1633-1643
- Burke, J.M. and Arnold, M.L. 1999 Isolation and characterisation of microsatellites in iris. *Molecular Ecology* **8**: 1091-1092
- Burke, J.M., Carney, S.E. and Arnold, M.L. 1998b Hybrid fitness in the Louisiana irises: analysis of parental and F<sub>1</sub> performance. *Evolution* **52**: 37-43
- Burke, J.M., Voss, T.J. and Arnold, M.L. 1998a Genetic interactions and natural selection in Louisiana iris hybrids. *Evolution* **52**: 1304-1310
- Butlin, R.K., Ritchie, M.G. and Hewitt, G.M. 1991 Comparisons among morphological characters and between localities in the *Chorthippus parallelus* hybrid zone (Orthoptera: Acrididae). *Philosophical Transactions of the Royal Society of London B* **334**: 297-308
- Caetano-Anollés, G. 1993 Amplifying DNA with arbitrary oligonucleotide primers. *PCR Methods and Applications* **3**: 85-94
-

- Caetano-Anollés, G., Bassam, B.J. and Gresshoff, P.M. 1992 Primer-template interactions during DNA amplification fingerprinting with single arbitrary oligonucleotides. *Molecular and General Genetics* **235**: 157-165
- Charlesworth, D. 1995 Evolution under the microscope. *Current Biology* **5**: 835-836
- Charters, Y.M., Robertson, A., Wilkinson, M.J. and Ramsay, G. 1996 PCR analysis of oilseed rape cultivars (*Brassica napus* L. ssp. *oleifera*) using 5'-anchored simple sequence repeat (SSR) primers. *Theoretical and Applied Genetics* **92**: 442-447
- Clegg, M.T. 1993 Chloroplast gene sequences and the study of plant evolution. *Proceedings of the National Academy of Sciences of the USA* **90**: 363-367
- Comes, H.P., Kadereit, J.W., Pohl, A., and Abbott, R.J. 1997 Chloroplast DNA and isozyme evidence on the evolution of *Senecio vulgaris* (Asteraceae). *Plant Systematics and Evolution* **206**: 375-392
- Cook, P.J. 1998 Road verge *Cochlearia danica* - salt and sulfur? *Botanical Society of the British Isles News* **79**: 52-53
- Cook, L.M., Soltis, P.S., Brunfeld, S.J., and Soltis, D.E. 1998 Multiple independent formations of *Tragopogon* tetraploids (Asteraceae): evidence from RAPD markers. *Molecular Ecology* **7**: 1293-1302
- Crisp, P. 1972 *Cytotaxonomic Studies in the Section Annuifolius of Senecio*. Unpublished PhD thesis, University of London.
- Dawson, I.K., Chalmers, K.J., Waugh, R. and Powell, W. 1993 Detection and analysis of genetic variation in *Hordeum spontaneum* populations from Israel using RAPD markers. *Molecular Ecology* **2**: 151-159
- Dawson, I.K., Simons, A.J., Waugh, R. and Powell, W. 1996 Detection and pattern of interspecific hybridisation between *Gliricidia sepium* and *G. maculata* in Meso-America revealed by PCR-based assays. *Molecular Ecology* **5**: 89-98
- De Greef, B. and Triest, L. 1999 The use of random amplified polymorphic DNA (RAPD) for hybrid detection in *Scirpus* from the river Schelde (Belgium). *Molecular Ecology* **8**: 379-386
- dePamphilis, C.W. and Wyatt, R. 1989 Hybridisation and introgression in Buckeyes (*Aesculus*: Hippocastanaceae): a review of the evidence and a hypothesis to explain long-distance gene flow. *Systematic Botany* **14**: 593-611
- dePamphilis, C.W. and Wyatt, R. 1990 Electrophoretic confirmation of interspecific hybridisation in *Aesculus* (Hippocastanaceae) and the genetic structure of a broad hybrid zone. *Evolution* **44**: 1295-1317
- Druce, G.C. 1927 *The Flora of Oxfordshire* (2<sup>nd</sup> edition). Clarendon Press, Oxford.
- Dunn, G. and Everitt, B.S. 1982 *An Introduction to Mathematical Taxonomy*. Cambridge University Press, Cambridge.
-

- Echt, C.S., Erdahl, L.A. and McCoy, T.J. 1992 Genetic segregation of random polymorphic DNA in diploid cultivated alfalfa. *Genome* **35**: 64-67
- Eckenwalder, J.E. 1998 Hybridisation as evolutionary creation. *American Journal of Botany* **85**: 1043-1045
- Ellstrand, N.C., Whitkus, R. and Rieseberg, L.H. 1996 Distribution of spontaneous plant hybrids. *Proceedings of the National Academy of Sciences of the USA* **93**: 5090-5093
- Emms, S.K. and Arnold, M.L. 1997 The effect of habitat on parental and hybrid fitness: transplant experiments with Louisiana irises. *Evolution* **51**: 1112-1119
- Escaravage, N., Questiau, S., Pornon, A. Doche, B. and Taberlet, P. 1998 Clonal diversity in a *Rhododendron ferrugineum* L. (Ericaceae) population inferred from AFLP markers. *Molecular Ecology* **7**: 975-982
- Esselman, E.J., Jianqiang, L., Crawford, D.J., Windus, J.L. and Wolfe, A.D. 1999 Clonal diversity in the rare *Calamagrostis porteri* ssp. *insperata* (Poaceae): comparative results for allozymes and random amplified polymorphic DNA (RAPD) and intersimple sequence repeat (ISSR) markers. *Molecular Ecology* **8**: 443-451
- Excoffier, L. 1993 WINAMOVA computer package. Current version of the program available from [http://anthropologie.unige.ch/~laurent/default.htm#Software programs](http://anthropologie.unige.ch/~laurent/default.htm#Software%20programs)
- Excoffier, L., Smouse, P.E. and Quattro, J.M. 1992 Analysis of molecular variance inferred from metric distances among DNA haplotypes: applications to human mitochondrial DNA restriction data. *Genetics* **131**: 479-491
- Felsenstein, J. 1993 PHYLIP (Phylogeny Inference Package). Version 3.5C. Distributed by the author. Department of Genetics, University of Washington, Seattle, WI.
- Ferris, C., Oliver, R.P., Davy, A.J. and Hewitt, G.M. 1993 Native oak chloroplasts reveal an ancient divide across Europe. *Molecular Ecology* **2**: 337-344
- Fitter, A.H. and Hay, R.K.M. 1981 *Environmental Physiology of Plants*. Academic Press, London.
- Fritsch, P. and Rieseberg, L.H. 1992 High outcrossing rates maintain male and hermaphrodite individuals in populations of the flowering plant *Datisca glomerata*. *Nature* **359**: 633-636
- Fuertes Aguilar, J.F., Rosselló, J.A., and Feliner, G.N. 1999 Nuclear ribosomal DNA (nrDNA) concerted evolution in natural and artificial hybrids of *Armeria* (Plumbaginaceae). *Molecular Ecology* **8**: 1341-1346
- Futuyma, D.J. 1998 *Evolutionary Biology* (3<sup>rd</sup> edition). Sinauer Associates, Sunderland, Massachusetts.
- Gabrielsen, T.M., and Brochmann, C. 1998 Sex after all: high levels of diversity detected in the arctic clonal plant *Saxifraga cernua* using RAPD markers. *Molecular Ecology* **7**: 1701-1708
-

- Gallez, G.P. and Gottlieb, L.D. 1982 Genetic evidence for the hybrid origin of the diploid plant *Stephanomeria diegensis*. *Evolution* **36**: 1158-1167
- Gillies, A.C.M., Cornelius, J.P., Newton, A.C., Navarro, C., Hernández, M. and Wilson, J. 1997 Genetic variation in Costa Rican populations of the tropical timber species *Cedrela odorata* L., assessed using RAPDs. *Molecular Ecology* **6**: 1133-1145
- Gillies, A.C.M., Navarro, C., Lowe, A.J., Newton, A.C., Hernández, M., Wilson, J. and Cornelius, J.P. 1999 Genetic diversity in Mesoamerican populations of Mahogany (*Swietenia macrophylla*), assessed using RAPDs. *Heredity* **83**: 722-732
- Glover, B.J., Gillies, A.C.M. and Abbott, R.J. 1994 Use of the polymerase chain reaction to investigate the delimitation of two agronomically important species of *Stylosanthes* (Aubl.) Sw. *Botanical Journal of Scotland* **47**: 83-96
- Golding, G.B. and Strobeck, C. 1983 Increased numbers of alleles found in hybrid populations due to intragenic recombination. *Evolution* **37**: 17-29
- Gomes, C., Dales, R.B.G. and Oxenford, H.A. 1998 The application of RAPD markers in stock discrimination of the four-wing flying fish, *Hirundichthys affinis* in the central western Atlantic. *Molecular Ecology* **7**: 1029-1039
- González-de-León, D., Hoisington, D., Hu, X., Jiang, C. and Khairallah, M. 1995 *Molecular Marker Applications to Plant Breeding*. Notes from a workshop offered by the Applied Molecular Genetics Laboratory, CIMMYT, Int. Mexico
- Goudet, J. 1995 FSTAT (version 1.2): a computer program to calculate *F*-statistics. *Journal of Heredity* **86**: 485-486
- Gower, J.C. 1966 Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* **53**: 325-338
- Grant, V. 1966a Selection for vigor and fertility in the progeny of a highly sterile species hybrid in *Gilia*. *Genetics* **53**: 757-775
- Grant, V. 1966b The origin of a new species of *Gilia* in a hybridisation experiment. *Genetics* **54**: 1189-1199
- Grant, V. 1981 *Plant Speciation* (2<sup>nd</sup> edition). Columbia University Press, New York.
- Gray, A.J. 1986 Do invading species have definable genetic characteristics? *Philosophical Transactions of the Royal Society of London B* **314**: 655-674
- Hadrys, H., Balick, M. and Schierwater, B. 1992 Applications of random amplified polymorphic DNA (RAPD) in molecular ecology. *Molecular Ecology* **1**: 55-63
- Haig, S.M., Rhymer, J.M. and Heckel, D.G. 1994 Population differentiation in randomly amplified polymorphic DNA of red-cockaded woodpeckers *Picoides borealis*. *Molecular Ecology* **3**: 581-595
-

- Hails, R.S. 2000 Genetically modified plants - the debate continues. *Trends in Ecology and Evolution* **15**: 14-18
- Hamrick, J.L. and Godt, M.J.W. 1990 Allozyme diversity in plant species. In: *Plant Population Genetics, Breeding, and Genetic Resources*. Brown, A.H.D., Clegg, M.T., Kahler, A.L. and Weir, B.S., eds. pp. 43-63. Sinauer Associates, Sunderland, Massachusetts.
- Harris, S.A. 1999 RAPDs in systematics - a useful methodology? In: *Molecular Systematics and Plant Evolution*. Hollingsworth, P.M., Bateman, R.M. and Gornall, R.J., eds. pp. 211-229. Taylor and Francis, London.
- Harris, S.A. and Ingram, R. 1992 Molecular systematics of the genus *Senecio* L. 1: Hybridisation in a British polyploid complex. *Heredity* **69**: 1-10
- Harrison, R.G. 1986 Pattern and process in a narrow hybrid zone. *Heredity* **56**: 337-349
- Harrison, R.G. 1990 Hybrid zones: windows on evolutionary process. *Oxford Surveys in Evolutionary Biology* **7**: 69-128
- Harrison, R.G. 1993 Hybrids and hybrid zones: historical perspective. In: *Hybrid Zones and the Evolutionary Process*. Harrison, R.G., ed. pp 3-12. Oxford University Press, Oxford.
- Harrison, R.G. and Rand, D.M. 1989 Mosaic hybrid zones and the nature of species boundaries. In: *Speciation and its Consequences*. Otte, D. and Endler, J.A., eds. Sinauer Associates, Sunderland, Massachusetts.
- Hauber, D.P. and Bloom, W.L. 1983 Stability of a chromosomal hybrid zone in the *Clarkia nitens* and *Clarkia speciosa* ssp. *polyantha* complex (Onagraceae). *American Journal of Botany* **70**: 1454-1459
- Heiser, C.B. 1951 Hybridisation in the annual sunflowers: *Helianthus annuus* x *H. debilis* var. *cucumerifolius*. *Evolution* **5**: 42-51
- Hewitt, G.M. 1989 The subdivision of species by hybrid zones. In: *Speciation and its Consequences*. Otte, D. and Endler, J.A., eds. Sinauer Associates, Sunderland, Massachusetts.
- Hollingsworth, M.L., Hollingsworth, P.M., Jenkins, G.I., Bailey, J.P. and Ferris, C. 1998 The use of molecular markers to study patterns of genotypic diversity in some invasive alien *Fallopia* spp. (Polygonaceae). *Molecular Ecology* **7**: 1681-1691
- Hong, Y.-P., Hipkins, V.D. and Strauss, S.H. 1993 Chloroplast DNA diversity among tress, populations and species in the California closed-cone pines (*Pinus radiata*, *Pinus muricata* and *Pinus attenuata*). *Genetics* **135**: 1187-1196
- Howard, D.J. 1993 Reinforcement: origin, dynamics, and the fate of an evolutionary hypothesis. In: *Hybrid Zones and the Evolutionary Process*. Harrison, R.G., ed. pp. 46-69. Oxford University Press, Oxford.
-

- Howard, D.J., Preszler, R.W., Williams, J., Fenchel, S. and Boecklen, W.J. 1997 How discrete are oak species? Insights from a hybrid zone between *Quercus grisea* and *Quercus gambelii*. *Evolution* **51**: 747-755
- Huff, D.R., Peakall, R. and Smouse, P.E. 1993 RAPD variation within and among natural populations of outcrossing buffalograss [*Buchloë dactyloides* (Nutt.) Engelm.]. *Theoretical and Applied Genetics* **86**: 927-934
- Huff, D.R., Quinn, J.A., Higgins, B. and Palazzo, A.J. 1998 Random amplified polymorphic DNA (RAPD) variation among native little bluestem [*Schizachyrium scoparium* (Michx.) Nash] populations from sites of high and low fertility in forest and grassland biomes. *Molecular Ecology* **7**: 1591-1597
- Ingram, R. and Noltie, H.J. 1995 *Senecio cambrensis* Rosser. *Journal of Ecology* **83**: 537-546
- Ingram, R., Weir, J. and Abbott, R.J. 1980 New evidence concerning the origin of inland radiate groundsel, *S. vulgaris* L. var. *hibernicus* Syme. *The New Phytologist* **84**: 543-546
- Irwin, J.A. and Abbott, R.J. 1992 Morphometric and isozyme evidence for the hybrid origin of a new tetraploid radiate groundsel in York, England. *Heredity* **69**: 431-439
- James, J.K.P. 1995 *Isozyme and morphological analysis of Italian Senecio squalidus populations*. B.Sc. thesis, University of St. Andrews.
- Jansen, R.K. and Palmer, J.D. 1987 Chloroplast DNA from lettuce and *Barnadesia* (Asteraceae): structure, gene localisation and characterisation of a large inversion. *Current Genetics* **11**: 553-564
- Jarne, P., Lagoda, P.J.L. 1996 Microsatellites, from molecules to populations and back. *Trends in Ecology and Evolution* **11**: 424-429
- Keim, P., Paige, K.N., Whitham, T.G. and Lark, K.G. 1989 Genetic analysis of an interspecific hybrid swarm of *Populus*: occurrence of unidirectional introgression. *Genetics* **123**: 557-565
- Kent, D. H. 1955. Scottish records of *Senecio squalidus* L. *Proceedings of the Botanical Society of the British Isles* **1**: 312-313.
- Kent, D. H. 1956. *Senecio squalidus* L. in the British Isles. 1. Early records (to 1877). *Proceedings of the Botanical Society of the British Isles* **2**: 115-118
- Kent, D. H. 1957. *Senecio squalidus* L. in the British Isles. 3. East Anglia. *Transactions of the Norfolk & Norwich Naturalists Society* **18**: 30-31.
- Kent, D. H. 1960. *Senecio squalidus* L. in the British Isles. 2. The spread from Oxford. *Proceedings of the Botanical Society of the British Isles* **4**: 375-379.
- Kent, D. H. 1963. *Senecio squalidus* L. in the British Isles. 7. Wales. *Nature in Wales* **8**: 175-178.
-

- Kent, D. H. 1964a. *Senecio squalidus* L. in the British Isles. 4. Southern England (1940->). *Proceedings of the Botanical Society of the British Isles* **5**: 210-213.
- Kent, D. H. 1964b. *Senecio squalidus* L. in the British Isles. 5. The Midlands (1940 ->). *Proceedings of the Botanical Society of the British Isles* **5**: 214-216.
- Kent, D. H. 1964c. *Senecio squalidus* L. in the British Isles. 6. Northern England (1940 ->). *Proceedings of the Botanical Society of the British Isles* **5**: 217-219.
- Kent, D. H. 1966. *Senecio squalidus* L. in the British Isles. 8. The recent spread in Scotland. *Glasgow Naturalist* **18**: 407-408.
- King, L.M and Schaal, B.A. 1989 Ribosomal-DNA variation and distribution in *Rudbeckia missouriensis*. *Evolution* **43**: 1117-1119
- King, R.A. and Ferris, C. 1998 Chloroplast DNA phylogeography of *Alnus glutinosa* (L.) Gaertn. *Molecular Ecology* **7**: 1151-1161
- Kölliker, R., Stadelmann, F.J., Reidy, B. and Nösberger, J. 1998 Fertilisation and defoliation frequency affect genetic diversity of *Festuca pratensis* Huds. In permanent grasslands. *Molecular Ecology* **7**: 1557-1567
- Kyhos, D.W., Clark, C. and Thompson, W.C. 1981 The hybrid nature of *Encelia laciniata* (Compositae: Heliantheae) and control of population composition by post-dispersal selection. *Systematic Botany* **6**: 399-411
- Levin, D.A. and Kerster, H.W. 1967 Natural selection for reproductive isolation in *Phlox*. *Evolution* **21**: 679-687
- Li, W-H. and Graur, D. 1991 *Fundamentals of Molecular Evolution*. Sinauer Associates, Sunderland, Massachusetts.
- Liao, L.C. and Hsiao, J.Y. 1998 Relationship between population genetic structure and riparian habitat as revealed by RAPD analysis of the rheophyte *Acorus gramineus* Soland. (Araceae) in Taiwan. *Molecular Ecology* **7**: 1275-1281
- Lorenz, M., Weihe, A., Börner, T. 1994 DNA fragments of organellar origin in random amplified polymorphic DNA (RAPD) patterns of sugar beet (*Beta vulgaris* L.). *Theoretical and Applied Genetics* **88**: 775-779
- Lowe, A.J. 1996 *The Origin and Maintenance of a New Senecio Hybrid in York, England*. Unpublished PhD thesis, University of St. Andrews.
- Lowe, A.J. and Abbott, R.J. 1996 Origins of the new allopolyploid species *Senecio cambrensis* (Asteraceae) and its relationship to the Canary Islands endemic *Senecio teneriffae*. *American Journal of Botany* **83**: 1365-1372
- Lynch, M. and Milligan, B.G. 1994 Analysis of population genetic structure with RAPD markers. *Molecular Ecology* **3**: 91-99
- Mabey, R. 1997 *Flora Britannica*. Chatto and Windus, London.
-

- Mallet, J. 1995a A species definition for the Modern Synthesis. *Trends in Ecology and Evolution* **10**: 294-299
- Mallet, J. 1995b Letter. *Trends in Ecology and Evolution* **10**: 490-491
- Manly, B.F.J. 1994 *Multivariate Statistical Methods, a primer* (2<sup>nd</sup> edition). Chapman and Hall, London.
- Marshall, D.F. and Abbott, R.J. 1980 On the frequency of introgression of the radiate ( $T_r$ ) allele from *Senecio squalidus* L. into *Senecio vulgaris* L.. *Heredity* **45**: 133-135
- McCarthy, E.M., Asmussen, M.A. and Anderson, W.W. 1995 A theoretical assessment of recombinational speciation. *Heredity* **74**: 502-509
- McCauley, D.E. 1995 The use of chloroplast DNA polymorphism in studies of gene flow in plants. *Trends in Ecology and Evolution* **10**: 198-202
- Micheli, M.R., Bova, R., Calissano, P. and D'Ambrosio, E. 1993 Randomly amplified polymorphic DNA fingerprinting using combinations of oligonucleotide primers. *BioTechniques* **15**: 388-390
- Michelmore, R.W., Paran, I. and Kesseli, R.V. 1991 Identification of markers linked to disease-resistance genes by bulked segregant analysis: A rapid method to detect markers in specific genomic regions by using segregating populations. *Proceedings of the National Academy of Sciences of the USA* **88**: 9828-9832
- Millar, C.I. 1983 A steep cline in *Pinus muricata*. *Evolution* **37**: 311-319
- Milne, R.I. 1997 *Molecular Systematics of Rhododendron ponticum L. and its close allies*. Unpublished PhD thesis, University of St. Andrews.
- Moore, W.S. 1977 An evaluation of narrow hybrid zones in vertebrates. *The Quarterly Review of Biology* **52**: 263-277
- Morrell, P.L. and Rieseberg, L.H. 1998 Molecular tests of the proposed diploid hybrid origin of *Gilia achilleifolia* (Polemoniaceae). *American Journal of Botany* **85**: 1439-1453
- Nei, M. 1978 Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**: 583-590
- Neuffer, B. and Jahncke, P. 1997 RAPD analyses of hybridisation events in *Cardamine* (Brassicaceae). *Folia Geobotanica et Phytotaxonomica* **32**: 57-67
- Neuffer, B., Auge, H., Mesch, H., Amarell, U and Brandl, R. 1999 Spread of violets in polluted pine forests: morphological and molecular evidence for the ecological importance of interspecific hybridisation. *Molecular Ecology* **8**: 365-377
- Novak, S.J., Soltis, D.E. and Soltis, P.S. 1991 Ownbey's *Tragopogon*'s: 40 years later. *American Journal of Botany* **78**: 1586-1600
-

- O'Hanlon, P.C., Peakall, R. and Briese, D.T. 1999 Amplified fragment length polymorphism (AFLP) reveals introgression in weedy *Onopordum* thistles: hybridisation and invasion. *Molecular Ecology* **8**: 1239-1246
- Padgett, D.J., Les, D.H. and Crow, G.E. 1998 Evidence for the hybrid origin of *Nuphar x rubrodisca* (Nymphaeaceae). *American Journal of Botany* **85**: 1468-1476
- Page, R.D.M. 1996 TREEVIEW: An application to display phylogenetic trees on personal computers. *Computer Applications in Biosciences* **12**: 357-358
- Paige, K.N., Capman, W.C. and Jennetten, P. 1991 Mitochondrial inheritance patterns across a cottonwood hybrid zone: cytonuclear disequilibria and hybrid zone dynamics. *Evolution* **45**: 1360-1369
- Palacios, C., Kresovich, S. and González-Candelas, F. 1999 A population genetic study of the endangered plant species *Limonium dufourii* (Plumbaginaceae) based on amplified fragment length polymorphism (AFLP). *Molecular Ecology* **8**: 645-657
- Pammi, S., Schertz, K., Xu, G., Hart, G. and Mullet, J.E. 1994 Random-amplified-polymorphic DNA markers in sorghum. *Theoretical and Applied Genetics* **89**: 80-88
- Peakall, R., Smouse, P.E. and Huff, D.R. 1995 Evolutionary implications of allozyme and RAPD variation in diploid populations of dioecious buffalograss *Buchloë dactyloides*. *Molecular Ecology* **4**: 135-147
- Pérez, T., Albornoz, J and Domínguez, A. 1998 An evaluation of RAPD fragment reproducibility and nature. *Molecular Ecology* **7**: 1347-1357
- Perron, M., and Bousquet, J. 1997 Natural hybridisation between black spruce and red spruce. *Molecular Ecology* **6**: 725-734
- Perron, M., Gordon, A.G., and Bousquet, J. 1995 Species-specific RAPD fingerprints for the closely related *Picea mariana* and *P. rubens*. *Theoretical and Applied Genetics* **91**: 142-149
- Phelan, P.L. and Baker, T.C. 1987 Evolution of male pheromones in moths: reproductive isolation through sexual selection? *Science* **235**: 205-207
- Poli Marchese, E. 1991 *Piante e Fiori dell'Etna*. Palermo.
- Poli Marchese, E. 1999 Vegetation and changing environment on Mt. Etna. In: *Recent Shifts in Vegetation Boundaries of Deciduous Forests, Especially due to Global Warming*. Klötzli, F. and Walther, G.-R., eds. pp 217-236. Birkhäuser Verlag, Basel.
- Potts, B.M. and Reid, J.B. 1988 Hybridisation as a dispersal mechanism. *Evolution* **42**: 1245-1255
- Potts, B.M. and Reid, J.B. 1990 The evolutionary significance of hybridisation in *Eucalyptus*. *Evolution* **44**: 2151-2152
-

- Quicke, D.L.J. 1993 *Principles and Techniques of Contemporary Taxonomy*. Blackie Academic and Professional, Glasgow.
- Rabouam, C., Comes, A.M., Bretagnolle V., Humbert, J.-F., Periquet, G. and Bigot, Y. 1999 Features of DNA fragments obtained by random amplified polymorphic DNA (RAPD) assays. *Molecular Ecology* **8**: 493-503
- Rafalski, J.A. and Tingey, S.V. 1993 Genetic diagnostics in plant breeding: RAPDs, microsatellites and machines. *Trends in Genetics* **9**: 275-280
- Raymond, M. and Rousset, F. 1995 GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86**: 248-249
- Reyment, R.A., Blackith, R.E. and Campbell, N.A. 1984 *Multivariate Morphometrics* (2<sup>nd</sup> edition). Academic Press, London.
- Rhymer, J.M. and Simberloff, D. 1996 Extinction by hybridisation and introgression. *Annual Review of Ecology and Systematics* **27**: 83-109
- Ridout, C.J. and Donini, P. 1999 Use of AFLP in cereals research. *Trends in Plant Science* **4**: 76-79
- Rieseberg, L.H. 1991 Homoploid reticulate evolution in *Helianthus* (Asteraceae): evidence from ribosomal genes. *American Journal of Botany* **78**:1218-1237
- Rieseberg, L.H. 1995 The role of hybridisation in evolution: old wine in new skins. *American Journal of Botany* **82**: 944-953
- Rieseberg, L.H. 1996 Homology among RAPD fragments in interspecific comparisons. *Molecular Ecology* **5**: 99-105
- Rieseberg, L.H. 1997 Hybrid origins of plant species. *Annual Review of Ecology and Systematics* **28**: 359-389
- Rieseberg, L.H., Archer, M.A. and Wayne, R.K. 1999 Transgressive segregation, adaptation and speciation. *Heredity* **83**: 363-372
- Rieseberg, L.H., Baird, S.J.E. and Desrochers, A.M. 1998 Patterns of mating in wild sunflower hybrid zones. *Evolution* **52**: 713-726
- Rieseberg, L.H., Beckstrom-Sternberg, S. and Doan, K. 1990a *Helianthus annuus* ssp. *texanus* has the chloroplast DNA and nuclear ribosomal RNA genes of *Helianthus debilis* ssp. *cucumerifolius*. *Proceedings of the National Academy of Sciences of the USA* **87**: 593-597
- Rieseberg, L.H., Carter, R. and Zona, S. 1990b Molecular tests of the hypothesised hybrid origin of two diploid *Helianthus* species (Asteraceae). *Evolution* **44**: 1498-1511
- Rieseberg, L.H., Choi, H., Chan, R. and Spore, C. 1993 Genomic map of a diploid hybrid species. *Heredity* **70**: 235-293
-

- Rieseberg, L.H. and Ellstrand, N.C. 1993 What can molecular and morphological markers tell us about plant hybridisation? *Critical Reviews in Plant Sciences* **12**: 213-241
- Rieseberg, L.H., Sinervo, B., Linder, C.R., Ungerer, M.C. and Arias, D.M. 1996 Role of gene interactions in hybrid speciation: evidence from ancient and experimental hybrids. *Science* **272**: 741-745
- Rieseberg, L.H. and Soltis, D.E. 1988 A molecular reexamination of introgression between *Helianthus annuus* and *H. bolanderi* (Compositae). *Evolution* **42**: 227-238
- Rieseberg, L.H. and Soltis, D.E. 1991 Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary Trends in Plants* **5**: 65-84
- Rieseberg, L.H., Van Fossen, C. and Desrochers, A.M. 1995 Hybrid speciation accompanied by genomic reorganisation in wild sunflowers. *Nature* **375**: 313-316
- Rieseberg, L.H. and Wendel, J.F. 1993 Introgression and its consequences in plants. In: *Hybrid Zones and the Evolutionary Process*. Harrison, R.G., ed. pp. 70-109. Oxford University Press, Oxford.
- Ritchie, M.G. 1998 Shocking hybrids. *Trends in Ecology and Evolution* **13**: 123-124
- Rohlf, F.J. 1998 NTSYS-pc: *Numerical taxonomy and multivariate analysis system*. Version 2.0. Exeter Software, New York.
- Roose, M.L. and Gottlieb, L.D. 1976 Genetic and biochemical consequences of polyploidy in *Tragopogon*. *Evolution* **30**: 818-830
- Russell, J.R., Hosein, F., Johnson, E., Waugh, R. and Powell, W. 1993 Genetic differentiation of cocoa (*Theobroma cacao* L.) populations revealed by RAPD analysis. *Molecular Ecology* **2**: 89-97
- Sang, T., Crawford, D.J. and Stuessy, T.F. 1995 Documentation of reticulate evolution in peonies (*Paeonia*) using internal transcribed spacer sequences of nuclear ribosomal DNA: Implications for biogeography and concerted evolution. *Proceedings of the National Academy of Sciences of the USA* **92**: 6813-6817
- Schaal, B.A., O'Kane, S.L.Jr., Rogstad, S.H. 1991 DNA variation in plant populations. *Trends in Ecology and Evolution* **6**: 329-333
- Schemske, D.W. and Morgan, M.T. 1990 The evolutionary significance of hybridisation in *Eucalyptus*. *Evolution* **44**: 2150-2151
- Shoemaker, D.D., Ross, K.G. and Arnold, M.L. 1994 Development of RAPD markers in two introduced fire ants, *Solenopsis invicta* and *S. richteri*, and their application to the study of a hybrid zone. *Molecular Ecology* **3**: 531-539
- Shoemaker, D.D., Ross, K.G. and Arnold, M.L. 1996 Genetic structure and evolution of a fire ant hybrid zone. *Evolution* **50**: 1958-1976
-

- Smith, R.L. and Sytsma, K.J. 1990 Evolution of *Populus nigra* (sect. Aigeiros): introgressive hybridisation and the chloroplast contribution of *Populus alba* (sect. *Populus*). *American Journal of Botany* **77**: 1176-1187
- Sneath, P.H.A. and Sokal, R.R. 1973 *Numerical Taxonomy, the principles and practices of numerical classification*. W.H. Freeman and Company, San Francisco.
- Soltis, D.E. and Soltis, P.S. 1989 Allopolyploid speciation in *Tragopogon*: insights from chloroplast DNA. *American Journal of Botany* **76**: 1119-1124
- Soltis, D.E. and Soltis, P.S. 1993 Molecular data and the dynamic nature of polyploidy. *Critical Reviews in Plant Sciences* **12**: 243-273
- Soltis, D.E. and Soltis, P.S. 1999 Polyploidy: recurrent formation and genome evolution. *Trends in Ecology and Evolution* **14**: 348-352
- Soltis, P.S. and Soltis, D.E. 1991 Multiple origins of the allotetraploid *Tragopogon mirus* (Compositae): rDNA evidence. *Systematic Botany* **16**: 407-413
- Song, K., Lu, P., Tang, K. and Osborn, T.C. 1995 Rapid genomic change in synthetic polyploids of *Brassica* and its implications for polyploid evolution. *Proceedings of the National Academy of Sciences of the USA* **92**: 7719-7723
- Stace, C.A. 1977 The origin of radiate *Senecio vulgaris* L. *Heredity* **39**: 383-388
- Stammers, M., Harris, J., Evans G.M., Hayward, M.D. and Forster, J.W. 1995 Use of random PCR (RAPD) technology to analyse phylogenetic relationships in the *Lolium/Festuca* complex. *Heredity* **74**: 19-27
- Stewart, C.N.Jr. and Excoffier, L. 1996 Assessing population genetic structure and variability with RAPD data: Application to *Vaccinium macrocarpon* (American Cranberry). *Journal of Evolutionary Biology* **9**: 153-171
- Swofford, D.L. and Selander, R.B. 1981 Biosys-1: a FORTRAN program for the comprehensive analysis for electrophoretic data in population genetics and systematics. *Journal of Heredity* **72**: 281-283
- Sytsma, K.J. 1990 DNA and morphology: inference of plant phylogeny. *Trends in Ecology and Evolution* **5**: 104-110
- Szymura, J.M. and Barton, N.H. 1991 The genetic structure of the hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata*: comparisons between transects and between loci. *Evolution* **45**: 237-261
- Taylor, L. 1984 *The Potential for Introgression in a British Polyploid Complex*. Unpublished PhD thesis, University of St. Andrews.
- Thomas, H.M., Morgan, W.G., Meredith, M.R., Humphreys, M.W., Thomas, H. and Leggett, J.M. 1994 Identification of parental and recombined chromosomes in hybrid derivatives of *Lolium multiflorum* x *Festuca pratensis* in genomic *in situ* hybridisation. *Theoretical and Applied Genetics* **88**: 909-913
-

- Thompson, J.D. and Lumaret, R. 1992 The evolutionary dynamics of polyploid plants: origins, establishment and persistence. *Trends in Ecology and Evolution* **7**:302-307
- Thormann, C.E., Ferreira, M.E., Camargo, L.E.A., Tivang, J.G. and Osborn, T.C. 1994 Comparison of RFLP and RAPD markers to estimating genetic relationships within and among cruciferous species. *Theoretical and Applied Genetics* **88**: 973-980
- Thorpe, R.S. 1984 Primary and secondary transition zones in speciation and population differentiation: a phylogenetic analysis of range expansion. *Evolution* **38**: 233-243
- Villani, F., Sansotta, A., Cherubini, M., Cesaroni, D. and Sbordoni, V. 1999 Genetic structure of natural populations of *Castanea sativa* in Turkey: evidence of a hybrid zone. *Journal of Evolutionary Biology* **12**: 233-244
- Walters, S.M. 1963 *Senecio rupestris* Waldst. and Kit. and *Senecio squalidus* L. *Proceedings of the Botanical Society of the British Isles* **5**: 382
- Wang, H. McArthur, E.D., Sanderson, S.C. and Graham, J.H. 1997 Narrow hybrid zone between two subspecies of big sagebrush (*Artemisia tridentata*: Asteraceae). IV. Reciprocal transplant experiments. *Evolution* **51**: 95-102
- Wang, H. McArthur, E.D. and Freeman, D.C. 1999 Narrow hybrid zone between two subspecies of big sagebrush (*Artemisia tridentata*: Asteraceae). IX. Elemental uptake and niche separation. *American Journal of Botany* **86**: 1099-1107
- Wang, X.-R., Szmidt, A.E., Lewandowski, A. and Wang, Z.-R. 1990 Evolutionary analysis of *Pinus densata* Masters, a putative Tertiary hybrid. 1. Allozyme variation. *Theoretical and Applied Genetics* **80**: 635-640
- Wang, X.-R. and Szmidt, A.E. 1990 Evolutionary analysis of *Pinus densata* (Masters), a putative Tertiary hybrid. 2. A study using species-specific chloroplast DNA markers. *Theoretical and Applied Genetics* **80**: 641-647
- Wang, X.-R. and Szmidt, A.E. 1994 Hybridisation and chloroplast DNA variation in a *Pinus* species complex from Asia. *Evolution* **48**: 1020-1031
- Warwick, S.I., Bain, J.F., Wheatcroft, R. and Thompson, B.K. 1989 Hybridisation and introgression in *Carduus nutans* and *C. acanthoides* reexamined. *Systematic Botany* **14**: 476-494
- Weir, B.S. and Cockerham, C.C. 1984 Estimating *F*-statistics for the analysis of population structure. *Evolution* **38**: 1358-1370
- Weising, K., Nybom, H., Wolff, K. and Meyer, W. 1995 *DNA Fingerprinting in Plants and Fungi*. CRC Press, Boca Raton, Florida.
- Welsh, J. and McClelland M. 1990 Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Research* **18**: 7213-7218
- Welsh, J. and McClelland M. 1991 Genomic fingerprinting using arbitrarily primed PCR and a matrix of pairwise combinations of primers. *Nucleic Acids Research* **19**: 5275-5279
-

- Wendel, J.F., Stewart, J. MCD. and Rettig, J.H. 1991 Molecular evidence for homoploid reticulate evolution among Australian species of *Gossypium*. *Evolution* **45**: 694-711
- Whittemore, A.T. and Schaal, B.A. 1991 Interspecific gene flow in sympatric oaks. *Proceedings of the National Academy of Sciences of the USA* **88**: 2540-2544
- Williams, J.G.K., Hanafey, M.K., Rafalski, J.A. and Tingey, S.V. 1993 Genetic analysis using random amplified polymorphic markers. *Methods in Enzymology* **218**: 704-740
- Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A., and Tingey, S.V. 1990 DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research* **18**: 6531-6535
- Wolfe, A.D. and Elisens, W.J. 1994 Nuclear ribosomal DNA restriction-site variation in *Penstemon* section *Peltanthera* (Scrophulariaceae): an evaluation of diploid hybrid speciation and evidence for introgression. *American Journal of Botany* **81**: 1627-1635
- Wolfe, A.D., Xiang, Q.-Y., and Kephart, S.R. 1998 Assessing hybridisation in natural populations of *Penstemon* (Scrophulariaceae) using hypervariable intersimple sequence repeat (ISSR) bands. *Molecular Ecology* **7**: 1107-1125
- Wolff, K., El-Akkad, S. and Abbott, R.J. 1997 Population substructure in *Alkanna orientalis* (Boraginaceae) in the Sinai Desert, in relation to its pollinator behaviour. *Molecular Ecology* **6**: 365-372
- Wolff, K. and Peters-Van Rijn, J. 1993 Rapid detection of genetic variability in chrysanthemum (*Dendranthema grandiflora* Tzvelev) using random primers. *Heredity* **71**: 335-341
- Young, N.D. 1996 An analysis of the causes of genetic isolation in two Pacific Coast iris hybrid zones. *Canadian Journal of Botany* **74**: 2006-2013
- Zietkiewicz, E., Rafalski, A. and Labuda, D. 1994 Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. *Genomics* **20**: 176-183
-

# Appendices

## **Appendix A**

### Recipes of solutions

#### *CTAB extraction buffer*

0.1 M tris base [tris(hydroxymethyl)methylamine]  
20 mM EDTA [ethylenediaminetetra-acetic acid disodium salt]  
1.4 M NaCl  
2% (w/v) CTAB [hexadecyl trimethylammonium bromide]  
1% (w/v) PVP [polyvinylpyrrolidone AMW 40000]  
pH 8.0 with HCl  
2% 2-mercaptoethanol (added to buffer just before use)

#### *Wet chloroform*

96% chloroform  
4% iso-Amyl alcohol

#### *Sodium acetate wash buffer*

76% ethanol  
0.2 M sodium acetate

#### *Ammonium acetate wash buffer*

76% ethanol  
10 mM ammonium acetate

#### *TE*

1 mM EDTA (disodium salt)  
10 mM tris base  
fixed to pH 7.6 with HCl

#### *RNase solution*

10 mg/ml RNase A (Boehringer Mannheim) in sterile distilled water  
boiled for 10 min

#### *5X Tris-borate-EDTA buffer (TBE) (used at 0.5x)*

0.45 M tris base  
0.445 M orthoboric acid  
12.5 mM EDTA (disodium salt)  
~ pH 8.0

#### *Loading buffer*

0.042% bromophenol blue (0.25% for RFLP analysis)  
60 mM EDTA (disodium salt)  
30% glycerol

#### *RFLP size marker*

75% (v/v) loading buffer (RFLP)  
25% (v/v) 0.2 µg/µl λ DNA digested with *Hind* III (nbl GENE SCIENCES)

---

*Southern denaturation buffer*

1.5 M NaCl  
0.5 M NaOH

*Southern neutralisation buffer*

1.5 M NaCl  
0.5 M tris base  
1 mM EDTA

*20X SSC*

3 M NaCl  
0.3M trisodium citrate

*Pre-hybridisation and hybridisation buffer*

5X sterile SSC  
0.1% sodium-Sarkosyl  
0.02% SDS [sodium dodecyl sulphate]  
0.5% blocking stock [see below]

*Low stringency solution*

2X SSC  
0.1% (w/v) SDS

*Medium stringency solution*

1X SSC  
0.1% SDS

*Buffer 1*

0.1 M maleic acid  
0.15 M NaCl  
fixed to pH 7.5 with NaOH

*Buffer 2*

Blocking stock (10% w/v blocking reagent (Boehringer Mannheim) in buffer 1, sterile)  
diluted 1:20 in buffer 1

*Buffer 3*

0.1 M tris-HCl, pH 9.5  
0.1 M NaCl  
50 mM MgCl<sub>2</sub>

*CSPD<sup>®</sup> solution*

25 ml buffer 3  
250 µl CSPD<sup>®</sup> (Applied Biosciences)

*10X Tris-acetate-EDTA buffer (TAE) (used at 1X)*

400 mM tris base  
10 mM EDTA  
fixed to pH 7.6 with glacial acetic acid

---

*Isozyme extraction buffer*

50 ml  $\text{LiBO}_3$  gel buffer (see below)  
37 mg KCl  
10 mg  $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$   
18 mg EDTA tetra-sodium salt  
25 mg PVP  
0.5 ml Triton X100  
~ 1 ml 2-mercaptoethanol  
fixed to pH 8.3

*Lithium borate ( $\text{LiBO}_3$ ) electrode buffer*

per l of buffer:  
11.9 g boric acid  
1.2 g lithium hydroxide  
fixed to pH 8.3 with dry constituents

*Lithium borate ( $\text{LiBO}_3$ ) gel buffer*

per l of buffer:  
100 ml  $\text{LiBO}_3$  electrode buffer  
5.45 g tris base  
1.28 g anhydrous citric acid  
fixed to pH 8.3 with citric acid or 1 M NaOH

*Tris citrate (TC) electrode buffer*

per l of buffer:  
16.35 g tris base  
6.1 g citric acid monohydrate  
fixed to pH 8.0 with 8 M NaOH or citric acid

*Tris citrate gel buffer*

1 : 14 dilution of TC electrode buffer

*Appropriate buffer systems and slices for isozyme stains*

$\text{LiBO}_3$  electrode/gel buffer

top slice - *Acp*

middle slice - *Aat*

bottom slice - *Pgm*

TC electrode/gel buffer

top slice - *Aco*

middle slice - *Pgi*

bottom slice - *Idh*

---

*Aat stain*

50 ml 0.1 M Tris HCl buffer pH 8.5  
18 mg  $\alpha$ -ketoglutaric acid  
65 mg aspartic acid  
250 mg PVP-40T  
25 mg EDTA  
710 mg  $\text{Na}_2\text{HPO}_4$   
5 mg pyridoxal-5-phosphate  
200 mg fast blue BB

*Aco stain*

50 ml 0.2 M Tris HCl buffer pH 8,0  
75 mg cis-aconitic acid  
0.2033 g  $\text{MgCl}_2$   
15 mg NADP }  
10 mg MTT } in 0.5 ml 10%  $\text{MgCl}_2$   
3 mg PMS }

*Acp stain*

Pre-soak: 50 ml 0.4 M acetate buffer pH 5.0 for 20 min at 4°C

50 ml 0.2 M acetate buffer pH 5.0  
50 mg Sodium  $\alpha$ -naphthyl acid phosphate  
0.5 ml 10%  $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$   
40 mg fast garnet

*Idh stain*

50 ml 0.1 M Tris HCl buffer pH 8.0  
75 mg isocitric acid (trisodium salt)  
10 mg NADP }  
15 mg MTT } in 0.5 ml 10%  $\text{MgCl}_2$   
3 mg PMS }

*Pgi stain*

50 ml 0.1 M Tris HCl buffer pH 8.5  
40 mg fructose-6-phosphate (disodium salt)  
12 units glucose-6-phosphate dehydrogenase  
10 mg NADP }  
15 mg MTT } in 0.5 ml 10%  $\text{MgCl}_2$   
3 mg PMS }

*Pgm stain*

50 ml 0.1 M Tris HCl buffer pH 7.5  
100 mg glucose-1-phosphate  
30 mg ATP  
25 units glucose-6-phosphate dehydrogenase  
10 mg NADP }  
15 mg MTT } in 0.5 ml 10%  $\text{MgCl}_2$   
1 mg PMS }

---

## **Appendix B**

Binary data matrix showing presence (1) and absence (0) of 26 RAPD/ISSR markers in *Senecio chrysanthemifolius*, *S. aethnensis* and hybrid populations from the Mt. Etna area of Sicily.

Twenty six characters (in order): A9-C, A11-C, A20-C, B15-C, C2-C, C5-C, F12-C, G5-C, G8-C, G10-C, G12-C, H7-C, 40-C, A11-A, B15-A, C2-A, C20-A, F12-A, G5-A, G6-A, G8-A, G11-A, G12-A, H13-A, H19-A, 18-A

*S. chrysanthemifolius* markers are identified by the suffix -C and *S. aethnensis* markers are identified by the suffix -A.

### **Popn. C0**

C0-01 10111111111110000000000000  
 C0-02 11111110111110000000001000  
 C0-03 10101111111110000100000000  
 C0-04 11111110111110000000000000  
 C0-05 11101111111110000000000000  
 C0-06 11111110111110000000100000  
 C0-07 11111111011110000000001000  
 C0-08 11111111111110000000000000  
 C0-90 11111111111100000000000000  
 C0-10 11111111111110000000000000  
 C0-11 11101111111110000000000000  
 C0-12 01101110101110000001101100  
 C0-13 11101111111110000000000000  
 C0-14 11111110111110000000000000  
 C0-15 10111011111110000000100000  
 C0-16 11010111011110000000100000

### **Popn. C1**

C1-01 11111111111010000000000000  
 C1-02 11111111111100000000000000  
 C1-03 01111111111100000000010000  
 C1-04 10111111111100000000001000  
 C1-05 11111111111110000000000000  
 C1-06 11111111011110000000000000  
 C1-07 11111111111000100000000000  
 C1-08 11111111111100000000000000  
 C1-09 11111111111100000000010000  
 C1-10 11111111011000010000000000  
 C1-11 10111111101100010000000001  
 C1-12 11111111011100000000000000  
 C1-13 10110111111110010000000000  
 C1-14 11111111111100000000000000

### **Popn. C9**

C9-01 01111111111110000000000000  
 C9-02 01111111110010000000000000  
 C9-03 10101111111010000000000000  
 C9-04 11110111111100010000000000  
 C9-05 11111111111010000000000000  
 C9-06 11111110111110000000000001  
 C9-07 11111111111010000000000000  
 C9-08 11111111111010000000000000  
 C9-09 10111111101110000000000000  
 C9-10 11111111111100000100000000  
 C9-11 11111110111100000000000001  
 C9-12 11111111111010010000000000  
 C9-13 01111111111100000000100000  
 C9-14 11111111100010000000001000  
 C9-15 11111111101110000000100000

### **Popn. S2**

S2-01 10101111101100000000000000  
 S2-02 00010010110110000000000001  
 S2-03 11111111011100000000000000  
 S2-04 10111111011100000000000000  
 S2-05 01111110111110001001001100  
 S2-06 00011111111110000000000000  
 S2-07 11111111011100100000000000  
 S2-08 11011110111110000000000000  
 S2-09 11111110111110000000010000  
 S2-10 11101111111100000000000010  
 S2-11 11111110101110000000010000  
 S2-12 11011111011100000001010000  
 S2-13 10111111111100000000000000  
 S2-14 11111111011110100000010000

**Popn. S5**

S5-01 11111111111110000000111010  
 S5-02 10111111111111000000000010  
 S5-03 11110111101100000000100000  
 S5-04 11111111101100000000011010  
 S5-05 11111111111110000000000000  
 S5-06 11111110111111000000011000  
 S5-07 011111111111110000000111000  
 S5-08 11111111111111000000000010  
 S5-09 11011111111111010000111000  
 S5-10 11011110101111000000000000  
 S5-11 11111111111101000000011000  
 S5-12 11111111101100000010011011  
 S5-13 01111111111101000010011000  
 S5-14 11111111101110000000010000

**Popn. S6**

S6-01 110111111111110000000001000  
 S6-02 111011111111110000000110000  
 S6-03 11111111111100000000000001  
 S6-04 11110111101110000000000010  
 S6-05 01111110111111000100001001  
 S6-06 01111111111110000100011001  
 S6-07 11101111001110000000010010  
 S6-08 01101111111110000100000000  
 S6-09 010011111011110000000000010  
 S6-10 11011111111110000101001100  
 S6-11 111111101111110000000001000  
 S6-12 01011110101111000000001000  
 S6-13 11111111101110000000001010  
 S6-14 01011110101101000000001000  
 S6-15 01111111101111000000110010

**Popn. S8**

S8-01 011101111111110000010110000  
 S8-02 11111111101111000010110000  
 S8-03 0111111111010010010101000  
 S8-04 1011111111110000000101010  
 S8-05 11111111001111001001101100  
 S8-06 10111111001111000010110000  
 S8-07 0111111111111000000010000  
 S8-08 10111111111110000010101001  
 S8-09 01111111001111010010110010  
 S8-10 11101111111010000000110001  
 S8-11 11111111111110000000000001  
 S8-12 00111111111111000000010000  
 S8-13 01111111111110000000000000  
 S8-14 01111111111110000000000000  
 S8-15 01111111111111010010101000

**Popn. S9**

S9-01 011111111011110011001001111  
 S9-02 011111111111101000000010000  
 S9-03 111111111011110000000101001  
 S9-04 101111011111010001001001110  
 S9-05 11111111101101000000100000  
 S9-06 10111111010101000010011000

**Popn. S10**

10-01 00111101011111101011111111  
 10-02 11111011011011101001111111  
 10-03 01011101101110001001011100  
 10-04 10111101101110001011111111  
 10-05 10111111101111000000000000  
 10-06 11111011111110000000000000  
 10-07 11111111101110000000010010  
 10-08 11111111101110001101111100

**Popn. S11**

11-01 11111111111111001011111110  
 11-02 00110111110100101111111101  
 11-03 11111110001100010010111010  
 11-04 00010011100101100101111101  
 11-05 10000011101001101111111110  
 11-06 01000001000101001101101101  
 11-07 01010101101111001101111111  
 11-08 00110000000011001111101111  
 11-09 11111101000001101111111110  
 11-10 00011111101000101110110111  
 11-11 01111111101101001001111100  
 11-12 01111110101111011011101101  
 11-13 00110000100010001101111101  
 11-14 00001001100001111101011111

---

**Popn. RD**

RD-01 01100000100001001111111110  
RD-02 00000100000101001001111111  
RD-03 00010000000011011101111111  
RD-04 00100001010001011101111111  
RD-05 01001010100001111001011111  
RD-06 00001000000001111101011111  
RD-07 00010000000011111101111111  
RD-08 01100001000011011101111111  
RD-09 00010011100011011111111111  
RD-10 00000101000101011101111111  
RD-11 00000000000011011001111110  
RD-12 10110011000011000000000011  
RD-13 01101001000011011001101111  
RD-14 00000100001101001101111111

**Popn. BB**

BB-01 00000000000001001101111110  
BB-02 00000000000001111001111111  
BB-03 00000000000001011111111111  
BB-04 00100000000001111011111111  
BB-05 00000000000001111101111111  
BB-06 00000000000001011111111111  
BB-07 00000000000001101111101111  
BB-08 00100000000001111001101111  
BB-09 00000000000001001111101111  
BB-10 00000011000011011101111111  
BB-11 00000000000001101111111111  
BB-12 00100000000001111011111111  
BB-13 00000000000001011111111111  
BB-14 00000000000001011101111111  
BB-15 00000000000001111111111111  
BB-16 00100000000001011011111111  
BB-17 00000000000001011001111111

**Popn. VB**

VB-01 00000000000001111111111111  
VB-02 00000010000001111111111111  
VB-03 00000000000001111101111111  
VB-04 00000000000001001111111111  
VB-05 00000010000001001111111111  
VB-06 00000000000001101011111111  
VB-07 00000000000001011111111111  
VB-08 00000000000001011101111111  
VB-09 00000000000001011011111111  
VB-10 00000000000001111111111111  
VB-11 00001010000001011111111111  
VB-12 00000000000001011111111111  
VB-13 10000000000001011111111111  
VB-14 00000000000001111111111111  
VB-15 00000100000001111101111110  
VB-16 00000001100001011111111111  
VB-17 00001000000001011001101110

**Popn. TC**

TC-01 00000000000001111001111111  
TC-02 00000001000001011101111111  
TC-03 00000000000001111111111111  
TC-04 00000000000001111101111111  
TC-05 00000000000001111101111111  
TC-06 00000000000001111100111111  
TC-07 00000001000001101101111111

**Popn.'s VO and UJ**

VO-01 00010100000101011001111110  
VO-02 00000000000001111111111111  
UJ-01 00000000000001111111111110  
UJ-02 00000000000001101101101111

**Popn. F1**

F1-01 00111111101011101101111110  
F1-02 10111110111111011101111111  
F1-03 11001110111011011101001111  
F1-04 11001110110101001101011111  
F1-05 11001110111111011001111110  
F1-06 01111110111111101101111111  
F1-07 11111111110110111011111111  
F1-08 11111110011111001101101110  
F1-09 11010110110101001001111110  
F1-10 11111111111111001001111111

---

Binary data matrix showing presence (1) and absence (0) of 26 RAPD/ISSR markers in 6 populations of British *S. squalidus*.

Twenty six characters (in order): A9-C, A11-C, A20-C, B15-C, C2-C, C5-C, F12-C, G5-C, G8-C, G10-C, G12-C, H7-C, 40-C, A11-A, B15-A, C2-A, C20-A, F12-A, G5-A, G6-A, G8-A, G11-A, G12-A, H13-A, H19-A, 18-A

*S. chrysanthemifolius* markers are identified by the suffix -C and *S. aethnensis* markers are identified by the suffix -A.

#### **Popn. PB**

PB-01 01101101110100011001001101  
PB-02 11101010110110010001111101  
PB-03 11001111110110011001111101  
PB-04 01101111110110001001101101  
PB-05 11001011100110010000111001  
PB-06 11001110010100010001111101  
PB-07 00101010110111010000001001  
PB-08 10001111110110011001101101  
PB-09 00101111110100011001011101  
PB-10 10101111010100010000010001  
PB-11 01101111110110001001111101  
PB-12 00001111110100011001111100

#### **Popn. OX**

OX-01 11001011100010011001001101  
OX-02 00001101110110011001101101  
OX-03 11001001010010011001001001  
OX-04 10001111110010011000100000  
OX-05 00001001100010011001001101  
OX-06 11001001110010011000110101  
OX-07 10001011110010010000001001  
OX-08 10001001110110011101101101  
OX-09 00001001100010001001001100  
OX-10 00101101010110011001011101  
OX-11 10101111100110011001101101  
OX-12 01101011110000011000001000

#### **Popn CA**

CA-01 10001111110111001001101100  
CA-02 00001011110100000000111001  
CA-03 01001011110110011001101101  
CA-04 10001011110101010000010001  
CA-05 11001011110110010001111001  
CA-06 11001011100100010000011001  
CA-07 10001011100110010001101001  
CA-08 10001011100100010000101001  
CA-09 1100111111010101000001101  
CA-10 00101010100110010000111001  
CA-11 10001011000110010000010001  
CA-12 00001111000110011001101100

#### **Popn. NO**

NO-01 01100111100110010000101001  
NO-02 11101011110110010000101001  
NO-03 01101101110100011001101101  
NO-04 00101111110110001000100001  
NO-05 01001111110110000000100001  
NO-06 00100111110111011001101101  
NO-07 10101111110111010000001001  
NO-08 11101011110110010000001001  
NO-09 11000110110111011000101001  
NO-10 11001101110110010000100001  
NO-11 11001111110110010000100001  
NO-12 11101111110110011001101101

#### **Popn. HU**

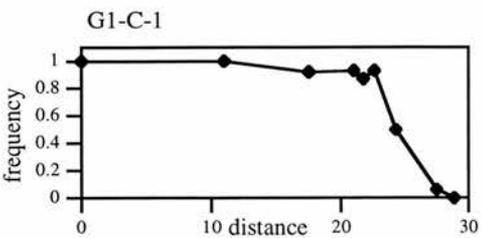
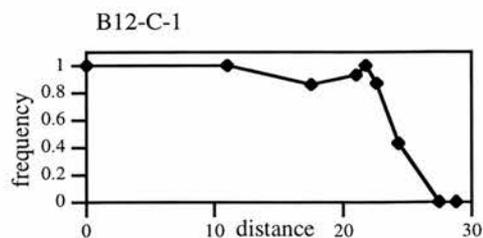
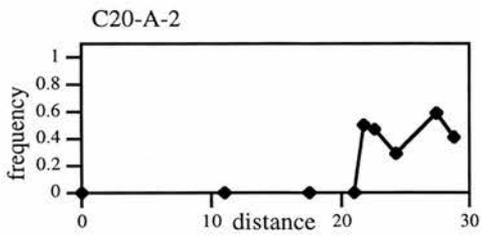
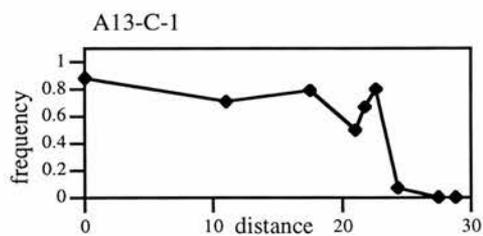
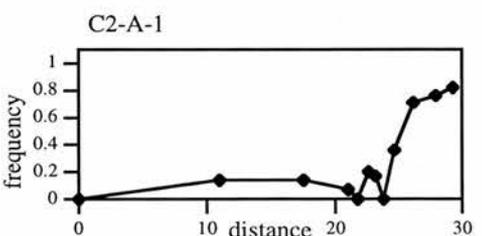
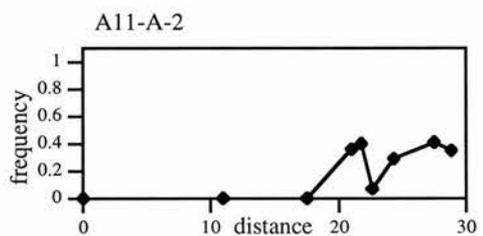
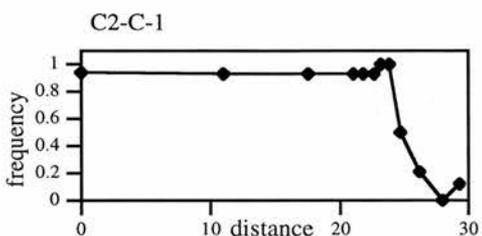
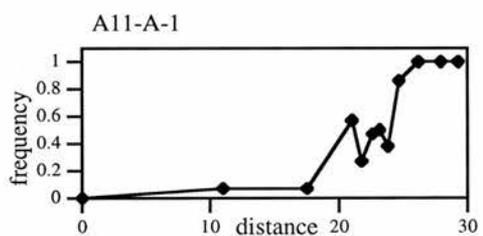
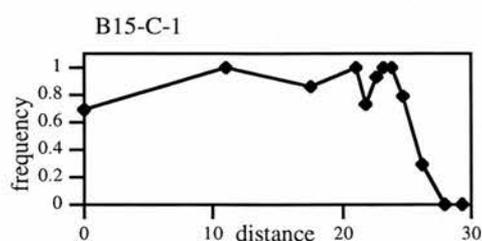
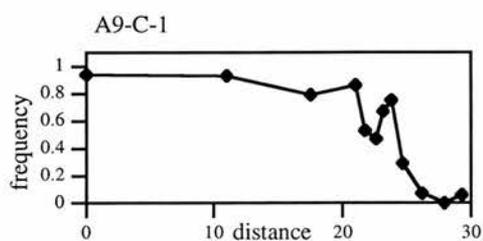
HU-01 01101111100110011001001101  
HU-02 10101111110100011001001101  
HU-03 10101111100100011001101101  
HU-04 10001110100111010001001101  
HU-05 00101110100110010000101001  
HU-06 11001111100110011001101101  
HU-07 11101111100110011000001101  
HU-08 11101110010110011001101101  
HU-09 11101110100110011001001101  
HU-10 00001011100111011001101101  
HU-11 10101110100110011001101101  
HU-12 00101011100101011001101101

#### **Popn. KY**

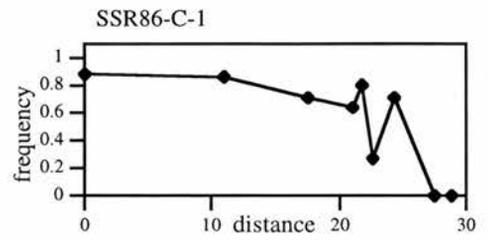
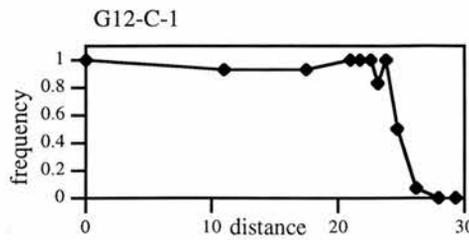
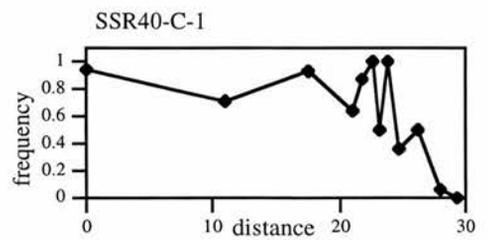
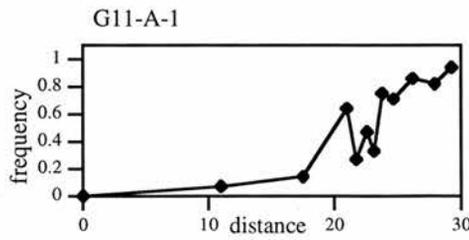
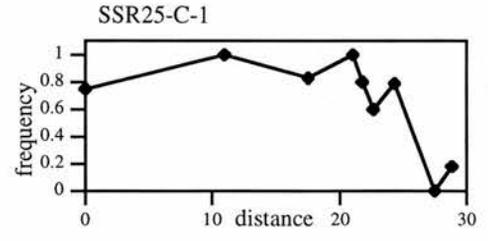
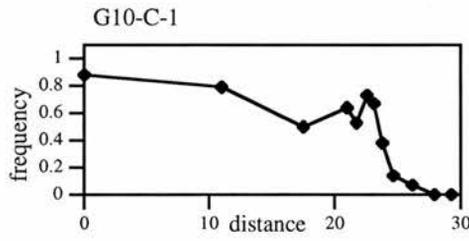
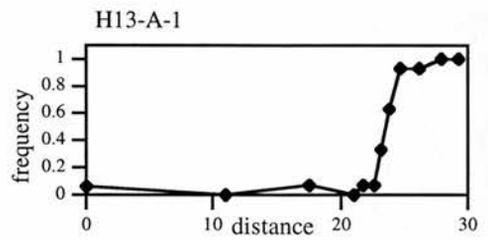
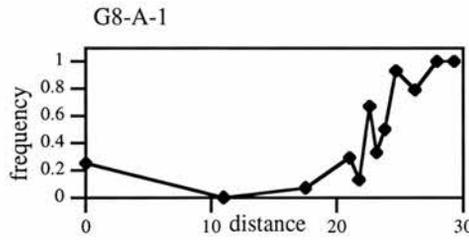
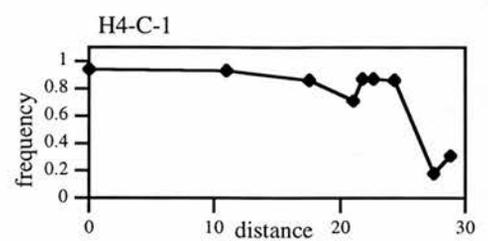
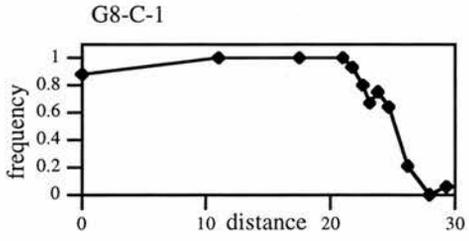
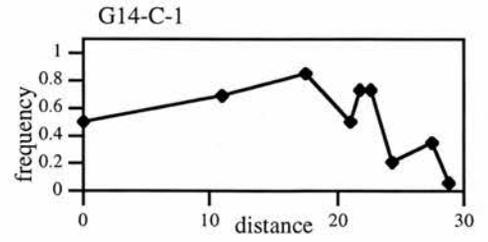
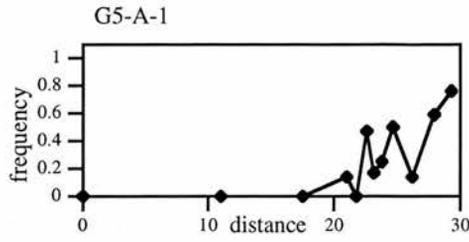
KY-01 11101110110111011001101101  
KY-02 11101111100111001000100001  
KY-03 01001111110111010000100001  
KY-04 1100111011011101100000001  
KY-05 1010111110011100100000001  
KY-06 1110111001011101100000001  
KY-07 00101110110100000000100000  
KY-08 11101110110111010000100001  
KY-09 0010111011011101100000001  
KY-10 10101010100001011000010000  
KY-11 11101110100111011001000101  
KY-12 11001111000111011000000000

## Appendix C

**Appendix C** Frequency of 22 *S. chrysanthemifolius* (C) and *S. aethnensis* (A) RAPD and ISSR (denoted SSR) markers in populations along transect S, through the postulated hybrid zone between *S. chrysanthemifolius* and *S. aethnensis* on Mt. Etna, plotted against cumulative distance of each population from the start of the transect (km).



**Appendix C**



**Appendix D****Table D1** Frequencies of taxon-specific RAPD/ISSR markers (bands) in populations along transect S, through the postulated hybrid zone between *Senecio chrysanthemifolius* and *S. aethnensis*, on Mt. Etna.

Population	Primer	'Pure' <i>S. chrysanthemifolius</i>		S2	S5	S6	S8	S11	'Pure' <i>S. aethnensis</i>	
		C0	C1						BB	VB
Altitude		150 m	600 m	950 m	1300 m	1400 m	1500 m	1800 m	2525 m	2600 m
Distance from popn. C0		0 km	11.00 km	17.50 km	21.00 km	21.75 km	22.60 km	24.30 km	27.45 km	28.80 km
		n = 16	n = 14	n = 14	n = 14	n = 15	n = 15	n = 14	n = 17	n = 17
<b>A11</b>	<b>AET-2</b>	0.00	0.00	0.00	0.36	0.40	0.07 <sup>14</sup>	0.29	0.41	0.35
<b>A13</b>	<b>CHR-1</b>	0.88	0.71	0.79	0.50	0.67	0.80	0.07	0.00	0.00
<b>A14</b>	<b>CHR-1</b>	1.00	0.85 <sup>13</sup>	0.93	0.93	0.80	0.80	0.21	0.00	0.00
	<b>CHR-2</b>	0.94	0.69 <sup>13</sup>	0.93	0.79	0.80	0.87	0.64	0.12	0.00
	<b>AET-1</b>	0.00	0.06 <sup>13</sup>	0.21	0.36	0.53	0.20	0.86	0.76	0.88
<b>B6</b>	<b>CHR-1</b>	1.00 <sup>15</sup>	0.69 <sup>13</sup>	0.57	0.71	0.53	0.53	0.50	0.29	0.06
	<b>AET-1</b>	0.00 <sup>15</sup>	0.08 <sup>13</sup>	0.00	0.14	0.07	0.13	0.64	0.65	0.53
<b>B12</b>	<b>CHR-1</b>	1.00	1.00	0.86	0.93	1.00	0.87	0.43	0.00	0.00
<b>C11</b>	<b>AET-1</b>	0.00 <sup>15</sup>	0.00	0.00	0.08 <sup>13</sup>	0.07 <sup>14</sup>	0.00	0.36	0.24	0.47
<b>C20</b>	<b>AET-2</b>	0.00 <sup>15</sup>	0.00	0.00	0.00	0.50 <sup>14</sup>	0.47	0.29	0.59	0.41
<b>G1</b>	<b>CHR-1</b>	1.00	1.00	0.92 <sup>13</sup>	0.93	0.87	0.93	0.50	0.06	0.00
<b>G14</b>	<b>CHR-1</b>	0.50	0.69 <sup>13</sup>	0.85 <sup>13</sup>	0.50	0.73	0.73	0.21	0.35	0.06
	<b>CHR-1</b>	0.94	0.93	1.00 <sup>13</sup>	0.93	0.80	0.73	0.79	0.29	0.12
<b>G15</b>	<b>AET-1</b>	0.06	0.00	0.14	0.43	0.33	0.14 <sup>14</sup>	0.64	1.00 <sup>16</sup>	1.00
	<b>AET-2</b>	0.00	0.21	0.29	0.21	0.40	0.21 <sup>14</sup>	0.50	0.56 <sup>16</sup>	0.65
<b>H4</b>	<b>CHR-1</b>	0.94	0.93	0.86	0.71	0.87	0.87	0.86	0.18	0.31 <sup>16</sup>
<b>H19</b>	<b>CHR-1</b>	1.00	1.00 <sup>13</sup>	1.00	1.00	1.00	1.00	0.29	0.00	0.06 <sup>16</sup>
<b>B2/B10</b>	<b>CHR-1</b>	0.73 <sup>15</sup>	0.64	0.64	0.57	0.62 <sup>13</sup>	0.62 <sup>13</sup>	0.38 <sup>13</sup>	0.12	0.12
<b>F2/G9</b>	<b>AET-1</b>	0.07 <sup>15</sup>	0.14	0.07	0.36	0.40	0.40	0.62 <sup>13</sup>	0.94	0.88
	<b>AET-2</b>	0.00 <sup>15</sup>	0.00	0.07	0.43	0.40	0.20	0.69 <sup>13</sup>	1.00	1.00
<b>18</b>	<b>CHR-1</b>	0.94	0.93	0.86	0.79	0.67	0.73	0.71	0.00 <sup>16</sup>	0.00
<b>25</b>	<b>CHR-1</b>	0.75	1.00	0.83 <sup>12</sup>	1.00	0.80	0.60	0.79	0.00	0.18
<b>26</b>	<b>CHR-1</b>	0.67 <sup>15</sup>	0.86	0.69 <sup>13</sup>	0.57	0.80	0.73	0.50	0.06	0.00
	<b>AET-1</b>	0.00 <sup>15</sup>	0.00	0.00 <sup>13</sup>	0.00	0.00	0.00	0.50	1.00	1.00
	<b>AET-2</b>	0.13 <sup>15</sup>	0.36	0.15 <sup>13</sup>	0.57	0.60	0.47	0.93	1.00	1.00
<b>57</b>	<b>CHR-1</b>	1.00 <sup>15</sup>	1.00	1.00	1.00	1.00	1.00	0.50	0.00	0.00
	<b>AET-1</b>	0.00 <sup>15</sup>	0.00	0.00	0.07	0.07	0.00	0.71	1.00	1.00
<b>86</b>	<b>CHR-1</b>	0.88	0.86	0.71	0.64	0.80	0.27	0.71	0.00	0.00
<b>88</b>	<b>AET-1</b>	0.06	0.07	0.21	0.36	0.20	0.21 <sup>14</sup>	0.43	0.82	0.71

**Table D2** Frequencies of taxon-specific RAPD/ISSR markers in populations along transect S, through the postulated hybrid zone between *Senecio chrysanthemifolius* and *S. aethnensis*, on Mt. Etna.

Population	'Pure' <i>S. chrysanthemifolius</i>										'Pure' <i>S. aethnensis</i>				
	C0	C1	S2	S5	S6	S8	S9	S10	S11	RD	BB	VB	F1		
Altitude	150 m	600 m	950 m	1300 m	1400 m	1500 m	1600 m	1650 m	1800 m	2175 m	2525m	2600m	-		
Distance from popn. C0	0 km	11.00 km	17.50 km	21.00 km	21.75 km	22.60 km	23.15 km	23.80 km	24.68 km	26.18 km	27.93 km	29.28 km	-		
	n = 16	n = 14	n = 14	n = 14	n = 15	n = 15	n = 6	n = 8	n = 14	n = 14	n = 17	n = 17	n = 10		
Primer	Band														
A9	CHR-1	0.94	0.93	0.79	0.86	0.53	0.47	0.67	0.75	0.29	0.07	0.00	0.06	0.80	
A11	CHR-1	0.81	0.79	0.64	0.93	1.00	0.73	0.67	0.63	0.50	0.29	0.00	0.00	0.80	
	AET-1	0.00	0.07	0.07	0.57	0.27	0.47	0.50	0.38	0.86	1.00	1.00	1.00	1.00	
A20	CHR-1	0.94	1.00	0.71	0.86	0.67	1.00	1.00	0.88	0.57	0.36	0.24	0.00	0.60	
B15	CHR-1	0.69	1.00	0.86	1.00	0.73	0.93	1.00	1.00	0.79	0.29	0.00	0.00	0.70	
	AET-1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.25	0.29	0.21	0.47	0.41	0.20	
C2	CHR-1	0.94	0.93	0.93	0.93	0.93	0.93	1.00	1.00	0.50	0.21	0.00	0.12	0.90	
	AET-1	0.00	0.14	0.14	0.07	0.00	0.20	0.17	0.00	0.36	0.71	0.76	0.82	0.40	
C5	CHR-1	0.94	1.00	0.93	1.00	1.00	1.00	1.00	0.75	0.57	0.21	0.00	0.06	1.00	
C20	AET-1	0.00	0.14	0.07	0.00	0.00	0.07	0.33	0.63	0.86	0.93	1.00	1.00	1.00	
F12	CHR-1	1.00	1.00	1.00	1.00	1.00	1.00	0.83	0.63	0.57	0.21	0.06	0.18	1.00	
	AET-1	0.06	0.00	0.00	0.00	0.27	0.00	0.00	0.13	0.71	0.64	0.65	0.82	0.70	
G5	CHR-1	0.75	1.00	0.64	0.86	0.73	1.00	1.00	1.00	0.71	0.43	0.06	0.06	0.30	
	AET-1	0.00	0.00	0.00	0.14	0.00	0.47	0.17	0.25	0.50	0.14	0.59	0.76	0.00	
G6	AET-1	0.06	0.00	0.07	0.00	0.07	0.07	0.33	0.63	0.93	0.93	1.00	1.00	1.00	

Table D2 continued.

Population	'Pure' <i>S. chrysanthemifolius</i>											'Pure' <i>S. aethnensis</i>			
	C0	C1	S2	S5	S6	S8	S9	S10	S11	RD	BB	VB	F1		
Primer	Band	n = 16	n = 14	n = 14	n = 15	n = 15	n = 6	n = 8	n = 14	n = 14	n = 17	n = 17	n = 10		
G8	CHR-1	0.88	1.00	1.00	0.93	0.80	0.67	0.75	0.64	0.21	0.00	0.06	0.90		
	AET-1	0.25	0.00	0.07	0.13	0.67	0.33	0.50	0.93	0.79	1.00	1.00	0.80		
G10	CHR-1	0.88	0.79	0.50	0.53	0.73	0.67	0.38	0.14	0.07	0.00	0.00	0.90		
G11	AET-1	0.00	0.07	0.14	0.27	0.47	0.33	0.75	0.71	0.86	0.82	0.94	0.80		
G12	CHR-1	1.00	0.93	0.93	1.00	1.00	0.83	1.00	0.50	0.07	0.00	0.00	0.80		
	AET-1	0.19	0.14	0.21	0.53	0.33	0.67	0.63	1.00	0.93	1.00	1.00	1.00		
H7	CHR-1	1.00	0.93	1.00	1.00	0.87	0.83	0.88	0.50	0.21	0.00	0.00	0.70		
H13	AET-1	0.06	0.00	0.07	0.07	0.07	0.33	0.63	0.93	0.93	1.00	1.00	1.00		
H19	AET-1	0.00	0.00	0.07	0.33	0.13	0.33	0.50	0.57	1.00	1.00	1.00	1.00		
18	AET-1	0.00	0.07	0.07	0.20	0.20	0.33	0.38	0.64	0.86	0.94	0.88	0.60		
40	CHR-1	0.94	0.71	0.93	0.87	1.00	0.50	1.00	0.36	0.50	0.06	0.00	0.80		