Evolution in the genus *Arum*: a comparative analysis of morphological and genetic variation

Robert Porteous

A thesis submitted in partial fulfilment of the requirements of Bournemouth University for the degree of Doctor of Philosophy

December 2005

Bournemouth University and the Centre for Ecology and Hydrology, Dorset

PAGE NUMBERING AS ORIGINAL

Abstract

Testing the correlation of morphological and genetic marker variation enables the investigation of evolutionary processes. Knowledge of evolutionary processes can be used to identify those morphological characters that could be used to produce evolutionary meaningful taxonomies. This thesis aims to test the correlation between morphological and genetic marker variation to further understand the evolution of species within the genus *Arum* and identify those morphological characters that correspond with evolutionary groups. The investigation is carried out at the intraspecific level, intrageneric level and in a putative hybrid zone.

At the intraspecific level, genetic (ISSR) and morphological variation was quantified in populations of the morphologically similar species *A. maculatum* and *A. italicum*. Populations of *A. maculatum* showed evidence of isolation by distance, presumably a result of pollinator behaviour and seed dispersal. Leaf patterning in *A. maculatum* did not correspond to evolutionary lineages. However, similar leaf patterning characters in *A. italicum* are used to classify the two subspecies *neglectum* and *italicum* and the ISSR analysis confirmed that these taxa are genetically distinct. These two subspecies were shown to be interbreeding in sympatric populations. The interbreeding has created a morphological and genetic difference between subsp. *neglectum* in sympatric populations compared with allopatric populations.

At the intrageneric level, a phylogenetic analysis of *Arum* (using *trnL* and ITS1 sequences) indicated that both vegetative and reproductive characters are convergent within the genus. The apparent convergent evolution of reproductive and vegetative characters indicates that both have been important during the diversification of the genus. These convergent characters are not useful for producing classifications that reflect evolutionary groups as the groups they produce are polyphyletic.

In the putative hybrid zone, ISSR markers confirmed the presence of *A. creticum* and *A. idaeum* hybrids. There appears to be introgression of the *A. idaeum* genome into *A. creticum*; this could have implications for the future genetic integrity of *A. creticum*. Within this hybrid zone, continuous characters were found to be representative of genetic variation, however categorical characters were not.

In conclusion, this thesis has shown that even within a single genus, the correlation between morphological and genetic marker variation is influenced by both the taxa being studied and the nature of the morphological trait. In particular, if morphological characters are found to be adaptively important, their correspondence to genetic groups should be tested before their use in taxonomies. The findings of this thesis also suggest there is great value in the complementary use of genetic and morphological analysis for taxonomic studies as well as evolutionary studies. For example, the importance of reproductive characters in the diversification of *Arum* species has produced a wide range of morphological variation, with limited taxonomic utility due to a tendency for homoplasy. Vegetative characters were also found to need careful testing before use in taxonomies as leaf patterning was found to correspond to sub-species status for one species of *Arum* but not another. Finally, this thesis has shown that, if closely related taxa are hybridising, variation of continuous reproductive characters may be used as an indicator of hybridisation, even if the morphological characters are potentially polygenic.

Table of Contents

List of Tables

List of Figures

Acknowledgements

This thesis is the result not only of my 'dedication' but also through the involvement and support of numerous other scientists and friends. I am indebted to the continued support of Dr Diaz, Dr Bond and Dr Raybould without which the PhD would not have been possible. Work regarding the phylogeny of *Arum* species was carried out in the Jodrell Laboratory, Kew with the kind permission and help from Professor Mark Chase. Dr Peter Boyce provided much needed knowledge of all things *Arum* and an excellent guided tour of Kew. Surveying a thousand individuals across Europe was enabled through the use of a semi-automated DNA extraction method at Exeter University in the Laboratory Dr Anne Smithson. Staff at the Centre of Ecology and Hydrology are all sincerely thanked, without whom the PhD would have been a joyless task. The members of the Molecular Ecology Group: E. Mitchell, R. Mogg, R Welters and A. Jones provided much needed technical advice and humour during the dark days of failed PCRs and poor gels. Mike Gardner deserves special thanks for his unquestioning help in the laboratory, data analysis and many dry runs of talks. I am indebted to Ralph Clarke for initial help with the required statistics and Dr Jeremy Thomas for enabling me to stay at CEH for as long as I did. My mum and Ruth had the arduous task of proofreading my thesis and are thanked wholeheartedly. Thanks go to AP, MM, HV, MB, JS, DH, and Meg for providing Marmite and excellent sailing lessons.

Chapter 1 Correlation of morphological and genetic characters: an introduction

'Whether he realises it or not, every biologist – even he who works on the molecular level – works with species or parts of species and his findings may be influenced decisively by the choice of a particular species' - E. Mayr (1957)

The taxonomic category of species is used at the forefront of biological research. It is at the species level that comparative biology operates, through understanding interactions among species (e.g. hybridisation: Cronberg and Natcheva, 2002), to evolutionary theory (e.g. testing species origins: Manos *et al.*, 1999) and conservation biology (e.g. population translocation: Knapp and Rice, 1998). To be able to understand processes such as these, it is important to understand what species are (Sites and Marshall, 2003). Species are expected to consist of single lineages of evolving populations (Simpson, 1961; Wiley, 1978, 1981; Avise and Wollenberg, 1997; Hey *et al.*, 2003; Sites and Marshall, 2003; for a review of species concepts see Mayden, 1997). If species do not correspond to evolutionary groups, they are misleading by suggesting relationships among organisms that have no evolutionary support (Hennig, 1966).

Species are commonly defined by morphological characters (e.g. Boyce, 1989; Stace, 1997). However, there are problems associated with comparing morphological characters among species. Firstly, variation of morphological characters can be non-heritable as the expression of morphological characters can be affected by the

environment (plasticity); for example, the morphology of both *Amblystegium* (Vanderpoorten and Jacquemart, 2004) and *Impatiens* (Donohue *et al.*, 2000) is plastic and varies depending on the environmental conditions. A second problem may arise through the convergence of morphological characters. Convergent evolution is when characters independently evolve on different lineages. Convergent characters can therefore cause relationships among species to be interpreted incorrectly. Using genetic characters (such as nucleotide sequences) to infer evolutionary relationships avoids the problems of non-heritable variation and convergence associated with morphological characters (Hillis, 1987).

Genetic characters have been used to infer evolutionary relationships at all taxonomic levels, e.g. among angiosperms (Chase *et al.*, 1993; Savolainen *et al.*, 2000), within families, e.g. Cunoniaceae (Bradford and Barnes, 2001), or within genera, e.g. *Begonia* (Plana, 2003). These examples used genetic markers to identify evolutionary relationships, so that taxonomic descriptions could be adjusted to correspond to evolutionary patterns.

When evolutionary relationships are known, a comparative analysis of morphological and genetic variation can identify the morphological characters that correspond to differences in evolutionary history. Convergence has been identified in the fruit and flower morphology of species in the family Cunoniaceae (Bradford and Barnes, 2001) and leaf morphology in the Fontinalaceae (Shaw and Allen, 2000), indicating that these characters should not be used taxonomically. Correlation of both genetic and morphological variation at the population level within a species' range can also be

informative. It is possible to infer migration patterns such as the postglacial migration routes of the genus *Dryas* (Philipp and Siegismund, 2003) and adaptation of morphological characters to different environments, such as in *Nassella pulchra* (Knapp and Rice, 1998), or selection within hybrid zones, for example in a *Polystichum* (Dryopteridaceae) hybrid zone (Ketner and Mesler, 2000).

Comparisons of morphological and genetic variation can be carried out at any taxonomic level but the age of any points of divergence are likely to differ with the taxonomic level. Intraspecific variation can be attributed to evolutionary and population processes that have occurred recently or are currently occurring. However, when interspecific comparisons are made among species, the differences attributed to each point of divergence are older (possibly ancient). Divergence of lineages can be attributed to barriers to reproduction reducing gene flow and allowing populations to change. Often, barriers to gene flow can break down (or have not yet fully evolved) and allow interspecific hybridisation between diverged lineages (Stace, 1989). Areas where hybridisation is occurring are called hybrid zones, and like intraspecific variation, evolutionary processes in these zones have occurred recently or are currently occurring. In hybrid zones, comparisons of morphological and genetic variation can also be used to identify recent patterns of evolution.

Published work looking at both morphological and genetic variation usually focuses on only one taxonomic level, such as variation at the population level typified by the population genetic study of *Oryza malampuzhaensis* in an ecological gradient (Thomas *et al.*, 2001) or diversity of different *Acorus gramineus* populations (Liao and Hsiao,

1998). Phylogenetic studies of genera also usually focus on relationships at the species level, for example *Bupleurum* (Neves and Watson, 2004) or *Begonia* (Plana, 2003), and usually do not include analyses at the population level. When studying hybrid zones phylogenetic relationships between hybridising species are often not included and only genetic relationships among hybrids compared with parental species are examined (e.g. Klier *et al.*, 1991; Ketner and Mesler, 2000; Gonzalez-Rodriguez *et al.*, 2004).

This thesis investigates the correlation of morphology and genotype (revealed using mocecular markers) at several taxonomic levels within a single genus. By investigating the relationship between morphological and genetic marker variation within a single genus inferences gained at the interspecific level can be cast in the broader context of past evolution of the genus. Similarly, variation within interspecific hybrid zones can also be placed within the context of evolution of the genus as a whole. The three areas, intraspecific variation, interspecific variation and interspecific hybridisation, offer the opportunity to study the use of morphological characters in different situations. Investigating which morphological characters correspond to current and recent patterns of evolution compared with old evolutionary differences allows us to test the usefulness of morphological characters for representing ancient and more recent evolutionary differences. The following sections investigate correlation of morphological and genetic characters, for the taxonomic purpose of interpreting evolutionary differences at three taxonomic levels.

Intraspecific variation

Species tend not be homogeneous but comprise an assemblage of populations that may vary in both morphological and genetic characters (Linhart and Grant, 1996). The variation among conspecific populations can be due to random processes such as genetic drift or mutation (Davis and Gilmartin, 1985). Intraspecific variation can also arise through adaptation of populations to localised conditions caused by environmental heterogeneity throughout a species' range (Linhart and Grant, 1996). When populations throughout a species' range experience different selection pressures, the genotype may change in response to these different evolutionary pressures (Wright, 1951). These processes in turn can create morphological and genetic differences that correspond to ecological or geographical races that can be recognised as subspecies (Grant, 1981). Different morphological groups are often recognised taxonomically: for example, 20% of the species in *Arum* L. (Araceae) have subspecific taxa delimited by morphological characters (Boyce, 1993).

Subspecific taxa, like species, should correspond to a single evolutionary lineage (e.g. Brunell and Whitkus, 1997), otherwise subspecific taxa are misleading by suggesting relationships that are not supported by evolutionary history. Comparative analysis of morphological and molecular marker variation in subspecific taxa can indicate if taxa are genetically supported. In *Elymus glaucus* three subspecies are identified morphologically, but genetic data support only two subspecies (Wilson *et al.*, 2001). Genetic data have also shown that two putative species of *Antirrhinum* are not genetically distinct and should be considered as two subspecific taxa of a single species

(Mateu-Andres and Segarra-Moragues, 2003). These examples show that once genetic lineages are known, morphological groups can be identified as those that have single evolutionary origins (monophyletic). Therefore by ensuring taxa correspond with a single genetic lineage, they conform with the widely established evolutionary species concept defined by Simpson (1961) and Wiely (1978, 1981).

Once descriptions of subspecies correspond to monophyletic lineages, comparative analyses of morphological and genetic marker variation can test hypotheses of how the interspecific variation has arisen. For example, morphological variation in extant populations can often be attributed to adaptation. The morphological and molecular marker variation of *Eriastrum densifolium* subspecies can be attributed to local adaptation to elevation and moisture levels in geographically isolated populations (Brunell and Whitkus, 1997). Similarly, in *Oryza malampuzhaensis* variation of molecular markers was correlated with morphological variation and elevation (Thomas *et al.*, 2001). These studies indicate that morphological change may be of adaptive significance and by comparing geographical (environmental), morphological and molecular marker variation, evolutionary patterns producing extant diversity can be understood.

In situations where genetic and morphological variation do not correlate, the cause of the variation may be that the morphological characters are plastic. Morphological variation was not correlated with molecular marker variation in *Carex pachystachya* (Whitkus, 1992) nor was it in *Gaillardia pulchella* (Heywood and Levin, 1984). Leaf variation and genotype were uncorrelated in *Myrceugenia fernandeziana* (Jensen *et al.*,

2002); additionally the geographical position of populations did not affect either morphology or genotype. The species *Begonia dregei* and *B. homonyma*, although morphologically distinct, are not genotypically different, questioning their species status (Matolweni *et al.*, 2000). These examples indicate that correlation of morphological and genetic characters is not always apparent. A non-correlation of morphological and genetic characters indicates that the morphological groups should not be recognised because they do not correspond to evolutionary lineages.

Interspecific variation

Genetic characters can also be used to estimate the phylogenetic relationships among taxa. Estimating evolutionary relationships has been carried out at many different taxonomic levels such as within genera, e.g. *Prunus* (Bortiri *et al.*, 2001), within families, e.g. Cunoniaceae (Bradford and Barnes, 2001), or among all major plant groups (Bowe *et al.*, 2000). Phylogenies can be used to identify if the morphological characters used to classify species correspond with evolutionary lineages and therefore to identify if species groups reflect real evolutionary differences. Phylogenetic information has been used to modify the taxonomies of the Cunoniaceae (Bradford and Barnes, 2001) and the genus *Prunus* (Bortiri *et al.*, 2001). These taxonomies were both changed so that taxa corresponded to monophyletic groups.

Molecular data can be used to infer processes of species divergence; for example, using molecular phylogenetic data it is thought that there have been two independent radiations of *Corynocarpus* (Wagstaff and Dawson, 2000). Genetic data have also been

used to confirm an Old World origin of the genus *Prunus* (Bortiri *et al.*, 2001). Comparisons of morphological characters with patterns of molecular marker variation have been carried out in many genera. In *Illicium* seed and leaf characters could distinguish species groups, but floral characters were homoplasious (Oh *et al.*, 2003). In the genus *Viola*, pansies (section *Melanium*) are monophyletic and distinct from *Viola*, a division corresponding with lateral petal orientation and pollen size (Yockteng *et al.*, 2003). Each lineage in *Corynocarrpus* is represented by a distinct fruit morphology (Wagstaff and Dawson, 2000), indicating that fruit morphology reflects evolutionary differences. By comparing morphology and evolutionary patterns in *Fontalis*, leaf characters were shown to be convergent (Shaw and Allen, 2000) and therefore cannot be used reliably to distinguish among species. Inflorescence morphology in the family Cunoniaceae was also shown to be convergent (Bradford and Barnes, 2001). Comparative analysis of morphological and genetic patterns identifies morphological characters, such as petal orientation in pansies (Yockteng *et al.*, 2003), that represent evolutionary differences. Such analyses can also identify characters that do not represent evolutionary differences among species, such as leaf morphology in *Fontalis* (Shaw and Allen, 2000).

The previous sections have indicated how a comparative analysis of genetic and morphological characters can advance our understanding of the evolution of variation within and among discrete independent evolutionary lineages, i.e. species. Species remain as independent units through limited gene flow; however, if barriers to gene flow break down, breeding between two species may occur, allowing hybridisation. Hybridisation is an important evolutionary process as it can lead to the formation of

new species such as in *Penstemon clevelandii* (Wolfe *et al.*, 1998). There is evidence to suggest that many angiosperm lineages have undergone hybridisation at some point (Ellstrand *et al.*, 1996).

Interspecific hybridisation

Morphological intermediacy is considered to be a primary diagnostic indicator of hybridisation (Gottlieb, 1972); an example of the use of morphological intermediacy is the analysis of the *Salvia apiana x S. mellifera* hybrid zone by Meyn and Emboden (1987). However, morphological intermediacy between taxa may not always indicate hybridisation. Morphological intermediacy can arise from processes other than hybridisation (Rieseberg, 1997). Similar selection pressures can cause morphological characters to converge (Rieseberg *et al.*, 1990, 1997), giving the appearance of morphological intermediacy. Morphologically intermediate characters could also be retained ancient characters that where present before the divergence of two lineages (Rieseberg, 1997). The use of morphological characters in hybridization studies is further complicated as morphological characters may be the products of many different genes (e.g. reviewed by Buzgo *et al.*, 2004; Gottlieb, 1984). Characters under multi-gene control may have complex and unpredictable expression patterns causing hybrids not to be morphologically intermediate (Rieseberg and Ellstrand, 1993).

Genetic markers can be used to identify hybrid taxa without the problems associated with morphological characters. A combination of morphological and genetic characters in *Sphagnum* species indicated hybridisation (Cronberg and Natcheva, 2002). A similar

approach was used to confirm introgression of characters from *Helianthus annus* to *H. debilis* (Rieseberg *et al.*, 1990). Comparative analysis of genetic and morphological characters can be used to asses the usefulness of morphological characters for diagnosing hybrids and can reveal biological processes such as selection and hybrid fitness occurring within hybrid zones (e.g. Ketner and Mesler, 2000).

Aims

Morphological characters are commonly used to define relationships between species and within species. If morphological characters do not correspond with evolutionary relationships the taxonomies will be misleading. The aim of the thesis is therefore to test the correlation between morphological and molecular marker variation. This test will achieve two goals:

1. Correlations between morphology and genetics can test which morphological characters are the most informative in creating a taxonomy that reflects evolutionary relationships.
2. The correlations between genetic and morphological variation will provide information on past species divergence and current directions of evolutionary change.

To achieve the aim at three taxonomic levels, a single genus with intraspecific variation, a recognised morphological taxonomy and putative hybrid zones is required. The genus *Arum* L. is ideal as it fits these three criteria. Of the *Arum* species, 20% have subspecific taxa, recognised using morphological characters (Boyce, 1993). *Arum* is also a small

genus with only 25 recognised species (Boyce, 1993), and a phylogenetic analysis is therefore a realistic aim in a PhD timescale. Also, the group is taxonomically complex, with subgenera, sections and subsections based on a mixture of reproductive and vegetative characters (Boyce, 1993). Although several species within the genus are thought to hybridise (Boyce, 1993), hybrid zones have yet to be proven.

References

Avise JC, Wollenberg K (1997) Phylogenetics and the origin of species. *Proceedings of the National Academy of Sciences USA* **94**, 7748–7755.

Bortiri E, Oh SH, Jiang JG, Baggett S, Granger A, Weeks C, Buckingham M, Potter D, Parfitt DE (2001) Phylogeny and systematics of *Prunus* (Rosaceae) as determined by sequence analysis of ITS and the chloroplast *trnL–trnF* spacer DNA. *Systematic Botany* **26**, 797–807.

Bowe LM, Coat G, dePamphilis CW (2000) Phylogeny of seedplants based on all three genomic compartments: Exant gymnosperms are monophylectic and Gnetales' closest relatives are conifers. *Proceedings of the National Accadamy of Sciences USA* **97**, 4092–4097.

Boyce P (1989) A new classification of *Arum* with keys to the infrageneric taxa. *Kew Bulletin* **44**, 383–395.

Boyce P (1993) *The genus Arum. A Kew Magazine Monograph*. HMSO, London.

Bradford JC, Barnes RW (2001) Phylogenetics and classification of Cunoniaceae (Oxalidales) using chloroplast DNA sequences and morphology. *Systematic Botany* **26**, 354–385.

Brunell MS, Whitkus R (1997) RAPD marker variation in *Eriastrum densifolium* (Polemoniaceae): Implications for subspecific delimitation and conservation. *Systematic Botany* **22**, 543–553.

Buzgo M, Soltis DE, Soltis PS, Ma H (2004) Towards a comprehensive integration of morphological and genetic studies of floral development. *Trends in Plant Science* **9**, 164–173.

Chase MW, Soltis DE, Olmstead RG, Morgan D, Les DH, Mishler BD, Duvall MR, Price RA, Hills HG, Qiu YL, Kron KA, Rettig JH, Conti E, Palmer JD, Manhart JR, Sytsma KJ, Michaels HJ, Kress WJ, Karol KG, Clark WD, Hedren M, Gaut BS, Jansen RK, Kim KJ, Wimpee CF, Smith JF, Furnier GR, Strauss SH, Xiang QY, Plunkett GM, Soltis PS, Swensen SM, Williams SE, Gadek PA, Quinn CJ, Eguiarte LE, Golenberg E, Learn GH, Graham SW, Barrett SCH, Dayanandan S, Albert VA (1993) Phylogenetics of seed plants - an analysis of nucleotide

sequences from the plastid gene Rbcl. *Annals of the Missouri Botanical Garden* **80**, 528–580.

Cronberg N, Natcheva R (2002) Hybridization between the peat mosses, *Sphagnum capillifolium* and *S. quinquefarium* (Sphagnaceae, Bryophyta) as inferred by morphological characters and isozyme markers. *Plant Systematics and Evolution* **234**, 53–70.

Davis JI, Gilmartin AJ (1985) Morphological variation and speciation. *Systematic Botany* **10**, 417–425.

Donohue K, Messiqua D, Pyle EH, Heschel MS, Schmitt J (2000) Evidence of adaptive divergence in plasticity: Density- and site-dependent selection on shade-avoidance responses in *Impatiens capensis*. *Evolution* **54**, 1956–1968.

Ellstrand NC, Whitkus R, Rieseberg LH (1996) Distribution of spontaneous plant hybrids. *Proceedings of the National Academy of Sciences USA* **93**, 5090–5093.

Gonzalez-Rodriguez A, Arias DM, Valencia S, Oyama A (2004) Morphological and RAPD analysis of hybridization between *Quercus affinis* and *Q. laurina* (Fagaceae), two Mexican red oaks. *American Journal of Botany* **91**, 401–409.

Gottlieb LD (1972) Levels of confidence in the analysis of hybridisation in plants. *Annals of the Missouri Botanical Garden* **59**, 435–446.

Gottlieb LD (1984) Genetics and morphological evolution in plants. *The American Naturalist* **123**, 681–709.

Grant V (1981) *Plant Speciation*, 2nd edn. Columbia University Press, New York.

Hennig W (1966) *Phylogenetic Systematics*. University of Illinois Press, Urbana.

Hey J, Waples RS, Arnold ML, Butlin RK, Harrison RG (2003) Understanding and confronting species uncertainty in biology and conservation. *Trends in Ecology & Evolution* **18**, 597–603.

Heywood JS, Levin DA (1984) Allozyme variation in *Gaillardia pulchella* and *Gaillardia amblyodon* (Compositae) - relation to morphological and chromosomal variation and to geographical isolation. *Systematic Botany* **9**, 448–457.

Hillis DM (1987) Molecular versus morphological approaches to systematics. *Annual Review of Ecology and Systematics* **18**, 23–42.

Jensen RJ, Schwoyer M, Crawford DJ, Stuessy TF, Anderson GJ, Baeza CM, O MS, Ruiz E (2002) Patterns of morphological and genetic variation among populations of *Myrceugenia fernandeziana* (Myrtaceae) on Masatierra island: Implications for conservation. *Systematic Botany* **27**, 534–547.

Ketner EK, Mesler MR (2000) Evidence for natural selection in a fern hybrid zone. *American Journal of Botany* **87**, 1168–1174.

Klier K, Leoschke MJ, Wendel JF (1991) Hybridization and introgression in white and yellow ladyslipper orchids (*Cypripedium candidum* and *C. pubescens*). *Journal of Heredity* **82**, 305–318.

Knapp EE, Rice KJ (1998) Comparison of isozymes and quantitative traits for evaluating patterns of genetic variation in purple needlegrass (*Nassella pulchra*). *Conservation Biology* **12**, 1031–1041.

Liao LC, Hsiao JY (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.

Linhart YB, Grant MC (1996) Evolutionary significance of local genetic differentiation. *Annual Review of Ecology and Systematics* **27**, 237–277.

Manos PS, Doyle JJ, Nixon KC (1999) Phylogeny, biogeography and processes of molecular differentiation in *Quercus* subgenus *Quercus* (Fagaceae). *Molecular Phylogenetics and Evolution* **12**, 333–349.

Mateu-Andres I, Segarra-Moragues JG (2003) Allozymic differentiation of the *Antirrhinum graniticum* and the *Antirrhinum meonanthum* species groups. *Annals of Botany* **92**, 647–655.

Matolweni LO, Balkwill K, McLellan T (2000) Genetic diversity and gene flow in the morphologically variable, rare endemics *Begonia dregei* and *Begonia homonyma* (Begoniaceae). *American Journal of Botany* **87**, 431–439.

Mayden RL (1997) A hierarchy of species concepts: the denouncement in the saga of the species problem. In: *Species: The Units of Biodiversity* (eds Claridge MF, Dawah HA, Wilson MR). Chapman and Hall, New York.

Mayr E (1957) Species concepts and definitions. In: *The Species Problem*, pp. 1–22. American Association for the Advancement of Science no. 50, Washington, DC.

Meyn O, Emboden WA (1987) Parameters and consequences of introgression in *Salvia apiana* X *S mellifera* (Lamiaceae). *Systematic Botany* **12**, 390–399.

Neves SS, Watson MF (2004) Phylogenetic relationships in *Bupleurum* (Apiaceae) based on nuclear ribosomal DNA ITS sequence data. *Annals of Botany* **93**, 379–398.

Oh IC, Denk T, Friis EM (2003) Evolution of *Illicium* (Illiciaceae): Mapping morphological characters on the molecular tree. *Plant Systematics and Evolution* **240**, 175–209.

Philipp M, Siegismund HR (2003) What can morphology and isozymes tell us about the history of the *Dryas integrifolia-octopetala* complex? *Molecular Ecology* **12**, 2231–2242.

Plana V (2003) Phylogenetic relationships of the Afro-Malagasy members of the large genus *Begonia* inferred from *trnL* intron sequences. *Systematic Botany* **28**, 693–704.

Rieseberg LH (1997) Hybrid origins of plant species. *Annual Review of Ecology and Systematics* **28**, 359–389.

Rieseberg LH, Beckstromsternberg S, Doan K (1990) *Helianthus annuus* ssp *texanus* has chloroplast DNA and nuclear ribosomal-RNA genes of *Helianthus debilis* ssp. *cucumerifolius*. *Proceedings of the National Academy of Sciences USA* **87**, 593–597.

Rieseberg LH, Ellstrand NC (1993) What can molecular and morphological markers tell us about plant hybridisation? *Critical Reviews in Plant Science* **12**, 213–241.

Savolainen V, Chase M, Hoot SB, Morton CM, Soltis DE, Bayer C, Fay M, DeBrujin AY, Sullivan S, Qiu YL (2000) Phylogenetics of flowering plants based on combined analysis of plastid *atp*B and *rbc*L gene sequences. *Systematic Biology* **49**, 306–362.

Shaw AJ, Allen B (2000) Phylogenetic relationships, morphological incongruence, and geographic speciation in the Fontinalaceae (Bryophyta). *Molecular Phylogenetics and Evolution* **16**, 225–237.

Simpson GG (1961) *Principals of Animal Taxonomy*. Columbia University Press, New York.

Sites JW, Marshall JC (2003) Delimiting species: a Renaissance issue in systematic biology. *Trends in Ecology & Evolution* **18**, 462–470.

Stace CA (1989) *Plant Taxonomy and Biosystematics*, 2nd edn. Cambridge University Press, Cambridge.

Stace CA (1997) *New Flora of the British Isles*, 2nd edn. Cambridge University Press, Cambridge.

Thomas G, Joseph L, Varghese G, Sreejayan, Kalyanaraman K, Kuriachan P (2001) Analysis of phenotypic and genetic variations among populations of *Oryza malampuzhaensis* show evidence of altitude-dependent genetic changes. *Canadian Journal of Botany-Revue Canadienne De Botanique* **79**, 1090–1098.

Vanderpoorten A, Jacquemart AL (2004) Evolutionary mode, tempo, and phylogenetic association of continuous morphological traits in the aquatic moss genus *Amblystegium*. *Journal of Evolutionary Biology* **17**, 279–287.

Wagstaff SJ, Dawson MI (2000) Classification, origin, and patterns of diversification of *Corynocarpus* (Corynocarpaceae) inferred from DNA sequences. *Systematic Botany* **25**, 134–149.

Whitkus R (1992) Allozyme variation within the *Carex-Pachystachya* complex (Cyperaceae). *Systematic Botany* **17**, 16–24.

Wiley EO (1978) The evolutionary species concept reconsidered. *Systematic Zoology* **41**, 17–26.

Wiley EO (1981) *Phylogenetics: The Theory and Practice of Phylogenetic Systematics*.

Wilson BL, Kitzmiller J, Rolle W, Hipkins VD (2001) Isozyme variation and its environmental correlates in *Elymus glaucus* from the California Floristic Province. *Canadian Journal of Botany* **79**, 139–153.

Wolfe AD, Xiang QY, Kephart SR (1998) Assessing hybridization in natural populations of *Penstemon* (Scrophulariaceae) using hypervariable inter simple sequence repeat (ISSR) bands. *Molecular Ecology* **7**, 1107–1125.

Wright S (1951) The genetical structure of populations. *Annals of Eugenics* **15**, 323–354.

Yockteng R, Ballard HE, Mansion G, Dajoz I, Nadot S (2003) Relationships among pansies (Viola section Melanium) investigated using ITS and ISSR markers. *Plant Systematics and Evolution* **241**, 153–170.

Chapter 2 General methods

This methods chapter describes the molecular markers and data analysis techniques suitable for population-level investigations that are used in Chapters 3 and 5. No growing or crossing experiments were carried out due to the ecology of *Arum* species. Species within *Arum* have can take up to 7 years to mature and reproduce (Boyce, 1993) and therefore can not be readily used in a common garden experiment or a quantitative trait loci study. Thus, morphological variation can not be attributed with confidence to extrinsic factors or heritable differences. The comparison between morphological and molecular marker variation is a simple correlation, such data analysis methods specific to a single chapter are included within a separate methods section in their respective chapter. The methods associated with phylogenetic analyses are presented in Chapter 4.

Molecular markers

In Chapter 3, genetic relationships among population locations of *Arum maculatum* and *A. italicum* from Spain, France and Britain were surveyed to identify correlations of morphology and molecular marker variation. In Chapter 5, putative hybrid zones of *A. idaeum* and *A. creticum* were investigated, again identifying any correlation of morphological and molecular marker variation. The molecular markers used in these chapters must have a proven ability to detect population differentiation and identify hybrids.

Allozymes were the first type of molecular markers to be used for population studies in humans (Harris, 1966) and *Drosophila* (Hubby and Lewontin, 1966). Allozymes have been successfully used to study variation over species ranges such as: *Carex pachystachya* (Whitkus, 1992); *Myrceugenia fernandeziana* (Jensen *et al.*, 2002); *Pinus rzedowskii* in Mexico (Delgado *et al.*, 1999); *Begonia* (Matolweni *et al.*, 2000). Hybrid zones of many species have also been investigated using allozymes: e.g. *Quercus* (Gomory *et al.*, 2001); *Cypripedium* (Klier *et al.*, 1991); *Iris nelsonii* (Arnold *et al.*, 1990) and *Sphagnum* (Cronberg and Natcheva, 2002). Although allozymes are useful at the population level they will not be used in the survey of *Arum* species for several reasons. Allozyme variation is caused by amino acid substitutions, which then alter the proteins' electrophoretic mobility. This has led to criticism because allozyme variation is an indirect method of assessing DNA variation (Schlötterer, 2004). Additionally, variation at allozyme loci can often be insufficient to distinguish between populations (Heywood and Levin, 1984). Lastly, the suitability of allozymes is limited because of the large distances involved in transporting samples from the field to the laboratory and any allozyme study requires fresh material.

Molecular markers screened using the polymerase chain reaction (PCR) detect higher levels of variation than allozymes and allow clearer resolution of genetic differences (Cruzan, 1998). This is because some PCR markers can amplify sites throughout the genome, allowing many loci to be compared simultaneously (Bachmann, 1994). Methods using the PCR will therefore be used to investigate genetic relationships in Chapters 3 and 5.

There are, however, a large number of PCR-based techniques available to choose from. The range of PCR-based methods can be broken down into two broad categories, those in which prior knowledge of the nucleotide sequence is necessary and those in which no sequence information is needed. The most common technique in the first category is the use of microsatellite markers, which are short tandemly repeated motifs (e.g. AACAACAAC) found throughout the genome (Gupta *et al.*, 1994). Variation at a microsatellite locus is generated by the addition or deletion of repeat units, although larger changes sometimes occur. To survey microsatellite variation, knowledge of the nucleotide sequences flanking either side of the microsatellite is required so that primers can be designed. Unless microsatellite markers are already isolated, the process of isolating microsatellites is relatively expensive and time-consuming (Wolfe and Liston, 1998). However, microsatellite markers are highly variable and therefore informative, and have been used successfully to investigate the correlation of morphological and molecular marker variation in *Quercus petraea* (Bruschi *et al.*, 2003), and diversification in *Eucalyptus* (Holman *et al.*, 2003). Even though microsatellites can be applied at the population level, they cannot be applied to the study of *Arum* populations because there are no published microsatellite sequences.

Sequencing nucleotides is another method that requires prior sequence knowledge to amplify the chosen section of the genome. Although sequencing is very expensive, it can be applied on any taxonomic level. Sequence variation has been applied to study populations over large geographical areas, such as for oak (Ferris *et al.*, 1993; Petit *et al.*, 1997; Samuel *et al.*, 1998) and holly (Rendell and Ennos, 2003). Nucleotide sequencing is therefore suitable for studying variation within and between populations

as in Chapter 3. There are also universal primers avalible (e.g. Taberlet *et al.*, 1991) that can be used to survey DNA variation in any species. However, because DNA sequencing of over 1500 individuals is required in Chapters 3 and 5 it was therefore prohibitively expensive to use it for the investigations described these chapters.

Since there is little sequence information available for *Arum* species the second type of marker, requiring no prior sequence information, is applicable. Molecular markers that require no sequence information include RAPD (randomly amplified polymorphic DNA), AFLP (amplified fragment length polymorphisms) and ISSR (inter simple sequence repeats). The RAPD technique uses a random oligonucleotide to amplify many different sites throughout the genome (Welsh and McClelland, 1990). The RAPD technique has been successfully used to investigate morphological and genetic marker variation at the species level, for example: investigating the correlation of molecular marker and morphological variation in *Oryza malampuzhaensis* (Thomas *et al.*, 2001); identifying high marker variation in morphologically similar *Acorus gramineus* populations (Liao and Hsiao, 1998); and finding no genetic support for subspecies of *Eriastrum* (Brunell and Whitkus, 1997). However, use of the RAPD technique has declined owing to inconsistent results (Schlötterer, 2004), which can arise from small variations in the PCR reaction mix (Ellsworth *et al.*, 1993). Alternative methods that are more reliable and variable than the RAPD technique are AFLP and ISSR markers (Wolfe and Liston, 1998). AFLP markers are produced from amplified restriction digests and have been used in population-level studies such as of *Ranunculus* (Schonswetter *et al.*, 2003), *Quercus* (Ishida *et al.*, 2003) and hybridisation in *Onopordum* (O'Hanlon *et al.*, 1999). The AFLP technique is technically demanding

(Nybom, 2004), and because of this complexity errors may be induced; therefore the less technically demanding ISSR technique will be used. ISSR markers amplify parts of the genome between adjacent microsatellites (Nagaoka and Ogihara, 1997; Wolfe *et al.*, 1998a, b; Zietkiewicz *et al.*, 1994). The primers contain a microsatellite motif with 6 repeat units and an anchor sequence usually 2–3 bases long. ISSR markers have been used to investigate species diversity over large geographical areas, as in *Cakile maritima* and *Eryngium maritima* (Clausing *et al.*, 2000), hybridisation in *Penstemon* (Wolfe *et al.*, 1998a, b) and introgression in *Fallopia* (Hollingsworth *et al.*, 1998).

Available laboratory equipment is a further factor influencing the choice of markers for Chapters 3 and 5. The laboratory at the Centre for Ecology and Hydrology, Dorset (CEH) did not, during the time allocated for Chapter 3 and 5 experimental work have the capacity for a large scale sequencing project. *Arum* species have no published microsatellites thus time-consuming and consequently expensive cloning would have to be carried out to use a microsatellite marker which was unfeasible at CEH in 2000-01 as a cloning licence had also not been granted. Therefore molecular markers that require no prior nucleotide sequence knowledge will be used. Of the techniques available, the CEH laboratory has both equipment and expertise for conducting experiments using both ISSR and AFLP protocols. After witnessing difficulties surrounding the complex nature of the AFLP protocol at the ligation stage and discussing the AFLP protocol with experienced Post-Doctoral researchers the ISSR markers were chosen as the best suited markers to this *Arum* study.

Practical Methods

DNA amplification

A subset of ISSR primers from the University of British Columbia, Microsatellite set #9, containing the primers 891, 888, 886, 848, 834, 822, 811 and 857 (Table 2.1) were tested for suitability. Optimisation of each primer was carried out by varying the amount of template DNA in each reaction and by altering the stringency of the PCR. PCR stringency was changed by varying the concentration of magnesium, the annealing temperature and number of cycles. Successfully amplified primers were selected on the basis of band clarity and amplification strength

Table 2.1 ISSR primer sequences

Primer	Repeat
811	5′gag aga gag aga gag ac-3′
822	5′-tct ctc tct ctc tct ca-3′
834	5′-aga gag aga gag aga g(ct)t-3′
848	5′-cac aca cac aca cac a(ag)-3′
857	5′-aca cac aca cac aca c(ct) g-3′
886	5′-(agc) (agt) (agc) ctc tct ctc tct ct-3′
888	5′-(gct) (agt) (gct) cac aca cac aca ca-3′
891	5′-(act) (agc) (act) tgt gtg tgt gtg tg-3′

Primers 811 and 834 were found to work reliably in *A. italicum*, *A. maculatum*, *A. creticum* and *A. idaeum* and were therefore chosen for screening all samples. The primers 811 and 834 are both based on AG repeats, but have different anchor sequences: the anchor sequence of primer 811 is AC whereas in primer 834 it is CT or TT. These two primers will both amplify DNA between AG microsatellites; however, the difference in the anchor sequence means that different sections of DNA will be

targeted and amplified. For example, a microsatellite ending in CT will be amplified by primer 834 but will not be amplified by primer 811. The two primers therefore amplify different regions of the genome and do not duplicate any information.

For both primers 811 and 834, the PCR reaction volumes were 10µl which consisted of 1x NH_4 reaction buffer (Bioline), 2mM $MgCl_2$, 2mM dNTPs, 0.5mM primer and 0.2U *Taq* DNA polymerase (Biotaq, Bioline). Amplifications were carried out using an MJ Research Tetrad Thermocycler, programmed as follows: denaturation at 94°C for 1 minute, annealing at 60°C for 2 minutes and extension at 72°C for 40 seconds, for two cycles; the annealing temperature was decreased by 2°C every two cycles until 56°C; at 56°C 22 cycles were carried out followed by a final 5-minute extension step at 72°C. Each reaction was performed twice to ensure that the amplified products were repeatable. ISSR primers are universal, and therefore eight reactions with sterile distilled water replacing the same volume of DNA (DNA-negative) were included throughout each 96-well PCR plate (every 10 wells). The DNA-negative reactions ensure any cross contamination can be identified and eliminated.

Visualisation of ISSR DNA fragments: polyacrylamide gel electrophoresis and silver staining

Polyacrylamide gel electrophoresis (PAGE) was carried out on 6% denaturing polyacrylamide gels, and 1.5µl of amplified PCR product was loaded onto pre-run gels at 50°C. The gels were then run at 50°C for 2.5 hours. The polyacrylamide gels contained urea and maintaining the gel at a minimum temperature of 50°C ensured that

DNA remains denatured. After electrophoresis DNA fragments were visualised using silver staining kits (Promega). All silver staining was carried out in large photographic trays within a fume hood. Water quality is critical for successful silver staining. Poor quality water can lead to dark and unscorable gels, consequently all silver staining solutions were made up with purified water at 18MΩ (USF-Elga Option 7 Water Purifier). Post electrophoresis, the gel was immersed in fix/stop solution for at least 30 minutes. The fix/stop solution was washed off by immersing the gel in two litres of deionised water for 10 minutes. This wash step was repeated three times. Immersing the gel in staining solution for 30 minutes stained the DNA fragments, which were then visualised after immersion in ice-cold developer solution. Development was stopped when sufficient bands were visualised. Long development times increase background staining and reduce the clarity of the gel. Ice-cold developer solution slows the speed at which DNA fragments become visible, allowing over development to be avoided. Immersing the gel for 1 minute in the fix/stop solution stopped the development reaction. The fix/stop solution was rinsed off with two litres of distilled water for 15 minutes; the gels were then dried overnight at 37°C. The silver staining method is highly sensitive to DNA concentration; smeared bands indicated too much DNA had been loaded, therefore by loading less PCR product clear bands were produced.

ISSR band scoring

Stained, dried gels were scanned and scored using Adobe Photoshop V7 and Cross Checker (Buntjer, 1999). Bands at the top of the gel were often poorly separated, while bands at the bottom of the gel were often faint, and in both cases these bands were

excluded from the analysis. Bands were only scored if they were present in duplicate PCR reactions. If the duplicate PCR reactions were inconsistent, the samples were excluded from analysis. If the DNA-negative PCRs produced amplified DNA fragments the PCR was performed again with fresh reagents. Results were entered onto a binary matrix, with '1' indicating presence and '0' absence of a band. ISSRs are scored as dominant markers similar to RAPDs (Wolfe *et al.*, 1998a, b). ISSR markers are dominant because a heterozygote and a homozygote cannot be distinguished.

Statistical analysis of ISSR bands

ISSR analysis cannot discriminate between a homozygous individual (AA) and a heterozygous individual (Aa) as both of these individuals would appear to have the same band. Absence of a band can result from either a mutation of primer binding site, chromosomal rearrangement or a change in the number of repeat units (Wolfe and Liston, 1998). The inability to identify why a band has gone missing means that absence of a band cannot be used to confer any information. Only shared presence of identically sized bands was used to calculate genetic distances between each individual. This inability to distinguish between homo- and heterozygotes precludes the use of Fstatistics as these require the studied populations to be in Hardy-Weinberg equilibrium. There is however an alternative, this is Φstatistics. Φstatistics can be used as an analogous test to Fstatistics however to use them the populations are assumed to be in Hardy-Weinberg equilibrium. Hardy-Weinberg equilibrium can be disturbed if individuals within populations inbreed, if there is assortative mating, or natural selection (Nei and Kumar 2000). This assumption of Hardy-Weinberg equilibrium in *Arum* popualtaions is

unrealistic as they are tuberous herbs that form clolonies by splitting tubers (Boyce 1993). Therefore, *Arum* populations may undergo some inbreeding also little is known regarding selection within the populations.

The specific aims of the theisis and each chapter require simple correlation analyses, adding further statistical tests like Fstats (or analogies) and analysis of molecular varience (ANOVA) although may provide interesting information pertaining to diversity partitioning whould not directly answer the chapter or theis aims. Thus, the genetic distance measure by Nei and Li (1979) for restriction digest fragments was used to calculate the genetic distance among samples from the ISSR data. The Nei and Li genetic distance is simply the number of shared DNA fragments between two individuals as a proportion of the total number of fragments in both tested individuals (Equation 2.1).

Equation 2.1 The proportion of shared DNA fragments $\hat{F}$, where n_x and n_y are the numbers of fragments in sample *x* and *y* respectively, and n_{xy} is the number of fragments shared by both sample *x* and *y*. Equation 2.1 from Nei and Li (1979).

$$\hat{F} = \frac{2n_{xy}}{\left(n_x + n_y\right)}$$

The Nei and Li (1979) genetic distance estimator was used in the analyses of *A. italicum* and *A. maculatum* data in Chapter 3 and the analysis of *A. idaeum* and *A. creticum* data in Chapter 5. For these chapters the Nei and Li (1979) genetic distance measure was estimated using two different computer programs. For the assessment of intraspecific differences among populations (Chapter 3) the computer program ARLEQUIN (Schneider *et al.*, 2000) was used to calculate the average genetic distance among

populations. To investigate the putative hybrid zone (Chapter 5) the program PAUP* 4.0b10 (Swofford, 1998) was used to calculate pairwise genetic distances between each individual. More detailed descriptions of the subsequent data analysis techniques are presented within Chapters 3 and 5.

Ordination of large datasets: non-metric multidimensional scaling

In all chapters non-metric multidimensional scaling (N-MDS) is used as a way of visualising genetic and morphological relationships among many individuals or groups. Ordination of samples is a convenient method to visualise and interpret large datasets, as patterns are easier to identify through spatial clustering than on a complex, many branched, tree. The MDS analysis uses any pairwise distance measure (Clarke and Warwick, 2001), allowing comparisons of morphological and genetic relationships. The MDS algorithm in the PRIMER-5 computer program (Clarke and Warwick, 2001) was used. The MDS algorithm configures the position of each individual in either two or three dimensions constrained by its rank within the distance matrix (Clarke and Warwick, 2001). If two individuals are close in the distance matrix a two- or three-dimensional plot will show these individuals as being spatially close. For N-MDS plots of molecular marker variation the Nei and Li genetic distance is used whereas Euclidean distances are used for the continuous morphological characters.

The reliability of the MDS ordination of the data points in *n* dimensions is measured by a *stress* value (Clarke and Warwick, 2001; Kruskal, 1964). As the *stress* value

decreases, the similarity of the ordination to the original data matrix increases, thus indicating how well the *n*-dimensional plot resembles the original data (Kruskal, 1964). A *stress* value >0.2 indicates that the relationships do not match the data well, whilst values in the range 0.05–0.1 indicate that the arrangement in *n*-dimensional space is satisfactory (Kruskal, 1964).

Statistical support for *a priori* groups was calculated using an analysis of similarities test (ANOSIM). The ANOSIM test statistic, *R*, is calculated by comparing the average rank of observed differences within each group to the average rank between groups as a function of the total number of samples. The null hypothesis is that there are no differences between the *a priori* groups. The statistic, *R*, ranges from –1 to 1, with values approaching 1 indicating that the individuals within the *a priori* groups are more similar to each other than they are to individuals from any other group. Significance of the test statistic is gained through a randomisation process where sample names are changed arbitrarily. If the null hypothesis is correct then randomisation of *a priori* group labels will have no effect on *R*. The significance of the *R* statistic is calculated using the distribution of the frequency of randomised *R* values compared with the observed value of *R*.

References

Arnold ML, Hamrick JL, Bennett BD (1990) Allozyme variation in Louisiana irises - a test for introgression and hybrid speciation. *Heredity* **65**, 297–306.

Bachmann K (1994) Molecular markers in plant ecology. *New Phytologist* **126**, 403–418.

Brunell MS, Whitkus R (1997) RAPD marker variation in *Eriastrum densifolium* (Polemoniaceae): Implications for subspecific delimitation and conservation. *Systematic Botany* **22**, 543–553.

Bruschi P, Vendramin G, Bussotti F, Grossoni P (2003) Morphological and molecular diversity among Italian populations of *Quercus petraea* (Fagaceae). *Annals of Botany* **91**, 707–716.

Buntjer J (1999) Cross Checker Fingerprint analysis software v2.9. Wageningen University and Research Centre, The Netherlands.

Clarke KR, Warwick RM (2001) *Change in Marine Communities: An Approach to Statistical Analysis and Interpretation*, 2nd. PRIMER-E, Plymouth.

Clausing G, Vickers K, Kadereit JW (2000) Historical biogeography in a linear system: genetic variation of Sea Rocket (*Cakile maritima*) and sea holly (*Eryngium maritima*). *Molecular Ecology* **9**, 1823–1833.

Cronberg N, Natcheva R (2002) Hybridization between the peat mosses, *Sphagnum capillifolium* and *S. quinquefarium* (Sphagnaceae, Bryophyta) as inferred by morphological characters and isozyme markers. *Plant Systematics and Evolution* **234**, 53–70.

Cruzan MB (1998) Genetic markers in plant evolutionary ecology. *Ecology* **79**, 400–412.

Delgado P, Pinero D, Chaos A, Perez-Nasser N, Alvarez-Buylla ER (1999) High population differentiation and genetic variation in the endangered Mexican pine *Pinus rzedowskii* (Pinaceae). *American Journal of Botany* **86**, 669–676.

Ellsworth DL, Rittenhouse KD, Honeycutt RL (1993) Artifactual variation in randomly amplified polymorphic DNA banding patterns. *Biotechniques*, 214–217.

Ferris C, Oliver RP, Davy AJ, Hewitt GM (1993) Native oak chloroplasts reveal an ancient divide across Europe. *Molecular Ecology* **2**, 337–344.

Gomory D, Yakovlev I, Zhelev P, Jedinakova J, Paule L (2001) Genetic differentiation of oak populations within the *Quercus robur/Quercus patrea* complex in Central and Eastern Eurpoe. *Heredity* **86**, 557–563.

Gupta M, Chyi YS, Romeroseverson J, Owen JL (1994) Amplification of DNA markers from evolutionarily diverse genomes using single primers of simple-sequence repeats. *Theoretical and Applied Genetics* **89**, 998–1006.

Harris H (1966) Enzyme polymorphisms in man. *Proceedings of the Royal Society London, Series B, Biological Sciences* **164**, 298–310.

Heywood JS, Levin DA (1984) Allozyme variation in *Gaillardia pulchella* and *Gaillardia amblyodon* (Compositae) - Relation to morphological and chromosomal variation and to geographical isolation. *Systematic Botany* **9**, 448–457.

Hollingsworth ML, Hollingsworth PM, Jenkings GI, Bailey JP, Ferris C (1998) The use of molecular markers to study patterns of genotypic diversity in some alien *Fallopia* spp. (Polygonaceae). *Molecular Ecology* **7**, 1681–1691.

Holman JE, Hughes JM, Fensham RJ (2003) A morphological cline in *Eucalyptus*: a genetic perspective. *Molecular Ecology* **12**, 3013–3025.

Hubby JL, Lewontin RC (1966) A molecular approach to the study of genetic variation in natural populations. I. The number of alleles at different loci in *Drosophila pseudoobscura*. *Genetics* **54**, 577–594.

Ishida TA, Hattori K, Sato H, Kimura MT (2003) Differentiation and hybridization between *Quercus crispula* and *Q. dentata* (Fagaceae): insights from morphological traits, amplified fragment length polymorphism markers, and leafminer composition. *American Journal of Botany* **90**, 769–776.

Jensen RJ, Schwoyer M, Crawford DJ, Stuessy TF, Anderson GJ, Baeza CM, O MS, Ruiz E (2002) Patterns of morphological and genetic variation among populations of *Myrceugenia fernandeziana* (Myrtaceae) on Masatierra island: Implications for conservation. *Systematic Botany* **27**, 534–547.

Klier K, Leoschke MJ, Wendel JF (1991) Hybridization and introgression in white and yellow ladyslipper orchids (*Cypripedium candidum* and *C. pubescens*). *Journal of Heredity* **82**, 305–318.

Kruskal JB (1964) Nonmetric multidimensional scaling: a numerical method. *Psychometrika* **29**, 115–129.

Liao LC, Hsiao JY (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.

Matolweni LO, Balkwill K, McLellan T (2000) Genetic diversity and gene flow in the morphologically variable, rare endemics *Begonia dregei* and *Begonia homonyma* (Begoniaceae). *American Journal of Botany* **87**, 431–439.

Nagaoka T, Ogihara Y (1997) Applicability of inter-simple sequence repeat polymorphisms in wheat for use as DNA markers in comparison to RFLP and RAPD markers. *Theoretical and Applied Genetics* **94**, 597–602.

Nei M, Li H (1979) Mathematical Model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences USA* **76**, 5269–5273.

Nybom H (2004) Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Molecular Ecology* **13**, 1143–1155.

O'Hanlon PC, Peakall R, Briese DT (1999) Amplified fragment length polymorphism (AFLP) reveals introgression in weedy *Onopordum* thistles: hybridization and invasion. *Molecular Ecology* **8**, 1239–1246.

Petit RJ, Pineau E, Demesure B, Bacilieri R, Ducousso A, Kremer A (1997) Chloroplast DNA footprints of postglacial recolonization by oaks. *Proceedings of the National Academy of Sciences USA* **94**, 9996–10001.

Rendell S, Ennos RA (2003) Chloroplast DNA diversity of the dioecious European tree *Ilex aquifolium* L. (English holly). *Molecular Ecology* **12**, 2681–2688.

Samuel R, Bachmair A, Jobst J, Ehrendorfer F (1998) ITS sequences from nuclear rDNA suggest unexpected phylogenetic relationships between Euro-Mediterranean, East Asiatic and North American taxa of *Quercus* (Fagaceae).

Schlötterer C (2004) The evolution of molecular markers - just a matter of fashion? *Nature Reviews Genetics* **5**, 63–69.

Schneider S, Roessli D, Excoffier L (2000) Arlequin: A software for population genetics data analysis. Ver 2.000. Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva.

Schonswetter P, Paun O, Tribsch A, Niklfeld H (2003) Out of the Alps: colonization of Northern Europe by East Alpine populations of the Glacier Buttercup *Ranunculus glacialis* L. (Ranunculaceae). *Molecular Ecology* **12**, 3373–3381.

Swofford DL (1998) PAUP*: phylogenetic analysis using parsimony (*and other methods) version 4. Sinauer, Sunderland, Massachusetts, USA.

Taberlet P, Gielly L, Pauto G, Bouvet J (1991) Universal primers for the amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology* **17**, 1105–1109.

Thomas G, Joseph L, Varghese G, Sreejayan, Kalyanaraman K, Kuriachan P (2001) Analysis of phenotypic and genetic variations among populations of *Oryza malampuzhaensis* show evidence of altitude-dependent genetic changes. *Canadian Journal of Botany-Revue Canadienne De Botanique* **79**, 1090–1098.

Welsh J, McClelland M (1990) Fingerprinting genomes using PCR arbitrary primers. *Nucleic Acids Research* **18**, 7213–7218.

Whitkus R (1992) Allozyme variation within the *Carex-Pachystachya* complex (Cyperaceae). *Systematic Botany* **17**, 16–24.

Wolfe AD, Liston A (1998) Contributions of PCR-based methods to plant systematics and evolutionary biology. In: *Molecular Systematics of Plants II: DNA Sequencing* (ed Soltis DE, Soltis PS, and Doyle JJ), pp. 43–86. Kluwer Academic Press, Boston.

Wolfe AD, Xiang QY, Kephart SR (1998a) Assessing hybridization in natural populations of *Penstemon* (Scrophulariaceae) using hypervariable inter simple sequence repeat (ISSR) bands. *Molecular Ecology* **7**, 1107–1125.

Wolfe AD, Xiang QY, Kephart SR (1998b) Diploid hybrid speciation in *Penstemon* (Scrophulariaceae). *Proceedings of the National Academy of Sciences USA* **95**, 5112–5115.

Zietkiewicz E, Rafalski A, Labuda D (1994) Genome Fingerprinting by simple sequence repeat (Ssr)-anchored polymerase chain-reaction amplification. *Genomics* **20**, 176–183.

Chapter 3 Intraspecific morphological and genetic marker variation over large geographical scales

Abstract

Morphological characters vary within and between conspecific populations; this variation is often formally recognised with the creation of subspecific taxa. Morphological character variation can be caused by differences in the evolutionary history of populations or be a plastic response by an individual in response to extrinsic factors. If the latter is the cause of variation, the subspecific taxa may not correspond to evolutionary genetic groups, and are therefore not real. By surveying variation of both genetic and morphological characters it is possible to identify those morphological characters that correspond to genetic differences among conspecific populations. Morphological characters that correspond to molecular marker variation can be used to define subspecific taxa.

This chapter investigates the correlation of morphological and genetic character variation within two *Arum* species: *A. maculatum* and *A. italicum*. *A. maculatum* has no recognised subspecific taxa but its leaves are polymorphic in both shape and pattern. *A. italicum* is separated into four subspecies. Two of the subspecies, subsp. *italicum* and subsp. *neglectum*, have overlapping ranges and identical inflorescence morphology but are differentiated by leaf lobe shape and leaf patterning.

In *A. maculatum* leaf lobe width correlated with genetic variation. However, when spotted and non-spotted plants were compared, the data indicated that they were not genetically different. Unlike the results in *A. maculatum*, the patterned leaves of *A. italicum* subsp. *italicum* and the non patterned leaves of *A. italicum* subsp. *neglectum* corresponded to two genetic groups, thus supporting the subspecies differences. No correlation between the leaf lobe characters and molecular marker variation in subsp.

italicum was found. In subsp. *neglectum* the lobe character and molecular marker did correlate, but the correlation was different in allopatric and sympatric populations. Lobe length correlated with molecular marker in allopatric populations of subsp. *neglectum*, whereas in sympatric populations (containing both subspecies) lobe angle variation correlated with molecular marker. The differences between the two population types could be through either competition or gene flow. The current dataset cannot be used to confirm if competition is causing the difference between sympatric and allopatric populations, it does however indicate that gene flow occurs between the subspecies in sympatric populations.

In conclusion, the results in this chapter show that similar characters in different species do not necessarily follow the same patterns or correspond to genetic differences. The correlation of morphological variation with molecular marker can also be affected by the type of populations studied. If a population contains more than one subspecific taxon, interbreeding between the subspecific taxa may affect the correlation of morphological and molecular marker. Therefore, no generalisations can be made *a priori* on the usefulness of a morphological character variation in representing underlying genetic variation estimated using molecular markers.

Introduction

Species are not homogeneous; they can vary in both morphological and genetic characters (Linhart and Grant, 1996). Therefore, species descriptions must encompass intraspecific variation. Such variation is usually recognised taxonomically through differences in morphology and distribution. For example, both morphology and geography are used to distinguish among four subspecies of *Arum italicum*. (Boyce, 1993). The evolutionary species concept defined by Simpson (1961) and Wiley (1978, 1981) states that species consist of single lineages of evolving populations. Theoretically, species and subspecies should follow the evolutionary species concept (e.g. Brunell and Whitkus, 1997), which states that taxa should correspond to a single evolutionary lineage.

Evolutionary lineages can be estimated by surveying molecular marker variation; however, many species and subspecies are only identified on the basis of morphology and distribution. The lack of genetic data means that some subspecies should not be recognised, such as the two intraspecific taxa of *Digitalis minor*. The two taxa *D. minor* var. *minor* and *D. minor* var. *palaui* are separated by differences in pubescence (hairs) but these taxa are not genetically distinct and therefore should not be recognised taxonomically (Sales *et al.*, 2001). In contrast, morphological variation of leaves and inflorescence in the species *Cardamine amara* is used to identify six subspecies, all of which correspond to genetically distinct groups, thus supporting the subspecies of *Cardamine amara* (Lihova *et al.*, 2000).

The formation of evolutionary lineages within species is dependent on reduced gene flow among populations, eventually allowing genetically isolated populations to become evolutionary distinct. For instance, geographical features like mountain ranges can reduce gene flow between populations, leading to genetic divergence through geographical or physical isolation. Genetic differentiation has been observed in *Eriastrum densifolium* populations in California either side of the Transverse Mountain Range (Brunell and Whitkus, 1997).

In tandem with geographical isolation, adaptation of populations to localised environmental conditions can also lead to differentiation of taxa. In populations of *Oryza malampuzhaensis* in southern India, molecular marker variation was significantly related to population location (Thomas *et al.*, 2001). In addition, morphological variation of culm, ligule and anther length in *Oryza malampuzhaensis* were correlated with altitude (Thomas *et al.,* 2001). This indicates that the populations are adapting to local environmental conditions associated with differences in altitude. The differences in morphology are reinforced through restricted gene flow between the populations. A similar situation in *Quercus patraea* has also been shown: populations are genetically differentiated through isolation by distance with leaf character variation corresponding to local moisture conditions (Bruschi *et al.*, 2003). In these examples, morphological differences between populations corresponded to differences in environmental conditions; genetic differentiation caused by restricted gene flow contributes to the divergence of the populations.

Gene flow between populations can reduce genetic differentiation, and can form a continuum with fewer genetic groups than morphologically recognised taxa. For example, the subspecies *Elymus glaucus* subsp. *jepsonii* is not genetically distinct from *E. glaucus* subsp. *glaucus* and is therefore not a separate taxon (Wilson *et al.*, 2001). Similarly, the taxonomic status of the species *Antirrhinum ambiguum* is questioned as it is not genetically distinct from *A. graniticum* and therefore should be considered as a subspecies of *A. graniticum* rather than a separate species (Mateu-Andres and Segarra-Moragues, 2003). Gene flow between taxa therefore limits genetic differentiation, creating a single taxon and preventing the differentiation of subspecific taxa.

This chapter will use a comparative analysis of morphological and molecular marker variation over a large geographical area to investigate intraspecific variation in *Arum maculatum* and *A. italicum*. These two species are morphologically similar and are classified taxonomically within the section *Arum* of the genus *Arum* (Boyce, 1993). However, intraspecific variation within these species is treated differently by Boyce (1993). In *A. italicum*, morphological characters and patterns of distribution have been used to delimit subspecies, whereas, in *A. maculatum*, similar morphological characters exist but they have not been used to delimit subspecies. A comparative approach using both morphological and molecular genetic charcters can be used to identify if the taxonomic situation in *A. maculatum* and *A. italicum* reflects the underlying evolutionary differences within these two species.

Arum maculatum

A. maculatum is distributed throughout Europe (Figure 3.1), occurring in oak woodlands (Harmes, 1982) and hedgerows (Boyce, 1993). Both Prime (1960) and Harmes (1982) have recorded polymorphic leaf shape and patterning; some individuals have leaves with black spotting whilst others have plain leaves (Figure 3.2). The current taxonomic classification of *A. maculatum* by Boyce (1993) does not identify any subspecific taxa, even though there is evidence of extensive polymorphism within the species. The polymorphism recorded by Prime (1960) and Harmes (1982) is similar to that used to delimit subspecies of *A. italicum.* Whether such polymorphisms should be used to detect subspecies of *A. maculatum* can be tested by investigating the genetic relationships between individuals with different leaf polymorphisms.

Figure 3.1 Map showing the distribution of *A. maculatum*. The green dashed line shows the extent of *A. maculatum* distribution from Map 1 of Boyce (1993, p. 64)

Figure 3.2 Two A. maculatum plants growing together. The plant on the left has spotted leaves (1) whilst the plant on the right has plain leaves (2).

Arum italicum

The current taxonomic classification of *A. italicum* separates the species into four subspecies, subsp. *italicum*, *neglectum*, *canariense* and *albispathum*. The latter two subspecies (subsp. *canariense* and *albispathum*) are differentiated on flower coloration, some anatomical floral characters, and by geographical range. *Arum italicum* subsp. *canariense* is endemic to the Canary Islands whilst subsp. *albispathum* occurs east of the Black Sea (Figure 3.3). The other two subspecies, subsp. *italicum* and subsp. *neglectum*, have indistinguishable inflorescence morphology and have overlapping distributions but differ in leaf morphology (Figures 3.3–3.5). Subspecies *neglectum* has shorter and wider posterior leaf lobes compared with subsp. *italicum*, in which the lobes

are longer and divergent (Boyce, 1993). Subspecies *italicum* has narrower leaves, often with silver venation that is absent in subsp. *neglectum* (Boyce, 1993)

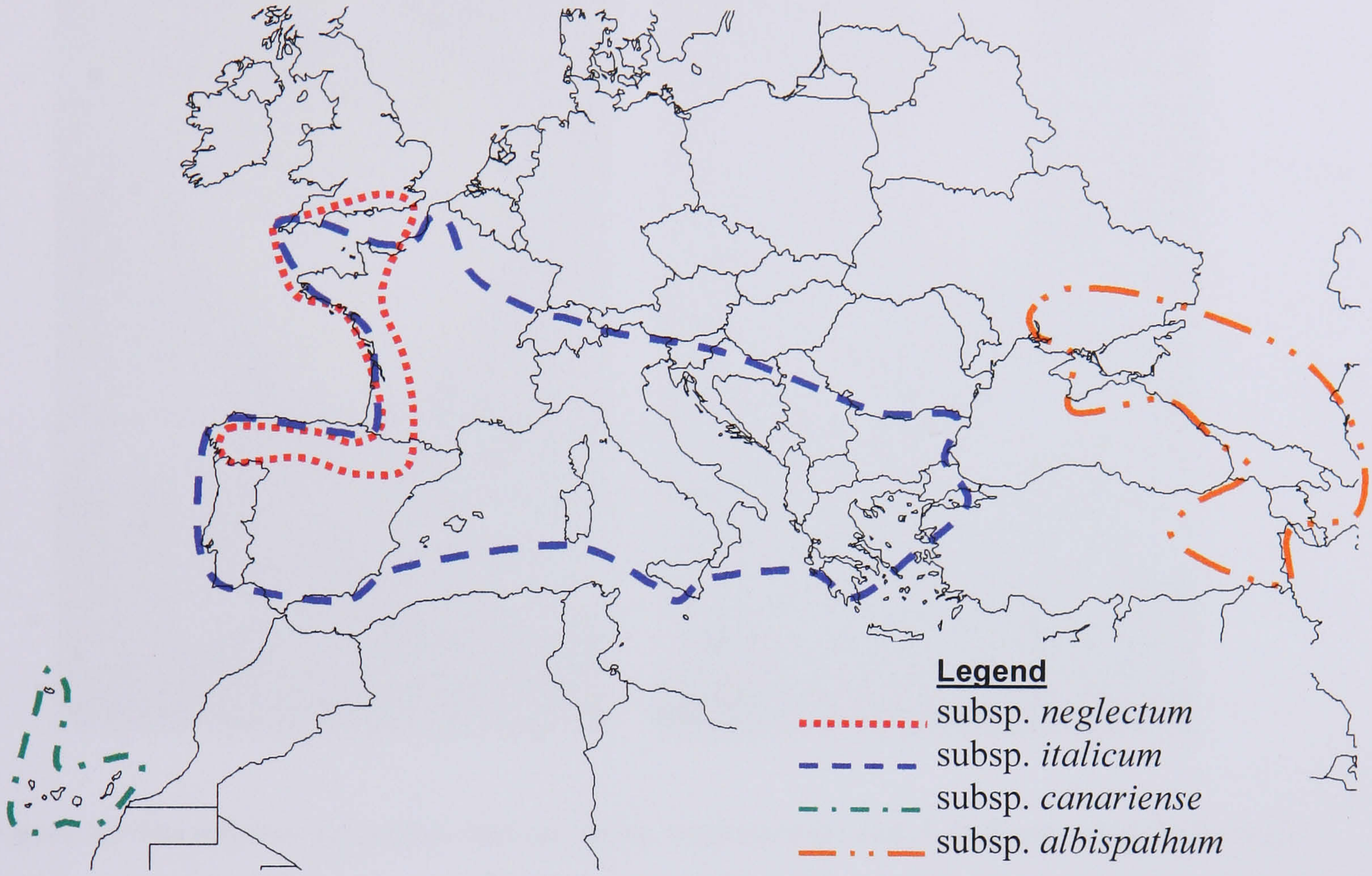


Figure 3.3 Map showing the distributions of the four *A. italicum* subspecies, subsp. *neglectum*, subsp. *italicum*, subsp. *canariense* and subsp. *albispathum*. Adapted from Map 3 a-d of Boyce (1993, p. 78).

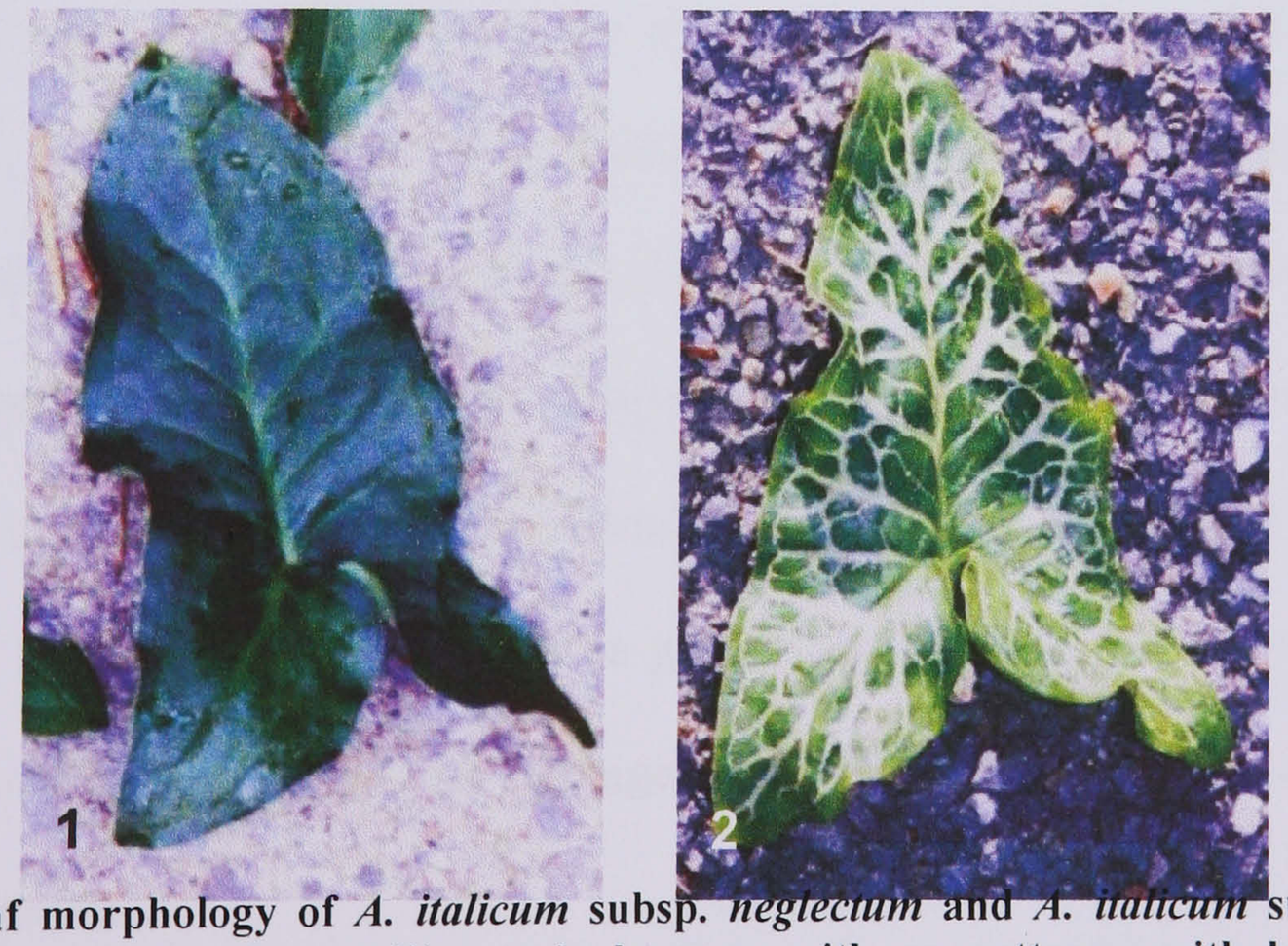


Figure 3.4 Leaf morphology of *A. italicum* subsp. *neglectum* and *A. italicum* subsp. *italicum*. *A. italicum* subsp. *neglectum* leaves (1) are dark green with no patterns, with broad overlapping posterior lobes. *A. italicum* subsp. *italicum* leaves (2) are a lighter green with patterned veins or blotching and narrow posterior lobes.

Figure 3.5 The inflorescences of *A. italicum* subsp. *neglectum* (1) and *A. italicum* subsp. *italicum* (2).

The inflorescences of subsp. *italicum* and subsp. *neglectum* are morphologically very similar. Both subspecies attract the same pollinators: midges from the genus *Psycoda* (Albre *et al.*, 2003; Diaz and Kite, 2002). The sympatric distribution and shared pollinators indicate that there is potential for gene flow between the two subspecies. If there is extensive gene flow between subspecies a genetic continuum may arise, similar to that found in *Eristatum densifolium* (Brunell and Whitkus, 1997) and *Digitalis minor* (Sales *et al.*, 2001). If there is extensive gene flow between the two subspecies they would not constitute two separate lineages and therefore should not be considered distinct subspecies.

This chapter investigates correlations of morphological and genetic marker variation within species over wide geographical scales. The aims are to identify which, if any, morphological characters should be used to distinguish intraspecific taxa and which processes are affecting the genetic and morphological variation. To accomplish these aims the following questions are asked: 1. What morphological variation is there within *A. maculatum* and *A. italicum*? 2. What molecular marker variation is there within these two species over the surveyed range? 3. Do any morphological characters correlate with the estimated genetic lineages? By answering these questions in this logical order, it is possible to identify those morphological characters that correspond to evolutionary lineages and the population processes that have contributed to patterns of intraspecific variation in both *A. maculatum* and *A. italicum*.

The two species are treated independently within this chapter. Study 1 concerns *A. maculatum* and Study 2 examines *A. italicum*. Both studies have a dedicated method, results and discussion section. A final discussion section is included, drawing together conclusions from both species. Comparing and contrasting the different results found in each species will enable a greater understanding in the utility of morphological characters at the intraspecific level.

General methods used in Chapter 3

The following section details the DNA extraction method common to both *A. maculatum* an *A. italicum* studies. An additional DNA extraction method was carried out on a subset of *A. maculatum* populations and is detailed in the *A. maculatum* section of this chapter.

Sampling design

All three studied taxa were sampled in an overlapping area including southern Britain western France and northern Spain. Two *A. italicum* subspecies occur in this area, and are differentiated on leaf morphology. In the same area similar leaf variation pattern exists in *A. maculatum* but is not used to define subspecific taxa. This sampling design enables a comparison of similar morphological characters between taxonomically similar species over the same geographical area.

DNA extraction

The Qiagen DNEasy 96 plant kit was used to extract DNA from 10mg of dried leaf tissue as per the manufacturer's instructions with the following amendments. Initial disruption was carried out for 1.5 minutes using a Mixer Mill MM300, with tungsten carbide beads as a grinding aid, in boxes of 96 2ml tubes. Disrupted plant material was centrifuged for 5 minutes at 5600***g***, the orientation of the sample boxes was reversed and centrifuged again for 5 minutes at 5600***g***. Reversing the orientation of the boxes

ensures that an even centrifugal force is applied to each sample. After the –20°C incubation, the samples were centrifuged again for two spins of 5 minutes at 5600***g***, with the box in the reverse orientation for the second spin. Eluted DNA samples were separated into two aliquots of 400µl. One set of aliquots was stored at –70°C for long-term storage and future work. The second aliquot was stored at –20°C for general use and short-term storage.

DNA amplification and visualisation

DNA was amplified using ISSR primers and visualised on silver-stained acrylamide gels. Chapter 2 contains a full description of the PCR and fragment visualisation methods.

Data analysis

All of the measured leaf characters were tested for normality using the Anderson–Darling Normality Test in MiniTab. All of the characters differed significantly from normal, and therefore non-parametric statistical tests were used. To test differences between the morphology of allopatric and sympatric populations, the Mann–Whitney *U* test was used to identify if the median values differed significantly between groups.

Study 1: *A. maculatum*

This chapter investigates correlations of morphological and genetic marker variation within species. The aim is to identify which, if any, morphological characters should be used to distinguish intraspecific taxa in the genus *Arum*. This section is concerned with morphological and genetic marker variation in *A. maculatum* and is divided into two parts. Part A investigates variation in *A. maculatum* populations in France and Britain. Part B investigates the correlation between leaf spotting and genetic marker variation in *A. maculatum* populations within Britain in more detail.

There are no subspecific taxa currently recognised in *A. maculatum* (Boyce, 1993), but polymorphisms in leaf shape and patterning have been recorded (Harmes, 1982; Prime, 1960). Investigating the genetic relationships between *A. maculatum* individuals will enable any subspecific taxa to be identified. Three questions are asked in this section: 1. What morphological variation is there within *A. maculatum*? 2. What genetic marker variation is there within *A. maculatum* over the surveyed range? 3. Do any morphological characters correlate with estimated genetic lineages?

Part A: Variation in *A. maculatum* across Europe

Methods

Arum maculatum plants were surveyed in locations in Britain and France during spring 2001 (Figure 3.6 and Table 3.1). The location of sampling sites was chosen to overlap with the distribution of *A. italicum* subsp. *neglectum*. The sampling strategy was adopted as similar leaf variation to that being studied in *A. maculatum* has been used to differentiate between two subspecies (subsp. *italicum* and subsp *neglectum*). The plants sampled were from the southern coast of Britain, the Channel Islands and western France (Figure 3.6). The distribution of *A. maculatum* published by Boyce (1993) indicates that the range of *A. maculatum* extends across the Pyrenees into Spain. However, no populations of *A. maculatum* were found in southern France as the soils are too acidic and northern Spain as these populations were found to be *A. alpinum* (A. Diaz, personal communication).

Ten randomly selected individual plants were sampled from each location, and a mature leaf with no obvious disease or damage was selected. To ensure no clones were sampled, sampled individuals were more than 1m apart as the tuber growth habit favours spreading colonies (Boyce 1993). Leaves were then scored for lobe variation: the left lobe (adaxial surface uppermost) was measured for length (LL), width (LW) and the angle between the midrib and the main vein on the lobe (LA). The amount of spotting was calculated using photographs of previously calibrated standard leaves. After the leaf characters were recorded, the leaf was rinsed in ethanol and

approximately $3cm^2$ was ripped off, avoiding the edge and midrib, and stored in a sealable plastic bag containing fine silica gel (pore diameter ca 6nm).

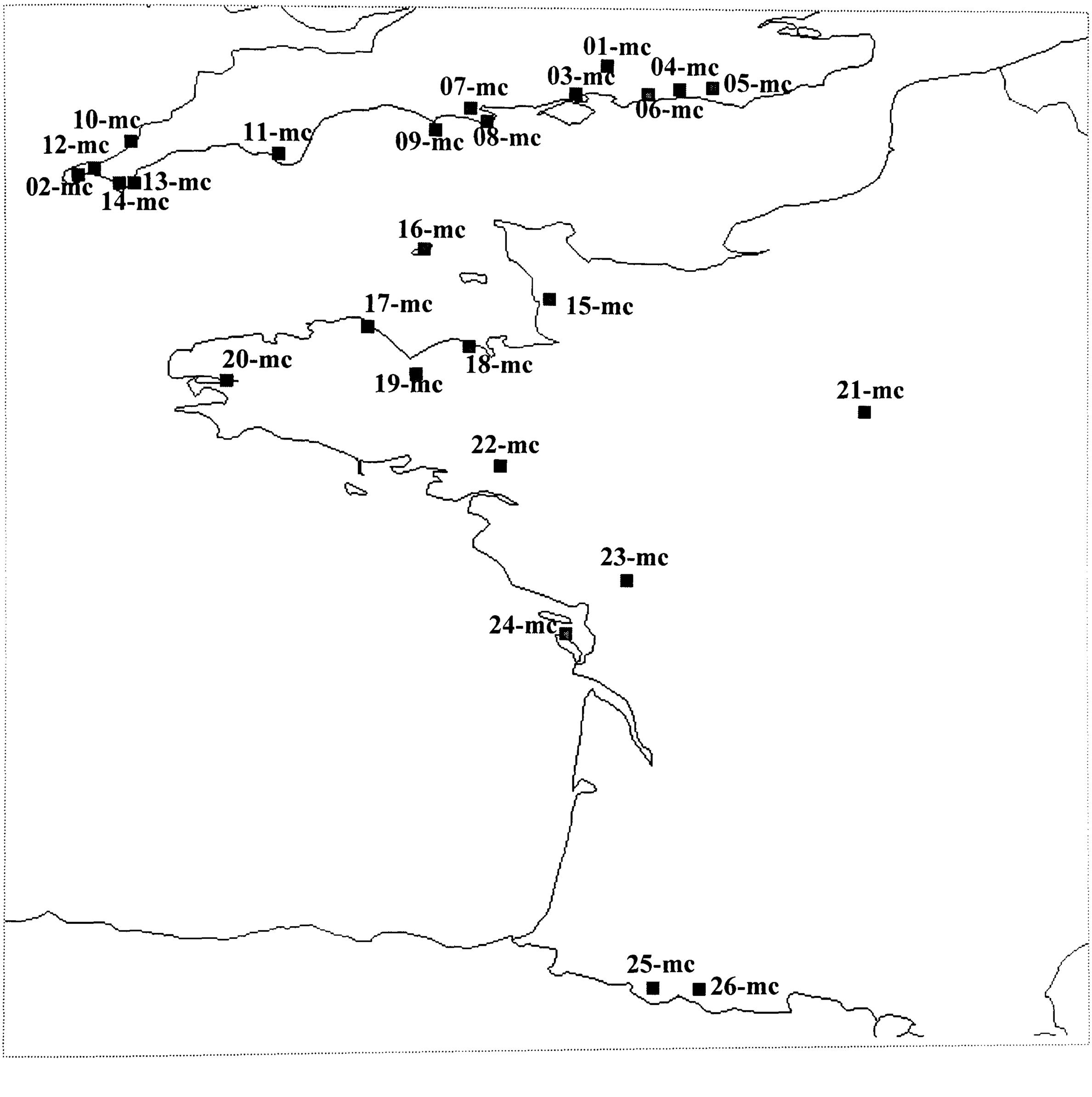


Figure 3.6 Map of *A. maculatum* sampling locations in spring 2001 for Part A.

Table 3.1 The sampling locations of *A. maculatum*. For each population the location, longitude and latitude are specified. Black squares in the column "Successful ISSR Profile" indicates that reliable amplified products were obtained for that population.

Location Code	Location	Longitude (°W)	Latitude (°N)	Successful ISSR profile
01-mc	Selbourne	0.940	51.090	
02-mc	Penzance	5.570	50.110	
03-mc	East Meon	1.220	50.840	■
04-mc	Steyning	0.300	50.880	■
05-mc	Offham	0.010	50.890	■
06-mc	Arundel	0.580	50.840	■
07-mc	Wareham	2.140	50.710	■
08-mc	Dancing	2.005	50.594	■
09-mc	Portland	2.460	50.520	■
10-mc	Crantock	5.100	50.400	■
11-mc	South Milton	3.830	50.290	
12-mc	Lelant	5.430	50.160	■
13-mc	St Kev	5.070	50.040	■
14-mc	Mullion	5.210	50.040	■
15-mc	P. de Countance	1.453	49.035	■
16-mc	Guernsey	2.553	49.469	
17-mc	Paimpol	3.054	48.777	■
18-mc	Lancieux	2.156	48.606	■
19-mc	Moncontour	2.633	48.361	■
20-mc	Landevennec	4.275	48.297	
21-mc	Nozay	-1.357	48.024	■
22-mc	Plesse	1.886	47.538	
23-mc	Mervant	0.757	46.523	■
24-mc	Chize	1.306	46.052	■
25-mc	Ferneres	0.530	42.907	
26-mc	Cauterets	0.116	42.886	■

Results

Morphological variation

The variation of each measured character in each sampled location of *A. maculatum*: lobe length (LL), lobe angle (LA), and lobe width (LW) is plotted on a series of box plots (Figure 3.7). The box plots indicate the range of variation within and among the populations. Each vertical bar represents a single location, and horizontal bars represent the range of data between the 25^{th} and 75^{th} percentiles. Comparison of the data range and medians (middle horizontal bar) gives an indication of variation among the sampling sites. The box plots indicate that there is variation within and among all of the sampled populations of *A. maculatum* studied.

The variation of the morphological characters differs significantly from normal (Anderson–Darling Normality Test). Therefore, Spearman's rank correlations were used to investigate relationships among characters. There are statistically significant trends within the morphological data both spatially and between the morphological characters (Table 3.2). The three lobe characteristics (LL, LA and LW) are all significantly correlated with each other, plants with long lobes having the widest lobes and the largest lobe angles. Lobe length significantly correlates with latitude (d.f.=189 r=-0.144, P<0.002) whereas spotting correlates individually with both longitude (d.f.=189 r=-0.247, P<0.001) and latitude (d.f.=189 r=0.193, P<0.002, Table 3.2). Therefore, plants in the south-east tended to have a higher proportion of spotting on their leaves. The

statistically significant correlations of spotting and lobe length with latitude and longitude indicate that there is spatial structure of the lobe characteristics among the sampled locations of *A. maculatum*.

A Mantel test was used to identify if morphologically similar individuals occur close together. It is possible that the simple correlation analysis (Table 3.2) could be affected by the uneven geographical distribution of the sampling sites. The effect of uneven sampling distribution is reduced by comparing geographical distance with morphological distance (Euclidean distance). All morphological characters were compared together and separately with geographical distance between the sampled sites. A correlation that is approaching significance (n=19, r=0.159, P=0.057) was found when all the morphological characters are tested together. However, when the leaf characters were tested individually, no statistically significant correlations with geographical distance were identified (Table 3.3). The morphological variation therefore shows no localised substructure as there is no relationship between sample location and morphological character variation.

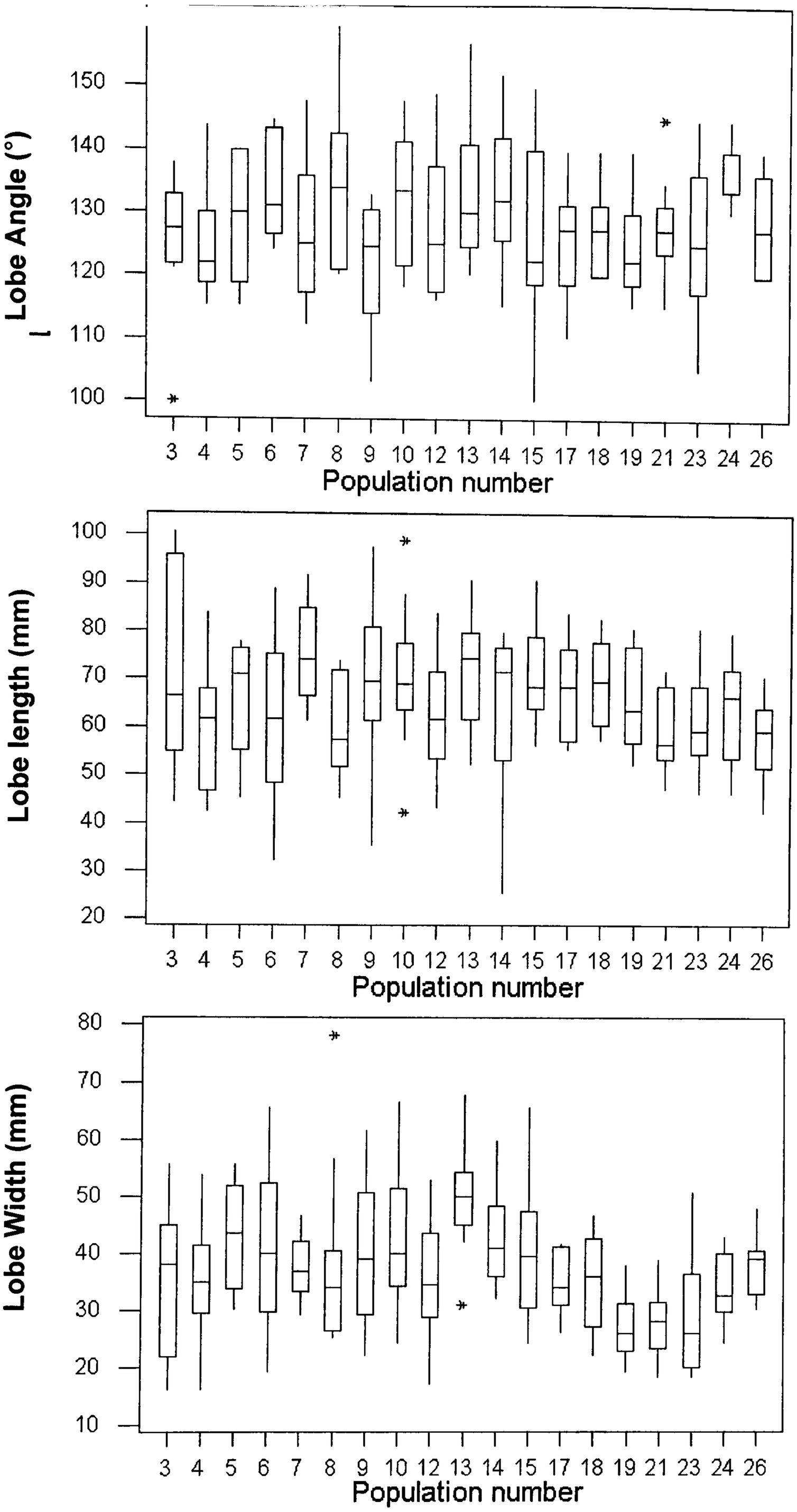


Figure 3.7 Box plots showing within- and among-sampled location variation of leaf characters in *A. maculatum*. The characters are posterior leaf lobe angle, length and width. The distances between the lower and upper percentiles (lower and upper limit of each box) from the median value (horizontal line) indicate within-location variation. Variation among locations is indicated by the median values. Total range of values (those lower and higher than the lower and upper quartiles) for each population is indicated by the solid lines. Asterisks (*) indicate outliers. Population number is equivalent to those found in Table 3.1, e.g. Population 9 is 09-mc from Portland.

Table 3.2 Correlation matrix of four morphological characters in *A. maculatum*. The characters are % spotting, lobe angle, length and width. Correlations of the morphological characters with longitude and latitude are also included. Upper value is the correlation value (*r*), the lower value is the *P*-value. A *P*-value greater than 0.05 is non-significant. Significant correlations are outlined in black. Degrees of freedom (*n*–1) for this test are 189.

	Longitude	Latitude	% Spotting	Lobe angle	Lobe length
% Spotting	–0.247 <0.001	0.193 0.002			
Lobe angle	0.106 0.088	–0.01 0.871	–0.054 0.39		
Lobe length	0.106 0.089	–0.144 0.021	0.008 0.902	–0.181 0.003	
Lobe width	0.115 0.064	–0.105 0.092	–0.103 0.098	–0.177 0.004	0.555 <0.001

Table 3.3 Mantel test result. Comparison of the Euclidean distance for each morphological character and geographical distance among all 19 sampled *A. maculatum* locations. *P*-values greater than 0.05 are non-significant and are indicated by the annotation n.s.

	r- Mantel correlation	*P*-value
Lobe length	0.167	0.076 n.s.
Lobe width	0.079	0.234 n.s.
Lobe angle	0.067	0.665 n.s.
% Spotting	0.121	0.098 n.s.

Genetic marker variation

For populations of *A. maculatum* the two different ISSR primers amplified 56 scorable bands, all of which exhibited polymorphism (Table 3.4). Primer 811 produced 26 scorable bands and primer 834, 30 scorable bands. Seven *A. maculatum* population locations, including the sole Channel Island samples (16-mc), failed to produce reliable ISSR profiles for both primers (Table 3.1).

Table 3.4 ISSR band frequencies in *A. maculatum* for primers 811 and 834.

	Primer band frequencies				
	811	**834**		**811**	**834**
1	0.29	0.88	**16**	0.56	0.07
2	0.46	0.48	**17**	0.34	0.06
3	0.31	0.14	**18**	0.01	0.10
4	0.19	0.87	**19**	0.32	0.29
5	0.62	0.91	**20**	0.76	0.20
6	0.46	0.11	**21**	0.09	0.23
7	0.55	0.38	**22**	0.86	0.36
8	0.99	0.65	**23**	0.86	0.19
9	0.99	0.16	**24**	0.40	0.15
10	0.43	0.37	**25**	0.18	0.02
11	0.90	0.07	**26**	0.15	0.26
12	0.42	0.06	**27**	-	0.03
13	0.33	0.16	**28**	-	0.03
14	0.90	0.30	**29**	-	0.80
15	0.47	0.13	**30**	-	0.72

The genetic relationships among the locations of sampled *A. maculatum* elucidated using ISSR markers are inferred in a neighbour-joining (NJ) tree (Figure 3.8). The populations cluster into two monophyletic clades, one containing British samples and the second containing French samples and a single British location (5-mc). The difference between samples form Britain and France was assessed using an ANOSIM

test. ANOSIM test results show there is a statistically significant difference between the two groups of populations (n=19, r=0.511, P<0.001). A Mantel test was used to identify if increasing geographical distance and genetic distance among sampled locations were correlated, and a significant positive relationship was identified (n=19, r=0.36, P=0.013). This suggests that there is isolation by distance among the *A. maculatum* samples.

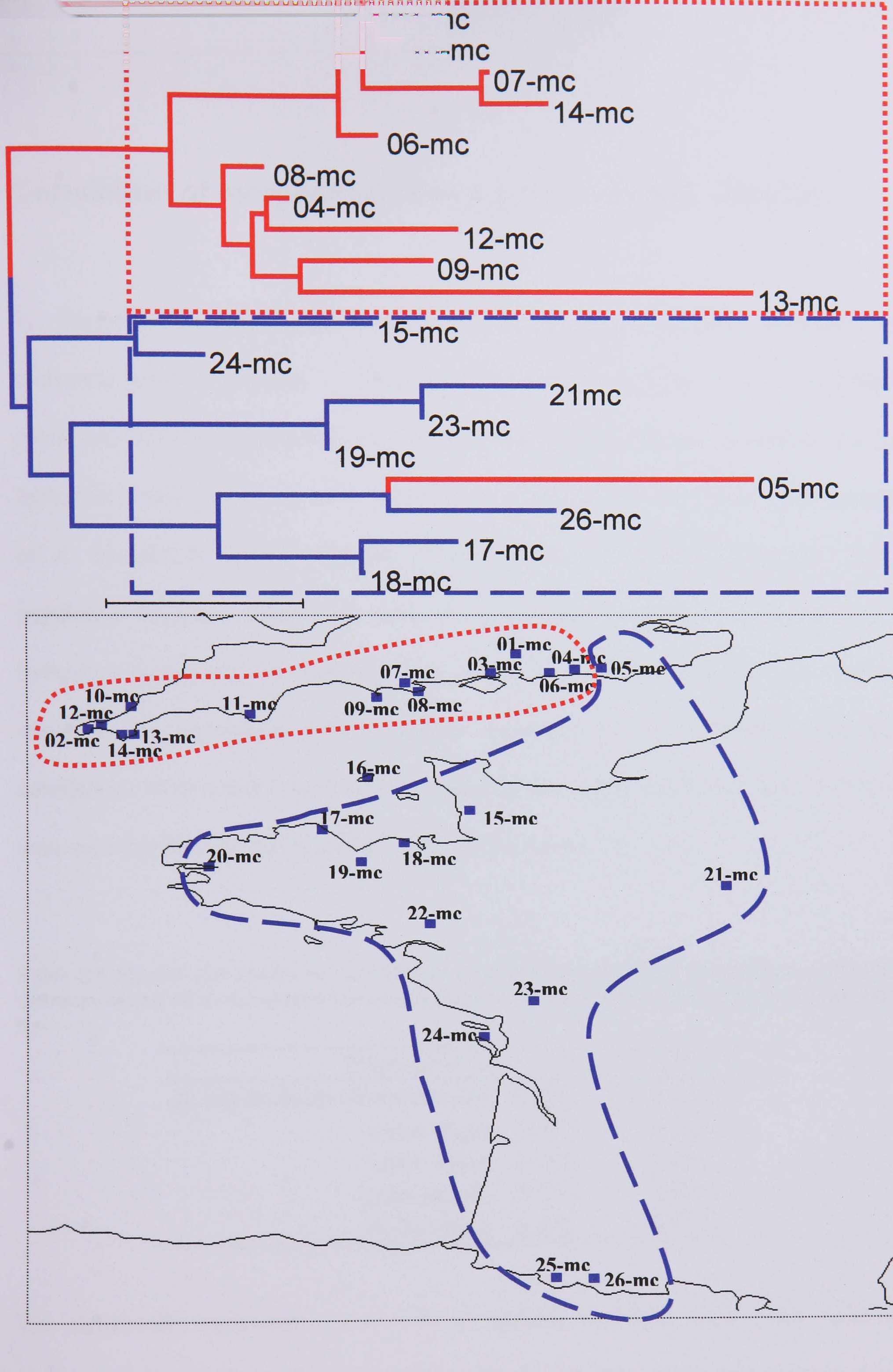


Figure 3.8 Neighbour-joining tree of pairwise genetic distances between 19 *A. maculatum* sampled populations using data from primers 811 and 834. The branch lengths are proportional to the genetic distances among the sampled locations (scale indicates genetic distance). Branch colour indicates the geographic location of each location: red branches are British sampled locations whilst blue are French.

Correlation of morphological and genetic marker variation

To identify if morphological variation of *A. maculatum* samples correlates with molecular marker variation, a Mantel test was used to compare Euclidean distance calculated from the morphological characters and genetic distance. Morphological and molecular marker variation were significantly correlated between population locations of *A. maculatum* (n=19, r=0.204, P=0.015, Table 3.5); morphologically similar population locations (therefore) tend to be genetically similar. To identify if an individual morphological character is significantly correlated with molecular marker variation the characters were analysed separately. Two statistically significant correlations were found (Table 3.5). Variation of lobe angle and the amount of spotting were also significantly correlated with molecular marker variation.

Table 3.5 Mantel test results for correlation of genetic distance and morphological Euclidean distances among all *A. maculatum* sampling locations. A *P*-value greater than 0.05 is not significant, n.s.

Comparison	*r*	*P*
***A. maculatum*: all characters**	0.204	0.015
Lobe angle	0.073	0.729 n.s.
Lobe width	0.368	0.002
Lobe length	0.019	0.402 n.s.
% Spotting	0.202	0.031

The statistically significant correlations of spotting and lobe width with molecular marker variation indicate that individuals with similar lobe width and amount of leaf spotting are genetically similar. As a tool to distinguish between taxa, lobe width does not unambiguously define any groups as there are similar amounts of variation within

all surveyed locations. However, leaf spotting could be used unambiguously to define taxa as it can be interpreted as a binary character: it is present or absent, there is no intermediate state.

These data suggests that there is a relationship between genetic similarity and spotting. However, use of spotting as a taxonomic character cannot be assessed with confidence in this survey because there are a disproportionate number of successful ISSR profiles for spotted than non-spotted samples (Table 3.6). Half of the non-spotted locations do not have an ISSR profile; of those that do three are located in Cornwall and one is located in Brittany. This uneven distribution pattern leads to a second problem. The British and French samples are statistically significantly genetically different (ANOSIM test result: n=19, r=0.511, P<0.001). Consequently any genetic difference between spotted and non-spotted could not be distinguished from the effect of geographical isolation. A second study where samples are distributed evenly was therefore carried out and is presented in Part B of this chapter.

Table 3.6 Amount of spotting in sampled *A. maculatum* populations and ISSR profile success.

Location code	% Spotting on leaves	Successful ISSR profile
01-mc	0.0	
02-mc	0.0	
03-mc	0.2	■
04-mc	8.0	■
05-mc	11.0	■
06-mc	3.5	■
07-mc	5.5	■
08-mc	3.0	■
09-mc	1.5	■
10-mc	0.0	■
11-mc	0.0	
12-mc	0.0	■
13-mc	0.0	■
14-mc	0.0	■
15-mc	9.2	■
16-mc	0.0	
17-mc	10.9	■
18-mc	0.0	■
19-mc	13.2	■
20-mc	7.7	
21-mc	7.7	■
22-mc	16.2	
23-mc	18.0	■
24-mc	12.0	■
25-mc	2.4	
26-mc	3.5	■

Part B: Variation in *A. maculatum* within Britain

A second round of sampling was required after the previous analysis, in which it was found that leaf patterning and successful ISSR profiles followed an uneven distribution. Therefore, these sample locations could not be used to test if leaf patterning was indicative of differences in evolutionary history because isolation by distance could also account for the observed variation. A further study was therefore required to achieve all the aims of this study. Part B is of a smaller scope than Part A, because the samples were collected in the final year of study, and time was limited. Only a single question is asked of the data: is the presence or absence of leaf spotting indicative of evolutionary history?

Methods

In the second *A. maculatum* study, plants from Britain were collected in spring 2003. This study had the explicit aim to compare spotted and non-spotted morphs of *A. maculatum* from Britain only. By using only British samples, the potential for skewed results caused by the large distances is reduced. In total, 24 new population locations from southern Britain were sampled: 27-mc to 50-mc (Table 3.7). Samples in this study were collected in the same manner as before with individuals separated by 1m. Twenty plants were sampled per location, 10 spotted and 10 non-spotted. Because of the limited time available only the presence or absence of spotting was recorded. DNA for the second collection of *A. maculatum* populations was extracted using the method of Mogg

and Bond (2003). This method is cheap and quick, enabling DNA from all of the samples to be extracted within 2 weeks. A single ISSR primer (811) was tested on these populations; variation in a second primer was not surveyed because of financial and time constraints.

Table 3.7 Locations of sampled *A. maculatum* populations sampled in the second season, spring 2003. For each population the location, longitude and latitude are specified. Black squares in the column "successful ISSR profile" indicate that reliable amplified products were obtained for that population.

Location code	Location	Longitude (°W)	Latitude (°N)	Successful ISSR profile
27-mc	St Aldhams	-2.0593	50.5785	■
28-mc	Wareham	-2.1124	50.6939	■
29-mc	Blandford Forum	-2.1696	50.8550	■
30-mc	Shaftesbury	-2.1214	51.0101	■
31-mc	Warminster	-2.1728	51.2041	■
32-mc	Devizes	-1.9938	51.3485	■
33-mc	Wantage	-1.4259	51.5920	■
34-mc	Oxford	-1.2780	51.6965	■
35-mc	Banbury	-1.3031	52.0433	■
36-mc	Daventry	-1.1843	52.2448	■
37-mc	Wellingbrough	-0.6810	52.3207	■
38-mc	Kettering	-0.7171	52.4063	■
39-mc	Stamford	-0.7116	52.5955	■
40-mc	Longbough	-1.1766	52.7653	■
41-mc	Canterbury	1.0929	51.2817	■
42-mc	Tonbridge	0.2579	51.1860	■
43-mc	Arundel	-0.7777	50.8263	
44-mc	Selbourne	-0.9764	51.1506	
45-mc	Salisbury	-1.7873	51.0691	■
46-mc	Dorchester	-2.4395	50.7144	■
47-mc	Crewkerne	-2.7814	50.8880	■
48-mc	Taunton	-3.1007	51.0215	■
49-mc	Exeter	-3.5160	50.6950	■
50-mc	Great Torrington	-4.1373	50.9598	■

Results

Genetic marker variation

The single ISSR primer (811) amplified 27 polymorphic bands in the British *A. maculatum* samples. The relationships among the sampled locations are shown in an NJ tree (Figure 3.9). There is a strong spatial structuring of the sampled locations, with three main groups corresponding to the East Midlands, South West and Kent. There is a statistically significant relationship between geographical and genetic distance among them (n=22 r=0.286, P=0.007), suggesting gene flow is limited by distance.

Correlation of leaf patterning and genetic marker variation

The genetic relationships among the tested individuals are shown in an MDS plot (Figure 3.10). It can be seen that there is no obvious clustering of the spotted and non-spotted *A. maculatum* varieties. The *stress* value of the MDS plot is higher than 0.2; a value higher than 0.2 indicates that the relationships among the individuals are not well represented in two dimensions (Clarke and Warwick, 2001; Kruskal, 1964). The ANOSIM test was used to identify if there is a statistically significant difference between the genetic relatedness of individual plants with spotted leaves and non-spotted plants. The ANOSIM test indicates that the spotted and non-spotted individuals are not significantly genetically different (n=440 r=0.003 P=0.12) these data indicate that leaf spotting in *A. maculatum* does not correspond to distinct genetic groups

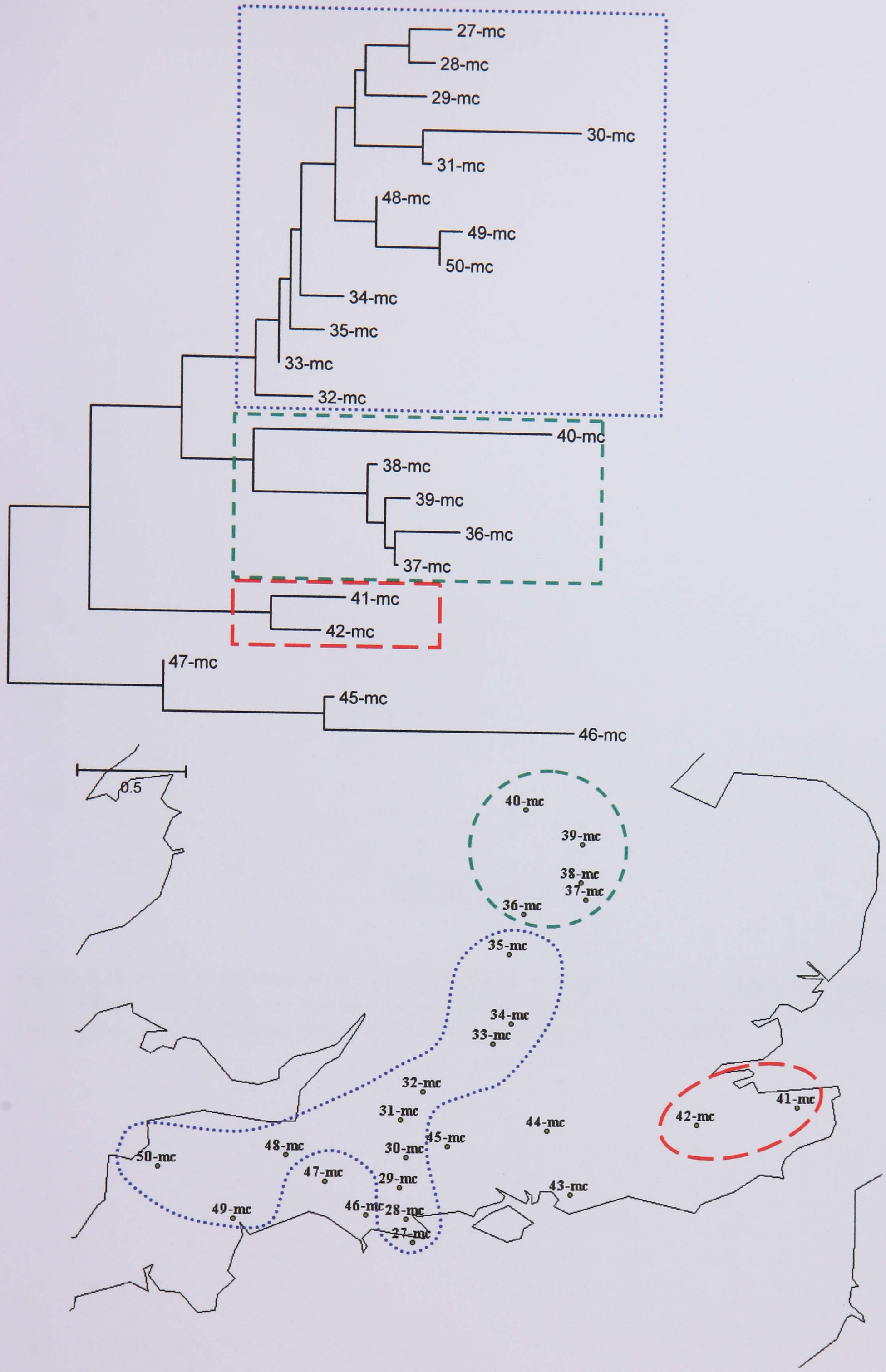


Figure 3.9 Neighbour-Joining tree of pairwise genetic distances among the sampled population locations; population groups are superimposed on a map of the British population locations. Branch lengths are proportional to genetic distance. Locations 43-mc and 44-mc are not represented as no ISSR profile was generated.

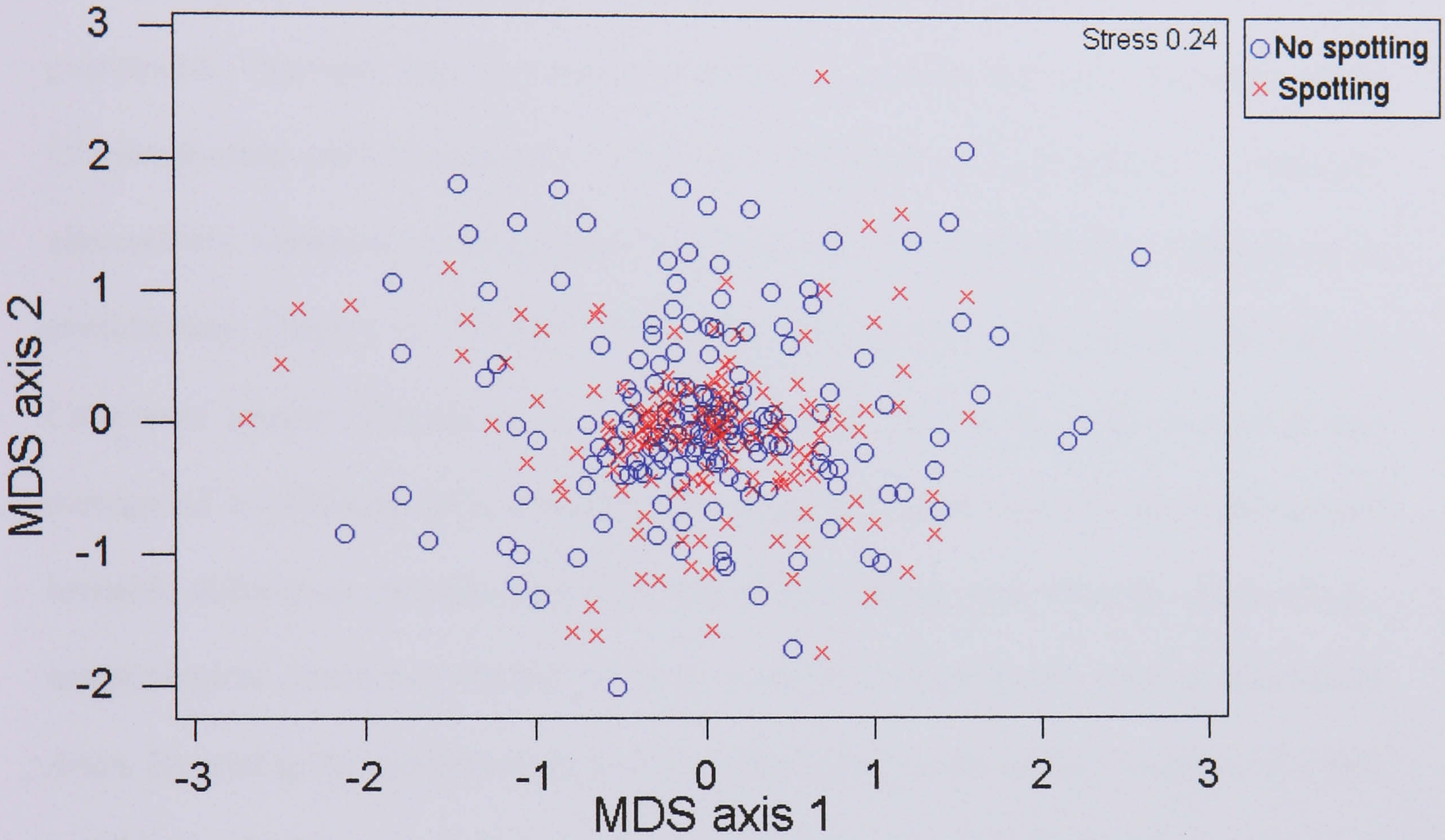


Figure 3.10 MDS ordination of the pairwise genetic distances among *A. maculatum* individuals from Britain sampled in the second, season summer 2003. The individuals are classed into non-spotted, indicated by the blue circles, and spotted, indicated by the red crosses.

Study 1: *A. maculatum* discussion

This chapter investigates the correlation between morphological and genetic marker variation within species. Since morphological variation can arise under different circumstances correlation between genetic and morphological variation is not guaranteed. Variation may be a result of plasticity, such as that seen in *Amblystegium* (Vanderpoorten and Jacquemart, 2004), and therefore no correlation is observed; alternatively variation of morphological characters can correspond to differences in evolutionary history, as has been found for inflorescence and leaf morphology in *Cardamini amara* (Lihova *et al.*, 2000). It is important to use only characters that correspond to differences in evolutionary history to define taxa, as these reflect real heritable differences between taxa. The aim of this chapter is to identify which, if any, morphological characters should be used to identify intraspecific taxa in species of *Arum*. By testing the correlation of morphological and genetic marker variation it is also possible to identify the population processes involved in producing patterns of genetic and morphological variation. In this section, within-species variation in *A. maculatum* is investigated. Throughout its range, *A. maculatum* has polymorphic leaf patterns and leaf shape (Harmes, 1982; Prime, 1960). Three questions were asked: 1. What morphological variation is there in *A. maculatum*? 2. What molecular marker variation is there in *A. maculatum*? 3. Do any morphological characters correlate with genetic lineages?

In Study A, the sampled populations of *A. maculatum* in Britain and France display considerable within- and among-population variation for leaf spotting and the lobe

characteristics length, width and angle. However, this variation is not correlated with geographical distance among the sampled locations as geographically close sampling sites are not necessarily morphologically similar. Other studies have shown no correlation between leaf variation and geographical distance, e.g. for *Myrceugenia fernandeziana* (Jensen *et al.*, 2002). The absence of a statistically significant correlation in *A. maculatum* populations suggests that the studied morphological characters are not responding to differences in environmental conditions.

It is known that morphological characters often change in response to environmental conditions, for example variation of leaf and flower morphology in *Nassella pulchra* (Knapp and Rice, 1998) and *Clarkia dudleyana* (Podolsky and Holtsford, 1995). The morphological variation observed in *A. maculatum* could be a response to environmental variation but it could also be plastic. There is currently a lack of information pertinent to the ecological and selective importance of lobe characters and spotting in *A. maculatum* to identify if the variation is plastic or a response to an environmental variable. A growing experiment similar to that of Donohue *et al.*, (2000) could be used to identify which morphological characters are plastic by crossing and growing subsequent progeny in a controlled environment. However *Arum* speices take between 4 to 5 years from sowing to flowering (Boyce 1993) thus this is not an experiment achievable in a Phd timescale.

Research outside the scope of this PhD is currently underway investigating spotting in *A. maculatum;* it is possible that spotting is an adaptive trait. Spotting in *A. maculatum* is caused by an anthocyanin like compound (A. Diaz pers. comm.) which could be

related to herbivore defence as, in *Eucalyptus* species, anthocyanin production is correlated with herbivore damage (Stone *et al*., 2001). However, in *A. maculatum* the spots are formed prior to leaf emergence and not change throughout the life of a plant. They are in therefore unlikely to be an induced response to herbivore attack as in *Eucalyptus*. An alternative explanation could be that leaf spotting is a sunscreen. It is known that in lines of *Zea mays* with low amounts of anthocyanin like compounds, photosynthesis genes are down regulated in high light (UV) environments compared to wild type plants (Casati and Walbot 2003) therefore, anthocyanins have a sunscreen affect. In *Quintinia serrata* anthocyanin distribution is associated with chlorophyllous cells (Gould *et al*., 2000) suggesting that these spots protect the sensitive chloroplasts from occasional but brief intense sun flecks (Gould *et al.,* 2000). A similar use of anthocyanin spotting could also be occurring in *A. maculatum* as it occurs in woody habitat where it is shaded by larger trees and shrubs and could experience occasional bright sunshine flecks.

The sampled population locations of *A. maculatum* sampled in both Part A and Part B show a distinct genetic structure that corresponds to geographical location. In the first study (Part A), population locations from Britain and France were significantly distinct. In the second study of British locations (Part B) groups corresponding to the east and west were found. In both Part A and Part B there was a statistically significant relationship between genetic distance and geographical distance. A significant relationship between geographical and genetic distances is considered to be an indicator of isolation by distance, in which increasing distance between sampled locations

decreases the frequency of gene flow. Genetic patterns observed in *A. maculatum* could be influenced by pollinator movement and seed dispersal.

Pollinator movement and seed dispersal of *Iris haynei* and *I. atrofusca* may have caused a significant relationship between geographical and genetic distances (Arafeh *et al.*, 2002). The geographical structuring of the *Iris* populations is probably caused by a combination of limited pollen dispersal by insects and seed dispersal by gravity and ants (Arafeh *et al.*, 2002). A combination of topography and pollinator behaviour could also account for the genetic structure of *Phyllodoce glandiflora* populations (Rochefort and Peterson, 2001). Populations of *P. glandiflora* are isolated by glaciers and forests which potentially limit pollinator dispersal, therefore reducing the amount of long-distance gene flow (Rochefort and Peterson, 2001). However, if dispersal occurs over a large scale populations can appear less differentiated, and this has been observed in Scandinavian populations of *Potamogeton pectinatus*. Dispersal of *Potamogeton* seeds in migrating wildfowl has resulted in reduced genetic differentiation in populations where there are large migrations as compared with areas where there is little bird migration (King *et al.*, 2002).

The genetic relationships among the sampled locations of *A. maculatum* could be a result of limited long-distance gene flow. It is known that *A. maculatum* is pollinated by flies from the genus *Psychoda* (Diaz and Kite, 2002) and that the seeds are dispersed by gravity and birds (A. Diaz, pers. comm.). The genetic structure observed in both parts A and B indicate that long-distance gene flow in *A. maculatum* is limited. The English Channel could be acting as a barrier to gene flow, as with glaciers and mountains in

Phylodoce (Rochefort and Peterson, 2001), and mountains in *Eriastrum* (Brunell and Whitkus, 1997). The behaviour of pollinator and seed dispersal vectors will also affect the genetic structure of the populations. This is a topic for further research to quantify the amount of seed that is dispersed by birds and by gravity, and how patterns of dispersal may affect *A. maculatum* populations.

Variation of lobe length and angle did not correlate with genetic marker variation in *A. maculatum*. A lack of correlation between morphological and genetic markers has been found in several species, for example leaf shape variation in *Begonia* (Matolweni *et al.*, 2000), *Myrceugenia fernandeziana* (Jensen *et al.*, 2002), and *Eriastrum densifolium* (Brunell and Whitkus, 1997). Lobe width variation in *A. maculatum* correlated with molecular marker variation. As previously indicated, the ecological significance of the morphological variation in *A. maculatum* is currently unknown. Research outside of this PhD is currently investigating the importance of the morphological variation.

In Part A the data showed that spotting was correlated with genetic marker variation, with spotted leaf individuals being more closely related than individuals with no spotting. However, most of the spotted plants occurred in France and the non-spotted plants in Britain, and these populations from either side of the English Channel have been shown to be genetically distinct. Therefore, the statistically significant relationship seen in *A. maculatum* Part A could be an artefact of sample distribution. The second study of British populations (Part B) had the sole aim to examine if spotted individuals were genetically distinct from non-spotted individuals. It was shown that spotted individuals were not genetically distinct from non-spotted individuals. Consequently,

the spotting variation cannot be used to define genetic groups. Spotting is therefore not a good indicator of evolutionary history in *A. maculatum*.

Study 2: *A. italicum*

The previous section investigated the relationship of morphology and genetics in *A. maculatum*, and found no evidence of subspecific taxa. This section investigates variation in two subspecies of *A. italicum*. There are four subspecies of *A. italicum*, two of which (*A. italicum* subsp. *canariense* and subsp. *albispathum*) are differentiated on the basis of inflorescence characters and geographical distribution. The remaining two subspecies, *A. italicum* subsp. *neglectum*, and *A. italicum* subsp. *italicum*, have overlapping ranges (Figure 3.3), identical inflorescences (Figure 3.5) and are only differentiated by leaf morphology (Figure 3.4). Correlation of morphological and genetic marker variation within populations of *A. italicum* subsp. *italicum* and *A. italicum* subsp. *neglectum* is investigated within this section and genetic support for the subspecies is quantified.

Three questions are investigated: 1. What morphological variation is there within the two *A. italicum* subspecies *italicum* and *neglectum*? 2. What genetic marker variation is there within these two *A. italicum* subspecies over the range surveyed? 3. Do any morphological characters correlate with genetic lineages?

A. italicum: Methods

Plant collection

The taxonomy of Boyce (1989) distinguishes between the two *A. italicum* subspecies based on leaf patterning and shape. Subspecies *italicum* leaves are patterned (lines or blotches) with narrow posterior lobes whereas subsp. *neglectum* leaves are unpatterned with broad posterior lobes. Of the taxonomic differences, leaf patterning is the most visually obvious character, and so to simplify identification in the field only leaf patterning was used to assign plants to one of the two *A. italicum* subspecies. Using leaf patterning removes any bias/errors that may arise through measuring leaf shape. By using leaf pattern, leaf shape characters can later be tested with no *a priori* assumptions.

Samples from population locations of both subspecies of *A. italicum* were sampled throughout southern Britain, France and northern Spain. Three types of *A. italicum* locations were defined: Type 1 contained only subsp. *italicum*; Type 2 contained only subsp. *neglectum* and Type 3 contained both subspecies (for locations see Figures 3.11 and 3.12; Tables 3.8 and 3.9). Ten randomly selected individual plants at least 1m apart were sampled from each sampled location: in total 690 plants were collected. Mature leaves with no obvious disease or damage were then scored for lobe variation: the left lobe (adaxial surface uppermost) was measured for length (LL), width (LW) and the angle between the midrib and the main vein on the lobe (LA). The amount of blotching and lines was scored as a percentage of leaf area using photographs of previously

calibrated standard leaves. After the leaf characteristics were recorded, the leaf was rinsed in ethanol and approximately $3cm^2$ ripped off, avoiding the edge and midrib, and stored in a sealable plastic bag containing fine silica gel (pore diameter ca 6nm).

DNA was extracted using a semi-automated method from Qiagen. DNA variation was surveyed using PCR amplification of ISSRs. The amplified DNA fragments were sorted by size on polyacrylamide gels and visualised by silver staining. A detailed method for DNA extraction is presented in the methods section earlier in this chapter; PCR amplification, fragment visualisation and scoring methods are presented in Chapter 2.

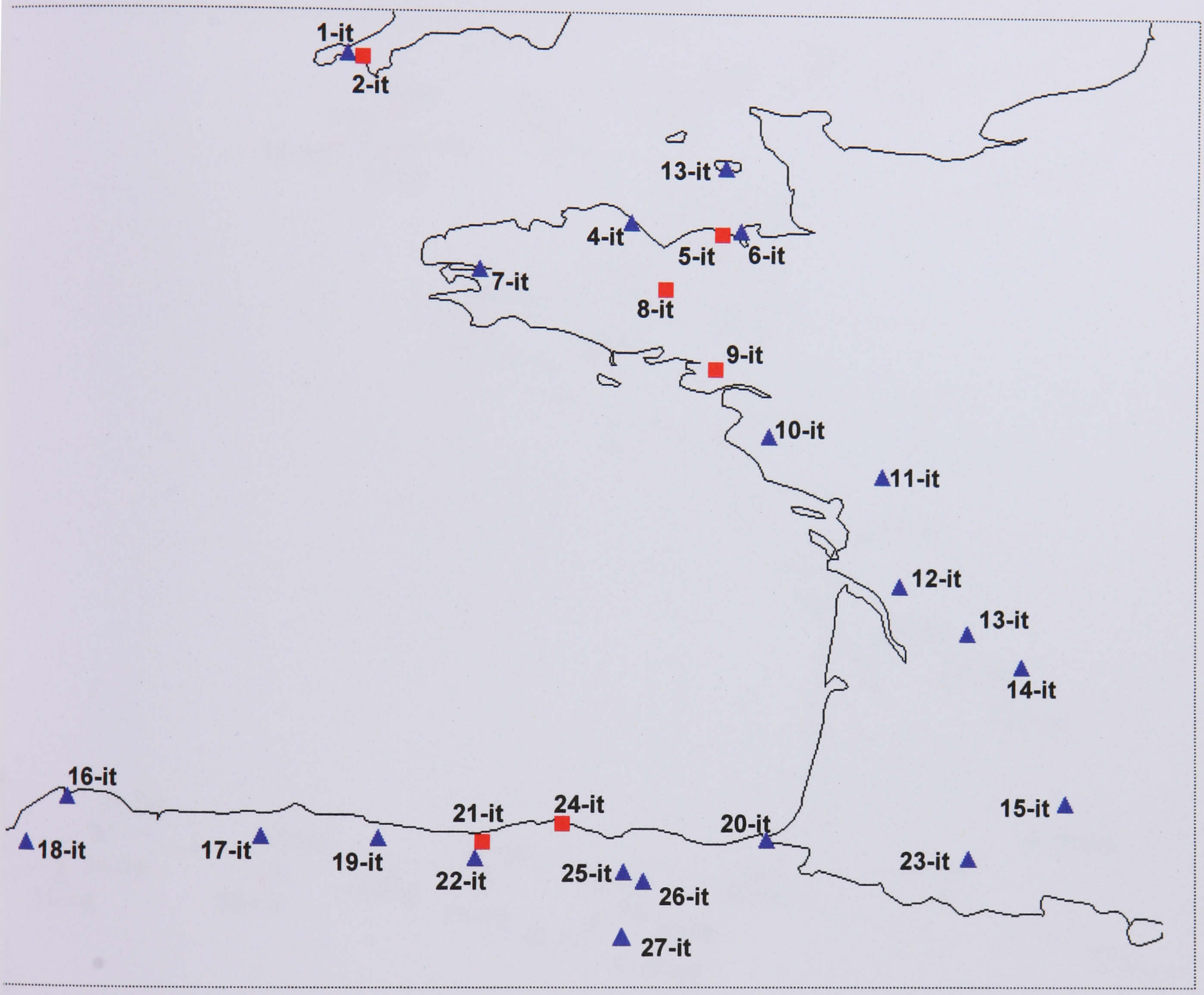


Figure 3.11 Locations of *A. italicum* subsp. *italicum* samples. Type 1 locations (only subsp. *italicum*) are indicated by red squares and Type 3 locations (both subspecies) by blue triangles.

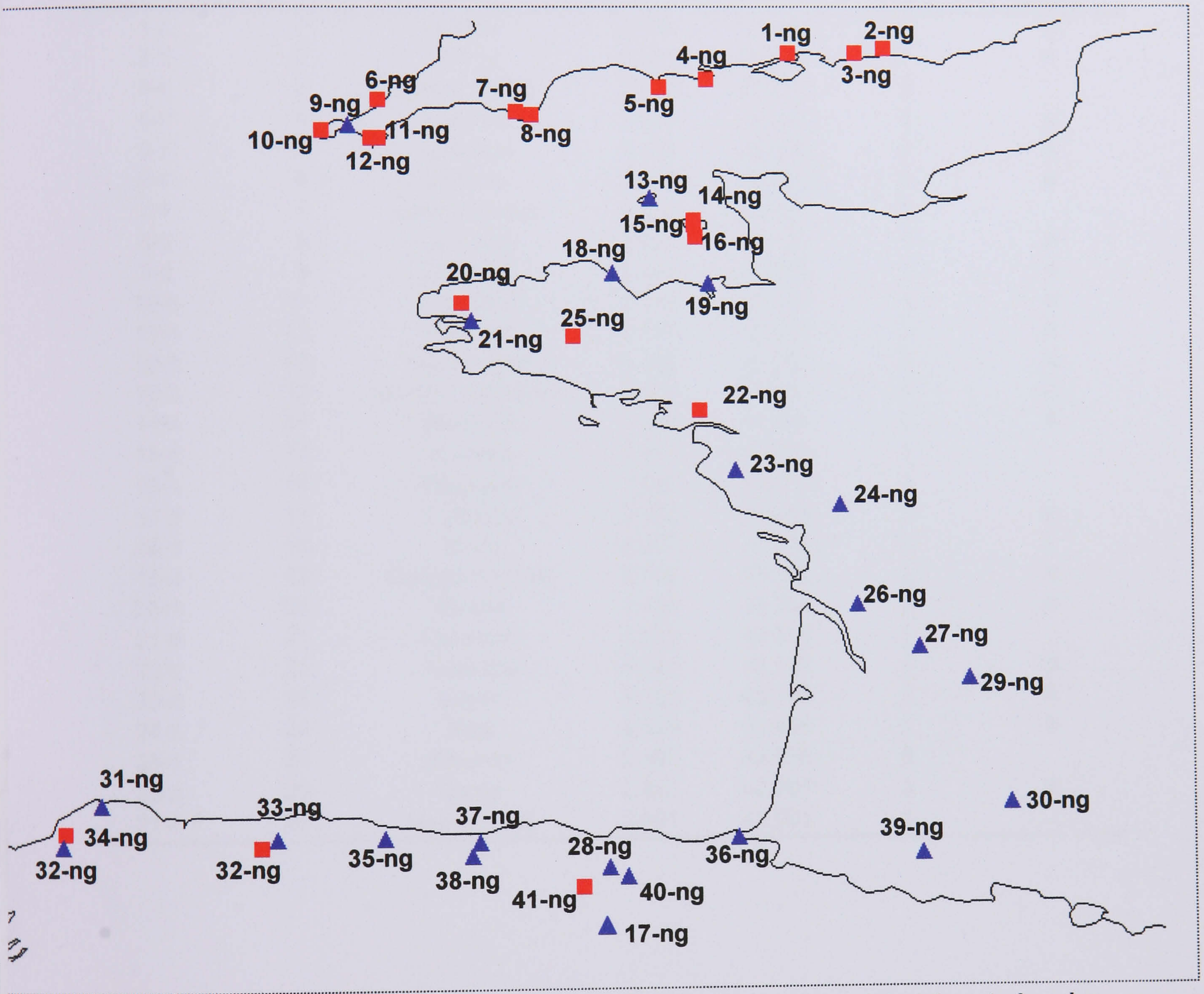


Figure 3.12 Locations of *A. italicum* subsp. *neglectum* samples. Type 2 locations (only subsp. *neglectum*) are indicated by red squares and Type 3 locations (both subspecies) by blue triangles.

Table 3.8 The location of *A. italicum* subsp. *italicum* samples. For each, the longitude and latitude are specified. The location type is either Type 1, containing only *A. italicum* subsp. *italicum*, or Type 3, containing both *A. italicum* subspecies. Black squares in the column "Successful ISSR Profile" indicate reliable amplified products were obtained for that location.

Location code	Location number	Location	Longitude (°W)	Latitude (°N)	Population type	Successful ISSR profile
1-it	1	Lelant	5.430	50.160	3	■
2-it	2	Sithny	5.289	50.123	1	■
3-it	3	Jersey South	2.139	49.202	3	
4-it	4	Lanloup	2.964	48.711	3	■
5-it	5	Lancieux	2.156	48.606	1	■
6-it	6	St Malo	2.010	48.651	3	■
7-it	7	Landevernnec	4.275	48.297	3	
8-it	8	La Cheze	2.657	48.132	1	■
9-it	9	Quebrite	2.223	47.436	1	■
10-it	10	Froidfond	1.751	46.867	3	■
11-it	11	Mervant	0.757	46.523	3	■
12-it	12	Machennes	0.590	45.563	3	■
13-it	13	La Roche Chalais	-0.011	45.151	3	■
14-it	14	Bergerac	-0.483	44.856	3	■
15-it	15	St Mere	-0.874	43.661	3	
16-it	16	Ortigueira	7.836	43.685	3	
17-it	17	Cutiellos	6.153	43.359	3	■
18-it	18	Souto	8.207	43.300	3	■
19-it	19	Cangas de Onis	5.132	43.351	3	■
20-i1	20	Biriatu	1.753	43.333	3	■
21-it	21	Cabezon	4.229	43.310	1	
22-it	22	Fresneda	4.299	43.181	3	■
23-it	23	Louey	-0.023	43.179	3	■
24-it	24	Noja	3.526	43.466	1	■
25-it	25	Amurrio	2.998	43.056	3	
26-it	26	Sarria	2.825	42.967	3	■
27-it	27	Herramelluri	3.021	42.501	1	

Table 3.9 The location of sampled of *A. italicum* subsp. *neglectum*. For each the longitude and latitude are specified. The location type is either Type 2, containing only *A. italicum* subsp. *neglectum*, or location Type 3, containing both *A. italicum* subspecies. Black squares in the column "Successful ISSR Profile" indicate that reliable amplified products were obtained for that location.

Location code	Location number	Location	Longitude (°W)	Latitude (°N)	Population type	Successful ISSR profile
1-ng	28	East Meon	1.220	50.840	2	■
2-ng	29	Steyning	0.300	50.880	2	
3-ng	30	Arundel	0.580	50.840	2	■
4-ng	31	Dancing	2.005	50.594	2	■
5-ng	32	Portland	2.460	50.520	2	■
6-ng	33	Treago	5.133	50.400	2	■
7-ng	34	South Milton	3.830	50.290	2	■
8-ng	35	Beeson	3.684	50.248	2	
9-ng	36	Lelant	5.432	50.164	3	■
10-ng	37	Penzance	5.680	50.115	2	■
11-ng	38	Trelanven	5.133	50.033	2	■
12-ng	39	Mullion	5.213	50.036	2	■
13-ng	40	Guernsey	2.553	49.469	2	■
14-ng	41	Jersey South	2.139	49.202	2	■
15-ng	42	Jersey Middle	2.117	49.083	2	■
16-ng	43	Jersey North	2.136	49.255	2	■
17-ng	44	Herramelluri	3.021	42.501	2	■
18-ng	45	Lanloup	2.919	48.748	3	■
19-ng	46	St Malo	2.010	48.651	3	■
20-ng	47	St Thuron	4.367	48.467	2	■
21-ng	48	Landevernnec	4.275	48.297	3	■
22-ng	49	Pontchateau	2.089	47.431	2	■
23-ng	50	Froidfond	1.751	46.867	3	■
24-ng	51	Mervant	0.757	46.523	3	■
25-ng	52	Loucon	3.304	48.150	2	■
26-ng	53	Machennes	0.590	45.563	3	■
27-ng	54	La Roche-Chalais	-0.011	45.151	3	■
29-ng	55	Bergerac	-0.483	44.856	3	■
30-ng	56	St Marie	-0.874	43.661	3	■
31-ng	57	Origueira	7.836	43.685	3	■
32-ng	58	Breives	6.300	43.270	2	
33-ng	59	Cutiellos	6.153	43.359	3	
34-ng	60	Pontedeume	8.175	43.409	2	■
35-ng	61	Cangas de Onis	5.132	43.351	3	
36-ng	62	Biriatu	1.753	43.333	3	
37-ng	63	Cabazon	4.229	43.310	3	■
38-ng	64	Fresneda	4.299	43.181	3	■
39-ng	65	Louey	-0.023	43.179	3	■
28-ng	66	Amurrio	2.998	43.056	3	■
40-ng	67	Sarria	2.825	42.967	3	
41-ng	68	Pena	3.244	42.874	2	
42-ng	69	Souto	8.207	43.300	3	

A. italicum: Results

Morphological variation in *A. italicum*

Morphological differences between sampled population locations of subsp. italicum and subsp. neglectum

To identify the extent of intraspecific morphological variation in *A. italicum,* the three continuous leaf characters: lobe length (LL), lobe width (LW), and lobe angle (LA), were measured. The variation of each lobe character is plotted on a series of box plots (Figures 3.13–3.15). Only the sampling locations with successful ISSR profiles are included in these box plots and subsequent morphological analysis. The box plots indicate the range of variation within a site and also allow the comparison of variation between sampling sites. Variation within a site is indicated by the distance between the lower and upper horizontal bars, representing the range of data between the 25th and the 75th percentiles. Comparison of the range of data and medians (middle horizontal bar) gives an indication of the variation between locations. Figures 3.13–3.15 illustrate the non-uniformity of the three lobe characters (LL, LW and LA) in both *A. italicum* subspecies. The characters vary between and within sampled locations, with some being very variable and others less so. Variation of all lobe characters differs significantly from normal. As the variation of the characters is significantly different from normal, relationships between the characters were investigated using Spearman's rank correlation. Trends between groups were investigated using the non-parametric Mann–Whitney test.

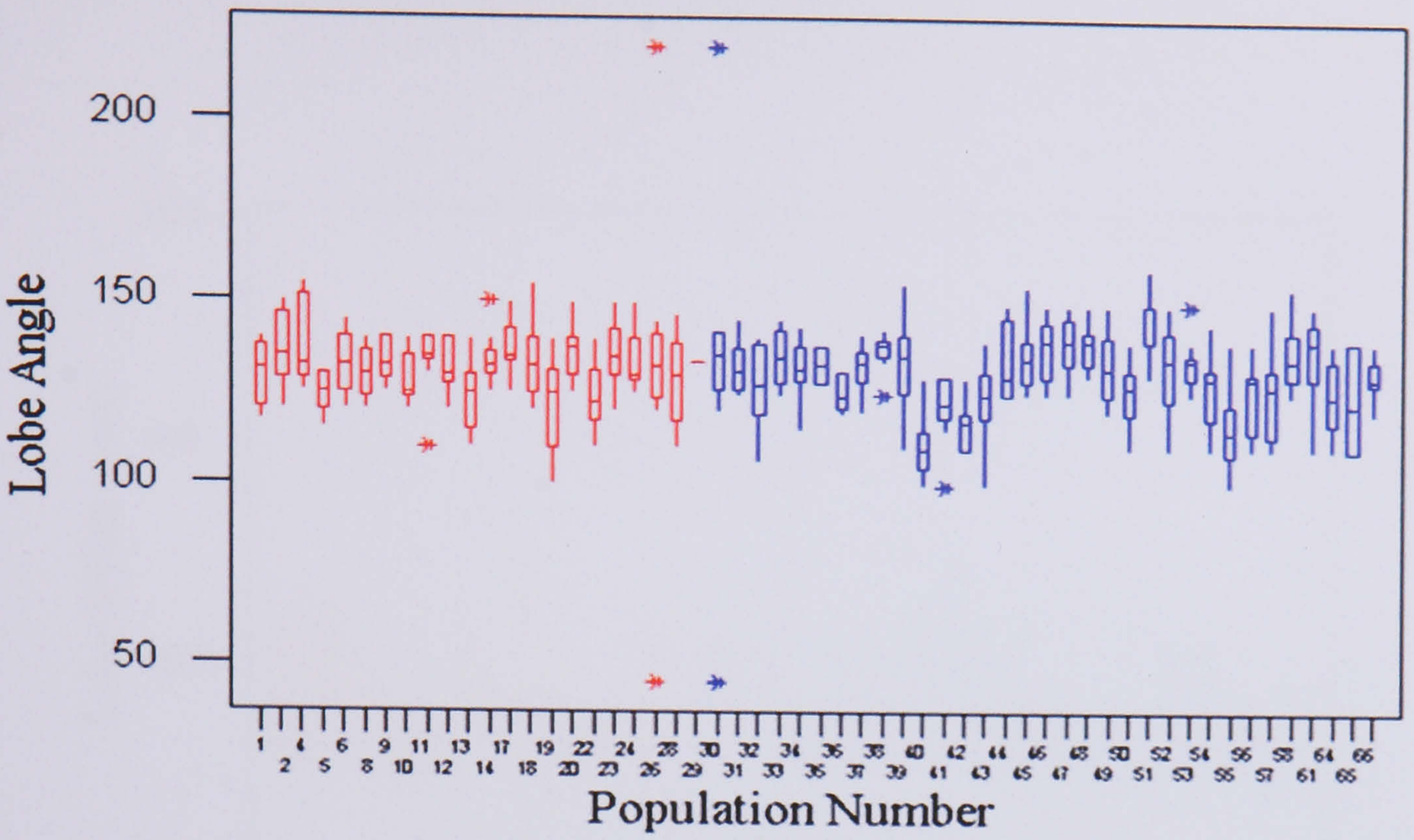


Figure 3.13 Box plots showing within- and among-location variation for leaf lobe angle (degrees). Asterisks (*) indicate outlying data points. Red populations are those assigned *a priori* as *A. italicum* subsp. *italicum* and blue those populations assigned *a priori* as *A. italicum* subsp. *neglectum*. Population location numbers are given in Tables 3.8 and 3.9.

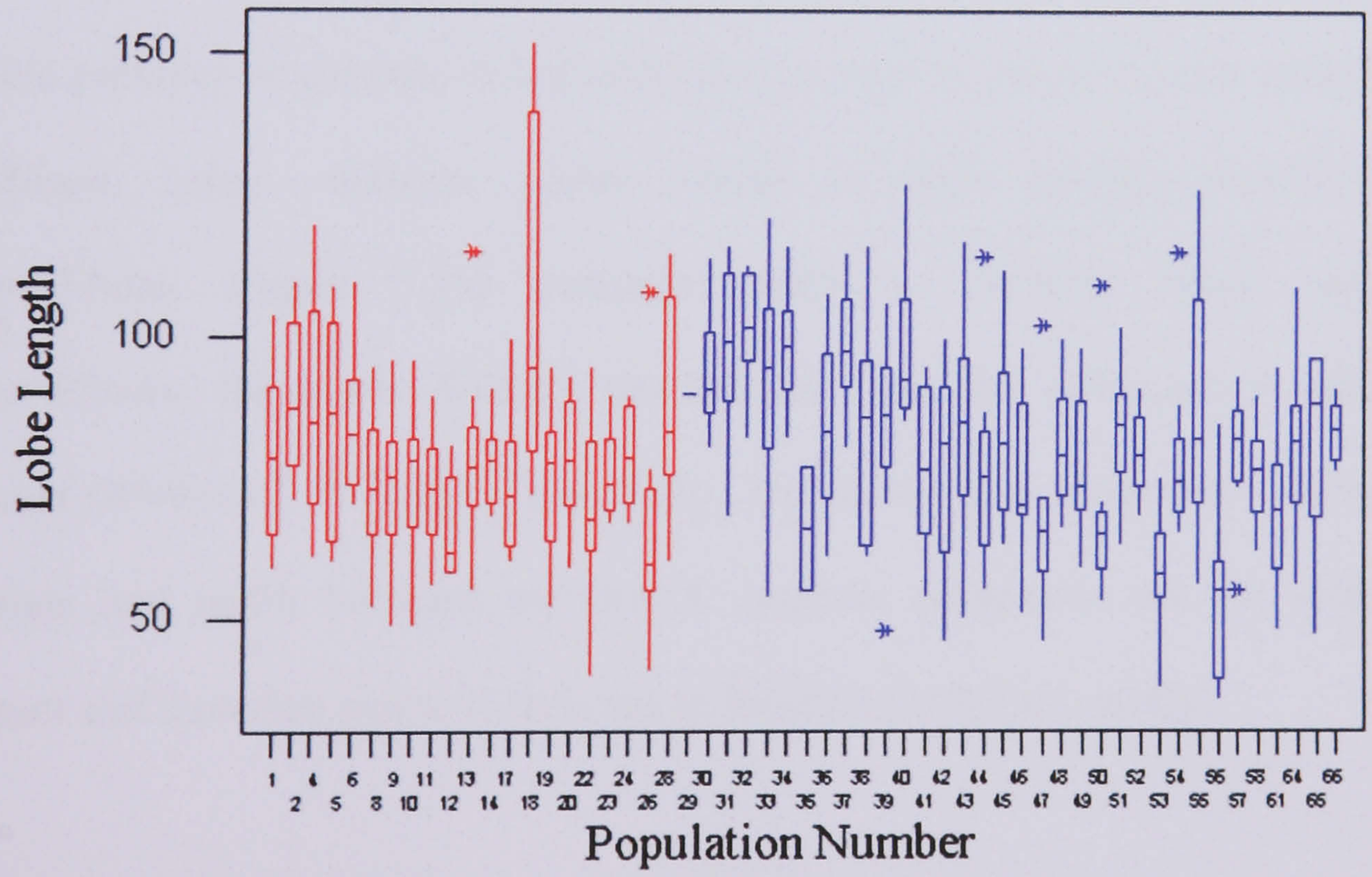


Figure 3.14 Box plots showing within- and among-location variation for leaf lobe length (mm). Blue populations are those assigned *a priori* as *A. italicum* subsp. *italicum* and red those locations assigned *a priori* as *A. italicum* subsp. *neglectum*. Asterisks (*) indicate outliers. Population location numbers are given in Tables 3.8 and 3.9.

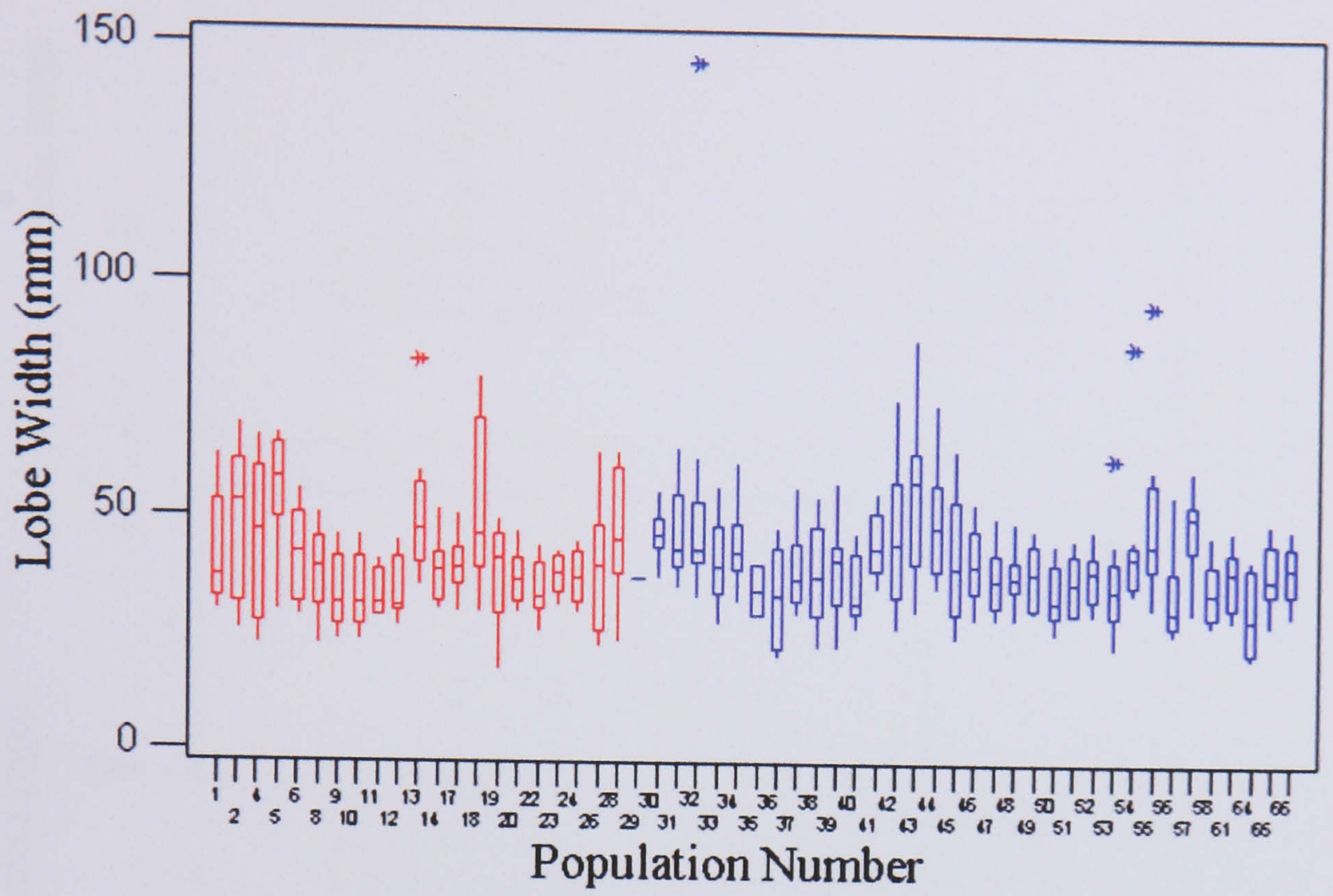


Figure 3.15 Box plots showing within- and among-location variation for leaf lobe width (mm). Red locations are those assigned *a priori* as *A. italicum* subsp. *italicum* and blue those locations assigned *a priori* as *A. italicum* subsp. *neglectum*. Asterisks (*) indicate outliers. Population location numbers are given in Tables 3.8 and 3.9.

When the presence or absence of leaf patterning is used to define the two subspecies of *A. italicum*, subsp. *italicum* plants tended to have shorter posterior lobes (median=77mm, Figure 3.16) compared with *A. italicum* subsp. *neglectum* (median=82mm); the Mann–Whitney test indicated that this difference is statistically significant (W=92615, P=0.003, Table 3.10). The results also show that differences in lobe angle and width between the two *A. italicum* subspecies are not statistically significant and therefore question their use in Boyce's (1993) taxonomy.

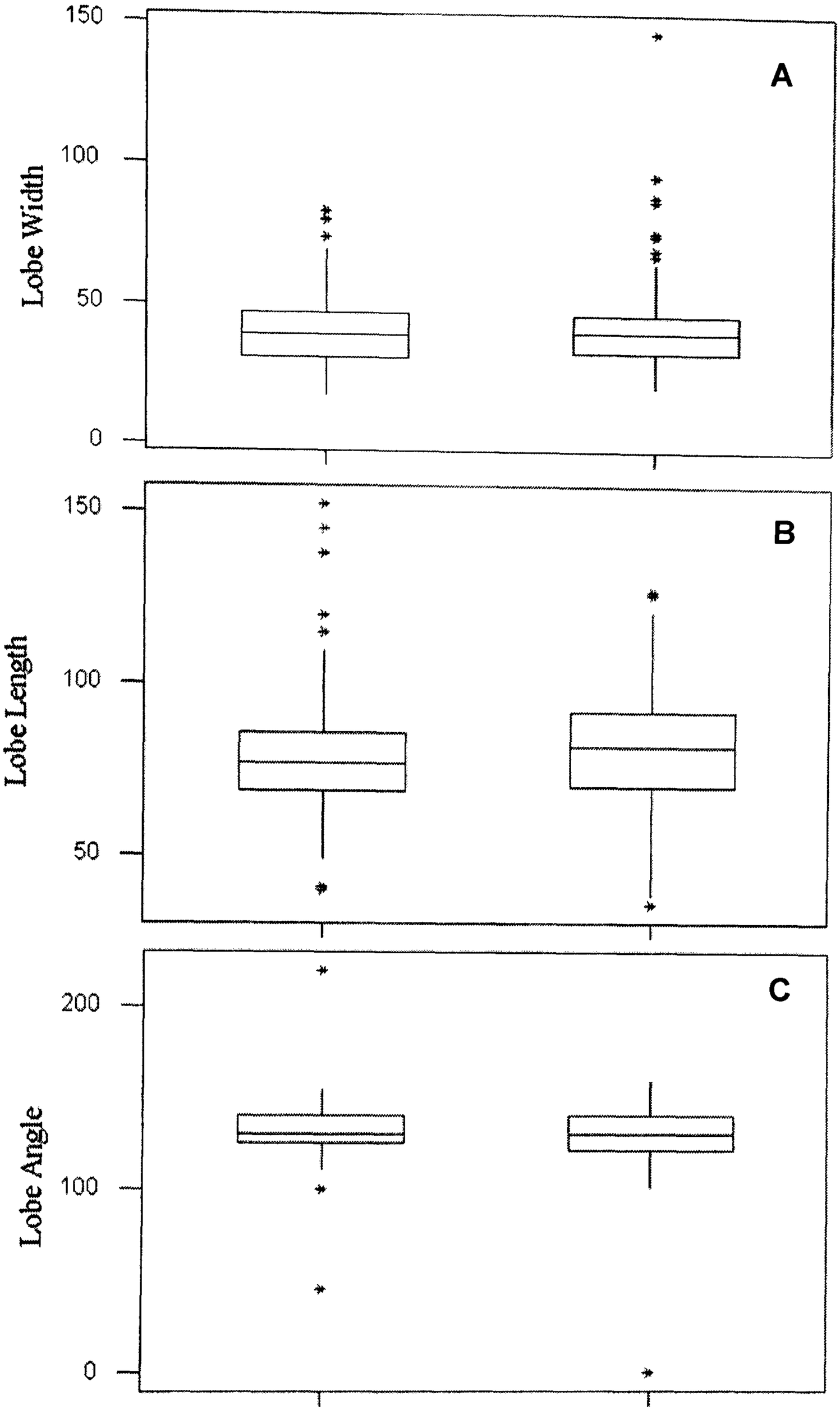


Figure 3.16 Box plots indicating the variation within, and the differences between, the two subspecies *A. italicum* subsp. *italicum* and *A. italicum* subsp. *neglectum*. Difference between and total variation in the subspecies of: (A) width of the posterior leaf lobe (mm); (B) length of the posterior leaf lobe (mm); (C) angle of the posterior leaf lobe (degrees). The distances of the lower and upper percentiles (lower and upper limit of each box) from the median value (horizontal line) indicate within-subspecies variation. Asterisks (*) indicate outlying data points.

Table 3.10 Mann–Whitney test results. Comparisons of the rank medians of lobe angle, lobe length and lobe width between the two subspecies *A. italicum* subsp. *italicum* and *A. italicum* subsp. *neglectum*. *W* is the Mann–Whitney test statistic. *P* indicates the significance level. A *P*-value greater than 0.05 is non-significant and is indicated by the annotation n.s. *P* (adjusted for ties) indicates the adjusted *P*-value discounting tied ranks.

	Subspecies	*n*	Median	*W*	*P*	*P* (adjusted for ties)
Lobe width	*italicum*	200	39.0	51241.5	0.140 n.s.	0.276 n.s.
	neglectum	340	40.0			
Lobe length	*italicum*	200	77	92615.5	0.003	0.003
	neglectum	340	82			
Lobe angle	*Italicum*	200	130	85112.0	0.143 n.s.	0.143 n.s.
	neglectum	340	130			

Morphological differences between allopatric and sympatric population locations

The composition of the sampled locations differs, some containing a single subspecies and others containing both subspecies. The locations have been classified depending upon whether they contain a single subspecies or both: those that contain only subsp. *italicum* are called Type 1, when only subsp. *neglectum* is present the location is termed Type 2, whilst locations containing both subspecies are called Type 3.

Two Mann–Whitney tests were used to investigate the significance of any differences in the morphology of individuals growing in allopatric locations (Type 1 or 2) compared with individuals growing in sympatric locations (Type 3). The first Mann–Whitney test compares all Type 1 and 2 individuals with corresponding subspecies in Type 3 locations. There is, however, an inherent problem with this test as the distribution of the allopatric and sympatric locations is uneven. To reduce the effect of uneven population location distribution a second Mann–Whitney test was carried out. This second test

compares only the geographically nearest locations, thus removing any bias of geographical location. A statistically significant result in the first test would indicate that there is a difference in morphology between individuals growing in sympatric and allopatric sites. If the second test is not statistically significant, the significant difference in the first test can be accounted for by the uneven sampling distribution. However, a statistically significant result in the second test would indicate that the difference between the locations is not as a result of the uneven distribution but a result of the subspecies growing sympatrically.

A. italicum subsp. italicum

In *A. italicum* subsp. *italicum* there are no significant differences in lobe morphology width, length and angle (LL, LW and LA) between allopatric (Type 1) of subsp. *italicum* and subsp. *italicum* individuals growing in sympatry with subsp. *neglectum* (Type 3, Table 3.11). However, the percentage of leaf area covered by white lines and blotches is statistically significantly larger in allopatric compared with sympatric locations (Table 3.11). The second Mann–Whitney test uses the subset of five allopatric locations and the five nearest sympatric locations. There were no statistically significant differences between sympatric and allopatric locations for any character using the second dataset (Table 3.12). This non-significant second result indicates that the uneven distribution of sampled locations was causing the significant differences in leaf lines and blotches. Therefore, the overall conclusion is that when subsp. *italicum* plants grow in sympatry with subsp. *neglectum* plants, the morphology (lobe characters and leaf

patterning) is not statistically significantly different from individuals growing in allopatry.

Table 3.11 Mann–Whitney test results. Comparison of leaf characters between all allopatric *A. italicum* subsp. *italicum* individuals (Type 1) and all *A. italicum* subsp. *italicum* individuals from sympatric Type 3 locations. *W* is the Mann–Whitney test statistic, and *P* indicates the significance level. A *P*-value greater than 0.05 is non-significant and is indicated by the annotation n.s. *P* (adjusted for ties) is the *P*-value discounting tied ranks.

	A. italicum subsp. *italicum*					
	Location type	***n***	**Median**	***W***	***P***	***P* (adjusted for ties)**
Lobe width	Allopatric Sympatric	50 150	39.00 38.00	5298.5	0.4412 n.s.	0.4409 n.s.
Lobe length	Allopatric Sympatric	50 150	79.00 76.50	5422.0	0.2633 n.s.	0.2630 n.s.
Lobe angle	Allopatric Sympatric	50 150	130.00 130.00	4807.5	0.5404 n.s.	0.5342 n.s.
Line	Allopatric Sympatric	50 150	70.00 30.00	6431.0	<0.001	<0.001
Blotch	Allopatric Sympatric	50 150	0.00 0.00	42550.0	0.0299	0.0081

Table 3.12 Mann–Whitney test results. Comparison of leaf characters between the five allopatric *A. italicum* subsp. *italicum* locations (Type 1) and the five nearest sympatric (Type 3) *A. italicum* subsp. *italicum* populations. *W* is the Mann–Whitney test statistic, and *P* indicates the significance level. A *P*-value greater than 0.05 is non significant and is indicated by the annotation n.s. *P* (adjusted for ties) is the *P*-value discounting tied ranks.

	A. italicum subsp. *italicum*					
	Locations type	***n***	**Median**	***W***	***P***	***P* (adjusted for ties)**
Lobe width	Allopatric Nearest sympatric	50 50	39.50 36.00	2669.5	0.2385 n.s.	0.2382 n.s.
Lobe length	Allopatric Nearest sympatric	50 50	80.00 78.50	2578	0.7148 n.s.	0.7147 n.s.
Lobe angle	Allopatric Nearest sympatric	50 50	130.00 130.00	25	0.6761 n.s.	No ties
Line	Allopatric Nearest sympatric	50 50	70.00 52.50	2583.5	0.7148 n.s.	0.7147 n.s.
Blotch	Allopatric Nearest sympatric	50 50	0.00 0.00	2444.5	0.5813 n.s	0.4103 n.s

A. italicum subsp. neglectum

In *A. italicum* subsp. *neglectum* there are significant differences in the lobe morphology (LL and LW) between individuals growing in allopatric locations (Type 2) compared with individuals growing in locations with subsp. *italicum* (Type 3, Table 3.13). Individuals of subsp. *neglectum* in allopatric locations (Type 2) have wider lobes (median=39.3mm) than subsp. *neglectum* individuals from Type 3 locations (median=37.25mm). This difference is statistically significant (P=0.0164, Table 3.13). The lobe length of subsp. *neglectum* in allopatric locations (Type 2) is statistically significantly smaller (median=36mm) than that of subsp. *neglectum* individuals from sympatric (Type 3) locations (median=76mm, Table 3.13). However, there are no statistically significant differences in the lobe angle between subsp. *neglectum* individuals from allopatric locations and sympatric locations. The geographically restricted subset of 13 allopatric locations compared with the closest 13 sympatric locations are also significantly different for lobe width and length (Table 3.14). The statistically significant difference observed between the sympatric and allopatric locations of subsp. *neglectum* for LW and LL are therefore not a result of uneven sampling distribution as the differences remain statistically significant when allopatric locations are compared with the nearest sympatric locations (Table 3.14). Therefore, the lobe morphology of subsp. *neglectum* plants growing in sympatry with subsp. *italicum* is significantly different from that of subsp. *neglectum* individuals growing allopatrically, when the geographically restricted subset of locations is examined.

Table 3.13 Mann–Whitney test results. Comparison of leaf characters between all allopatric *A. italicum* subsp. *neglectum* (Type 2) and all *A. italicum* subsp. *neglectum* individuals from Type 3 locations. *W* is the Mann–Whitney test statistic, and *P* indicates the significance level. A *P*-value greater than 0.05 is non-significant and is indicated by the annotation n.s. *P* (adjusted for ties) is the *P*-value discounting tied ranks.

	A. italicum subsp *neglectum*					
	Location type	***n***	**Median**	***W***	***P***	***P* (adjusted for ties)**
Lobe Width	Allopatric	180	42.00	17213.0	<0.001	<0.001
	Sympatric	130	37.50			
Lobe Length	Allopatric	180	86.00	16678.0	<0.001	<0.001
	Sympatric	130	76.00			
Lobe Angle	Allopatric	180	130.00	212770.0	0.1729 n.s.	0.1710 n.s.
	Sympatric	130	130.00			

Table 3.14 Mann–Whitney test results. Comparison of leaf characters between the 13 allopatric *A. italicum* subsp. *neglectum* (Type 2) and *A. italicum* subsp. *neglectum* individuals from the 13 nearest Type 3 locations. *W* is the Mann–Whitney test statistic, and *P* indicates the significance level. A *P*-value greater than 0.05 is non-significant as indicated by the annotation n.s. *P* (adjusted for ties) is the *P*-value discounting tied ranks.

	A. italicum subsp. *neglectum*					
	Location type	***n***	**Median**	***W***	***P***	***P* (adjusted for ties)**
Lobe width	Allopatric	130	37.5	14892.5	<0.001	<0.001
	Nearest sympatric	130	42.0			
Lobe length	Allopatric	130	88.0	14038.0	<0.001	<0.001
	Nearest sympatric	130	76.0			
Lobe angle	Allopatric	130	131.0	17090.0	0.8373 ns	0.8366 n.s.
	Nearest sympatric	130	130.0			

As there are no significant differences between the morphology of *A. italicum* subsp. *italicum* from allopatric and sympatric locations, no distinction will be made between allopatric and sympatric locations in subsequent analyses. The significant difference between sympatric and allopatric locations of subsp. *neglectum* shows that the morphology of subsp. *neglectum* growing in sympatry with subsp. *italicum* differs from that of allopatric subsp. *neglectum* locations. Therefore, all further analyses involving subsp. *neglectum* will be split according to the location type.

Morphological character intercorrelations and spatial patterns in A. italicum subsp. italicum

The morphological characters were tested for intercorrelation and for clinal variation in the two *A. italicum* subspecies. In locations of subsp. *italicum* significant correlations exist between lobe length and width (r=0.568, P <0.001), lobe angle and length (r=0.173, P<0.001), and lobe angle and width (r=−0.138, P<0.001, Table 3.15). Each morphological character correlated with either longitude, latitude or both (Table 3.15); for example, lobe angle correlated significantly with both latitude (r=0.114, P=0.003) and longitude (r=0.193, P<0.001) and therefore suggests that the characters are clinally variable.

Table 3.15 Spearman's correlation matrix of three leaf characters, latitude and longitude for all tested individuals of *A. italicum* subsp. *italicum*. The upper number indicates the correlation value whilst the lower number indicates the *P*-value. A *P*-value greater than 0.05 indicates a non-significant correlation; significant correlations (*P*>0.05) are outlined in black. Degrees of freedom for this test are 199.

	Longitude	Latitude	Lobe angle	Lobe length	Lobe width	% Line
Lobe angle	0.155	−0.035				
	<0.001	0.419				
Lobe length	0.121	−0.282	−0.163			
	0.005	<0.001	<0.001			
Lobe width	−0.043	−0.104	−0.136	0.523		
	0.323	0.016	<0.001	<0.001		
% Line	0.080	0.066	0.006	0.052	0.024	
	0.065	0.126	0.886	0.229	0.583	
% Blotching	0.006	0.186	0.046	0.066	0.047	0.162
	0.886	<0.001	0.294	0.130	0.274	<0.001

The correlation analysis is not ideally suited for investigating a relationship between spatial data and morphological characters because the tested locations are distributed unevenly. To establish if there is a statistically significant relationship between geographical location and morphological variation a Mantel test was used. The results indicate that there is no statistically significant correlation between morphological and genetic marker variation in locations of *A. italicum* subsp. *italicum* (n=20, r=0.066, P=0.216). However, when the characters are analysed separately the Mantel test results indicate that there is a significant correlation between lobe length variation and geographical distance among the populations (n=20, r=0.994, P=0.005). Percentage of lines was also significantly related to geographical distance (n=20, r=0.205, P=0.019). Lobe width, lobe angle and blotching were not significantly correlated with geographical distance (r=0.117, P=0.119; r=0.066, P=0.776 and r=0.018, P=0.364 respectively).

Morphological character intercorrelations and spatial patterns in A. italicum subsp. neglectum

As the morphology of subsp. *neglectum* differed between allopatric (Type 2) and sympatric (Type 3) populations, the intercorrelations were carried out separately and significant differences were found between the two location types (Table 3.16). In allopatric locations, lobe angle, length and width were all significantly intercorrelated. Lobe angle and lobe length correlated significantly with longitude, whilst lobe length was the only character to correlate significantly with latitude. For sympatric locations, however, no significant correlations were found between any lobe character and latitude (Table 3.16) whereas lobe length correlated significantly with longitude. The significant correlations of morphological characters with latitude or longitude must be treated with caution as the locations surveyed are not evenly geographically distributed. Furthermore, Mantel tests (Table 3.17) provide no support for spatial structuring of any lobe character in either Type 2 or Type 3 locations.

Table 3.16 Spearman's correlation matrix of three leaf morphological characters, latitude and longitude for *A. italicum* subsp. *neglectum* individuals in Type 2 populations and *A. italicum* subsp. *neglectum* individuals in Type 3 populations. Upper number: correlation value, lower number: *P*-value. Significant correlations (*P*>0.05, 95% confidence interval) are outlined in black. The degrees of freedom are 179 for Type 2 populations and 129 for Type 3 populations.

	A. italicum subsp. *neglectum* Type 2				*A. italicum* subsp. *neglectum* Type 3			
	Longitude	Latitude	Lobe angle	Lobe length	Longitude	Latitude	Lobe angle	Lobe length
Lobe angle	0.468 <0.001	−0.052 0.483			−0.007 0.929	−0.120 0.131		
Lobe length	0.126 0.088	-0.331 <0.001	−0.220 <0.003		0.215 0.006	0.036 0.650	−0.130 <0.100	
Lobe width	−0.289 <0.001	−0.002 0.973	−0.294 <0.001	0.554 <0.001	−0.145 0.067	−0.122 0.126	−0.014 0.860	0.339 <0.001

Table 3.17 Mantel test results. Correlation of Euclidean distances between *A. italicum* subsp. *neglectum* individuals for lobe length, lobe width and lobe angle with distances between locations. Tests are divided into two groups. Type 2 populations contained only *A. italicum* subsp. *neglectum* (*n*=18) and Type 3 locations contained both *A. italicum* subspecies (*n*=13). The Mantel test statistic is the *r*-value; the *P*-value is significant at *P*<0.05, and non-significant results are indicated by the annotation n.s.

	Type 2 location	Type 3 location
Lobe length	r=0.122 P=0.177	r=0.047 P=0.615 n.s.
Lobe width	r=0.002 P=0.467	r=0.077 P=0.229 n.s.
Lobe angle	r=0.210 P=0.970	r=0.010 P=0.486 n.s.

Intraspecific genetic marker variation in *A. italicum*

Genetic differences between subsp. italicum and subsp. neglectum

To test the genetic support for the subspecies *italicum* and *neglectum*, the extent of genetic differentiation was measured using ISSR markers. The two primers 811 (Table 3.18) and 834 (Table 3.19) produced 59 polymorphic bands in total, none of which was diagnostic for either subspecies. Some bands, such as number 1 from primer 811 and number 30 from primer 834, were more frequent in subsp. *italicum* than in subsp. *neglectum*. These two bands are exceptions, and most bands were found at similar frequencies in both subspecies.

Table 3.18 ISSR band frequencies for primer 811 in *A. italicum* subsp. *italicum* and *A. italicum* subsp. *neglectum*.

Primer 811					
Band	***italicum***	***neglectum***	**Band**	***italicum***	***neglectum***
1	0.93	0.00	**16**	0.78	0.91
2	1.00	0.99	**17**	0.98	0.92
3	0.96	0.72	**18**	0.23	0.31
4	1.00	0.97	**19**	0.89	0.81
5	0.97	1.00	**20**	0.80	0.90
6	0.99	0.99	**21**	0.66	0.78
7	0.89	0.92	**22**	0.14	0.30
8	0.77	0.85	**23**	0.80	0.71
9	0.92	0.82	**24**	0.18	0.33
10	0.90	0.87	**25**	0.12	0.08
11	0.84	0.83	**26**	0.00	0.02
12	0.69	0.68	**27**	0.02	0.06
13	0.65	0.57	**28**	0.97	0.93
14	0.99	0.84	**29**	1.00	0.99
15	0.32	0.59			

Table 3.19 ISSR band frequencies for primer 834 in *A. italicum* subsp. *italicum* and *A. italicum* subsp. *neglectum*.

Primer 834					
Band	***italicum***	***neglectum***	**Band**	***italicum***	***neglectum***
1	0.33	0.53	**16**	0.36	0.10
2	0.90	0.71	**17**	0.38	0.32
3	0.15	0.11	**18**	0.15	0.12
4	0.03	0.05	**19**	0.17	0.21
5	0.88	0.95	**20**	0.24	0.25
6	0.35	0.29	**21**	0.22	0.14
7	0.23	0.24	**22**	0.24	0.26
8	0.49	0.28	**23**	0.39	0.28
9	0.58	0.64	**24**	0.42	0.51
10	0.26	0.43	**25**	0.52	0.29
11	0.25	0.33	**26**	0.48	0.58
12	0.24	0.22	**27**	0.25	0.35
13	0.59	0.51	**28**	0.28	0.34
14	0.77	0.81	**29**	0.38	0.59
15	0.81	0.77	**30**	0.80	0.54

The genetic relationships among sampled locations, elucidated using ISSR markers, are shown in an NJ tree (Figure 3.17). The tree comprises two main clades: clade A contains mainly subsp. *neglectum* and clade B contains predominantly subsp. *italicum*. The difference between the two clades (A and B) of subsp. *italicum* and subsp. *neglectum* is statistically significant (ANOSIM; r=0.107, P=0.02). These data consequently support the subspecies *italicum* and *neglectum*.

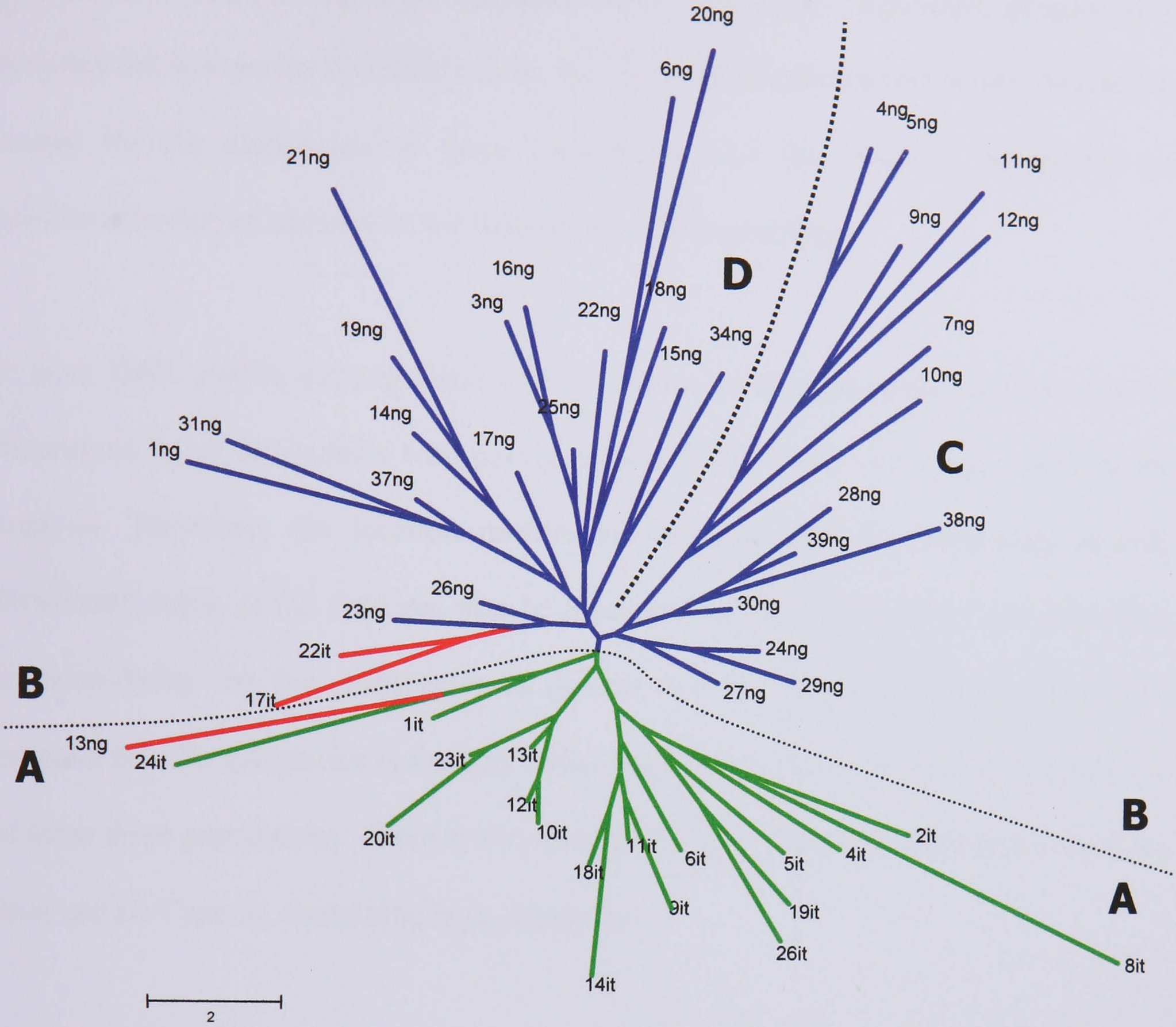


Figure 3.17 Unrooted neighbour_joining tree of *A. italicum* subspecies using pairwise genetic distances generated from primers 811 and 834. Location numbers are indicated as 'it' for subsp. *italicum* locations and 'ng' for subsp. *neglectum* locations. Blue branches indicate locations of subsp. *neglectum.* Green branches indicate populations of subsp. *italicum.* The three Red branches indicate locations not associated with the correct subspecies. Clade A are *A. italicum* subsp. *italicum*; Clade B are *A. italicum* subsp. *neglectum.* Clade C are *A. italicum* subsp. *neglectum* from Britain (west of the Isle of Wight), southern France and Spain. Clade D are subsp. *neglectum* from Britain (east of the Isle of Wight), northern France and Spain.

Three sampled locations are not associated with the correct subspecies. Two subsp. *italicum* locations (17-it and 22-it) are positioned within the subsp. *neglectum* clade (B) and one subsp. *neglectum* location (13-ng) is found within the subsp. *italicum* clade (A). A potential cause for the close association of locations 17-it and 22-it with the subsp. *neglectum* is gene flow between the subspecies; both 17-it and 22-it are sympatric

locations. Location 13-ng is an allopatric location of subsp. *neglectum*, so gene flow between the subspecies is unlikely to be the cause of this anomalous result. Alternative causes for the association of these populations with the incorrect subspecies are misidentification of samples in the field or poor ISSR profiling.

A poor ISSR profile causing these three locations to be genetically different can be discounted as all individuals with poorly defined ISSR bands were excluded from the analysis. Therefore, the location profiles are accurate and the misplacing is real. Misidentification in the field can also be discounted as the character of leaf patterning for identifying the two subspecies is distinct and not easily confused. Gene flow between the two subspecies is the only remaining likely cause of the genetic misplacing of these three populations. There is the potential for gene flow within the three locations (they are all Type 3), containing both subspecies.

In the clade containing the majority of subsp. *italicum* locations (A) the two British subsp. *italicum* locations (1-it and 2-it) are genetically distinct from each other; location 1it is closer to 24-it (Spanish) and 2-it is closer to 8-it (French). These two subsp. *italicum* locations are both located in Cornwall 10km apart. The populations of subsp. *neglectum* separated into two distinct clades, one containing locations from Britain (west of the Isle of Wight), the south of France and north-eastern Spain (clade C), and the second (clade D) containing locations from Britain (east of the Isle of Wight), the Channel Islands, north-western France and northern Spain (Figure 3.17). This second clade also contains two subsp. *italicum* locations from Spain.

Genetic differences between allopatric and sympatric sampling locations

To determine if there are significant genetic differences between the allopatric and sympatric locations an ANOSIM test was carried out. The genetic differences between the Type 1 and 3 locations and Type 2 and 3 locations are all statistically significant between all comparisons (Table 3.20). However, as previously discussed these significant results could be an artefact produced by the uneven distribution of allopatric (Type 1 and 2) and sympatric (Type 3) populations. To test if the uneven distribution is affecting the result a second comparison was made. The second comparison tested the similarity of allopatric populations compared with the nearest sympatric population. No statistically significant difference was found when comparing allopatric locations of subsp. *italicum* with the nearest sympatric populations of subsp. *italicum*. This result shows that the genetic differences between allopatric and sympatric subsp. *italicum* found in the initial test are likely to be caused by the uneven distribution of the sampled populations.

A statistically significant difference was found when comparing allopatric subsp. *neglectum* (Type 1) with individuals growing in sympatry with subsp. *italicum* (Type 3, Table 3.20). The test was also statistically significant when comparing the nearest allopatric and sympatric sampled locations. The significant difference between allopatric and sympatric locations of subsp. *neglectum* is therefore not caused by the uneven distributions of sites. The significant differences are potentially caused by gene flow from subsp. *italicum* to subsp. *neglectum* in sympatric locations. This finding corresponds with the morphological analysis in which subsp. *neglectum* individuals

from sympatric locations were shown to be morphologically different compared with allopatrically distributed individuals (see Tables 3.13 and 3.14).

Table 3.20 ANOSIM test results. Comparison of genetic distances between allopatric locations and sympatric locations. *r* is the ANOSIM *r*-value; *P*-value is significant at $P<0.05$, and non-significant results are indicated by the annotation n.s.

Comparison	*r*	*P*
Type 1 vs. Type 3 (subsp. *italicum*)	0.392	0.011
5 Type 1 vs. nearest 5 Type 3 (subsp. *italicum*)	0.048	0.643 n.s.
Type 2 vs. Type 3 (subsp. *neglectum*)	0.080	0.016
15 Type 2 vs. nearest 15 Type 3 (subsp. *neglectum*)	0.133	0.005

Genetic marker variation and population location

A Mantel test indicates that there was no significant correlation between geographical and genetic distance for either subspecies (*A. italicum* subsp. *italicum*: $r=0.048$, $P=0.31$; *A. italicum* subsp. *neglectum:* $r=0.048$ $P=0.293$). When analysed separately, genetic and geographical distance between sympatric and allopatric locations of subsp. *italicum* were also not correlated (Type 2: $r=0.042$, $P=0.260$; Type 3: $r=0.064$, $P=0.308$). This result indicates that geographical distance among locations of subsp. *italicum* does not affect genetic distance.

When subsp. *neglectum* are divided into sympatric and allopatric locations the genetic distance between the allopatric locations is not significantly correlated with geographical distance ($r=0.037$, $P=0.432$). However, the genetic distance and geographical distance among sympatric subsp. *neglectum* locations were significantly

correlated (r=0.210, P=0.045). Inferences taken from this significant result must, however, be treated with caution, as the sympatric locations of subsp. *neglectum* are not evenly distributed throughout the sampled range.

Evidence for gene flow between subsp. italicum and subsp. neglectum

To identify if there is gene flow between the two *A. italicum* subspecies in sympatric locations, genetic distances between the two subspecies in sympatric locations (Type 3) were compared with distances among the nearest allopatric locations (Type 1 vs. Type 2). To minimise any effect of uneven sampling distribution, only the allopatric locations closest to the sympatric locations were used. There are only five Type 1 locations so the five nearest Type 2 locations were used. The genetic distance between the two subspecies in sympatric locations is significantly smaller than the genetic distance between nearest allopatric locations (Table 3.21). This result indicates that there is gene flow between the two subspecies, which significantly reduces the genetic distance between them when the subspecies exist sympatricaly.

Table 3.21 Mann–Whitney test. Comparison of the genetic distances between *A. italicum* subsp. *neglectum* and *A. italicum* subsp. *italicum* individuals growing in sympatry, to individuals growing in allopatric locations. *W* is the Mann–Whitney test statistic; the *P*-value is significant at $P<0.05$.

	n	Median	*W*	*P*
Sympatric populations	10	6.141	60.0	0.0169
Nearest allopatric populations	5	10.340		

Correlation of morphological and genetic marker variation within the species *A. italicum*

The reliability of separating the *A. italicum* subspecies using leaf characters was tested by assigning each plant to a subspecies based on the presence or absence of leaf patterns. Two genetically distinct groups corresponding to the two subspecies were found (Figure 3.17). A comparison of the Euclidean distances derived from all leaf characters with genetic distances indicates that there is no statistically significant relationship between morphological and genetic similarity among locations of subsp. *italicum* (r=0.066, P=0.216) or locations of subsp. *neglectum* (r=−0.215 P=0.968).

The previous Euclidian distance was calculated from all morphological characters combined; here each character is examined individually to determine if any trends have been masked. Variation of the lobe and leaf pattern characters in subsp. *italicum* is not statistically significantly related to genetic marker variation (Table 3.22). However, a different pattern was observed in subsp. *neglectum*. Locations of subsp *neglectum* were divided into allopatric (Type 2) and sympatric (Type 3) as these locations were found to be genetically different. Only lobe length correlated significantly with molecular marker variation in Type 2 locations. However, when subsp. *neglectum* grows in sympatry with subsp. *italicum* (Type 3 sympatric locations) lobe length does not correlate with molecular marker variation but lobe angle does (Table 3.22). These results confirm that leaf shape patterns are not reliable in distinguishing between the two subspecies *italicum* and *neglectum* as the correlation of morphological and molecular marker variation changes depending upon whether subsp. *neglectum* plants are growing in

sympatric locations with subsp. *italicum*. Leaf patterning can be used to distinguish between two genetically distinct groups that correspond to the *A. italicum* subspecies. Overall variation of lobe characters is not useful as an indicator of molecular variation as it does not consistently correlate with genetic marker variation.

Table 3.22 Mantel tests. Correlation of genetic distance and Euclidean distance among all *A. italicum* subsp. *neglectum* and *A. italicum* subsp. *italicum* individuals for leaf characters. Tests are divided into location types: Type 1= *A. italicum* subsp. *italicum* only; Type 2= *A. italicum* subsp. *neglectum* only; Type 3= both subspecies growing in sympatry. *P*-value is significant at $P<0.05$.

Comparison	*r*	*P*
***A. italicum* subsp. *italicum* both Type 1 and 3: all characters**	0.066	0.216 n.s.
Lobe angle	0.025	0.545 n.s.
Lobe length	0.105	0.724 n.s.
Lobe width	0.202	0.947 n.s.
% Line cover	0.007	0.453 n.s.
% Blotching cover	0.158	0.936 n.s.
***A. italicum* subsp. *neglectum* both Type 2 and 3: all characters**	0.215	0.968 n.s.
***A. italicum* subsp. *neglectum* Type 2 only: all characters**	0.086	0.683 n.s.
Lobe angle	0.041	0.564 n.s.
Lobe length	0.370	0.006
Lobe width	0.075	0.674 n.s.
***A. italicum* subsp. *neglectum* Type 3 only: all characters**	0.062	0.592 n.s.
Lobe angle	0.331	0.020
Lobe length	0.024	0.436 n.s.
Lobe width	0.256	0.940 n.s.

Study 2: *A. italicum* discussion

Intraspecific morphological variation is often used to define subspecific taxa but this variation does not necessarily correspond to differences in evolutionary history. For example, in *Amblystegium* variation of leaf characters used to differentiate between species was shown to be plastic (Vanderpoorten and Jacquemart, 2004). The aim of this chapter is to identify which, if any, morphological characters can be used to delimit intraspecific taxa. By testing the correlation of morphological and molecular marker variation it is also possible to identify the population processes involved in producing patterns of genetic and morphological variation.

This section focuses on the species *A. italicum*, which has four subspecies, two of which (*A. italicum* subsp. *canariense* and subsp. *albispathum*) are differentiated based on distribution and inflorescence morphology. The remaining two subspecies, subsp. *italicum* and subsp. *neglectum*, have identical inflorescence morphology and overlapping ranges (Boyce 1993), and are differentiated based on leaf morphology. Three questions are asked to identify if the differences in leaf morphology between *A. italicum* subsp. *italicum* and *A. italicum* subsp. *neglectum* reliably reflect differences in evolutionary history. 1. What morphological variation is there within the two *A. italicum* subspecies? 2. What molecular marker variation is there within the two *A. italicum* subspecies? 3. Do any morphological characters correlate with genetic lineages?

Subspecies *italicum* and *neglectum* both have polymorphic posterior leaf lobes and show intra- and interpopulation (location) variation. As the population sites are surveyed over a wide geographical area this variation is not unexpected. An association of phenotypes and environmental/geographical location is a common occurrence and has been observed in several species with regard to flower morphology in *Nassella pulchra* (Knapp and Rice, 1998), and *Clarkia dudleyana* (Podolsky and Holtsford, 1995), anther and culm length in *Oryza* (Thomas *et al.*, 2001), and stomata number in *Quercus* (Bruschi *et al.*, 2003). However, the two *A. italicum* subspecies show conflicting sets of results when comparing morphological variation with sampling site location. In subsp. *italicum* there was a statistically significant relationship between the geographical location of the sampling sites with two morphological characters, posterior lobe length and percentage of lines on the leaves. Conversely, variation of morphological characters in subsp. *neglectum* is not related to the sampling site location. Therefore, the morphological characters in the two subspecies may be responding to the environment differently. Unlike the study on *Oryza*, in which population elevation was correlated with morphological variation (Thomas *et al.*, 2001), the extrinsic factors affecting the phenotype of *A. italicum* are currently not known.

The genetic data correspond with the classification by Boyce (1993) supporting both subsp. *italicum* and subsp. *neglectum*. Differences in leaf patterning were used to distinguish between the two subspecies; as patterned and non-patterned individuals are genetically distinct, the character patterning can be used to represent a real historical evolutionary difference between these two taxa. The classification by Boyce (1993) also used lobe characteristics to distinguish between the subspecies *italicum* and *neglectum*.

Although subsp. *italicum* lobes tended to be significantly shorter, this character cannot unambiguously distinguish the two subspecies as the range of variation for both subspecies overlaps. The data presented here indicate that the two subspecies should be distinguished on leaf patterning alone and not lobe size characters.

The variation of leaf morphology in populations of subsp. *italicum* was not correlated with molecular marker variation. A lack of concordance between morphological characters and genetic markers has also been observed in other species. A study of two *Begonia* species did not find a correlation between leaf shape and genetic marker variation (Matolweni *et al.*, 2000). Leaf shape variation in *Myrceugenia fernandeziana* was also not correlated with genetic marker variation (Jensen *et al.*, 2002). Conversely, variation of morphology in subsp. *neglectum* was statistically significantly correlated with genetic marker variation but the correlation differed depending on the type of population examined. In allopatric populations of subsp. *neglectum*, lobe length variation significantly correlated with genetic marker variation, whereas, in sympatric populations, lobe angle variation significantly correlated with genetic marker variation.

Gene flow between the two *A. italicum* subspecies is possible as they have identical inflorescences and share pollinators in Europe (Albre *et al.*, 2003) and Britain (Diaz and Kite, 2002). Gene flow certainly occurs; the two subspecies are genetically more similar when they grow in sympatry than when they are allopatrically distributed. Furthermore, sympatric and allopatric subsp. *neglectum* individuals are morphologically and genetically different, but allopatric and sympatric subsp. *italicum* individuals are not significantly different. The difference between the allopatric and sympatric locations of

subsp. *neglectum* adds further weight to the conclusion that gene flow between the species is occurring. These differences between the allopatric and sympatric locations also suggest that gene flow is not bi-directional, as only subsp. *neglectum* locations are different. Gene flow between subspecies is not unusual: sufficient gene flow occurs between *Elymus glaucus* subsp. *glaucus* and *Elymus glaucus* subsp. *jepsonii* to warrant the removal of the second subspecies (Wilson *et al.*, 2001). Such a situation currently does not exist between the two *A. italicum* subspecies as they are genetically distinct. However, in future, it is possible that extensive gene flow would lead to genetic homogenisation between subsp. *italicum* and subsp. *neglectum* in sympatric locations.

The different leaf patterns of the two subspecies could be caused by historic population factors such as different phylogeographic histories or adaptation to different environments. Subspecies *italicum* has a wide distribution from Greece to the Atlantic coast of France, whereas the distribution of subsp. *neglectum* is restricted, being only found on the west coast of France, northern coast of Spain and in southern Britain (Figure 3.18). The distribution pattern suggests different phylogeographical histories. The flora of Europe contains many examples of species such as *Quercus* (Ferris *et al.*, 1993; Petit *et al.*, 1997), *Betula* (Palme *et al.*, 2003) and *Ilex* (Rendell and Ennos, 2003) that occupied glacial refugia during the last glacial maximum. During glaciation, refugia populations diverged; post glaciation the refugia populations migrated north, where a mixing of genotypes in northern Europe (France, Germany and Britain) is now found. The distribution of the two *A. italicum* subspecies suggests that they occupied different refugia, subsp. *neglectum* in Spain and subsp. *italicum* in the Balkans or Italy (Figure 3.18). It is possible that after migration north, the subspecies genotypes are currently

mixing in France, northern Spain and Britain. Mixing of different glacial refugia populations in France also occurs in silver birch (Palme *et al.*, 2003) and holly (Rendell and Ennos, 2003). To test this hypothesis, a further study *of A. italicum* subsp. *italicum* populations is needed. These additional populations would need to be collected throughout its eastern range in Italy and the Balkans (Figure 3.18).

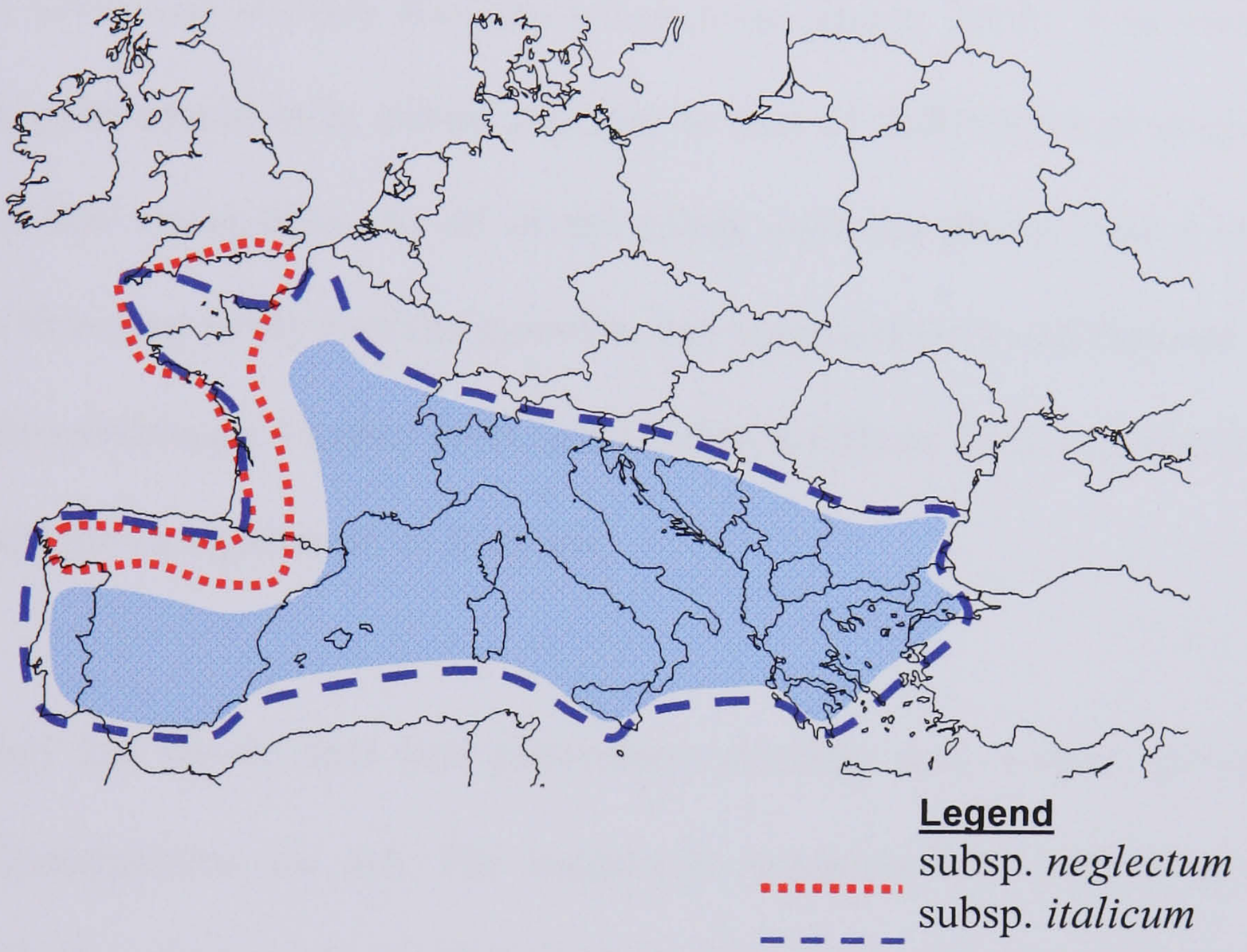


Figure 3.18 The distributions of *A. italicum* subsp. *italicum* and *A. italicum* subsp. *neglectum*. The shaded area indicates where further samples would need to be collected to test a post glacial refugia hypothesis.

Other than differences in phylogeography, the different leaf morphologies could reflect adaptations to varying light levels in different environments. Variegation in *Hedera helix* affects the light compensation points (Yeh and Hsu 2004). The compensation point is the amount of light needed for a plant to survive, in *Hedera*, variegated leaves need more light to survive and therefore grow best in full sun environments. A similar

scenario could exist in *A. italicum* where the variegation on the subsp. *italicum* leaves is an adaptation to a high light environment and the plain leaves of subsp. *neglectum* an adaptation to more shaded environments. However, it is also possible that the variegation in subsp. *italicum* is the result of a gene mutation. It has been shown that a recessive mutation in the IMMUTANS gene in *Arabidopsis* affects the correct function of thylakoid membranes, surrounding normal green leaf tissue has characteristics of high light stress and actively feed the white areas (Baerr, 2005). It is possible that an analogous gene mutation in subsp. *italicum* to that of IMUTANS in *Arabidopsis*, has caused the leaf variegation. As all of the subsp. *italicum* plants form a monophyletic clade it is therefore likely that the mutation has happened once and become fixed in the subsp. *italicum* lineage. Further work is needed to test these three hypothetical causes of differential leaf variegation in *A. italicum*.

This section has shown that leaf patterning correlates with genetic groups, but leaf shape polymorphisms do not. For taxonomic purposes leaf patterning is a useful character as it corresponds to evolutionary groups as determined by the molecular marker data. The leaf lobe characters are therefore not useful as they cannot distinguish between evolutionary groups. In conclusion, morphological characters, once tested, can be used to define subspecific taxa that correspond to genetic lineages; evolutionary processes that occur in very closely related lineages are likely to be dissimilar, leading to patterns of variation specific to each lineage.

Discussion

By comparing variation in morphological characters and genetic markers it is possible to identify which morphological characters correspond to differences in evolutionary history, such as leaf and inflorescence characters in *Cardamine* (Lihova *et al.*, 2000), and those characters, such as pubescence in *Digitalis* (Sales *et al.*, 2001), that do not.

This chapter has investigated morphological and genetic marker variation in two morphologically similar species *A. maculatum* and *A. italicum*. The results highlight the importance of testing the correspondence of morphological characters' to evolutionary groups.

Superficially similar characters (e.g. leaf patterning) were found to correspond with evolutionary groups in one species (*A. italicum*) but not in another (*A. maculatum*). In both *A. maculatum* and *A. italicum*, continuous morphological characters such as lobe length and width did not correspond to taxonomic groups. In *A. italicum* the variation of the continuous leaf lobe characters overlapped between the two subspecies so that neither was distinct. In contrast, the character of leaf patterning in *A. italicum* could unambiguously be used to define subspecific taxa.

The lack of concordance in the correspondence of morphological characters with genetic groups in the two species *A. maculatum* and *A. italicum* parallels other species. Variation of morphological and genetic characters in *A. maculatum* did not correspond to different evolutionary lineages as has been shown in *Eristrum* (Brunell and Whitkus, 1997), *Digitalis* (Sales *et al.*, 2001) and *Elymus glaucus* (Wilson *et al.*, 2001). There are

also just as many species like *A. italicum* in which morphologically defined subspecies correspond to evolutionary lineages (e.g. *Cardamine amora*: (Lihova *et al.*, 2000).

The conflicting result indicates that a morphological character's utility in differentiating between taxa cannot be made without first estimating evolutionary relationships. If the evolutionary relationships among populations are not known, there can be no confidence that morphological groups correspond to evolutionary groups. Opposing results for different species are a direct consequence of what species are according to the evolutionary species concept. In this concept, species are units with "their own evolutionary tendencies and historical fate" (Wiley, 1978, 1981). This means that no two species will respond to selection in the same way. Therefore, generalisations like that by Stebbins (1970) and used by Boyce (1993) regarding the paramount importance of certain characters (in this instance reproductive characters) for distinguishing species should not be made. It has been shown here that in morphologically similar species, and even in subspecies, the usefulness of morphological characters varies. Evolutionary lineages do not terminate at the species level but are continuous down to the population level. However, intraspecific taxa are a grey area with regard to the species concept (Lee, 2003). This is because gene flow between lineages within species may be possible, as observed here between the two *A. italicum* subspecies. As there is gene flow, the lineages are not independent.

This work has shown that different morphological characters correlated with genetic lineages in the two tested species, which indicates that the species are responding differently to selection. It has also shown that dissimilar correlations between

morphological and genetic variation can also extend to lineages within a species, again indicating different responses to extrinsic factors. However, these subspecific lineages were shown not to be completely isolated, but are currently distinct. They therefore satisfy only one criterion of the evolutionary species concept, that of having their own "evolutionary tendencies", but are currently not independent throughout their range. Rigid application of concepts should therefore be treated with caution at the subspecific level; each subspecific taxon needs to be rigorously tested with genetic data before acceptance.

References

Albre J, Quilichini A, Gibernau M (2003) Pollination ecology of *Arum italicum* (Araceae). *Botanical Journal of the Linnean Society* **141**, 205–214.

Arafeh RMH, Sapir Y, Shmida A, Iraki N, Fragman O, Comes HP (2002) Patterns of genetic and phenotypic variation in *Iris haynei* and *I atrofusca* (*Iris* sect. Oncocyclus = the royal irises) along an ecogeographical gradient in Israel and the West Bank. *Molecular Ecology* **11**, 39–53.

Baerr JN, Thomas JD, Taylor BG, Rodermel SR, Gray GR (2005) Differential photosynthetic compensatory mechanisms exist in the immutans mutant of Arabidopsis thaliana *Physiologia Plantarum* **124,** 390-402

Boyce P (1989) A new classification of *Arum* with keys to the infrageneric taxa. *Kew Bulletin* **44**, 383–395.

Boyce P (1993) *The genus Arum. A Kew Magazine Monograph*. HMSO, London.

Brunell MS, Whitkus R (1997) RAPD marker variation in *Eriastrum densifolium* (Polemoniaceae): Implications for subspecific delimitation and conservation. *Systematic Botany* **22**, 543–553.

Bruschi P, Vendramin G, Bussotti F, Grossoni P (2003) Morphological and molecular diversity among Italian populations of *Quercus petraea* (Fagaceae). *Annals of Botany* **91**, 707–716.

Casati P, Walbot V (2003) Gene Expression Profiling in Response to Ultraviolet Radiation in Maize Genotypes with Varying Flavonoid Content. *Plant Physiology* 132: 1739–1754.

Clarke KR, Warwick RM (2001) *Change in Marine Communities: An approach to Statistical Analysis and Interpretation*, 2nd edn. PRIMER-E, Plymouth.

Diaz A, Kite G (2002) A comparison of the pollination ecology of *Arum maculatum* and *Arum italicum* in England. *Watsonia* **24**, 171–181.

Ferris C, Oliver RP, Davy AJ, Hewitt GM (1993) Native oak chloroplasts reveal an ancient divide across Europe. *Molecular Ecology* **2**, 337–344.

Harmes P (1982) Leaf polymorphism in *Arum maculatum*. *Watsonia* **14**, 70.

Gould K, Markham KR, Smith RH, Goris JJ (2000) Functional role of anthocyanins in the leaves of *Quintinia serrata* A. Cunn. *Journal of Experimantal Botany* **51**, 1107–1115.

Jensen RJ, Schwoyer M, Crawford DJ, Stuessy TF, Anderson GJ, Baeza CM, O MS, Ruiz E (2002) Patterns of morphological and genetic variation among populations of *Myrceugenia fernandeziana* (Myrtaceae) on Masatierra island: Implications for conservation. *Systematic Botany* **27**, 534–547.

King RA, Gornall RJ, Preston CD, Croft JM (2002) Population differentiation of *Potamogeton pectinatus* in the Baltic Sea with reference to waterfowl dispersal. *Molecular Ecology* **11**, 1947–1956.

Knapp EE, Rice KJ (1998) Comparison of isozymes and quantitative traits for evaluating patterns of genetic variation in purple needlegrass (*Nassella pulchra*). *Conservation Biology* **12**, 1031–1041.

Kruskal JB (1964) Nonmetric multidimensional scaling: a numerical method. *Psychometrika* **29**, 115–129.

Lee MSY (2003) Species concepts and species reality: salvaging a Linnaean rank. *Journal of Evolutionary Biology* **16**, 179–188.

Lihova J, Marhold K, Neuffer B (2000) Taxonomy of *Cardamine amara* (Cruciferae) in the Iberian Peninsula. *Taxon* **49**, 747–763.

Linhart YB, Grant MC (1996) Evolutionary significance of local genetic differentiation. *Annual Review of Ecology and Systematics* **27**, 237–277.

Mateu-Andres I, Segarra-Moragues JG (2003) Allozymic differentiation of the *Antirrhinum graniticum* and the *Antirrhinum meonanthum* species groups. *Annals of Botany* **92**, 647–655.

Matolweni LO, Balkwill K, McLellan T (2000) Genetic diversity and gene flow in the morphologically variable, rare endemics *Begonia dregei* and *Begonia homonyma* (Begoniaceae). *American Journal of Botany* **87**, 431–439.

Mogg RJ, Bond JM (2003) A cheap, reliable and rapid method of extracting high-quality DNA from plants. *Molecular Ecology Notes* **3**, 666–668.

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

Nei M, Kumar S (2000) *Molecular Evolution and Phylogenetics*. Oxford University Press, Oxford.

Palme AE, Su Q, Rautenberg A, Manni F, Lascoux M (2003) Postglacial recolonization and cpDNA variation of silver birch, *Betula pendula*. *Molecular Ecology* **12**, 201–212.

Petit RJ, Pineau E, Demesure B, Bacilieri R, Ducousso A, Kremer A (1997) Chloroplast DNA footprints of postglacial recolonization by oaks. *Proceedings of the National Academy of Sciences of the USA* **94**, 9996–10001.

Podolsky RH, Holtsford TP (1995) Population-structure of morphological traits in *Clarkia dudleyana* 1. Comparison of F_{ST} between allozymes and morphological traits. *Genetics* **140**, 733–744.

Prime CT (1960) *Lords and Ladies*. Collins, London.

Rendell S, Ennos RA (2003) Chloroplast DNA diversity of the dioecious European tree *Ilex aquifolium* L. (English holly). *Molecular Ecology* **12**, 2681–2688.

Rochefort RM, Peterson DL (2001) Genetic and morphologic variation in *Phyllodoce empetriformis* and *Phyllodoce glanduliflora* (Ericaceae) in Mount Rainier National Park, Washington. *Canadian Journal of Botany-Revue Canadienne De Botanique* **79**, 179–191.

Sales E, Nebauer SG, Mus M, Segura J (2001) Population genetic study in the Balearic endemic plant species *Digitalis minor* (Scrophulariaceae) using RAPD markers. *American Journal of Botany* **88**, 1750–1759.

Simpson GG (1961) *Principals of Animal Taxonomy*. Columbia University Press, New York.

Stebbins GL (1970) Adaptive radiation of reproductive characteristics in angiosperms, I: Pollination mechanisms. *Annual Review of Ecology and Systematics* **1**, 307–326.

Stone C, Chisholm L, Coops N (2001). Spectral reflectance characteristics of eucalypt foliage damaged by insects. *Australlan Journal of Botany* 49: 687–698.

Thomas G, Joseph L, Varghese G, Sreejayan, Kalyanaraman K, Kuriachan P (2001) Analysis of phenotypic and genetic variations among populations of *Oryza malampuzhaensis* show evidence of altitude-dependent genetic changes. *Canadian Journal of Botany-Revue Canadienne De Botanique* **79**, 1090–1098.

Vanderpoorten A, Jacquemart AL (2004) Evolutionary mode, tempo, and phylogenetic association of continuous morphological traits in the aquatic moss genus *Amblystegium*. *Journal of Evolutionary Biology* **17**, 279–287.

Wiley EO (1978) The evolutionary species concept reconsidered. *Systematic Zoology* **41**, 17–26.

Wiley EO (1981) *Phylogenetics: The Theory and Practice of Phylogenetic Systematics*.

Wilson BL, Kitzmiller J, Rolle W, Hipkins VD (2001) Isozyme variation and its environmental correlates in *Elymus glaucus* from the California Floristic Province. *Canadian Journal of Botany* **79**, 139–153.

Yeh DM, Hsu PY 2004 Differential growth and photosynthetic response of selected cultivars of English ivy to irradiance. *Journal of Horticultural Science & Biotechnology* **79**, 633-637

Chapter 4 Intrageneric morphological and DNA sequence variation

Abstract

Species are the fundamental units of biology; if evolutionary relationships are unknown for any given group of species, the correspondence of the taxonomy to evolutionary lineages cannot be known. Evolutionary relationships among extant taxa can be estimated using molecular genetic techniques. Knowing phylogenetic relationships among species firstly allows classifications to be changed so that they reflect underlying patterns of evolution. Secondly, when a phylogeny is known it can be used in conjunction with morphological characters to understand how characters have evolved within a taxon. This chapter uses molecular sequence data from the plastid and nuclear genomes to infer evolutionary relationships among the taxa of the genus *Arum*. These data are then used to identify support for the taxonomy.

The molecular data support the division of *Arum* into two subgenera *Arum* and *Gymnomesium*. Sections based upon tuber morphology are not supported by the molecular data. Flower height, a character used at the subsection level, was shown to have arisen independently on more than one evolutionary lineage. The data indicate that *a priori* assumptions of character usefulness can lead to classifications that do not reflect evolutionary lineages.

Introduction

Phylogenies can be used both to estimate the relatedness among species and to infer evolutionary processes that have produced variation among extant species (Pagel, 1999). The first outcome of phylogenies, estimating the evolutionary relatedness between species, allows taxonomies to be modified so that species and groups of species correspond to evolutionary groups. Secondly, phylogenetics can be used to identify the evolutionary processes such as convergence, hybridisation or patterns of migration that have influenced the development of variation within and between species.

Phylogenetic relationships among species can be identified because all species share common ancestors. Phylogenetic systematics uses this phenomenon of shared ancestry to examine how one taxon is related to another when compared to a third (Kitching *et al.*, 1998). Relationships are elucidated in a relative manner; for example, taxon A is more closely related to taxon B than either is to taxon C. This relative concept was first mooted by Hennig (1965). For example, humans and chimpanzees are more closely related to each other than either is to gorillas (Figure 4.1). This statement indicates that humans and chimpanzees share a common ancestor that is more recent than their shared common ancestor with gorillas. The phylogenetic relationships among the great apes (Figure 4.1) by Hayasaka *et al.* (1988) were elucidated by comparing nucleotide sequences from the mitochondria; there are however many more sources of phylogenetic information.

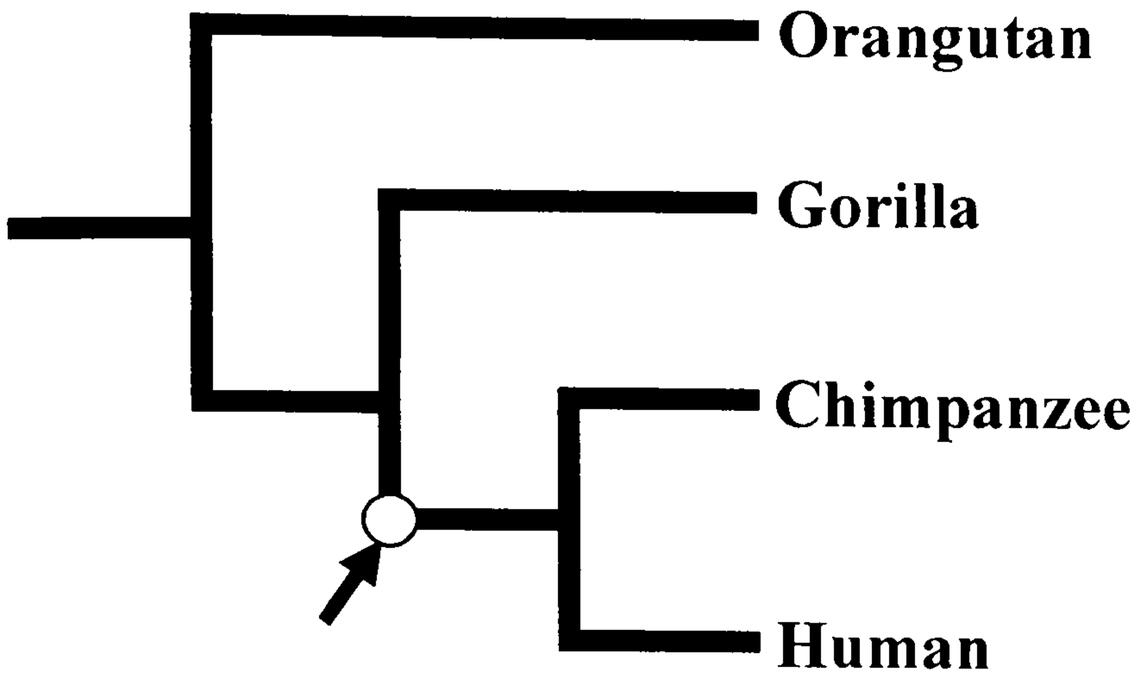


Figure 4.1 The phylogenetic relationship among the great apes inferred using mitochondrial DNA sequences. Figure adapted from Figure 3 of Hayasaka *et al.* (1988). Humans and chimpanzees are more closely related to each other than they are to gorillas, as the most recent common ancester shared by chimpanzees (arrow) and humans is not shared by gorillas.

Many recent reviews (e.g. Soltis and Soltis, 1998; Stace, 1989; Wolfe and Liston, 1998) highlight the large number of laboratory techniques available for phylogenetic systematics. Inferring relationships among species can be carried out using techniques such as karyology, biochemical variation, DNA variation and morphological variation. Morphological characters have been used to understand evolutionary relationships since the publication of Darwin's ideas (Patterson *et al.*, 1993). For example, morphological characters were used to elucidate the phylogenetic relationships among subfamilies of the Araceae (Grayum, 1990). Recently, genetic characters, especially those derived from PCR markers, dominate phylogenetic studies (Savolainen and Chase, 2003) and are typified by studies in the genera *Geum* (Smedmark and Eriksson, 2002) and *Viola* (Yockteng *et al.*, 2003). Techniques utilising the PCR have revolutionised systematic biology as any region of the genome, of any organism, can now be sequenced and compared (Wolfe and Liston, 1998).

Phylogenetics can test *a priori* morphological groups (e.g. subgenera, genera and families) with an evolutionary hypothesis, which allows monophyletic taxa to be identified. It is monophyletic taxa that are important (Hey *et al.*, 2003; Mayden, 1997; Sites and Marshall, 2003) as these taxa correspond with the evolutionary species

concept outlined by Simpson (1961) and Wiley (1978). The evolutionary species concept dictates that species (and in fact any taxon) comprise independent evolutionary lineages, and therefore only groups that are monophyletic should be recognised. Molecular data can be used to identify groups of taxa, such as those defined by leaf morphology in the Fontinalaceae (Shaw and Allen, 2000), that do not correspond to monophyletic evolutionary groups.

For a species or a higher taxon to be monophyletic, all taxa within the group must have only one ancestor and the group must contain all the taxa (lineages) that are derived from the closest shared ancestor (Figure 4.2). Only shared derived characters (**synapomorphy**) unite taxa in this manner. Often, similar characters can arise independently during the evolution of two different lineages. When this occurs the characters are termed homoplasious and, if used to infer taxonomic groups, they produce polyphyletic groups (Figure 4.2). Polyphyletic groups can often arise through convergence or parallelism of characters. If a group is defined with an ancestral character (**plesiomorphic**) and not a synapomorphy, a paraphyletic group is formed. Paraphyletic groups, like polyphyletic groups, do not indicate the real groups of organisms formed by evolution and contain only part of a monophyletic clade (Figure 4.2). When phylogenetic relationships have been estimated among organisms, the taxonomic characters used initially to group them can be tested to identify which characters correspond with monophyletic groups. Those characters that correspond with polyphyletic and paraphyletic groups can be identified and removed.

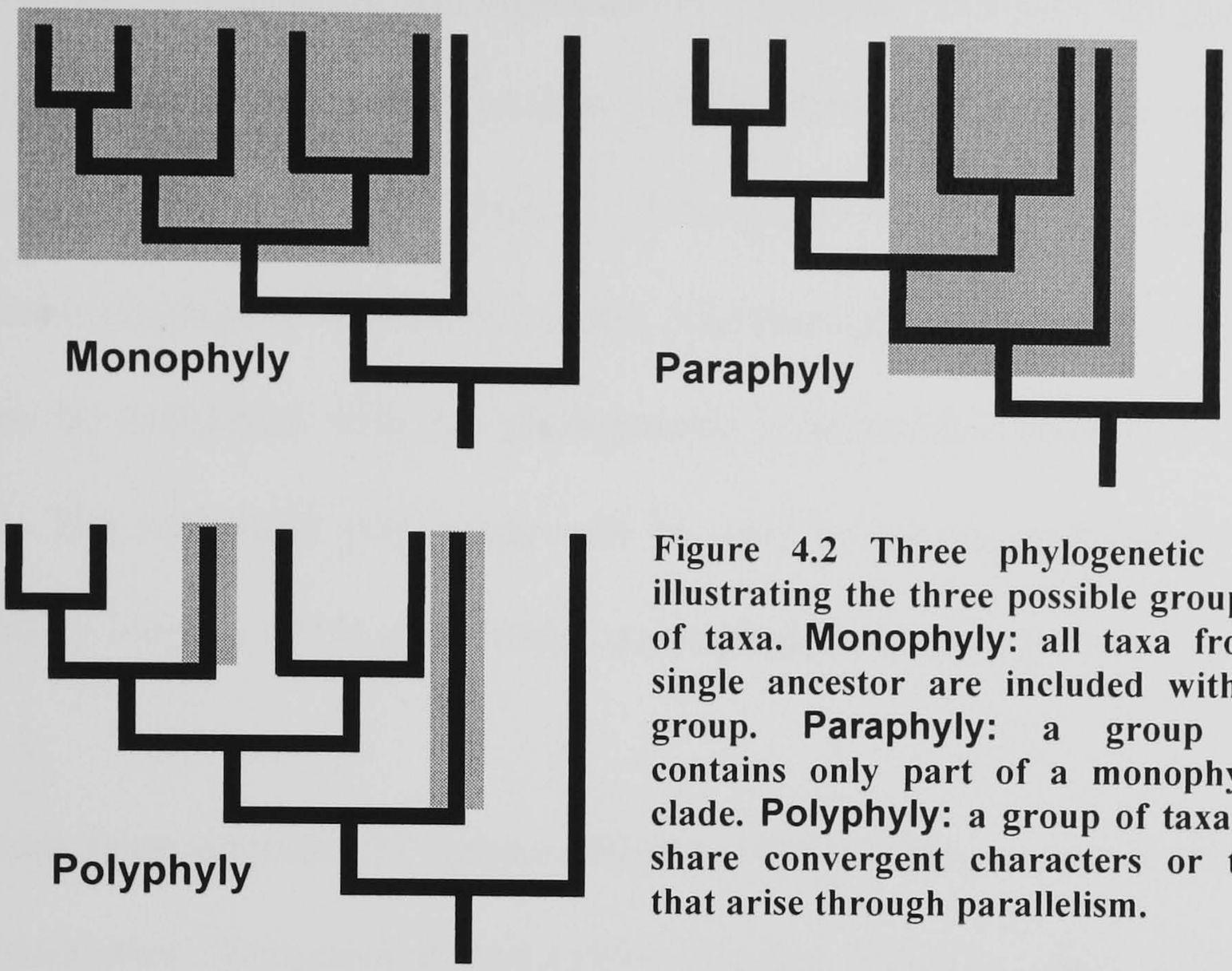


Figure 4.2 Three phylogenetic trees illustrating the three possible groupings of taxa. Monophyly: all taxa from a single ancestor are included within a group. Paraphyly: a group that contains only part of a monophyletic clade. Polyphyly: a group of taxa that share convergent characters or those that arise through parallelism.

Correlating morphological characters with evolutionary relationships can identify convergent characters. By comparing molecular phylogenies and leaf morphology in the family Fontinalaceae, Shaw and Allen (2000) showed that leaf morphology was convergent. The evolution of dioecy (male and female flowers are on different individuals) in the Bencomia alliance was shown to have occurred in three independent evolutionary events (Helfgott *et al.*, 2000). Patterns of migration can also be inferred by comparing the distribution and evolutionary relationships of taxa such as the two radiations of *Corynocarpus* from the tropics to Australia and New Zealand (Wagstaff and Dawson, 2000).

The aim of this chapter is to identify what morphological characters correlate with evolutionary lineages. This question is investigated using the genus *Arum*. It is an ideal genus to use as there is a recent morphological taxonomy by Boyce (1989, 1993) that

uses a variety of vegetative and reproductive characters to divide the genus into several sections and subsections. To examine which characters correlate with evolutionary lineages, and if the taxonomic structure of the genus reflects the evolutionary processes, a molecular phylogeny will be estimated. Morphological characters and species groups will then be compared with the phylogenetic relationships estimated using molecular markers. The molecular phylogeny will be used to test the monophyly of the groups proposed by Boyce (1993) and identify monophyletic species groups.

The genus *Arum* contains 25 species (Boyce, 1993). Using a combination of vegetative and reproductive characters Boyce (1993) divided the genus into two subgenera, two sections and six subsections, as outlined in Table 4.1. The division of the genus *Arum* into two subgenera is based upon differences in leaf maturation time, the presence of cataphylls and some floral anatomical characters. The monotypic subgenus *Gymnomesium* contains only *Arum pictum*, which is characterised by the leaves maturing after emergence of the inflorescence, a peduncle that does not enlarge during inflorescence maturation and two or three cataphylls (Table 4.1). The subgenera *Arum* contains the remaining 24 species and has leaves that mature before the inflorescence, a peduncle that extends greatly during maturation and no cataphylls at the inflorescence base (Table 4.1). The 24 species of subgenus *Arum* are divided into two sections (Table 4.1), section *Arum* and section *Dioscoridea*, based upon differences in tuber morphology (Figure 4.3). Section *Arum* contains four species with horizontal rhizomatous tubers (Figure 4.3, A) whilst the section *Dioscoridea* contains 20 species all with vertical discoid tubers (Figure 4.3, B). The 20 species in the section

Dioscoridea are further divided into six subsections based on differences in inflorescence morphology (Table 4.1, Figure 4.3).

The six subsections of the section *Dioscoridea* are subsect. *Alpina, Dioschroochiton*, *Tenuifila*, *Hygrophilum*, *Poeciloporphyochiton* and *Cretica* (Table 4.1). The subsection *Dioschroochiton* contains 10 species, characterised by a cryptic flower display (Table 4.1). Species with a cryptic flower display have a peduncle that is shorter than the petiole; this causes the inflorescence to grow below the height of the leaves. If the peduncle is larger than the petioles the flower is taller than the leaves and is called a flag. The subsection *Tenuifila* has three species that all have a flag display type and a purple spathe tube (Table 4.1). The subsection *Poeciloporphyochiton* contains two cryptic species. The final subsection *Cretica* contains two species that lack staminodes and pistillodes, but *A. idaeum* is cryptic and *A. creticum* is a flag species.

Table 4.1 Classification and defining characters of the subgenera, sections and subsections of the genus *Arum* from Boyce (1993).

Classification	Defining character(s)
Subgenus *Arum*	Inflorescence matures after leaf maturation
Section *Arum*	Horizontally orientated rhizomatous tuber
A. maculatum	
A. byzantinum	
A. italicum	
A. concinnatum	
Section *Dioscoridea*	Vertically or horizontally orientated discoid tuber
Subsection *Alpina*	Scentless inflorescence; peduncle longer than the petioles
A. alpinum	
Subsection *Dioschroochiton*	Peduncles shorter than petioles; fetid, oily, or sweet smell; pistillodes long, flexuous bristles, staminode bristles similar or shorter stiff bristles
A. orientale	
A. gratum	
A. lucanum	
A. apulum	
A. nigrum	
A. cyrenaicum	
A. pupureospathum	
A. balansanum	
A. hainesii	
A. elongatum	
Subsection *Tenuifila*	Peduncles equal or exceeding petioles; scentless inflorescence; pistillodes and staminodes long flexuous bristles; purple spathe-tube interior
A. rupicola	
A. jacquemontii	
A. korolkowii	
Subsection *Hygrophilum*	Peduncles equal or longer than petioles; spadix-appendix less than 4mm in diameter; scentless inflorescence; purple spathe-tube interior
A. euxinum	
A. hygrophilum	
Subsection *Poeciloporphyochiton*	Peduncles much shorter than petioles; fetid smell (fermenting apples)
A. dioscoridis	
A. palaestinum	
Subsection *Cretica*	Peduncles longer or shorter than petioles; staminodes and pistillodes absent
A. idaeum	
A. creticum	
Subgenus *Gymnomesium*	Inflorescence matures before leaf maturation
A. pictum	

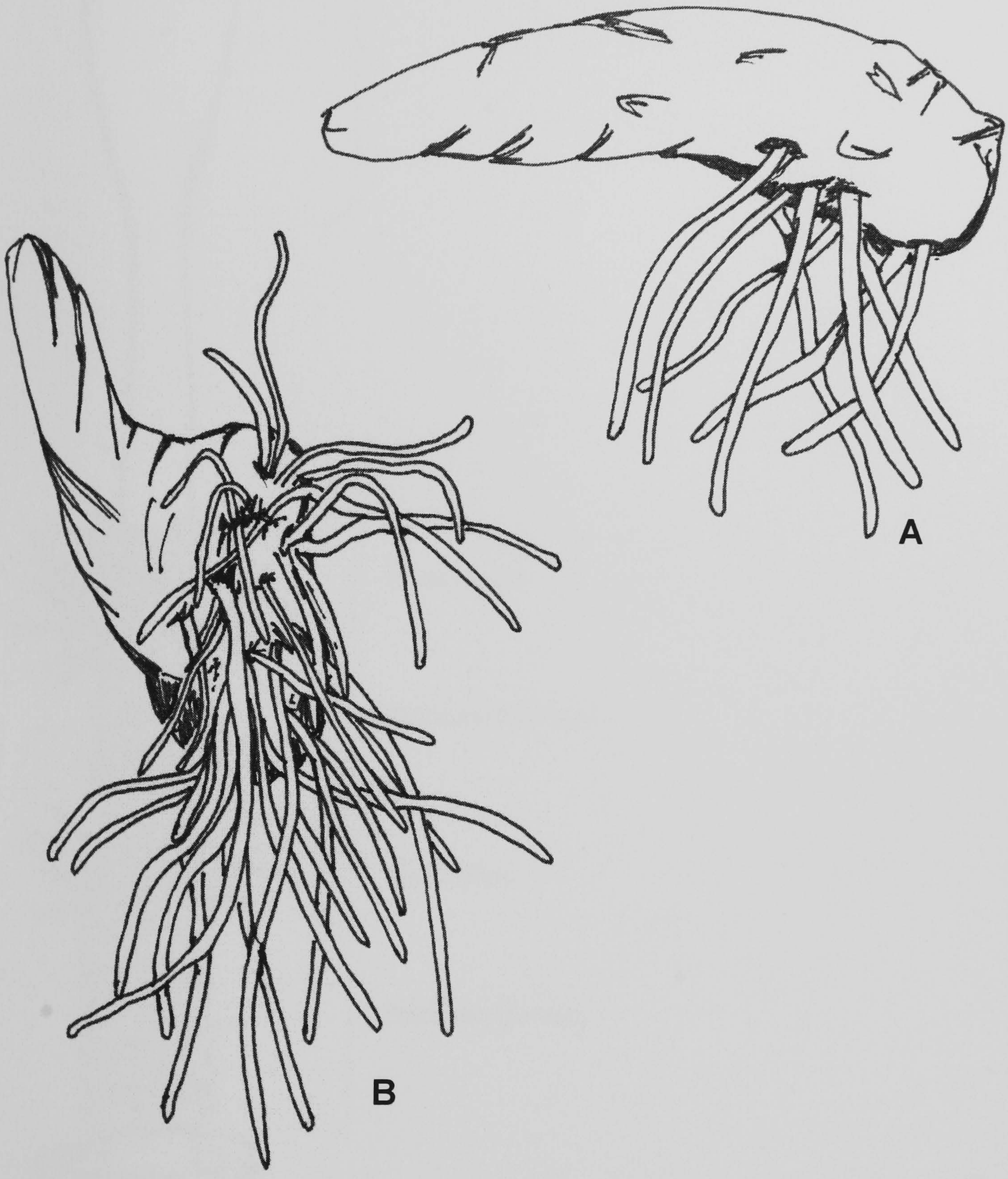


Figure 4.3 Diagrams showing the two types of tuber morphology in *Arum* species. Tuber A is horizontal rhizomatous and is indicative of section *Arum*. Tuber B is vertical discoid type and is indicative of section *Dioscoridea*.

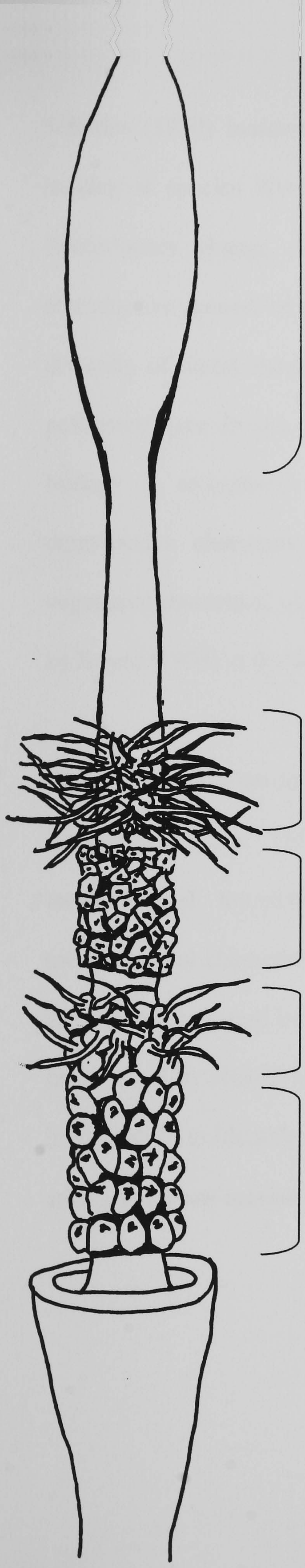


Figure 4.4 Generalised spadix–appendix morphology

Stebbins (1970) indicated that reproductive characters are the most important when looking at species diversification as these are the characters that affect gene flow. Evolutionary change of a reproductive character could potentially influence the reproductive success of an individual or population. In the genus *Arum* there is a huge diversity of floral morphologies, presumably, as a result of adaptation to different pollinators (see Boyce, 1993). The importance of reproductive characters in species biology is recognised by Boyce (1993) and he places primary importance on reproductive characters for forming species groups within the genus. However, vegetative characters, which do not directly affect reproductive biology, have been used by Boyce (1993) at the subgenus and sectional levels.

In this chapter molecular markers are used to estimate phylogenetic relationships among the *Arum* species. This chapter looks at the following questions: 1. Does the morphological taxonomy of *Arum* correlate with the phylogeny? 2. Which morphological characters correlate with genetic lineages? 3. If the morphological data from Boyce are used in a phylogenetic context are they congruent with the molecular phylogeny? By identifying morphological characters that correlate with genetic lineages it is possible to identity characters that accurately reflect the evolutionary relationships among the *Arum* species.

Phylogenetic methods

Phylogenetic methods can be subdivided into those that use algorithms to define trees and those that use optimality criteria to rank trees based on an *a priori* criterion (reviewed in: Hillis *et al.*, 1998; Page and Holmes, 1998; Swofford and Sullivan, 2003). Both types of methods have advantages and disadvantages. Algorithm methods infer phylogenetic trees with a single mathematical statement that produce a single tree (Hillis *et al.*, 1998; Swofford *et al.*, 1996). For example, UPGMA (unweighted-pair group method with arithmetic means) or NJ are algorithm-based methods that produce a single tree. UPGMA and NJ methods are computationally very fast (Swofford and Sullivan, 2003) but can potentially be misleading as only one tree is inferred and this may not be the optimum (Swofford *et al.*, 1996). There are two commonly used criterion-based methods: maximum parsimony (MP) and maximum likelihood (ML). Both MP and ML methods produce a score for each tree that is used to rank multiple trees in order of the optimality criterion (Swofford *et al.*, 1996; Swofford and Sullivan, 2003). Criterion-based methods (MP and ML) are computationally slower compared with UPGMA and NJ, but have the advantage of testing many arrangements of taxa, thereby producing optimum trees (Swofford and Sullivan, 2003).

Criterion-based methods will be used in this thesis to estimate the phylogenetic relationships among *Arum* taxa. The two main criterion methods, ML and MP, differ in their optimality criteria. The ML method infers phylogenetic relationships based upon assumptions of an explicit evolutionary model, whereas the MP method infers relationships that require the fewest changes (Kitching *et al.*, 1998; Kluge and Farris,

1969). The evolutionary model used in an ML analysis is a potential problem as it is difficuld to know what model of evolution to use. The MP method will be used to infer phylogenetic relationships as it does not require the use of an evolutionary model.

Parsimony analysis

Parsimony analysis infers phylogenetic relationships by arranging taxa so that the resulting tree minimises the number of evolutionary changes needed to explain patterns of synapomorphies (Kitching *et al.*, 1998; Kluge and Farris, 1969). For example, the data matrix in Table 4.2 contains four taxa with six binary characters. The most parsimonious tree for the sample data is tree number one (Figure 4.5). Tree number one has the fewest number of character changes (steps) that explain the data and is therefore the shortest (most parsimonious). Trees two and three are rejected as they are longer; they require more character changes to explain the data than tree number one. Characters used in reconstructions can have different attributes. In this sample dataset (Table 4.2), characters one and four convey no phylogenetic information as they are only present in one taxon and will only occur at the end of each branch (the terminal node). For the most parsimonious reconstruction (tree 1, Figure 4.5), characters two, three and five are shared derived characters (synapomorphies) and form monophyletic clades. Character six is homoplasious, occurring independently in taxa B and C.

Table 4.2 Sample data matrix modified from Kitching *et al.* (1998). The matrix contains four taxa (A–D), and six characters (1–6). Each character is binary with 1 indicating presence and 0 indicating absence.

Taxa	1	2	3	4	5	6
A	0	0	0	0	0	0
B	1	1	0	0	0	1
C	0	1	1	1	1	1
D	0	1	1	0	1	0

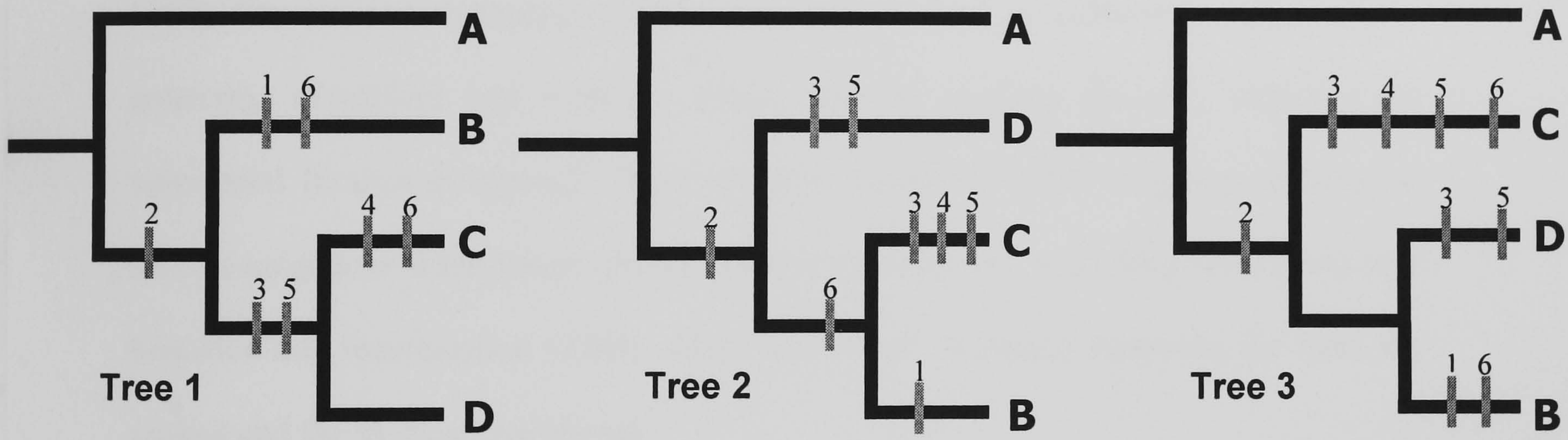


Figure 4.5 Three trees from the four hypothetical taxa. Vertical bars indicate presence of one of the six characters. Tree 1 is the shortest with seven changes (number of vertical bars), tree 2 has eight changes and tree 3 has nine changes. Adapted from Kitching *et al.* (1998).

Searching for MP trees is a two-step process: 1, calculating the number of changes to explain the pattern of data; and 2, searching for the tree that minimises the number of changes. Determining the optimal tree is straightforward if there are 11 or fewer taxa in the analysis, as an exhaustive search can be used (Swofford and Sullivan, 2003). An

exhaustive search generates all possible trees, which are then ranked based on length, and the shortest is chosen. When the number of taxa increases avove 11 the total number of trees becomes very large (see Table 3.2 in Kitching *et al.*, 1998). If there is a large number of taxa another exact procedure, the branch-and-bound method, can be used for up to 25 taxa (Swofford and Sullivan, 2003). This method calculates all possible trees; however, if the tree rearrangement produces a suboptimal tree then that search is stopped and another search started. Since the phylogenetic analysis, including outgroups and species, from the genus *Arum* will have over 25 taxa these exhaustive methods are not applicable and an approximate method will be used. Approximate methods use hill-climbing algorithms to search for optimum trees (Page and Holmes, 1998). The tree search strategy used is a two-step process; an initial suboptimum tree is generated (Swofford and Sullivan, 2003), then to improve the tree, branches are rearranged (branch-swapping). There are three kinds of branch-swapping techniques: nearest-neighbour interchange (NNI), subtree pruning and regrafting (SPR) and tree bisection and reconnection (TBR). After each round of branch swapping the trees are ranked and the shortest tree chosen.

Practical methods

Plant sampling

Plant samples for the phylogenetic analysis were collected from the alpine collection at Royal Botanic Gardens, Kew (RBG Kew). The collection contains the majority of the 25 species of *Arum* including some duplicate accessions (Table 4.3). Two leaves of each sampled plant were collected; one leaf was used for DNA extraction, the second was dried and pressed as a voucher specimen in the Kew Herbarium. The alpine collection at RBG Kew does not contain *A. byzantium* or *A. hainseii*, which were only available from dried material. In addition to the samples in the genus *Arum*, samples from five other genera were also obtained to be used as outgroups (Table 4.4)

Table 4.3 The species sampled from the genus *Arum* at RGB Kew. RGB Kew is plant accession number. The Kew Gardens Gene bank accession number is in the column labelled 'MWC'. The numbers in parentheses indicate the species sample number for the phylogenetic trees. Only species with two or more accessions are numbed.

Species	RBG Kew	MWC
A. maculatum (1)	Garden Wild	11161
A. maculatum (2)	1990-2018	12031
A. maculatum (3)	1990-475	12029
A. byzantinum	Dry material	12100
A. italicum subsp. *albispathum* (1)	1992-3622	11020
A. italicum subsp. *canariense* (2)	1994-3369	11031
A. italicum subsp. *italicum* (3)	1978-3353	11021
A. concinnatum	1994-3353	11014
A. alpinum	1992-3623	11013
A. orientale subsp. *orientale*	1990-1344	12024
A. gratum	Dry material	12101
A. lucanum	1987-1133	12025
A. nigrum	1992-2083	12020
A. nigrum	1978-2628	12027
A. cyrenaicum	1996-3897	11030
A. cyrenaicum	1993-1893	11033
A. cyrenaicum	1991-1047	11035
A. pupureospathum	1987-2032	11023
A. balansanum	1990-2456	11009
A. hainesii	Type Specimen	12034
A. elongatum subsp. *elongatum* (1)	1990-2019	12098
A. elongatum subsp. *alpinariae* (2)	1999-2014	12032
A. rupicola var. *virescens* (1)	1979-2206	11034
A. rupicola var. *rupicola* (2)	1993-1554	11036
A. jacquemontii	1969-5385	12030
A. korolkowii (1)	1988-2853	12022
A. korolkowii (2)	1994-3354	12028
A. euxinum	1982-2903	11019
A. hygrophilum	1969-3057	11027
A. dioscoridis var. *philistaeum* (1)	1979-2042	11010
A. dioscoridis var. *philistaeum* (2)	1978-4602	11011
A. dioscoridis var. *dioscoridis* (3)	1986-6222	11015
A. dioscoridis var. *dioscoridis* (4)	1979-2190	11018
A. dioscoridis var. *cyprium* (5)	1967-22354	11021
A. dioscoridis var. *dioscoridis* (6)	1986-6221	11026
A. dioscoridis var. *dioscoridis* (7)	1984-3705	11028
A. dioscoridis var. *dioscoridis* (8)	1985-5514	11029
A. palaestinum	1984-4185	11016
A. idaeum	1993-1895	12023
A. creticum	1953-39806	11037
A. pictum (1)	1967-64101	11012
A. pictum (2)	1985-1888	11024
A. pictum (3)	1982-2699	11025

Table 4.4 Outgroup species and accession numbers. RGB Kew is plant accession number. The Kew Gardens Gene bank accession number is in the column labelled 'MWC'.

Species	RBG Kew	MWC
Dracunculus canariensis	1989-2008	11159
Typhonium bulumeri	1997-6851	11160
Biarum tenuifolium	1972-6295	11162
Biarum tenuifolium subsp. *abbreviatum*	1974-3478	11163
Arisarum vulgare	1972-10427	11164
Eminium		11806

Molecular methods

DNA extraction

To extract DNA from the leaf samples the CTAB (hexadecyltrimethylammonium bromide) method outlined by Doyle and Doyle (1987) was used with the following amendments. Fresh leaf material was left for 24 hours in an airtight plastic bag at 4°C in the dark. This forces the leaf to metabolise stored compounds, such as starch, which can often hamper DNA extraction. Immediately before DNA extraction, leaves were washed with 96% ethanol to remove any surface contaminants. One gram of leaf tissue was ground with fine sand in a preheated pestle and mortar (at 60°C). A smaller amount of leaf tissue (20mg) was used from the two dried samples. The ground leaf tissue was incubated with 10ml of CTAB buffer at 60°C for 30 minutes. Preheating to 60°C causes proteins to denature, limiting any damage to the DNA by proteins. Proteins and polysaccharides were then removed by adding 10ml of SEVEG buffer (24:1 chloroform:isoamyl alcohol) and incubating for 45 minutes at room temperature with gentle agitation. After incubation, cell debris was pelleted by centrifuging at 8000rpm for 10 minutes and the supernatant was transferred into a clean tube. Adding an equal volume of pre-chilled (–20°C) isopropanol and placing the samples at –20°C for at least 24 hours precipitated the DNA. The precipitated DNA was stable enough at this stage to be left for 1 month before further purification.

To purify the DNA further, a caesium chloride gradient centrifugation was carried out.

The precipitated DNA was dissolved in 3ml of 1.5g/ml caesium chloride and ethidium

bromide by gentle agitation at room temperature. Once the DNA was dissolved, a process that took between 24 hours and 1 month (this depended on each sample), gradient centrifugation was carried out using 8g of dissolved DNA caesium chloride mix. This was centrifuged at 58,000rpm for 5 hours. Under UV illumination, the DNA appears as a discrete band in the middle of the centrifuge tube. The DNA band was removed using a pipette and mixed for 15 minutes with an equal volume of butanol to remove the ethidium bromide. Samples were further cleaned through dialysis. The samples were transferred into a 10cm length of dialysis tubing and immersed in 4 litres of milli-Q water (per 24 samples) for a minimum of 2 hours and no longer than 4 hours. The dialysis tubing was then removed from the water and placed flat on a tray containing a layer of sucrose (Tate and Lyle) and covered by a further layer of sucrose for 30 minutes. The sucrose was used to concentrate the DNA in the dialysis tubing. Sucrose was washed off by immersing the dialysis tubing in milli-Q water after which they were immersed in 1×TE buffer for 4 hours. Quantification of the DNA concentration of each sample was carried out on an agarose gel. In total, DNA was extracted from 46 samples representing each *Arum* species and in some duplicates a 100µl aliquot from each sample was transported on ice to the Molecular Laboratory at the Centre for Ecology and Hydrology, Dorset, and stored at 20°C for experimental use. The remaining DNA is stored at RBG Kew in the gene bank at –70°C (accession numbers are given in Table 4.3).

PCR amplification

The suitability of four different genomic regions for a phylogenetic analysis of *Arum* was investigated. Three of the regions: the *matK* gene, *atpB-rbcL* spacer and the *trnL*(UAA) intron, are from the chloroplast genome, and the fourth region was a section of the nuclear rDNA cistron containing: ITS1 (Internal Transcribed Spacer 1), the 5.8s gene and ITS2. These regions were chosen as candidates for genetic analysis as they are suitable for resolving relationships at the genus level. The intron *trnL*(UAA) has been used to investigate relationships within several genera, e.g. *Corynocarpus* (Wagstaff and Dawson, 2000) and *Geum* (Smedmark and Eriksson, 2002). The ITS1 and ITS2 regions are suitable for resolving relationships within genera e.g. *Draba* (Beilstein and Windham, 2003) and *Bupleurum* (Neves and Watson, 2004). In addition to these two regions, the chloroplast regions *matK* and *atpB-rbcL* spacer were also included as they are potentially informative for distinguishing among species within genera (Wolfe and Liston, 1998). For a genomic region to be deemed suitable it first must produce consistently amplified DNA fragments in all *Arum* species and secondly the DNA fragments should be of sufficient quality to sequence reliably.

Amplification of DNA from the *matK* gene was carried out using the primers trnKF and trnKR (Cronn *et al.*, 2002). PCR fragments were not of sufficient quality for reliable sequencing and consequently could not be used. Amplification of the *atpB-rbcL* spacer used primers *atpB*-1 and *rbcL*-1 (Chiang *et al.*, 1998); like the *matK* gene these amplified fragments were also unsuitable for sequencing. These primers were not

specifically developed for *Arum* species, which may account for their failure to amplify a PCR product under a range of conditions.

Amplification of the plastid *trnL* (UAA) intergenic spacer was carried out using primers c and d from Taberlet *et al.* (1991). Amplification of the *trnL* region generated DNA fragments of sufficient quality for sequencing in the subsequent phylogenetic analysis. DNA was amplified in 40µl reactions using the following conditions: 1×NH_4 based reaction buffer (Bioline), 1.5mM $MgCl_2$, 2mM dNTPs, 0.5mM of each primer and 0.2U *Taq* DNA polymerase (Biotaq, Bioline). DNA was amplified on an MJ Research DNA Engine Tetrad thermal cycler, programmed as follows: 94°C for 1 minute, 66°C for 2 minutes, and 72°C for 40 seconds repeated for 24 cycles with a final extension at 72°C for 5 minutes. The PCR product was separated according to size with the size standard Hyperladder 1 (Bioline) on a 1% agarose gel and visualised, post electrophoresis, using ethidium bromide. Successful PCR reactions were purified using QIAquick spin columns (Qiagen), which remove primers and excess dNTPs, and eluted in a volume of 30µl. Purified DNA was quantified by running a 5µl aliquot on an agarose gel and comparing with 5µl of Hyperladder 1 (Bioline).

Amplification of the nuclear rDNA cistron was initially pursued using the primers ITS1 and ITS4 (Gardes and Burns, 1993), but these primers failed to produce a consistent PCR product. The primers from Gardes and Burns (1993) were designed using fungal sequences, which probably explains their unsuitability. A second set of PCR primers developed for the genus *Gossypium* from Cronn *et al.* (2002) called 18s and 26s were used successfully to amplify the ITS1, 5.8s and ITS2 regions in a single fragment. Only

sequences from ITS1 could be used for phylogenetic analysis due to a high G+C content in the ITS2 region reducing the sequencing reliability. The 5.8s gene was found to be invariable among all of the species tested, whilst the ITS1 region was found to be variable. The rDNA cistron was amplified in 40µl reactions using the following conditions: 1× NH_4 based reaction buffer (Bioline), 1.5mM $MgCl_2$, 2mM dNTPs, 0.5mM of each primer and 0.2U *Taq* DNA polymerase (Biotaq, Bioline). DNA was amplified on an MJ Research DNA Engine Tetrad thermal cycler, programmed as follows: 94°C for 1 minute, 60°C for 2 minutes, and 72°C for 40 seconds repeated for 30 cycles with a final extension at 72°C for 5 minutes. The PCR product was separated according to size with the size standard Hyperladder 1 (Bioline) on a 1% agarose gel and visualised, post electrophoresis, using ethidium bromide. Successful PCR reactions were purified using Qiaquick spin columns (Qiagen), and eluted in a volume of 30µl. Purified DNA was quantified by running a 5µl aliquot on an agarose gel and comparing with 5µl of Hyperladder 1 (Bioline).

DYEnamic terminator cycle sequencing

Sequencing was carried out using the DYEnamic Terminator Sequencing Kit (Amersham Biosciences) in 20µl reaction volumes containing: ~40ng of purified PCR product, 3µl of DYEnamic terminator mix, 5µl DYEnamic dilution buffer and 5pmol of either the reverse (primer d for *trnL*/26s for the nuclear rDNA cistron) or the forward primer (primer c for *trnL*/18s for nuclear rDNA cistron). The cycle sequencing reaction was carried out as follows: 95°C for 20 seconds, 50°C for 15 seconds and 60°C for 1

minute for 25 cycles. Sequenced products were precipitated following the DYEnamic kit protocol. Precipitated PCR products were air-dried overnight and resuspended in 4µl of formamide loading buffer (DYEnamic kit). Two microlitres of the sequencing product were run on an MJ Research Basestation.

Sequence analysis and phylogenetic inference

Sequence base calling was checked by eye in SeqManII (DNA Star) and sequences were aligned using the ClustalX function in BioEdit (Hall, 1999; Thompson *et al.*, 1997). The computer-generated alignments were checked and, if necessary, the alignments were optimised by eye. Aligned sequences were exported into PAUP*4.0b10 (Swofford, 1998) for parsimony analysis. All parsimony analyses were performed with heuristic searches with the following parameters: ACCTRAN (accelerated transformation), TBR branch swapping and 1000 random addition sequences. Branch support was investigated using a bootstrap analysis with 1000 replicates. To test for congruence between the chloroplast and nuclear datasets a partition homogeneity test (Farris *et al.*, 1995) was performed in PAUP*4.0b10 (Swofford, 1998); if incongruence between the two datasets is not significant the datasets can be combined and analysed together.

For each tree the consistency index (CI) and retention index (RI) were calculated using PAUP*4.0b10 (Swofford, 1998). These statistics give an indication of how much homoplasy there is within the dataset. The consistency index indicates how well a

character fits a tree (Kitching *et al.*, 1998). It is calculated as the number of steps a character exhibits (m) as a proportion of the minimum number of steps it could show on any tree (s). The ensemble consistency index is the total number of steps exhibited by all characters ($M = \sum m$) as a proportion of the minimum number of steps all the characters could show on a tree ($S = \sum s$) (Kitching *et al.*, 1998). For example if the CI=0.800 the characters are 80% consistent with the phylogenetic tree. The retention index is a measure of the amount of synapomorphy within a tree. It is calculated as $(g - s)/(g - m)$ where m is the number of steps a character exhibits on a tree, s is the minimum number of steps a character could show on any tree and g is the greatest number of steps for a character on any tree (Kitching, 1998). The retention index for the whole dataset is calculated as the ensemble retention index in the same manner as CI, hence $(\sum G - \sum s) / (\sum G - \sum m)$. A value of RI=1 indicates that the similarity between taxa for the chosen character on a particular tree can be interpreted as a synapomorphy (shared derived characters) and hence the character has no homoplasy for the arrangement of taxa in the tree.

Morphological data

Morphological data were taken for each species in the genus *Arum* from the monograph by Boyce (1993) for reanalysis. The taxonomy of Boyce (1993) *a priori* assumes that different characters are more suitable at different taxonomic levels; the reanalysis of the morphological data was done without any assumptions of a character's usefulness. By not having *a priori* assumptions any bias towards a particular character is removed. There are 28 morphological characters (both vegetative and reproductive) used in the reanalysis (Table 4.5). The characteristics of the ovary, fruit and fruiting spike were not used as these characters are unknown for some species. Relationships among the species were visualised on a hierarchical UPGMA tree. The UPGMA tree was calculated using squared Euclidean distances among the individuals in PAUP*4.10b (Swafford 1998). Visualising the relationships among the individuals with a tree aids comparisons of species relationships with molecular data.

Table 4.5 Morphological characters from Boyce (1993).

Morphological character
Tuber morphology
Tuber length
Tuber width
Petiole length
Petiole width
Leaf length
Leaf width
Flag or cryptic
Peduncle length
Peduncle width
Spathe width
Spathe length
Spathe tube width
Spathe tube length
Spathe limb width
Spadix reflexed
Spadix length
Appendix length
Appendix length
Staminode width
Staminode whorl size
Staminode bristle length
Staminate flower length
Staminate flower width
Pistillodes width
Pistillodes zone size
Pistillate flower length
Pistillode bristle length

Results

Morphological relationships among *Arum* taxa

When the morphological relationships among all *Arum* species are plotted on a UPGMA tree (Figure 4.6), the topography corresponds to the two *Arum* subgenera but does not reflect the sectional classification by Boyce (1993). Two sections corresponding to the sections *Arum* and *Dioscoridea* are not found in this reanalysis. Three species from section *Arum*: *A. maculatum*, *A. italicum* and *A. concinnatum*, are clustered within the same clade as *A. dioscoridis*, *A. rupicola*, *A. purpureospathum* and *A. palaestinum* which according to Boyce (1993) belong in the section *Dioscoridea*. This clustering groups species with rhizomatous and discoid tubers together rather than separating them into discrete sections.

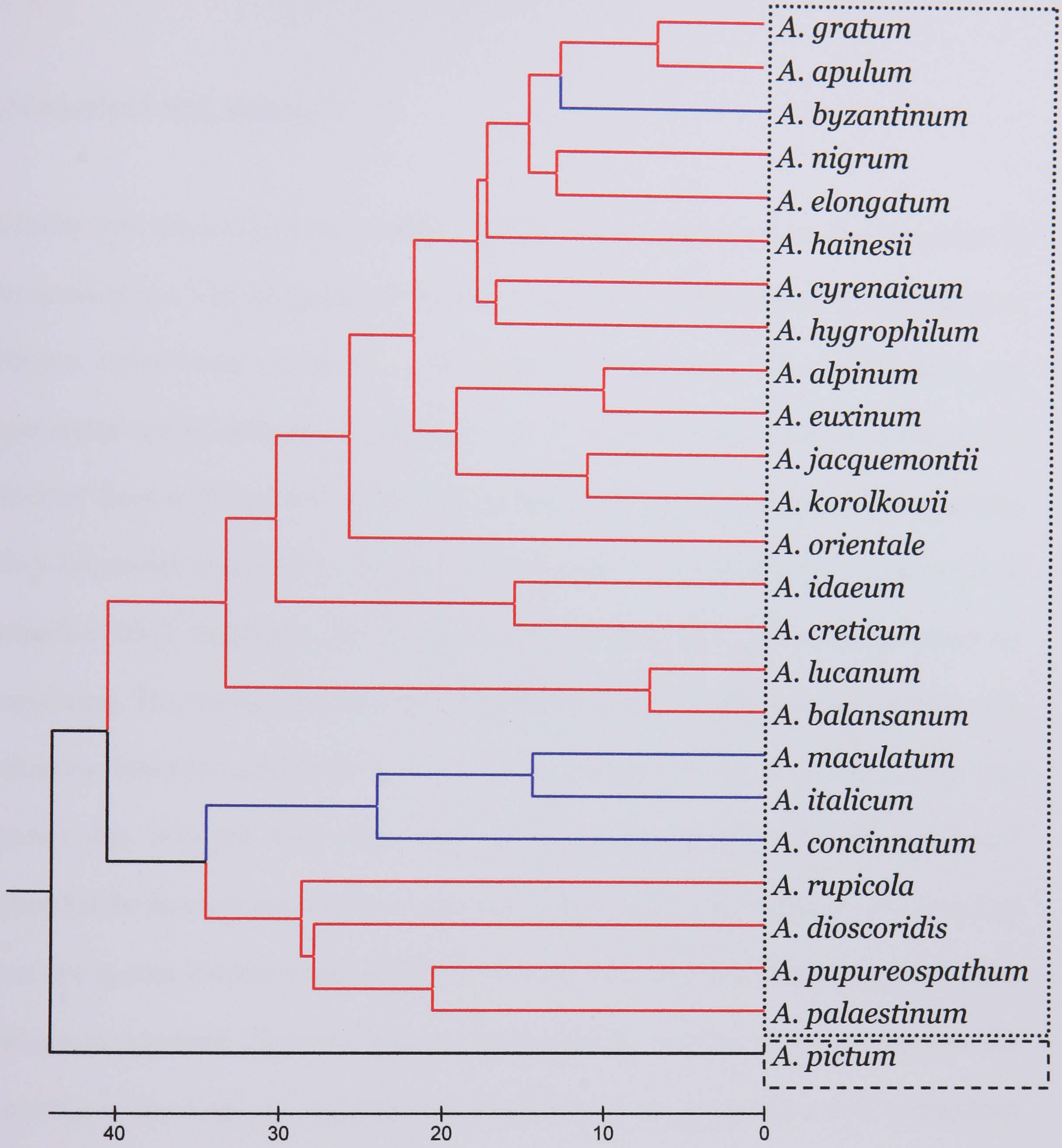


Figure 4.6 UPGMA tree of Euclidean distances calculated from 28 morphological characters for 25 *Arum* species. The boxes surrounding the species names indicate the two subgenera: subgenus *Gymnomesium* indicated with the dashed box; subgenus *Arum* indicated by the dotted box. The colour of the branches indicates the sectional separation in the subgenus *Arum*: red indicates section *Dioscoridea*; blue indicates section *Arum*. The scale bar indicates Euclidean distance.

Genetic relationships among *Arum* taxa

Chloroplast *trnL* dataset

Reliable *trnL* sequences were produced for 22 different species from the 25 species in the genus *Arum*. The alignment of the *trnL* sequence is 630bp with 63 indels (insertion deletion events) (see Appendix A for sequence alignment). The three species not represented are *A. hainesii*, *A. lucanum* and *A. nigrum.* DNA from *A. hainesii* was obtained from a 30-year-old herbarium specimen and failed to produce any amplified DNA fragments in any PCR. The two other species, *A. lucanum* and *A. nigrum*, both amplified DNA fragments, but the quality of the amplified DNA was too poor for sequencing. The reason why these two latter species failed to produce good sequences is unknown, but poor quality-DNA is an unlikely reason as the DNA from these two species was collected from fresh material. For the outgroup, *trnL* sequences were generated for *Biarum* spp. *Eminium* spp. and *Thyphonium* spp. However, the sequences from the genera *Eminium* and *Thyphonium* were excluded as inclusion resulted in an ambiguous sequence alignment because the sequences were too diverse. Therefore, for the analysis, the outgroup contained two accessions of the genus *Biarum*: *B. tenuifolium* and *B. tenuifolium* subsp. *abbreviatum.*

Parsimony analysis of *trnL* sequences yielded 980 equally parsimonious trees with a length of 203. The strict consensus tree (only clades shown in all of the 980 trees are included) is shown in Figure 4.7. The trees have a CI of 0.7438, HI of 0.2562, and RI of 0.686. The bootstrap support for clades is generally low, but all *Arum* taxa are

contained within a monophyletic clade with 100% bootstrap support. The *Arum* clade is further separated into two monophyletic clades: one monotypic clade contains *A. pictum*, and the second clade contains the remaining *Arum* species (Figure 4.7); this separation has 100% bootstrap support. Relationships within the clade corresponding to the subgenus *Arum* are not supported with bootstrap values greater than 50%. Many species, e.g. *A. dioscoridis* and *A. palaestinum*, could not be differentiated.

A 50% majority rule tree for the *trnL* data is shown in Figure 4.8. The 50% majority rule tree indicates how often clades occur when a phylogenetic analysis produces many parsimonious trees. In the 50% majority rule tree, clades that occur in 50% or more of the 980 equally parsimonious trees from the *trnL* analysis are drawn. In the tree, there is a greater resolution of the placing of the species *A. dioscoridis* and *A. palaestinum* as these species occur in a clade together in 62% of the 980 equally parsimonious trees; these species are also present in a clade with *A. byzantinum, A. italicum* and *A. concinnatum* in 56% of the 980 equally parsimonious trees. However these groups are not supported by high bootstrap values (>50%) but the consensus trees indicate that there is some affinity between these five species, which reflects the morphological relationships in Figure 4.6.

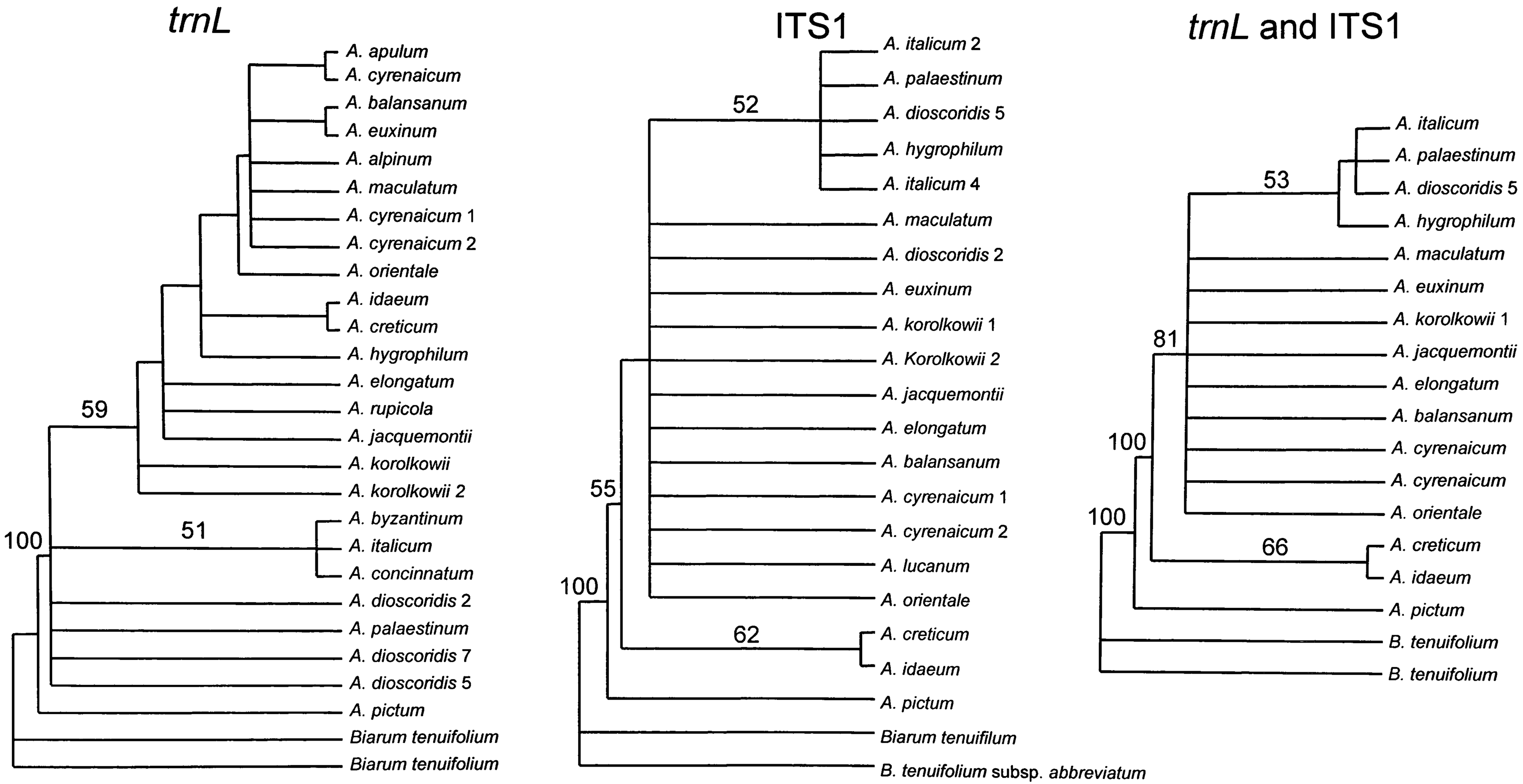


Figure 4.7 Strict consensus trees from the phylogenetic analysis of *Arum* species for *trnL*, ITS1 and combined *trnL* and ITS1 sequences. Values above the branches are bootstrap values, and only bootstrap values greater than 50% are shown.

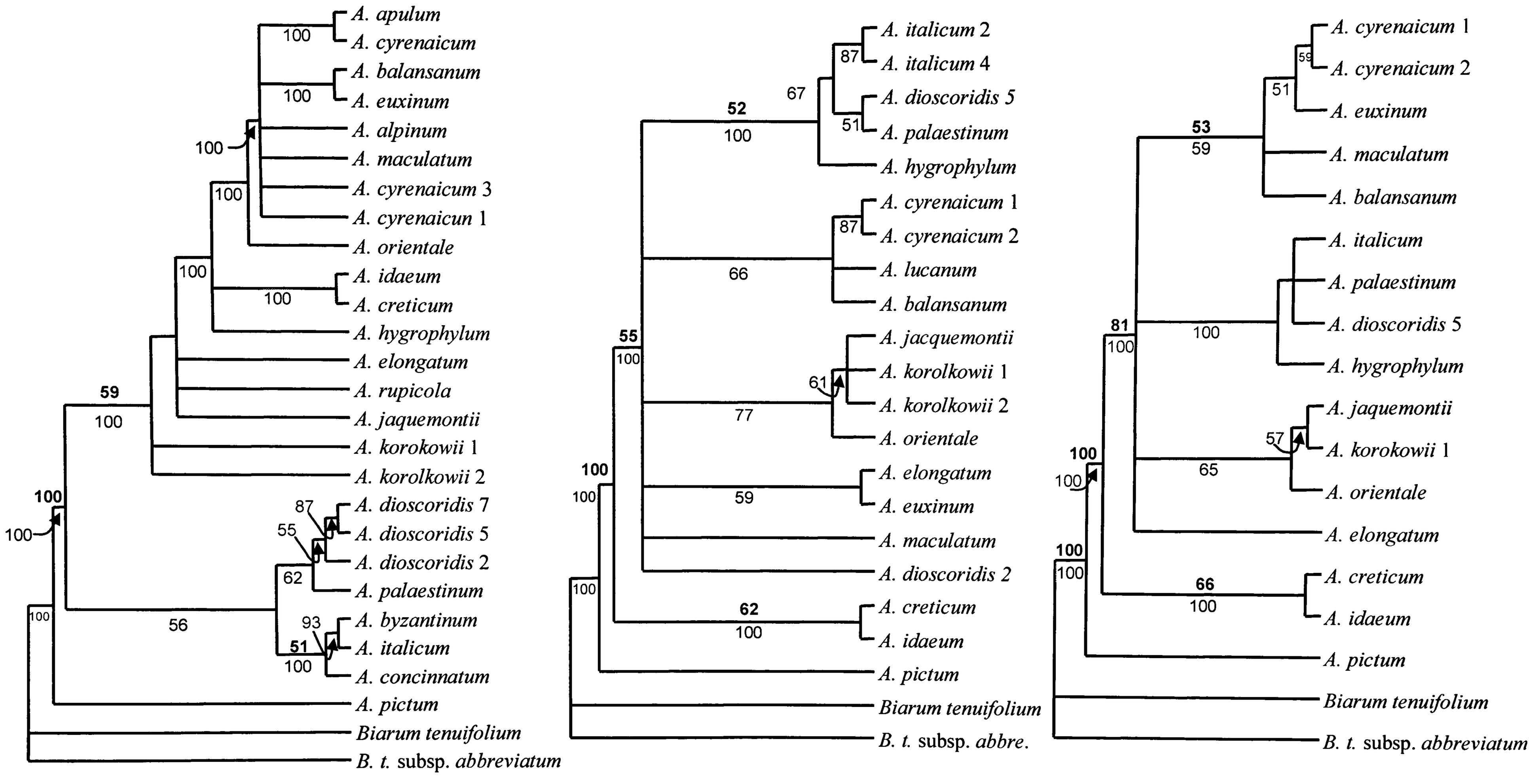


Figure 4.8 Majority rule consensus trees, only for clades present in 50% or more of the phylogenetic trees from analyses of the *trnL*, ITS1 and combined *trnL* and ITS datasets. Values above the branches are bootstrap support values. Values below the branches indicate the percentage of trees in which the clade occurs.

Nuclear ITS1 dataset

Reliable sequences of ITS1 were produced for 17 *Arum* species (see Appendix A for sequence alignment). Sequences of ITS2 were not of sufficient quality to include in the analysis due the high G+C content rendering the sequences unreadable. Technical problems with the automated sequencer towards the end of the experimental period also resulted in the lack of ITS1 sequence for seven of the eight missing species. The missing seven species are *A. concinnatum*, *A. alpinum*, *A. gratum*, *A. lucanum*, *A. nigrum*, *A. pupureospathum* and *A. rupicola*. *Arum hainesii* and *A. nigrum* failed to produce any amplified ITS DNA as was also found in the *trnL* amplification. The outgroup taxon *Biarum* was used in the phylogenetic analysis as sequences were not obtained for *Eminium*, *Arisarum* and *Dracunculus*.

The MP analysis of the ITS1 sequences produced 1000 equally parsimonious trees with a length of 91, CI=0.835, HI=0.165, RI=0.839. The CI and RI values for the ITS tree are larger than those produced from the *trnL* dataset. The higher values of CI and RI indicate that the ITS dataset has lower amounts of homoplasy than the *trnL* dataset. A strict consensus tree of the 1000 equally parsimonious trees is shown in Figure 4.7: the clades seen in this tree are present in all 1000 trees. The monophyletic *Arum* clade is supported by 100% bootstrap value. There are 10 species that cannot be differentiated with these data. The species *A. italicum*, *A. hygrophilum*, *A. palaestinum* and *A. dioscoridis* are contained within a clade supported by a 61% bootstrap value. In the 50% majority rule consensus tree (Figure 4.8), nine of the 10 undifferentiated species form

three clades found in 66%, 77% and 59% of the 1000 equally parsimonious trees. *A. maculatum* and a single accession of *A. dioscoridis* remain undifferentiated.

Combined *trnL* and ITS1 dataset

In total 16 different *Arum* species and the outgroup taxa *Biarum* have both ITS1 and *trnL* available for analysis (Table 4.6). The partition homogeneity test (Farris *et al.*, 1995) implemented within PAUP*4.0b10 (Swofford, 1998) indicated that the two sequences are not significantly incongruent (P=0.926) and therefore the two datasets (ITS1 and *trnL*) can be analysed together as there are no significant conflicts between them. The joint MP analysis yields 316 equally parsimonious trees with a length of 173, CI=0.878, HI=0.121, RI=0.812 (Figure 4.7). This indicates that analysing both sequence regions together has been accompanied by a small decrease in homoplasy which can be accounted for by the removal of some species. All of the *Arum* taxa form a monophyletic clade with 100% bootstrap support. *A. pictum* is distinct from the other *Arum* taxa with 100% bootstrap support. Within the *Arum* clade, a monophyletic clade containing the species *A. idaeum* and *A. creticum* is supported by a bootstrap value of 66%, and the remaining *Arum* species form a monophyletic clade with 81% bootstrap support. Within this clade the species *A. italicum*, *A. palaestinum*, *A. dioscoridis* and *A. hygrophilum* are distinct from the other species in a clade with 53% bootstrap support, and the remaining species are undifferentiated. In the 50% majority consensus tree (Figure 4.8) a clade containing *A. jacquemontii*, *A. korolkowii* and *A. orientale* is found in 65% of the 316 most parsimonious trees. A clade containing *A. cyrenaicum*, *A.*

euxinum, *A. maculatum* and *A. balansanum* is found in 59% of the 316 most parsimonious trees.

Table 4.6 Species of *Arum* with *trnL* and ITS1 sequences. The black box indicates good quality sequence obtained for that species.

Species	*trnL*	ITS1
A. maculatum	■	■
A. byzantinum	■	■
A. italicum	■	■
A. concinnatum	■	
A. alpinum	■	
A. orientale	■	■
A. gratum	■	
A. lucanum		■
A. nigrum		
A. cyrenaicum	■	■
A. pupureospathum	■	
A. balansanum	■	■
A. hainesii		
A. elongatum	■	■
A. rupicola	■	
A. jacquemontii	■	■
A. korolkowii	■	■
A. euxinum	■	■
A. hygrophilum	■	■
A. dioscoridis	■	■
A. palaestinum	■	■
A. idaeum	■	■
A. creticum	■	■
A. pictum	■	■

Correlation of morphological and phylogenetic relationships

The topologies of the re-analysed morphological data from Boyce (1993) and the new molecular data presented herein show both congruence and incongruence (Figure 4.9). Both datasets indicate that *A. pictum* is distinct from all other *Arum* taxa, thus supporting the two subgenera *Gymnomesium* and *Arum*. *Arum idaeum* and *A. creticum* also form a monophyletic clade in both the genetic data and the morphological data from Boyce (1993). The molecular dataset places *A. creticum* and *A. idaeum* together, agreeing with the taxonomy of Boyce (1993). However, the molecular dataset places *A. creticum* and *A. idaeum* basal to all *Arum* species with the exception of *A. pictum*; this relationship is not supported by the morphological data as *A. creticum* and *A. idaeum* are placed within the *Arum* species (Figure 4.9). The molecular data indicate that the species *A. hygrophilum*, *A. italicum*, *A. dioscoridis* and *A. palaestinum* form a monophyletic group, whereas these species form a paraphylectic group using the morphological data (Figure 4.9). There is also discordance in the placing of *A. maculatum* and *A. hygrophilum*. The molecular dataset places *A. maculatum* within a monophyletic group consisting of the species *A. balansanum*, *A. euxinum* and *A. cyrenaicum*, whereas *A. maculatum* and *A. italicum* are placed in a monophyletic group using the morphological dataset. The molecular data indicate that *A. hygrophilum* has a closer affinity with *A. italicum*, *A. dioscoridis* and *A. palaestinum* (Figure 4.9), whereas the morphological dataset places *A. hygrophilum* with *A. cyrenaicum*. However, the molecular dataset has poor statistical support for the placing of species and therefore no firm conclusions on the relationships among the species can be made. However, both re-analysed morphological and molecular data have clades that do not correspond to the

traditional sections *Arum* and *Dioscoridea.* The incongruence between the datasets and the classification increases the likelihood that these sections do not correspond with evolutionary history and therefore questions their continued use.

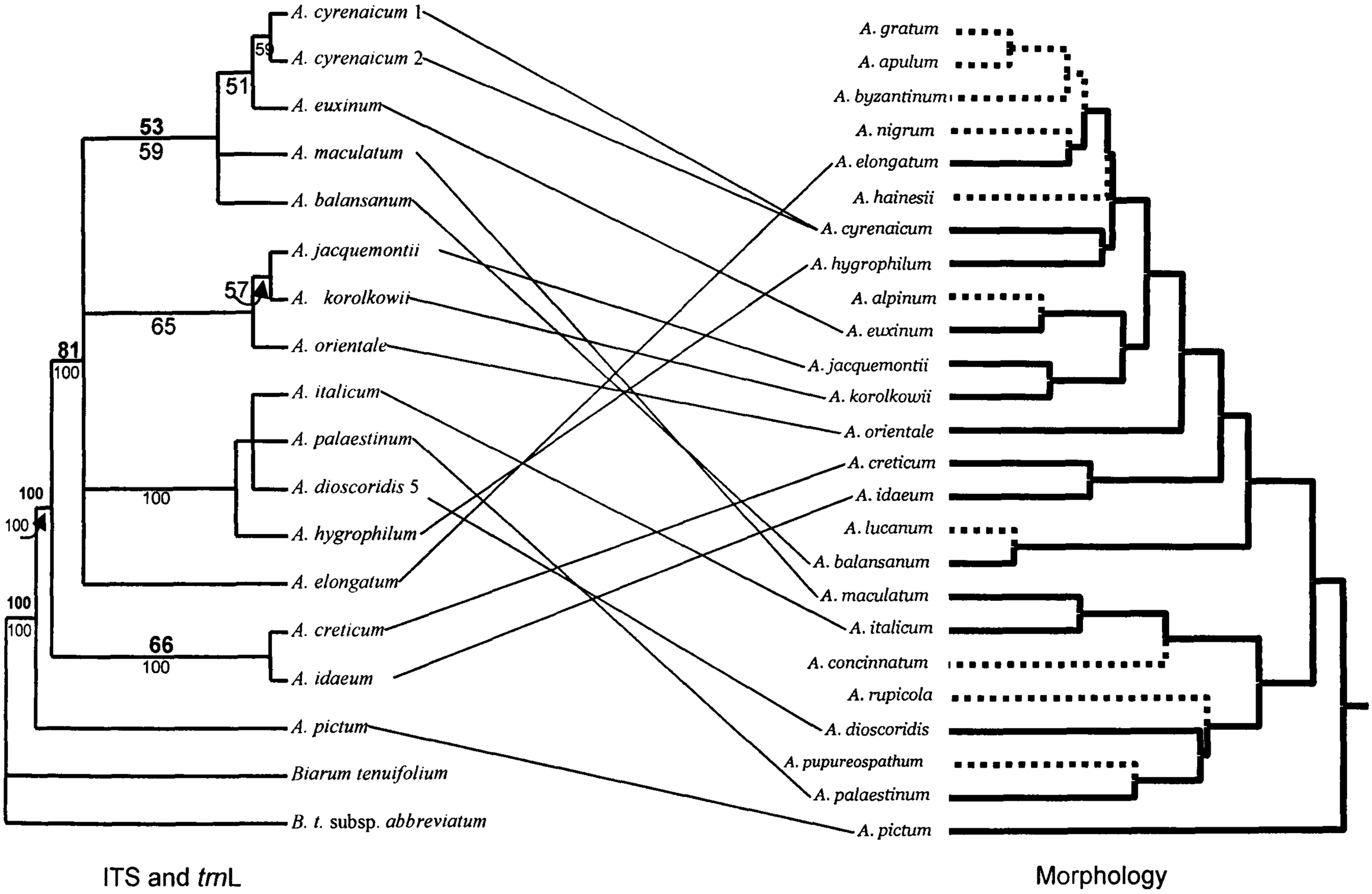


Figure 4.9 Topological incongruence between the reanalysed morphology data of Boyce (1993) and the combined analysis of *trnL* and ITS1 data. Dashed lines in the morphology tree indicate that these species were not present in the molecular analysis.

Comparing morphological characters that are used in creating the taxonomy to the molecular phylogeny can identify characters that correlate with real evolutionary groups and those that do not (Figure 4.10). The character of leaf growth time is used to differentiate between two subgenera within the *Arum* genus. The subgenus *Gymnomesium* is monotypic containing a single species *A. pictum* in which the inflorescence matures before the leaves, whilst in all other *Arum* species the leaves mature before the emergence of the inflorescence. The character state of inflorescence maturation after leaf maturation is a synapomorphy that unites all other *Arum* species in a monophyletic clade (Figure 4.10).

The sections defined with tuber morphology by Boyce (1993) are not supported by the molecular data or the reanalysed morphological data. If tuber morphology is used to group species the section *Arum* recognised by Boyce (1993) (containing *A. maculatum*, *A. italicum*, *A. concinnatum* and *A. byzantinum*) would be polyphyletic (Figure 4.10) and the section *Dioscoridis* (20 species) paraphyletic (Figure 4.10). The genetic data support two sections, a first containing the species *A. idaeum* and *A. creticum* and a second containing all other members of the subgenus *Arum*. This new sectional split is supported by the characters of pistillodes and staminodes as *A. idaeum* and *A. creticum* both lack the derived characters of staminodes.

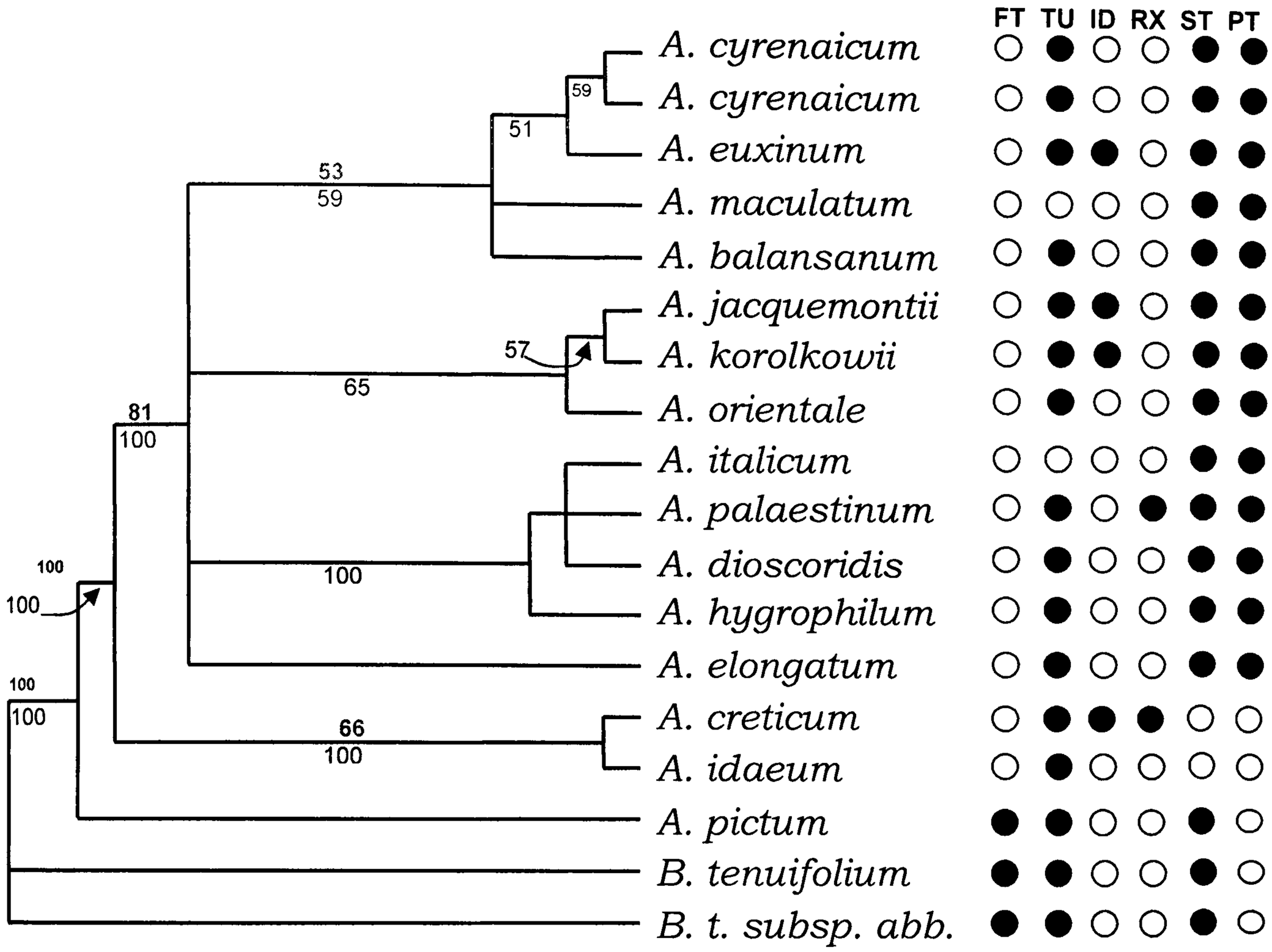


Figure 4.10 The MP tree produced from the combined *trnL* and ITS datasets. The distributions of six morphological characters used in the classification of *Arum* are indicated. The characters are FT: flowerng time, solid black circles indicate that the inflorescences mature before leaf maturation whereas a white circle indicates that the inflorescence matures after the leaves have matured. TU: tuber character state, a solid black circle indicates that the tuber state is discoid whereas a white circle indicates the tuber is horizontal rhizomatous. ID: inflorescence display mode, a solid black circle indicates that the peduncle is longer than the petiole (flag display) whereas a white circle indicates that the peduncle is shorter than the petiole (cryptic display). RX: reflexed spathe, a solid black circle indicates that the spathe is reflexed whereas a white circle indicates that the spathe is erect. ST: staminodes, a solid black circle indicates that staminodes are present whereas a white circle indicates staminodes are absent. PT: pistillodes, a solid black circle indicates that pistillodes are present whereas a white circle indicates pistillodes are absent.

Discussion

Molecular and morphological characters can both be used to infer phylogenetic relationships among organisms. When phylogenetic relationships are known, taxonomies can be tested. This chapter investigates the correlation of morphology and genetic data among the species in the genus *Arum* by asking the following questions: 1. Does the morphological taxonomy correspond with the molecular phylogeny? 2. Which morphological characters correlate with genetic lineages? 3. If the morphological data are analysed with no *a priori* assumptions are they congruent with the genetic data?

The taxonomy of the genus *Arum* described by Boyce (1993) is generally supported by the genetic data but there are some notable exceptions. The genetic data support the division of *Arum* into two subgenera: *Arum* and *Gymnomesium*, as defined by flowering time and other vegetative characters (Boyce, 1993). The morphological differences between the subgenera *Arum* and *Gymnomesium* have arisen only once within the genus and therefore flowering time and presence of cataphylls are useful morphological characters. Using molecular data, divisions within the subgenus *Arum* are, however, less well resolved. The statistical support for the groups within the subgenus *Arum* are not adequate as bootstrap values are low and fall outside the 95% (0.05) confidence level and are therefore not considered statistically significant (Felsenstein, 1985).

The low level of bootstrap support for the phylogenetic relationships could be due to low levels of parsimony-informative sites (sequence differences that are informative because they are neither homoplasious nor uninformative by occurring within a single

taxon) within the dataset. The proportion of parsimony-informative sites is 16% for ITS1 sequences and 4.6% for *trnL*. The levels of parsimony-informative sites revealed here are low compared with a recent review of phylogenies which calculated an average of 24% parsimony-informative characters for ITS (ITS1 and ITS2) and 14.8% for *trnL* studies (Alvarez and Wendel, 2003). The low variation in the molecular dataset can account for the poor differentiation among some species.

The poor differentiation among the species could also be compounded by homoplasy. If homoplasious characters are shared by taxa they will make taxa appear more closely related than they actually are. High levels of homoplasy could potentially swamp useful information from synapomorphic characters (shared derived characters). The amount of homoplasy within both datasets is, however, low as both the chloroplast *trnL* dataset and the ITS1 datasets have high CI and RI values. It is therefore likely the small numbers of variable sites are affecting the bootstrap values. Even though bootstrap values are low, there is general congruence among the datasets (*trnL* and ITS1), as the clades corresponding to the two subgenera are retained in the strict consensus tree, and supported by bootstrap values of 100% in the *trnL*, ITS1 and combined analyses. The difference in morphology recognised in the taxonomy between the two subgenera *Arum* and *Gymnomesium* reflects real evolutionary differences in both the chloroplast and the nuclear genomes; therefore the molecular data support the taxonomic difference between the two subgenera described by Boyce (1993).

The differences between the classification by Boyce (1993) and the reanalysis of the morphological data from Boyce's (1993) monograph can be accounted for by two subtle

differences in each treatment. The point of Boyce's (1993) monograph was to produce a hierarchical classification of the *Arum* taxa. Such an endeavour orders the species based upon morphological resemblance. To produce a taxonomic treatment Boyce (1993) decided *a priori* that some characters are more useful at a particular taxonomic level than others. For instance, tuber morphology is used at the sectional level whereas, at the subsectional level, differences in floral morphology are used as a primary tool to order taxa. Conversely, for the reanalysis of the data no *a priori* purpose was present, and in addition there was no weighting of the relative importance of the morphological characters. These fundamental differences in the analysis methods have resulted in the differences in topology between reanalysed morphological data and the taxonomic treatment.

Phylogenies can be used to identify convergent characters such as fruit morphology in the polyphyletic genus *Tournefortia* (Diane *et al.*, 2002) and produce a classification that reflects the evolution of the studied taxa, e.g. *Veratrum* (Zomlefer *et al.*, 2003). The phylogenetic relationships based on both DNA regions and the reanalysis of Boyce's (1993) morphological data do not support the taxonomic differences in the subgenus *Arum*. The monophyly of the sections *Arum* and *Dioscoridea* is also not supported by the molecular genetic data. There is limited statistical support for relationships among the species within the subgenus *Arum* but the molecular data suggest that the characters tuber shape and inflorescence display (flag/cryptic) have arisen more than once within the subgenus.

The data indicate that the ancestral tuber character state is discoid; all basal members of the genus, *A. pictum*, *A. creticum* and *A. idaeum*, have discoid tubers (Figure 4.10). The horizontal rhizomatous tuber character state has independently evolved in two lineages within the genus. In the molecular tree a single character state change from discoid to rhizomatous can group the three species *A. italicum*, *A. byzantinum* and *A. concinnatum*, whilst a second independent development occurred in the *A. maculatum* lineage. As tuber shape has potentially arisen more than once within the genus, it is not a useful character for creating a classification that reflects evolutionary history in *Arum*.

The combined molecular dataset indicates that *A. creticum* and *A. idaeum* are basal to all species within the subgenus *Arum*. The position of *A. creticum* and *A. idaeum* is supported by two morphological characters, pistillodes and staminodes, that are absent in *A. idaeum* and *A. creticum*. The absence of pistillodes and staminodes is an ancestral character state, as these characters are also absent in *A. pictum* (basal to all *Arum* species in all three datasets) and the genus *Biarum*. Pistillodes and staminodes are shared derived characters (synapomorphy) uniting all other members of genus *Arum*. The presence of pistillodes and staminodes could be used to create two new sections, the first containing *A. idaeum* and *A. creticum* and the second containing the remaining members of the section *Dioscoridea* and *Arum*. Interestingly, the morphological and *trnL* data place *A. creticum* and *A. idaeum* within the subgenus *Arum* not basal in the subgenus *Arum*.

The incongruent placing of *A. creticum and A. idaeum* among the individual trees can occur for several reasons, such as sampling error, lineage sorting and reticulation.

Sampling error can introduce conflicts between gene trees (trees based in a single locus, .g. the ITS1 tree) through errors in sequencing (data collection) or misidentification of samples (Wendle and Doyle, 1998). Sampling error is hard to identify, but appropriate steps were taken to ensure that plants sampled were not misidentified and that the sequence data from the samples were accurate. Inaccurate sequence data were excluded from the study, which resulted in discarding ITS1 sequences from eight species and not using ITS2 sequences at all.

Lineage sorting occurs when gene lineages diverge prior to speciation (Wendle and Doyle, 1998). Phylogenies based on two or more different gene sequences can display incongruence if the genes diverged at different times. For example, monocots are paraphyletic in 18s gene trees (e.g. Bharthan and Zimmer, 1995) due to the position of the Acoraceae and the Ceratophyllaceae (Duvall and Ervin, 2004). However, Duvall and Ervin show that the cause of the incongruence in placing the Acoraceae and the Ceratophyllaceae is likely to be due to lineage sorting affecting the 18s gene trees. Incongruence between trees can also be caused by reticulation. Reticulation is hybridisation between two diverged lineages, often leading to chloroplast capture (Rieseberg and Soltis, 1991). When this occurs it can result in a chloroplast genome that is different from the nuclear genome. Distinguishing between the effects of reticulation and lineage sorting requires species relationships to be statistically supported and have adequate sampling of species (Rokas *et al.*, 2003; Wendle and Doyle, 1998). The *Arum* data do not resolve differences among species with a high statistical confidence, and therefore it is impossible to identify the reasons for incongruence between the chloroplast and nuclear datasets.

As morphological characters are potentially under selection, it is possible that similar biological solutions can arise independently from one another, such as convergent evolution of asymmetric flowers and leaf rosettes in Commelinidae (Givnish *et al.*, 1999). This is especially true for reproductive characters, as those individuals with the best pollinator attraction characters will be the most successful, consequently producing more offspring (Lovett-Doust and Lovett-Doust, 1988). For example, floral guides have evolved independently in many species (Lunau, 2004). If convergent evolution of successful reproductive characters is likely, taxonomists should be careful when using reproductive characters.

Reproductive characters are considered the most important for distinguishing among species (Stebbins, 1970; Boyce, 1993). Because reproductive characters are involved with gene flow, any differences in their morphology are assumed to represent a change in reproductive behaviour and therefore an evolutionary change that may lead to lineages diverging. If convergent characters are used as a tool to group lineages together, a classification is produced that does not reflect the real evolutionary differences between two lineages. This investigation has shown that there are several convergent reproductive characters in the genus *Arum.* The characters flowering height and flower colour are shown to have arisen several times within the genus and are therefore convergent. This indicates that character usefulness in indicating evolutionary differences should not be assumed *a priori*, and no character type can be considered to be the best until it has been tested.

Some but not all, of the morphological characters used in the taxonomy of *Arum* by Boyce (1993) correlate with genetic lineages. It has been shown that flowering chronology, which is used to distinguish between the two *Arum* subgenera, correlates with two genetic lineages. Morphological characters within the subgenus *Arum* have been shown generally not to correlate with evolutionary lineages. The data also suggest that some of the morphological characters (e.g. flowering height) have arisen more than once and are convergent. It is likely that convergence of the floral characters within the genus *Arum* is through competition for pollinators, a conclusion supported by the widespread convergence of flower characters in angiosperms.

References

Alvarez I, Wendel JF (2003) Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution* **29**, 417–434.

Beilstein MA, Windham MD (2003) A phylogenetic analysis of Western North American *Draba* (Brassicaceae) based on nuclear ribosomal DNA sequences from the ITS region. *Systematic Botany* **28**, 584–592.

Bharthan G, Zimmer EA (1995) Early Branching events in monocotyledons partial 18s ribosomal DNA sequence analysis. In: *Monocotyledons: Systematics and evolution* (eds Rudall PJ, Cribb PJ, Cutler DF, Humphries CJ), pp. 81–107. Royal Botanic Gardens, Kew.

Boyce P (1989) A new classification of *Arum* with keys to the infrageneric taxa. *Kew Bulletin* **44**, 383–395.

Boyce P (1993) *The genus Arum. A Kew Magazine Monograph*. HMSO, London.

Chiang T-Y, Schaal B, Peng C-I (1998) Universal rrimers for amplification and sequencing a noncoding spacer between the *atpB* and *rbc*L genes of chloroplast DNA. *Botanical Bulletin of the National Academy Sincia* **39**, 245–250.

Cronn RC, Small B, Haselkron T, Wendel JF (2002) Rapid diversification of the cotton genus *Gossipium* (Malvaceae): revealed by sixteen nuclear and chloroplast genes. *American Journal of Botany* **89**, 707–725.

Diane N, Forther H, Hilger HH (2002) A systematic analysis of *Heliotropium tournefortia,* and allied taxa of the Heliotropiaceae (Boraginales) based on ITS1 sequences and morphological data. *American Journal of Botany* **89**, 287–295.

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

Duvall MR, Ervin AB (2004) 18s gene trees are positively misleading for monocot/dicot phylogenetics. *Molecular Phylogenetics and Evolution* **30**, 97-106.

Farris SJ, Kallersjo M, Kluge AG, Bult C (1995) Testing significance of incongruence. *Cladistics* **10**, 315–319.

Felsenstein J (1985) Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39**, 783–791.

Gardes M, Burns TD (1993) ITS primers with enhanced specificity for Basidiomycetes - application to the identification of mycorrhizae and rusts. *Molecular Ecology* **2**, 113–118.

Givnish TJ, Evans TM, Pires JC, Sytsma KJ (1999) Polyphyly and convergent morphological evolution in Commelinales and Commelinidae: Evidence from rbcL sequence data. *Molecular Phylogenetics and Evolution* **12**, 360–385.

Grayum MH (1990) Evolution and phylogeny of the Araceae. *Annals of the Missouri Botanical Garden* **77**, 628–697.

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

Hayasaka K, Gojobori T, Horai S (1988) Molecular phylogeny and evolution of primate mitochondrial DNA. *Molecular Biology and Evolution* **5**, 626–644.

Helfgott DM, Francisco-Ortega J, Santos-Guerra A, Jansen RK, Simpson BB (2000) Biogeography and breeding system evolution of the woody Bencomia alliance (Rosaceae) in Macaronesia based on ITS sequence data. *Systematic Botany* **25**, 82–97.

Hennig W (1965) Phylogenetic systematics. *Annual Review of Entomology* **10**, 97–116.

Hey J, Waples RS, Arnold ML, Butlin RK, Harrison RG (2003) Understanding and confronting species uncertainty in biology and conservation. *Trends in Ecology & Evolution* **18**, 597–603.

Hillis DM, Mable BK, Moritz C (1998) Applications of molecular systematics: The state of the field and a look to the future. In: *Molecular Systematics* (eds Hillis DM, Moritz C, Mable BK), pp. 515–543. Sinauer Associates.

Kitching IJ, Forey PL, Humphries CJ, Williams DM (1998) *Cladistics: The Theory and Practice of Parsimony Analysis*, 2nd edn. Oxford University Press, Oxford.

Kluge AG, Farris FJ (1969) Quantitative phylectics and the evolution of Anurans. *Systematic Zoology* **18**, 1–32.

Lovett-Doust J, Lovett-Doust L (1988) *Plant Reproductive Biology: Patterns and Strategies*. Oxford University Press Inc, USA.

Lunau K (2004) Adaptive radiation and coevolution - pollination biology case studies. *Organisms Diversity and Evolution* **4**, 207–224.

Mayden RL (1997) A hierarchy of species concepts: the denounement in the saga of the species problem. In: *Species: The Units of Biodiversity* (eds Claridge MF, Dawah HA, Wilson MR). Chapman and Hall, New York.

Neves SS, Watson MF (2004) Phylogenetic relationships in Bupleurum (Apiaceae) based on nuclear ribosomal DNA ITS sequence data. *Annals of Botany* **93**, 379–398.

Page RD, Holmes EC (1998) *Molecular Evolution: A Phylogenetic Approach.* Blackwell Science Ltd, Oxford.

Pagel M (1999) Inferring the historical patterns of biological evolution. *Nature* **401**, 877–884.

Patterson C, Williams DM, Humphries CJ (1993) Congruence between molecular and morphological phylogenies. *Annual Review of Ecology and Systematics* **24**, 153–188.

Rieseberg LH, Soltis DE (1991) Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary Trends in Plants* **5**, 65–84.

Rokas A, Melika G, Abe Y, Nieves-Aldrey J, Cook JM, Stone G (2003) Lifecycle closure, lineage sorting and hybridization revealed in a phylogenetic analysis of European oak gallwasps (Hymenoptera: Cynipidae: Cynipini) using mitochondrial sequence data. *Molecular Phylogenetics and Evolution* **26**, 36–45.

Savolainen V, Chase M (2003) A decade of progress in plant molecular phylogenetics. *Trends in Genetics* **19**, 717–724.

Shaw AJ, Allen B (2000) Phylogenetic relationships, morphological incongruence, and geographic speciation in the Fontinalaceae (Bryophyta). *Molecular Phylogenetics and Evolution* **16**, 225–237.

Simpson GG (1961) *Principals of Animal Taxonomy*. Columbia University Press, New York.

Sites JW, Marshall JC (2003) Delimiting species: a Renaissance issue in systematic biology. *Trends in Ecology & Evolution* **18**, 462–470.

Smedmark JEE, Eriksson T (2002) Phylogenetic relationships of Geum (Rosaceae) and relatives inferred from the nrITS and trnL-trnF regions. *Systematic Botany* **27**, 303–317.

Soltis DE, Soltis PS (1998) Choosing an approach and an appropriate gene for phylogenetic analysis. In: *Molecular Systematics of Plants II: DNA Sequencing* (ed Soltis DE, Soltis, PS and Doyle, JJ), pp. 1–42. Kluwer Academic Publishers, Boston.

Stace CA (1989) *Plant Taxonomy and Biosystematics*, 2nd edn. Cambridge University Press, Cambridge.

Stebbins GL (1970) Adaptive radiation of reproductive characteristics in angiosperms, I: pollination mechanisms. *Annual Review of Ecology and Systematics* **1**, 307–326.

Swofford DL (1998) PAUP*: phylogenetic analysis using parsimony (*and other methods) version 4. Sinauer, Sunderland, Massachusetts, USA.

Swofford DL, Olsen GJ, Waddell PJ, Hillis DM (1996) Phylogenetic inference. In: *Molecular Systematics* (eds Hillis DM, Moritz C, Mable BK), pp. 407–514. Sinauer Associates.

Swofford DL, Sullivan J (2003) Phylogeny inference based on parsimony and other methods using PAUP*. In: *The Phylogenetic Handbook: A Practical Approach to DNA and Protein Phylogeney* (eds Salemi M, Vandamme AM), pp. 160–206. Cambridge University Press.

Taberlet P, Giellt L, Pauto G, Bouvet J (1991) Universal primers for the amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology* **17**, 1105–1109.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **24**, 4876–4882.

Wagstaff SJ, Dawson MI (2000) Classification, origin, and patterns of diversification of *Corynocarpus* (Corynocarpaceae) inferred from DNA sequences. *Systematic Botany* **25**, 134–149.

Wendle JF, Doyle JJ (1998) Phylogenetic incongruence: window into genome history and molecular evolution. In: *Molecular Systematics of Plants II* (eds Soltis DE, Soltis PS, Doyle JJ), pp. 265–296. Chapman and Hall.

Wiley EO (1978) The evolutionary species concept reconsidered. *Systematic Zoology* **41**, 17-26.

Wolfe AD, Liston A (1998) Contributions of PCR-based methods to plant systematics and evolutionary biology. In: *Molecular Systematics of Plants II: DNA Sequencing* (ed Soltis DE, Soltis, PS and Doyle. JJ), pp. 43–86. Kluwer Academic Press, Boston.

Yockteng R, Ballard HE, Mansion G, Dajoz I, Nadot S (2003) Relationships among pansies (Viola section Melanium) investigated using ITS and ISSR markers. *Plant Systematics and Evolution* **241**, 153–170.

Zomlefer WB, Whitten WM, Williams NH, Judd WS (2003) An overview of *Veratrum* s.l. (Liliales: Melanthiaceae) and an infrageneric phylogeny based on ITS sequence data. *Systematic Botany* **28**, 250–269.

Chapter 5 Interspecific morphological variation and genetic marker variation in hybrid zones

Abstract

Intermediate morphological characters are often used as diagnostic characters for investigating patterns of hybridisation. However, morphologically intermediate characters can arise through processes other than hybridisation and may therefore be misleading. Molecular characters can be used to identify occurrences of hybridisation without the limitations of morphological characters. This chapter investigates the correlation of morphological and genetic characters in a putative hybrid zone between two *Arum* species. The two species, *A. creticum* and *A. idaeum*, coexist on the island of Crete, and individuals with intermediate morphology have been found. ISSR markers are used to identify if the morphologically intermediate individuals are hybrids and if the intermediate morphological characters correlate with molecular marker variation.

Arum creticum and *A. idaeum* are shown to be genetically distinct, whilst the morphological intermediates are genetically intermediate, confirming the presence of *A. creticum* × *A. idaeum* hybrids. The data indicate that introgression of *A. idaeum* genetic characters into *A. creticum* is occurring and is supported by the different reproductive strategies of these species.

There is a statistically significant correlation between continuous morphological character variation and molecular marker variation, but categorical characters did not reflect genetic patterns successfully. In conclusion, continuous morphological characters are a reliable indicator of hybridisation in *A. creticum* and *A. idaeum* as they

follow patterns of molecular marker variation. Categorical characters are, however, less reliable for indicating genetic variation in hybrid zones. When morphological characters are used in hybrid zones they first must be tested for correlation with molecular markers variation before using them as an indicator of hybridisation.

Introduction

Intermediacy of morphological characters has been used as an indicator of hybridisation in many taxa, e.g. *Salvia apiana* and *S. mellifera* (Meyn and Emboden, 1987), *Cypripedium candidum* and *C. pubescens* (Klier *et al.*, 1991) and *Polystichum munitum* and *P. imbricans* (Mullenniex *et al.*, 1998). However, a review of 46 hybrid studies by Rieseberg and Ellstrand (1993) showed that hybrids are no more likely to express intermediate characteristics than they are to express parental characteristics, and therefore morphological intermediacy may not be a reliable indicator of hybridisation.

There are several factors that reduce the utility of morphological intermediacy as a tool for diagnosing hybridisation. Morphological intermediacy can arise through processes other than hybridisation, such as convergence or retention of ancestral characters (Gottlieb, 1972; Rieseberg, 1997). Therefore, if intermediate characters can arise without hybridisation, intermediate characters cannot be used to diagnose hybrid zones with certainty. The expression of morphological characters in hybrids is not fully understood and further complicates their use. Morphological characters that have two states (presence/absence) may be controlled by a single gene (Gottlieb, 1984) and inheritance patterns may be more predictable. This is in contrast to quantitative morphological characters, which are often products of several genes (Gottlieb, 1984; Rieseberg and Ellstrand, 1993). In quantitative morphological characters the interaction between genes from different species is unknown and consequently inheritance and expression patterns are complex.

Natural selection can also limit the utility of morphological characters in the identification of hybrids. Selection can result in genetically intermediate populations that are dominated by a parental phenotype. Such a situation exists in *Quercus affinis* × *Q. laurina*, in which molecular markers were used in addition to morphological characters to show that genetically intermediate and introgressed *Q. affinis* individuals occur in the area where the species overlap. But the leaf morphology of the genetic intermediates was predominantly of *Q. laurina* type (Gonzalez-Rodriguez *et al.*, 2004). The results of the *Q. affinis* × *Q. laurina* study suggest that natural selection on leaf morphology has caused the incongruence between morphology and genetic markers. Similarly, molecular markers and morphological characters were used to investigate a *Polystichum* hybrid zone. Hybrids of *Polystichum minitum* and *P. imbricans* with intermediate morphology were found, but only in intermediate habitats (Ketner and Mesler, 2000). These studies indicate that selection can cause morphological characters to display a pattern that is not congruent with the amount of hybridisation or direction of introgression.

Molecular markers are able to identify hybrids unambiguously and are not subject to problems associated with morphological characters (Rieseberg and Ellstrand, 1993). A range of hybrids have been identified using molecular markers, e.g. introgression of molecular characters from *Helianthus debilis* subsp. *cucumerifolius* into *H. annus* subsp. *texanus* (Rieseberg *et al.*, 1990) and conformation of the hybrid origin of species from the genus *Penstemon* (Wolfe *et al.*, 1998). After screening hybrid zones with genetic markers, additional information provided by morphological characters can clarify any patterns observed. For example, morphological characters in a *Sphagnum*

hybrid zone allowed Cronberg and Natcheva (2002) to estimate sex ratios which would otherwise been impossible using genetic markers.

In this chapter, a comparative analysis of evolutionary neutral genetic markers and morphological characters will attempt to identify if morphological intermediacy can be used as an indicator of hybridisation in the genus *Arum*. The species studied are *A. creticum* and *A. idaeum*. These two species have a close affinity (confirmed in Chapter 4), and form a separate subsection within the subgenus *Arum*: subsection *Cretica* (Boyce, 1989, 1993). *A. creticum* and *A. idaeum* are distinct from the other members of the subgenus *Arum* by their shared absence of sterile male and female flowers (Boyce, 1989, 1993). The sterile flowers in other *Arum* species produce bristles, which may have an important role in pollinator selection (A. Diaz pers. comm.). *A. idaeum* is endemic to the island of Crete, whereas *A. creticum* is found in both the south-western corner of Turkey and in Crete (Boyce 1993). The distributions of both species only overlap on the island of Crete. The two species are morphologically very different (Figure 5.1). The colour of the inflorescence is different, yellow in *A. creticum* but cream in *A. idaeum* (Figure 5.1). The length and colour of the appendix also differs between the two species: in *A. creticum* it is yellow and exposed as the spathe hood is reflexed, whereas in *A. idaeum* the appendix is dark purple and enclosed by an erect spathe hood (Figure 5.1). There are also differences in the shape of the inflorescence: the spathe tube in *A. creticum* has straight sides with little or no constriction, whilst *A. idaeum* has an obvious constriction (Figure 5.1)

There are two known sympatric populations of *A. creticum* and *A. idaeum* on the island of Crete (A. Diaz, pers. comm.), the first on Mount Psilorities and the second on Mount Kedros. On these two mountains morphologically intermediate individuals exist (Boyce, 1993). The morphologically intermediate plants potentially represent hybrid individuals *A. creticum* × *A. idaeum*. To understand patterns of morphological and genetic variation in the putative hybrid zones, the following questions will be investigated: 1. What morphological variation exists in the putative hybrid zones? 2. Are *A. idaeum* and *A. creticum* genetically distinct? 3. Are morphologicaly intermediate plants also genetically intermediate? 4. Can morphological characters be used as an indicator of hybridisation?

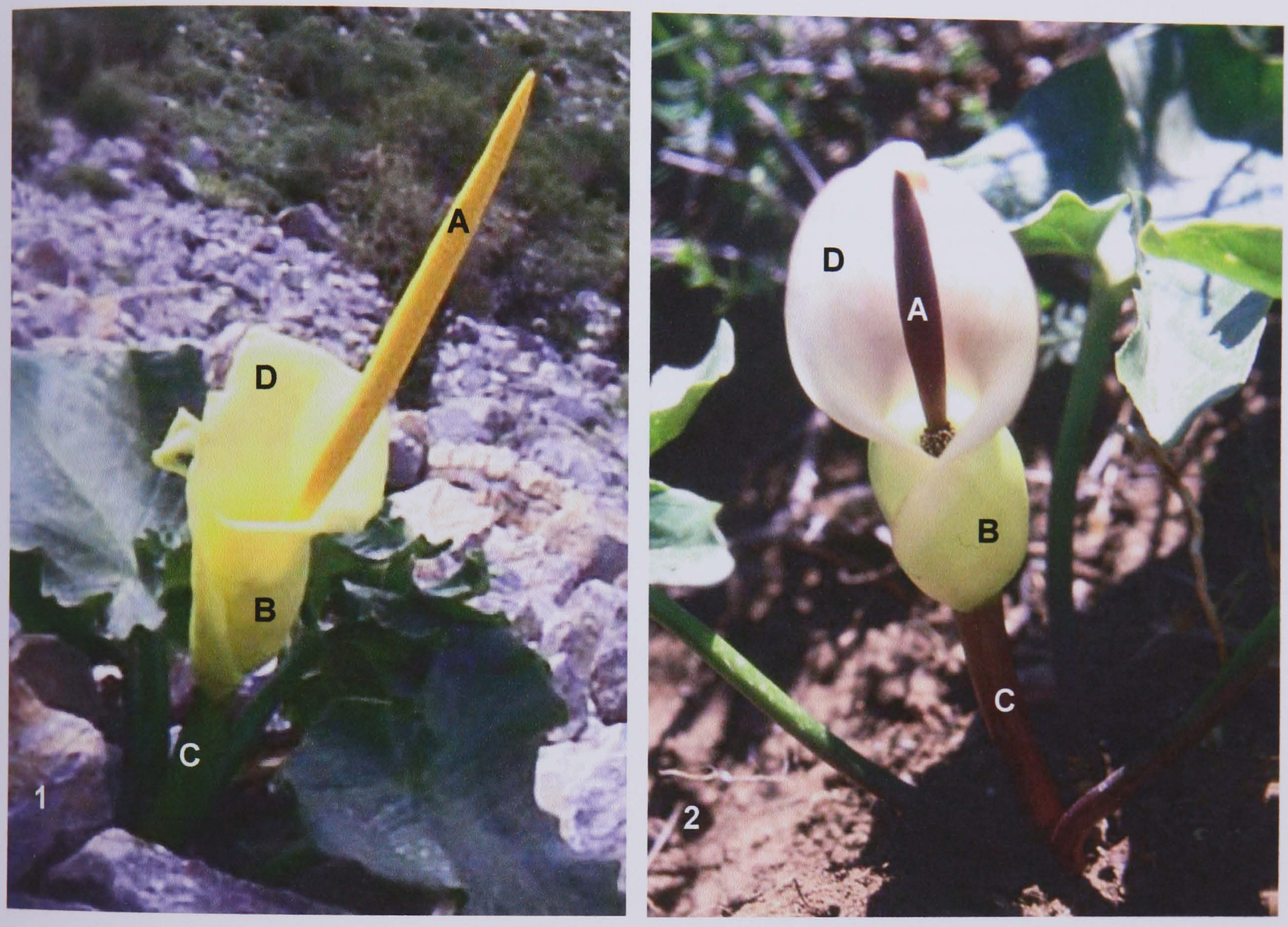


Figure 5.1 Photographs of *A. creticum* (1) and *A. idaeum* (2) inflorescences. Four inflorescence characters are indicated on each picture. A: Spadix-appendix; B: Spathe tube; C: Peduncle. Note that in *A. creticum* (1) the peduncle is large enough so that the inflorescence is higher than the leaves whereas in *A. idaeum* (2) the peduncle (C) is short so that the inflorescence is at the same height as the leaves. D: Spathe hood; in *A. creticum* (1) it is reflexed away from the appendix (A) whereas in *A. idaeum* (2) it (D) is erect over the appendix (A).

Methods

Plant collection

A total of 160 plants were collected from Crete in April 2001, from the two mountains where both *A. idaeum* and *A. creticum* individuals and morphologically intermediate plants exist. In addition to these putative hybrid populations, 38 individual plants from allopatric populations of *A. creticum* (10 individuals) and *A. idaeum* (28 individuals) were also collected (Table 5.1). Using individuals from allopatric populations minimises any potential inclusion of hybrid genotypes and maximises the purity of the *A. creticum* and *A. idaeum* genotype when assessing the genetic distinctness between the two speices.

Classification of individual plants with pure *A. idaeum* or *A. creticum* morphology follows the description by Boyce (1993). Those individuals from the hybrid populations that displayed a mixture of characteristics indicative of both *A. idaeum* and *A. creticum* were classified as putative hybrids and were divided *a priori* into three taxonomic groups (Figure 5.2 and Table 5.1). The three taxonomic groups of morphologically intermediate individuals are: **MH:** these individuals have a morphology midway between *A. creticum* and *A. idaeum* and are potentially F_1 hybrids; **CL** and **IL**: groups of individuals that are also morphologically intermediate and represent introgressed individuals. In group **CL**, *A. creticum* like characters that are indicative of pure *A. creticum* are present but there are also intermediate characters. Group **IL** contains individuals with an *A. idaeum* like morphology; characters that are typical of pure *A.*

idaeum are present but there are also intermediate characters. The individuals in groups CL and IL are not morphologically midway as characters from one parent dominate, so these individuals represent putative backcrosses (Table 5.1).

Table 5.1 Taxonomic descriptions of the seven morphological categories of individuals collected on Crete, including the number of samples amplified per taxonomic group. Letters in parentheses indicate representative inflorescence of each category shown in Figure 5.2. [$]The number of samples with successfully PCR amplified DNA.

Code	Gross morphology	Number of samples[$]
ALCR	Isolated **AL**lopatric populations of typical *A. **CR**eticum* morphology; isolated populations. (A)	10
CR	Typical *A. **Cr**eticum* morphology. (A)	22
CL	*A. **Cr**eticum* **L**ike: Partially reflexed spathe hood; yellow spadix appendix; flower higher than leaves. (B and C)	28
MH	**M**iddle **H**ybrid type: Partially erect spathe hood; yellow-purple spadix; flower above leaves. (D)	21
IL	*A. **Id**aeum* **L**ike: erect spathe hood; yellow-purple spadix appendix; flower same height as leaves. (E)	27
ID	Typical *A. **ID**aeum* morphology. (F)	22
ALID	Isolated **AL**lopatric populations of typical *A. **ID**aeum* morphology. (F)	28

Figure 5.2 Range of morphologically intermediate inflorescences found on Crete. The individual on the extreme left (A) is typical of the species *A. creticum*, whilst the individual on the extreme right (F) is typical of the species *A. idaeum*, as described by Boyce (1993).

For each individual plant 10 inflorescence characters were recorded from mature inflorescences (Table 5.2). The mode of inflorescence display was also recorded, as either flag (peduncle much larger than the petiole) or cryptic (peduncle same size or smaller than the petiole). After the morphological characters were recorded a leaf sample of approximately $3cm^2$ was collected from the centre of a mature leaf, avoiding the edges and the midrib, and rinsed in distilled water. To speed up the drying process each $3cm^2$ sample was ripped into smaller $0.5–1cm^2$ pieces and placed into a re-sealable sample bag containing fine silica gel (pore diameter ca 6nm).

Table 5.2 Morphological characters recorded for each individual

Code	Continuous character	Code	Categorical character
STLE	Stalk length (cm).	FLG	Flag or cryptic inflorescence display mode
CHLE	Chamber length (cm).	SML	Smell of *A. creticum* or *A. idaeum*
SPLE	Spadix length, not including flowers (cm).	SCOL	Spadix colour
CHWI	Chamber width (cm).		
SPWI	Spathe width (cm).		
SXWI	Spadix width (cm).		
MFLE	Length of male flowers (cm).		
MFDI	Diameter of male flowers (cm).		
FELE	Length of female flowers (cm).		
FEDI	Diameter of female flowers (cm).		

DNA extraction

PhytoPure extraction kits (Amersham) were used to extract DNA from 20mg of dried plant material. The dry leaf tissue was ground, by hand, in 1.5ml microcentrifuge tubes with 300µl of reagent 1. Once the tissue was ground up a further 300µl of reagent 1 was added. RNase was added to the samples to a final concentration of 20µg/ml and the samples were gently agitated for 1 hour at 37°C. After the RNase incubation, 200µl of reagent 2 was added, and the samples were incubated for 10 minutes at 65°C with agitation. The samples were then placed on ice for 20 minutes, after which 500µl of chloroform at –20°C and 40µl of PhytoPure resin were added and the mixture incubated for 10 minutes at room temperature with continual agitation. The samples were centrifuged for 10 minutes at 1300***g*** and the upper DNA-containing phase was transferred to a clean tube. DNA was precipitated by adding an equal volume of –20°C isopropanol and incubating for 1 hour at 4°C. The samples were centrifuged at 4000***g*** for 5 minutes to pellet the DNA. The isopropanol was aspirated and the sample washed with 300µl of 70% ethanol (at –20°C). The DNA pellets were air dried overnight and resuspended in 100µl of TE buffer and stored at –20°C.

The methods for PCR amplification, separation, visualisation and recording of PCR-amplified products are described in Chapter 2.

ISSR data analysis

The genetic relationship between *A. creticum* and *A. idaeum* was tested using only the allopatric populations: ALCR and ALID (Table 5.1). Using the presence and absence data from the ISSRs, genetic distances were calculated among all allopatric *A. creticum* (ALCR) and allopatric *A. idaeum* (ALID) individuals using the Nei and Li (1979) genetic distance in PAUP*4.0b (Swofford, 1998). The genetic distances were then used to construct an NJ tree. Support for the relationship was gained through a bootstrap test with 1000 replicates carried out in PAUP*4.10b (Swofford, 1998).

Visualisation of the relationships among all the tested individuals (n=158) utilised the N-MDS technique (see Chapter 2) using pairwise genetic distances. Statistical support for any morphological groups was carried out using the ANOSIM test as described in Chapter 2. The N-MDS method allows the genetic relationships among the sampled individuals to be visualised in a two-dimensional plot. The proportion of the genome associated with a specific taxon was assessed using Structure 2 (Pritchard *et al.*, 2000). This program assigns the proportion of each individual's genome (Q) to k clusters. In this analysis, k was set to 2 as there are two putative parental genotypes. Runs of the program consisted of a 10,000 burnin period and 100,000 iterations. The probability of an individual being a hybrid is not calculated by the Structure 2 program. However, the program New Hybrids (Anderson and Thompson, 2002) assigns a probability of an individual being a hybrid. The New Hybrid program estimates the probability of an individual's genotype being one of six genetic types (Table 5.3) that correspond to the

proportion of the genome attributable to either parental species. The rationale behind the genetics is that the nuclear genome of the initial F_1 hybrid will have an equal contribution from each parental species. Thereafter, every subsequent backcross halves the number of nuclear markers from one parent species and increases the relative proportion of the other parent species (Rieseberg *et al.*, 1990). The proportion of the genome attributed to one parental taxa or another can therefore give an indication of the level of introgression. The six genotype categories (Table 5.3) correspond to two parental types, initial F_1 hybrids, backcrosses of the initial hybrids to both parental species, and crosses between two F_1 hybrids (F_2).

Table 5.3 Six genotype categories with the expected frequencies of alleles corresponding to species A (*A. creticum*) or B (*A. idaeum*). The expected proportion of alleles corresponding to either of the two parental species A and B is shown for F_1 and F_2 hybrids, backcross of hybrid to *A. creticum* and backcross of hybrid to *A. idaeum*. $G_{g,2}$ = the proportion of loci where both alleles are attributed to species A; $G_{g,1}$ = the proportion of loci where alleles can be attributed to either species A or B; $G_{g,0}$ = the proportion of alleles that are attributed to species B. The correspondence of the six genotype groups used in the New Hybrids program to the morphological categories: ALCR, CR, CL, MH, IL, ID and ALID are shown in the column experimental group. Backcross is abbreviated to Bx.

g	$G_{g,2}$ (A, A)	$G_{g,1}$ (A, B) or (B, A)	$G_{g,0}$ (B, B)	Group	Experimental group
1	1	0	0	*A. creticum*	CR, ALCR
2	0	0	1	*A. idaeum*	ID, ALID
3	0	1	0	F_1	MH
4	0.25	0.50	0.25	F_2	MH
5	0.5	0.5	0	Bx *A. creticum*	CL
6	0	0.5	0.5	Bx *A. idaeum*	IL

Results

Range of morphological variation in the putative hybrid zones

The amount of morphological variation in the putative hybrid populations is shown in a series of box plots (Figures 5.3 and 5.4). Variation of the male and female flower characters is shown in Figure 5.3, whilst inflorescence character variation is shown in Figure 5.4. All of the morphological characters show a similar pattern: *A. creticum* (ALCR and CR) individuals are larger than both *A. idaeum* (ALID and ID) and individuals from the putative hybrid morphological groups (CL, MH and IL). *A. creticum* like (CL), middle hybrid (MH) and *A. idaeum* like (IL) are of intermediate size to *A. creticum* and *A. idaeum*. A Kruskal–Wallis test was used to identify if the median value of each morphological group was significantly different. This test showed that the medians of each group are significantly different for flower anatomy (Table 5.4) and inflorescence morphology (Table 5.5). There are statistically significant relationships among some of the morphological characters (Table 5.6). Female flower length, however, is unique among the characters as it does not significantly correlate with chamber length, spathe length, chamber width or male flower length.

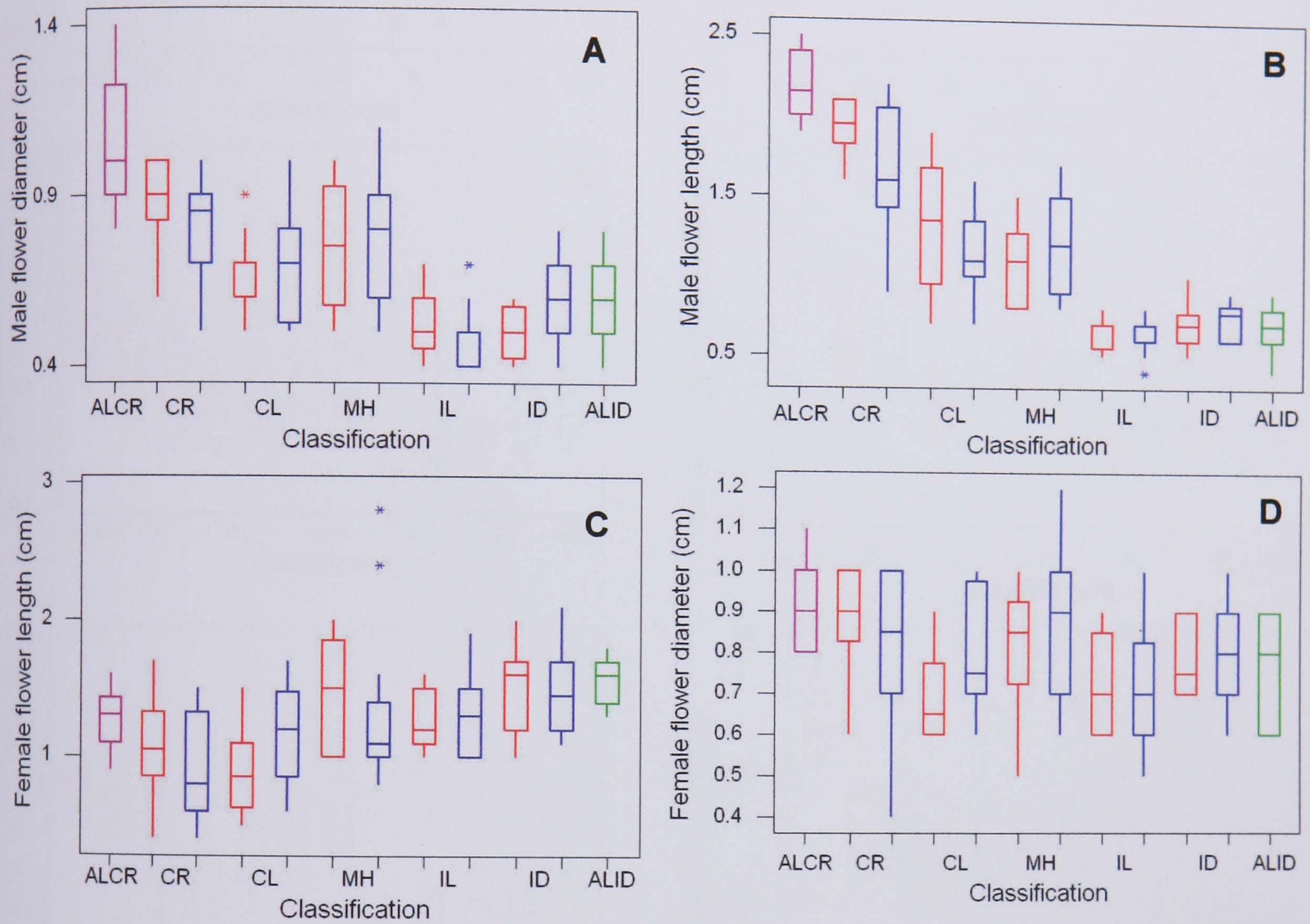
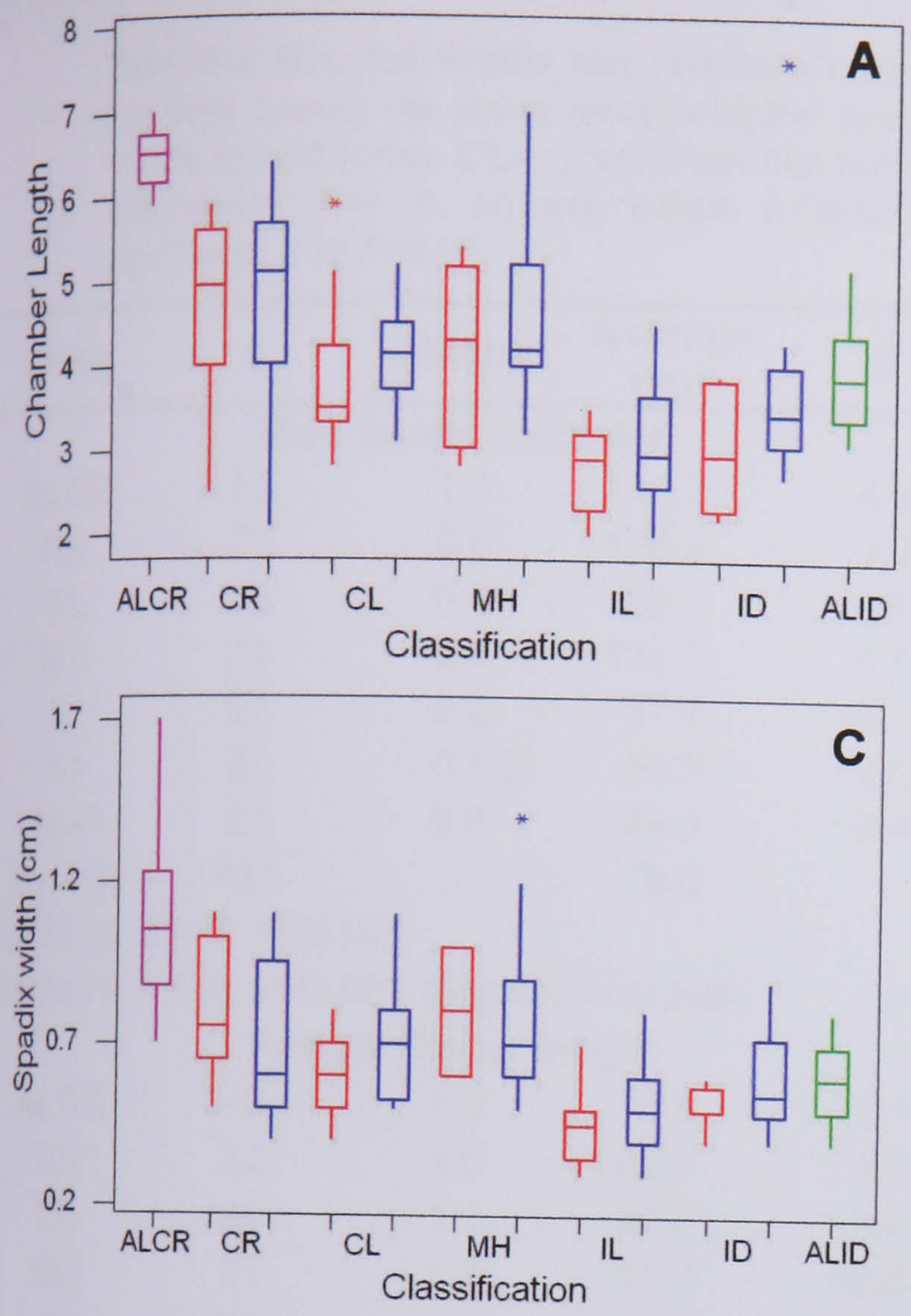
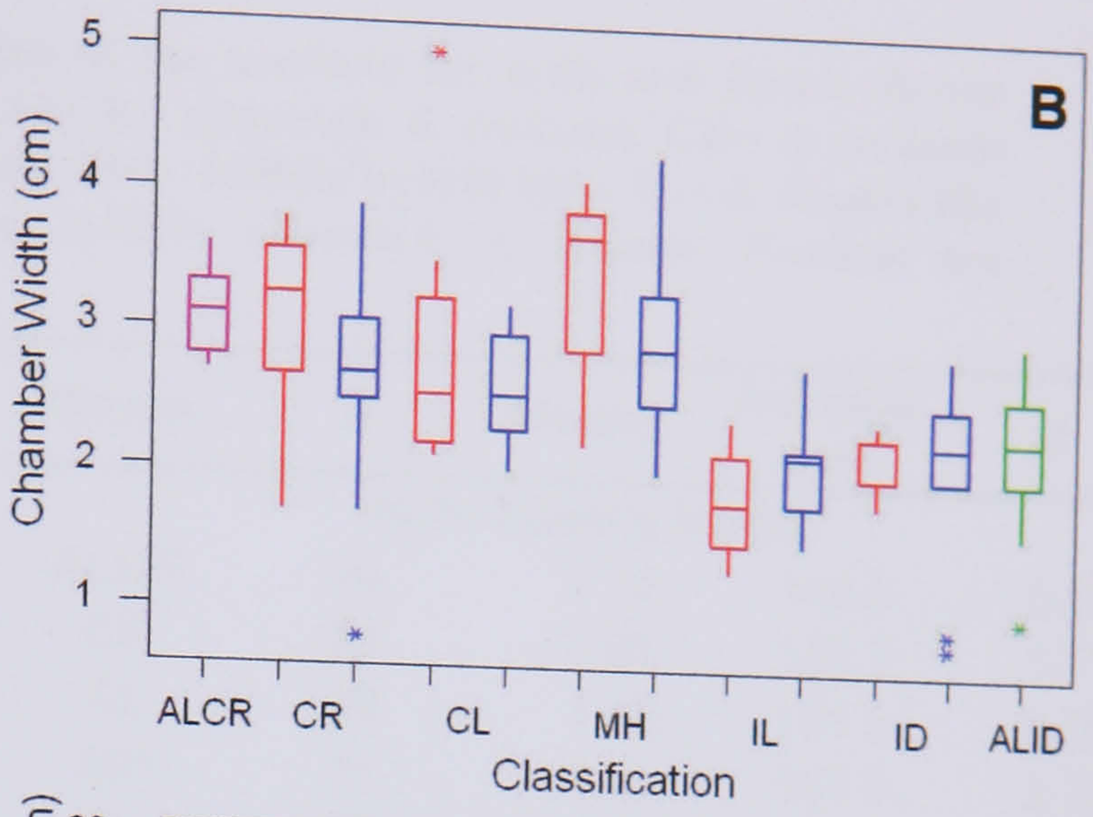
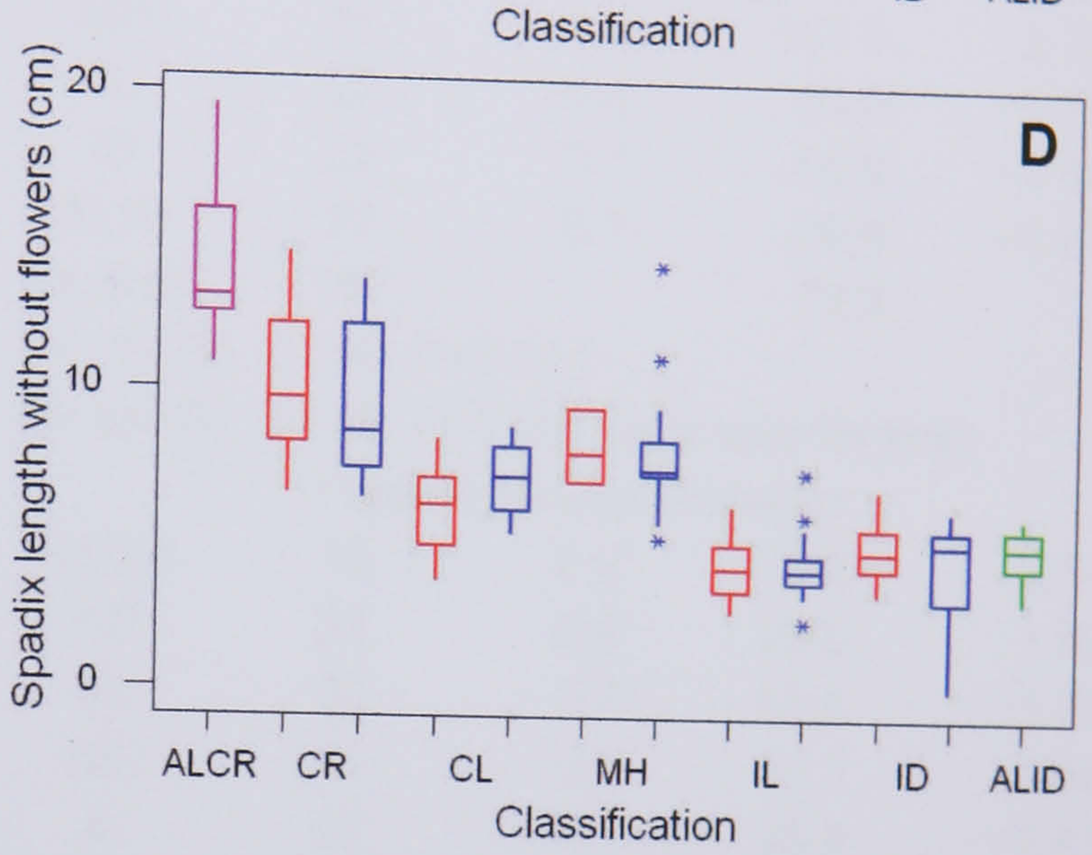
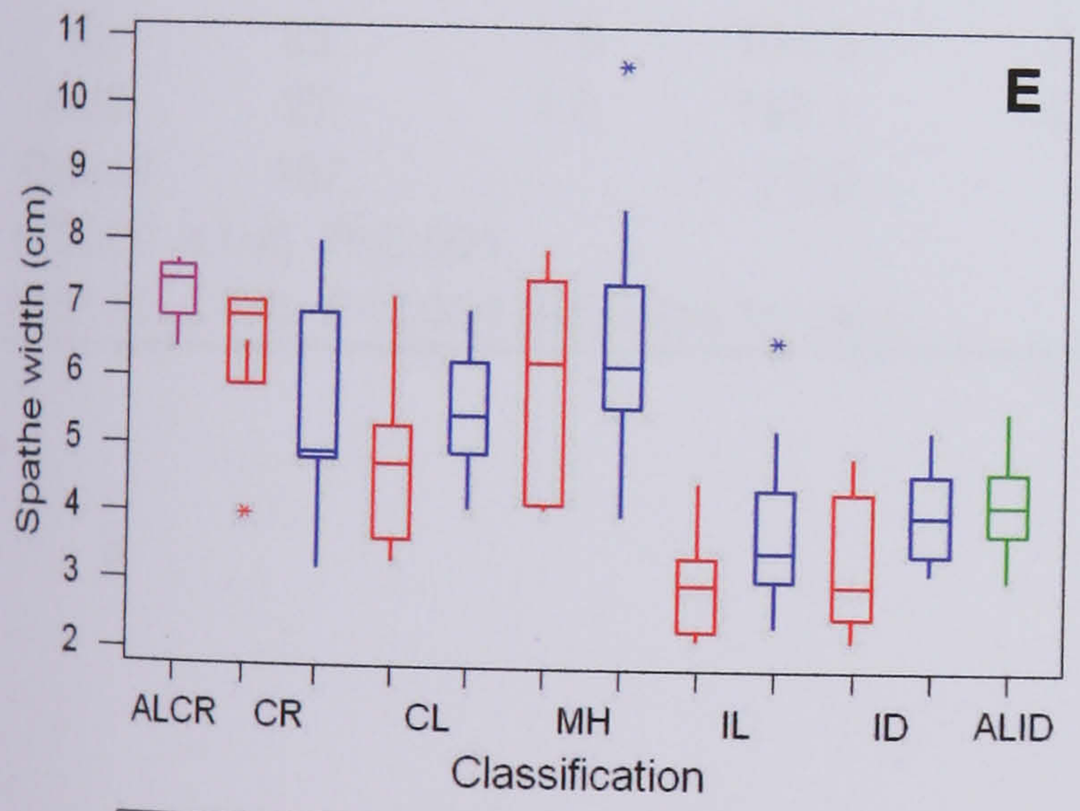


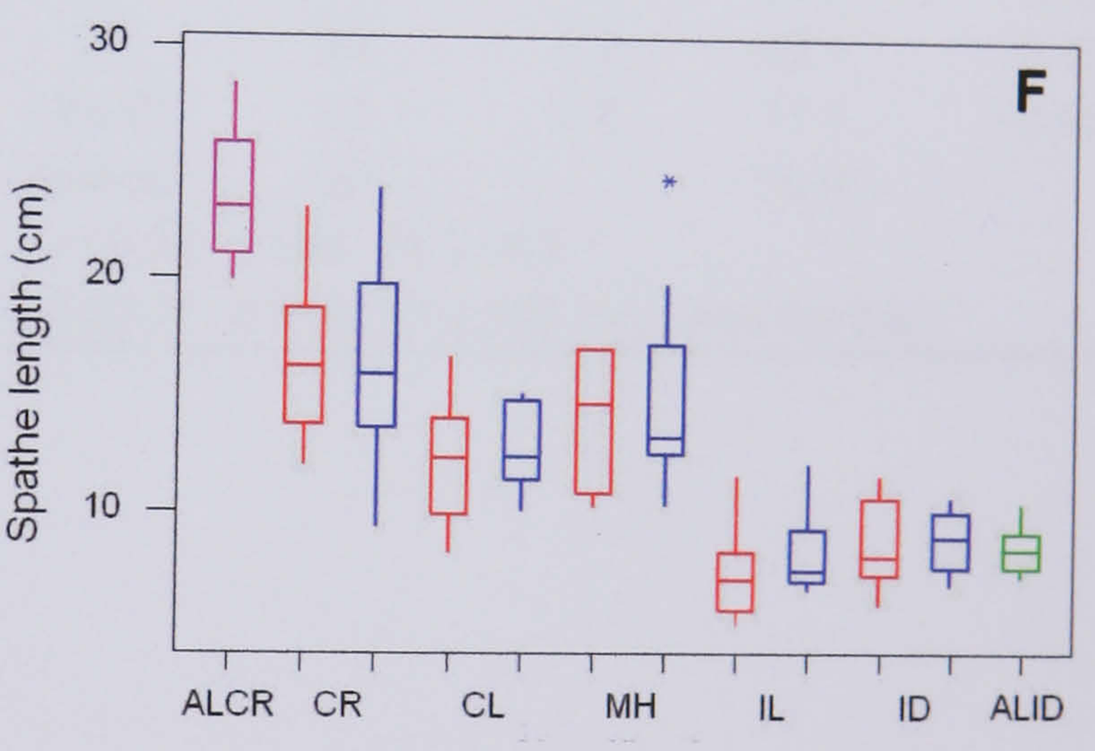
Figure 5.3 Box plots showing the range of morphological variation between populations from Mt Psilorities and Mt Kedros for the two male and two female flower anatomical traits in the seven *a priori* morphological groups. Colours. Purple: allopatric populations of *A. creticum*; Red: Individuals from Mt Kedros; Blue: Individuals from Mt Psilorities; Green: allopatric populations of *A. idaeum*. Classification. ALCR: allopatric *A. creticum*; CR: *A. creticum* in sympatric populations; CL: *A. creticum* like; MH: Middle Hybrid; IL: *A. idaeum* like; ID: *A. idaeum* in sympatric populations; ALID: allopatric *A. idaeum*. Graphs. A: male flower diameter; B: male flower length; C: female flower length; D: female flower diameter.











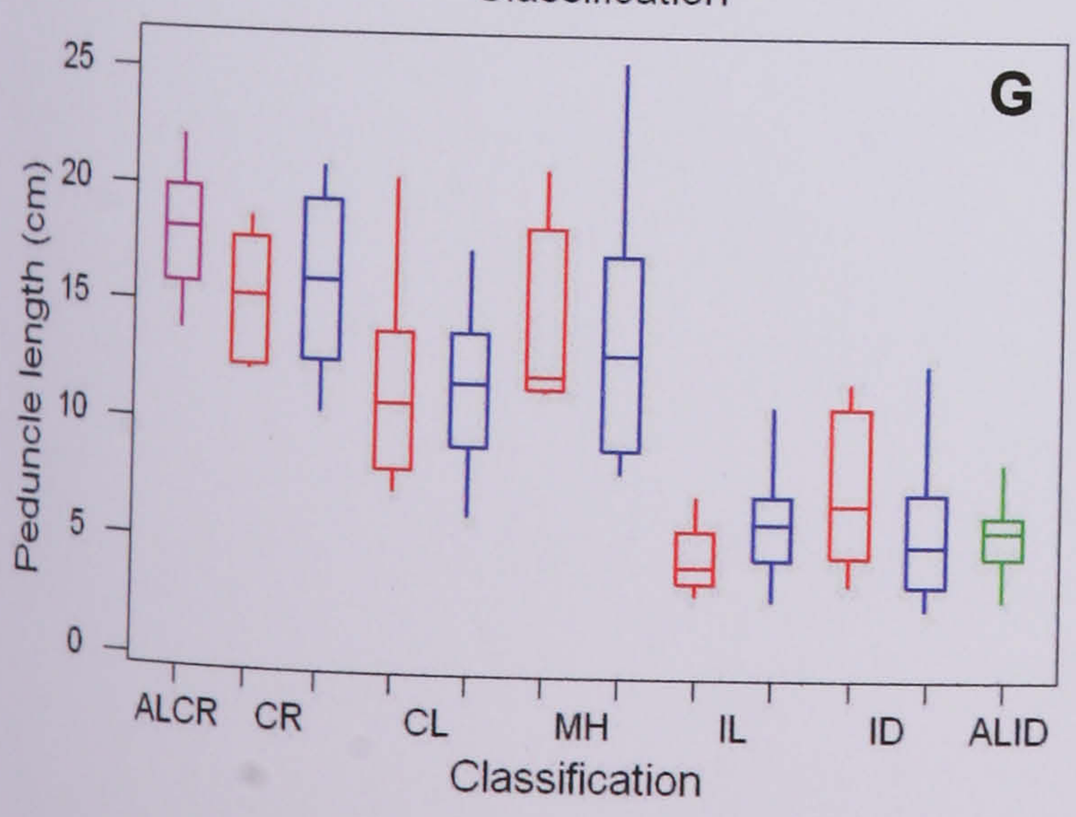


Figure 5.4 Box plots showing the range of morphological variation between populations from Mt Psilorities and Mt Kedros for seven inflorescence traits in the seven *a priori* morphological groups. Colours. Purple: allopatric populations of *A. creticum*; Red: Individuals from Mt Kedros; Blue: Individuals from Mt Psilorities; Green: allopatric populations of *A. idaeum*. Classification. ALCR: allopatric *A. creticum*; CR: *A. creticum* in sympatric populations; CL: *A. creticum* like; MH: Middle Hybrid; IL: *A. idaeum* like; ID: *A. idaeum* in sympatric populations; ALID: allopatric *A. idaeum*. Graphs. A: chamber length; B: chamber width; C: spadix width; D: spadix length not including flowers; E: spathe width; F: spathe length; G: peduncle length.

Table 5.4 Kruskal–Wallis test results. Comparisons of the medians for male and female flower anatomy among the seven morphological groups. ALCR= allopatric *A. creticum*. CR= *A. creticum* within hybrid zones. CL= *A. creticum* like individuals. MH= Middle hybrid type. IL= *A. idaeum* like individuals. ID= *A. idaeum* within hybrid zones. ALID= allopatric *A. idaeum*. *P*-values are significant if at $P<0.05$.

Group	*n*	Median	Average rank	*Z*	Group	*n*	Median	Average rank	*Z*
	Male flower diameter					**Male flowers length**			
ALCR	10	1.0	143.7	4.65	**ALCR**	10	2.15	149.5	5.07
CR	22	0.9	114.3	3.93	**CR**	22	1.65	130.3	5.71
CL	28	0.65	86.1	0.91	**CL**	28	1.15	101.3	2.86
MH	21	0.8	103.2	2.62	**MH**	21	1.2	101.1	2.39
IL	27	0.5	37.9	−5.16	**IL**	27	0.6	33.6	−5.71
ID	22	0.525	56.3	−2.53	**ID**	22	0.7	50.6	−3.16
ALID	27	0.6	59.8	−2.42	**ALID**	27	0.7	39.4	−4.97
Overall	157		79.0		**Overall**	157		79.0	
H=72.56 d.f.=6 *P*<0.001					H=119.78 d.f.=6 *P*<0.001				
H=74.79 d.f.=6 *P*<0.001 (adjusted for ties)					H=120.94 d.f.=6 *P*<0.001 (adjusted for ties)				
	Female flower length					**Female flower diameter**			
ALCR	10	1.3	76.6	−0.18	**ALCR**	10	0.9	117.2	2.74
CR	22	1.0	42.7	−4.04	**CR**	22	0.9	93.5	1.62
CL	28	1.1	52.2	−3.44	**CL**	28	0.7	65.4	−1.75
MH	21	1.2	78.4	−0.07	**MH**	21	0.9	94.3	1.65
IL	27	1.3	78.1	−0.12	**IL**	27	0.7	59.8	−2.41
ID	22	1.5	104.3	2.82	**ID**	22	0.8	83.1	0.45
ALID	27	1.6	118.1	4.91	**ALID**	27	0.8	71.1	−0.99
Overall	157		79.0		**Overall**	157		79.0	
H=50.60 d.f.=6 *P*<0.001					H=19.96 d.f.=6 *P*=0.003				
H=51.05 d.f.=6 *P*<0.001 (adjusted for ties)					H=20.75 d.f.=6 *P*=0.002 (adjusted for ties)				

Table 5.5 Kruskal–Wallis test results. Comparisons of the medians for inflorescence morphology among the seven morphological groups. ALCR= allopatric *A. creticum*. CR= *A. creticum* within hybrid zones. CL= *A. creticum* like individuals. MH= Middle hybrid type. IL= *A. idaeum* like individuals. ID= *A. idaeum* within hybrid zones. ALID= allopatric *A. idaeum*. *P*-values are significant at $P<0.05$.

Group	*n*	Median	Av.Rank	*Z*	Group	*n*	Median	Av.Rank	*Z*
		Chamber length					**Chamber width**		
ALCR	10	6.5	149.1	5.04	ALCR	10	3.1	126.9	3.44
CR	22	5.1	106.5	3.05	CR	22	2.9	104.1	2.79
CL	28	3.95	82.4	0.43	CL	28	2.55	92.5	1.73
MH	21	4.2	98.2	2.08	MH	21	2.9	115.1	3.91
IL	27	2.9	33.1	−5.77	IL	27	2.0	36.6	−5.32
ID	22	3.3	52.2	−2.99	ID	22	2.1	52.2	−2.98
ALID	27	3.9	80.0	0.13	ALID	27	2.3	62.9	−2.02
Overall	157	79.0			Overall	157	79.0		
H=70.88 d.f.=6 *P*<0.001					H=68.01 d.f.=6 *P*<0.001				
H=70.97 d.f.=6 *P*<0.001 (adjusted for ties)					H=68.25 d.f.=6 *P*<0.001 (adjusted for ties)				
		Spadix width					**Spadix length**		
ALCR	10	1.05	141.2	4.47	ALCR	10	13.15	149.5	5.07
CR	22	0.7	92.8	1.54	CR	22	9.1	127.4	5.38
CL	28	0.65	83.6	0.59	CL	28	6.75	89.9	1.41
MH	21	0.7	104.9	2.8	MH	21	7.4	111.6	3.54
IL	27	0.5	39.4	−4.98	IL	27	4.1	34.1	−5.64
ID	22	0.5	60.9	−2.02	ID	22	4.8	44.5	−3.84
ALID	27	0.6	74.2	−0.6	ALID	27	4.9	49.8	−3.66
Overall	157	79.0			Overall	157	79.0		
H=52.15 d.f.=6 *P*<0.001					H=111.58 d.f.=6 *P*<0.001				
H=53.49 d.f.=6 *P*<0.001 (adjusted for ties)					H=111.62 d.f.=6 *P*<0.001 (adjusted for ties)				
		Spathe width					**Spathe length**		
ALCR	10	7.4	141.5	4.49	ALCR	10	23.1	149.9	5.1
CR	22	6.15	105.0	2.9	CR	22	16.05	123.3	4.93
CL	28	5.0	90.9	1.52	CL	28	12.35	96.6	2.27
MH	21	6.2	117.6	4.18	MH	21	14.1	113.7	3.76
IL	27	3.4	34.6	−5.57	IL	27	7.4	31.1	−6.01
ID	22	3.95	47.4	−3.51	ID	22	8.65	47.7	−3.49
ALID	27	4.2	62.4	−2.08	ALID	27	8.4	44.7	−4.3
Overall	157	79.0			Overall	157	79.0		
H=83.10 d.f.=6 *P*<0.001					H=117.39 d.f.=6 *P*<0.001				
H=83.17 d.f.=6 *P*<0.001 (adjusted for ties)					H=117.42 d.f.=6 *P*<0.001 (adjusted for ties)				
		Peduncle length							
ALCR	10	18.2	138.2	4.26					
CR	22	16	126.3	5.26					
CL	28	11.55	98.6	2.51					
MH	21	12.5	112.5	3.63					
IL	27	5.1	38.3	−5.11					
ID	22	6.4	46.8	−3.58					
ALID	27	5.8	39.2	−5.0					
Overall	157	79.0							
H=110.68 d.f.=6 *P*<0.001									
H=110.71 d.f.=6 *P*<0.001 (adjusted for ties)									

Table 5.6 Spearman's correlation matrix of the nine continuous inflorescence characters: stalk length (peduncle length), chamber length, spathe length, chamber width, spathe width, spadix width, male flower length, male flower diameter and female flower diameter. Upper value: correlation value, lower value: *P*-value. The *P*-value is significant at $P<0.05$; these values are outlined in black.

	Stalk length	Chamber length	Spathe length	Chamber width	Spathe width	Spadix width	Male flower length	Male flower diameter	Female flower diameter
Chamber length	0.529 <0.001								
Spathe length	0.826 <0.001	0.761 <0.001							
Chamber width	0.665 <0.001	0.723 <0.001	0.802 <0.001						
Sapthe width	0.701 <0.001	0.79 <0.001	0.873 <0.001	0.825 <0.001					
Spadix width	0.485 <0.001	0.703 <0.001	0.704 <0.001	0.685 <0.001	0.773 <0.001				
Male flower length	0.781 <0.001	0.673 <0.001	0.865 <0.001	0.677 <0.001	0.729 <0.001	0.567 <0.001			
Male flower diameter	0.611 <0.001	0.725 <0.001	0.776 <0.001	0.708 <0.001	0.752 <0.001	0.686 <0.001	0.751 <0.001		
Female flower length	−0.31 <0.001	0.141 0.078	−0.133 0.097	0.086 0.286	0.05 0.535	0.263 <0.001	−0.299 <0.001	0.023 0.777	
Female flower diameter	0.247 0.002	0.525 <0.001	0.427 <0.001	0.465 <0.001	0.528 <0.001	0.563 <0.001	0.388 <0.001	0.588 <0.001	0.342 <0.001

Genetic distinction between *A. creticum* and *A. idaeum*

To identify if *A. idaeum* and *A. creticum* are genetically distinct, the genetic relationships among the allopatric populations of *A. idaeum* (ALID) and *A. creticum* (ALCR) were calculated using 53 ISSR bands from two primers. There are three diagnostic bands for either *A. creticum* or *A. idaeum* out of the 53 ISSR bands (Table 5.7). There are also a further seven bands occurring in 90% of either *A. idaeum* or *A. creticum*. The diagnostic bands are also present in individuals with intermediate morphology, suggesting gene flow and therefore hybridization has occurred. The ISSR data clearly show that *A. creticum* (ALCR) and *A. idaeum* (ALID) form two distinct monophyletic clades (Figure 5.5). The two clades are supported by a bootstrap value of 90%. In addition, an ANOSIM randomisation also indicates these two groups are significantly different (R=1.0, P<0.001).

Table 5.7 Band frequencies in each morphological group for primers 811 (Bands 1-38) and 834 (bands 39-53), a total of 53 bands. ALID= allopatric *A. idaeum*. ID= *A. idaeum* within hybrid zones. IL= *A. idaeum* like individuals. MH= Middle hybrid type. CL= *A. creticum* like individuals. CR= *A. creticum* within hybrid zones. ALCR= allopatric *A. creticum*. The type column indicates if the band is diagnostic for either *A. creticum* or *A. idaeum* (highlighted in green). The bands highlighted in red occur in 90% of the allopatric populations but are not totally diagnostic.

	Band frequency in each morphological group							
Band	ALID	ID	IL	MH	CL	CR	ALCR	Type
1	1.00	0.95	0.93	0.86	0.90	1.00	1.00	
2	1.00	0.95	0.93	0.86	0.90	1.00	1.00	
3	1.00	1.00	1.00	0.86	0.72	0.52	0.40	
4	1.00	1.00	1.00	0.86	0.72	0.52	0.30	
5	0.00	0.09	0.15	0.81	0.72	1.00	0.90	*A. creticum*
6	0.00	0.09	0.15	0.62	0.48	0.81	0.90	*A. creticum*
7	0.79	0.50	0.52	0.24	0.28	0.00	0.00	
8	0.79	0.50	0.52	0.24	0.28	0.00	0.00	
9	0.00	0.14	0.19	0.67	0.62	0.90	1.00	
10	0.00	0.05	0.11	0.67	0.62	0.90	1.00	
11	0.04	0.05	0.15	0.29	0.31	0.43	0.40	
12	0.04	0.05	0.15	0.29	0.28	0.38	0.30	
13	1.00	1.00	0.85	0.90	0.62	0.52	0.40	
14	1.00	1.00	0.85	0.90	0.62	0.52	0.40	
15	0.00	0.05	0.15	0.71	0.62	0.86	0.70	
16	0.00	0.05	0.15	0.71	0.62	0.86	0.70	
17	0.00	0.14	0.22	0.62	0.41	0.9	1.00	*A. creticum*
18	0.04	0.09	0.11	0.67	0.48	0.62	0.70	
19	0.04	0.09	0.11	0.67	0.48	0.62	0.70	
20	1.00	1.00	1.00	1.00	1.00	1.00	0.90	
21	0.04	0.59	0.52	0.95	0.90	1.00	0.90	
22	1.00	1.00	1.00	0.81	0.59	0.43	0.10	*A. idaeum*
23	1.00	1.00	1.00	0.71	0.41	0.38	0.10	*A. idaeum*
24	0.96	1.00	0.93	0.81	0.66	0.24	0.10	*A. idaeum*
25	0.96	1.00	1.00	0.95	1.00	1.00	1.00	
26	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
27	0.07	1.00	0.93	0.86	0.86	0.62	0.30	
28	0.79	0.95	0.96	0.95	0.62	0.62	0.80	
29	0.14	0.86	0.96	0.90	0.90	0.95	0.90	
30	0.50	0.18	0.26	0.48	0.52	0.71	0.90	
31	0.89	0.68	0.89	0.86	0.66	0.81	1.00	
32	0.79	0.82	0.96	0.86	0.76	1.00	0.90	
33	0.29	0.86	0.96	1.00	0.93	0.95	0.70	
34	0.07	0.32	0.26	0.38	0.45	0.62	0.70	
35	0.89	1.00	1.00	1.00	1.00	1.00	0.90	
36	0.63	0.91	1.00	1.00	0.93	0.95	1.00	

Table continued overleaf.

Table 5.7 continued. Band frequencies in each morphological group for primers 811 (bands 1–38) and 834 (bands 39–53), a total of 53 bands.

	Band frequency in each morphological group							
Band	**AL ID**	**ID**	**IL**	**MH**	**CL**	**CR**	**ALCR**	**Type**
37	0.89	0.95	0.96	1.00	0.97	0.95	1.00	
38	0.59	0.91	0.89	0.76	0.76	0.57	0.90	
39	0.00	0.00	0.04	0.24	0.17	0.62	0.50	
40	1.00	1.00	0.96	0.86	0.72	0.05	0.00	*A. idaeum*
41	1.00	1.00	1.00	1.00	1.00	1.00	0.20	
42	0.04	0.55	0.48	0.71	0.52	0.48	0.80	
43	0.86	0.64	0.70	0.67	0.79	0.57	0.90	
44	0.96	0.95	0.93	1.00	0.97	0.90	0.90	
45	0.96	0.95	0.93	1.00	0.97	0.90	0.90	
46	1.00	0.95	0.93	1.00	1.00	0.95	1.00	
47	0.00	0.68	0.59	0.81	0.76	0.76	1.00	*A. creticum*
48	0.75	0.91	0.89	0.62	0.79	0.48	0.00	
49	0.00	0.05	0.15	0.52	0.79	0.81	0.9	*A. creticum*
50	0.11	0.00	0.11	0.76	0.69	0.90	1.00	
51	0.32	0.18	0.37	0.57	0.62	0.71	0.80	
52	1.00	1.00	0.78	0.95	0.97	1.00	1.00	
53	0.00	0.00	0.19	0.71	0.79	0.86	0.90	*A. creticum*

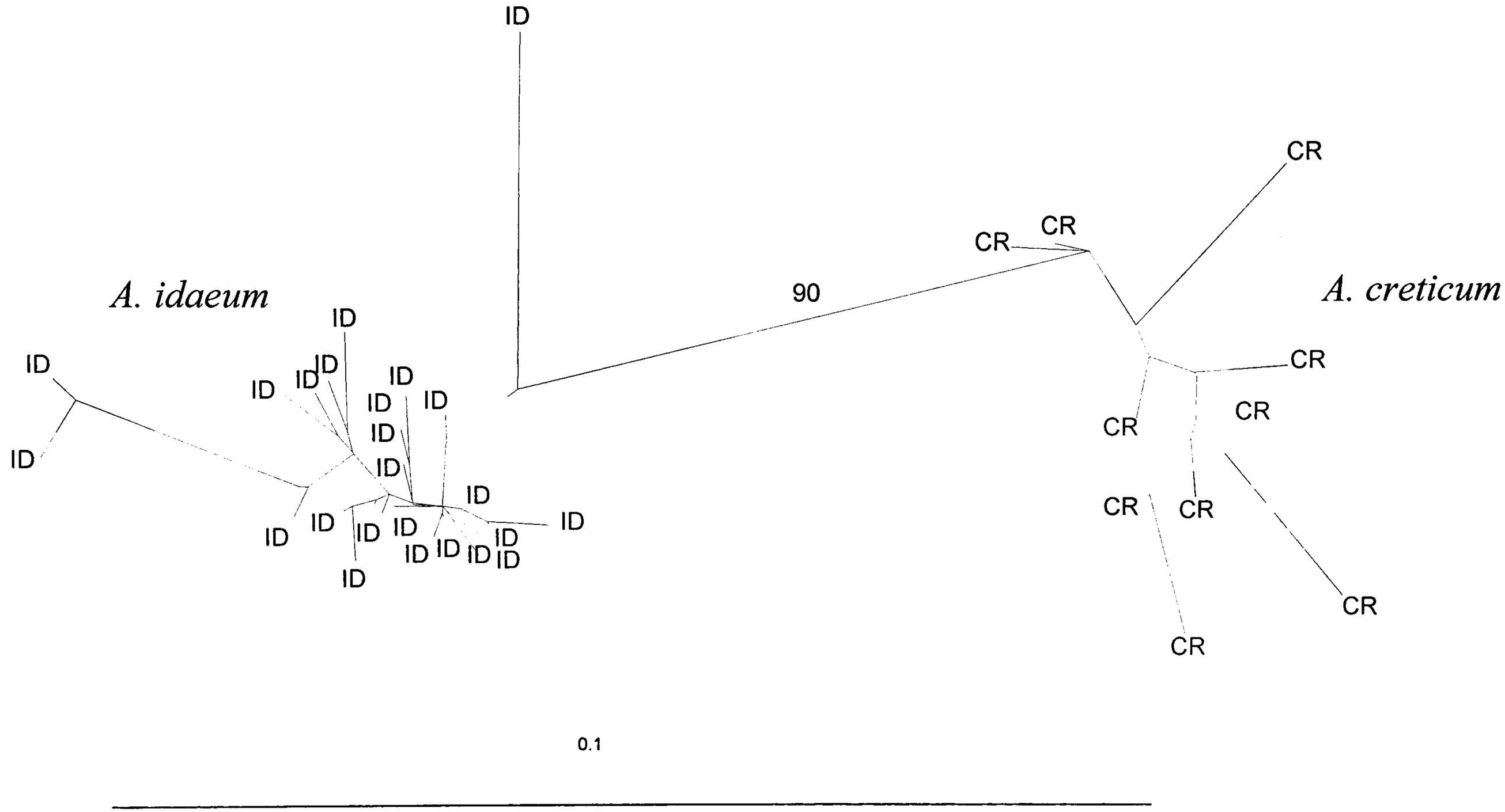


Figure 5.5 Neighbour-joining tree of Nei and Lei (1978) genetic distances among individuals from allopatric populations of *A. creticum* and *A. idaeum*. Bootstrap values greater than 50% are displayed above the branches. ID= *A. idaeum* individuals and CR= *A. creticum* individuals.

Genetic relationships among the morphologically intermediate individuals

The N-MDS ordination of the genetic data (Figure 5.6) shows that, in general, morphologically intermediate individuals are genetically intermediate compared with allopatric *A. creticum* and *A. idaeum* individuals. The genetic distinctiveness of each morphological group (see Table 5.1 for descriptions of the morphological groups) was tested using an ANOSIM test (Table 5.8). The allopatric populations of *A. creticum* (ALCR) were not significantly genetically different from the individuals with pure *A. creticum* (CR) morphology within the hybrid zones (Table 5.8). Conversely, individuals from the allopatric populations of *A. idaeum* (ALID) and individuals with pure *A. idaeum* (ID) morphology in the hybrid zones were significantly different (Table 5.8). The morphological class ID and IL were not significantly different from each other (Table 5.8). The MH classes was not significantly different from the CL class, but was significantly different to all other groups ($P<0.001$). The CL morphological group was not significantly genetically different from the CR group. However, the morphological groups ALCR, CR, ALID and ID were distinct from each other at a significance level of $P<0.001$ (Table 5.8).

The groups ALCR, CR, CL and MH all form a genetic continuum; the two extremes ALCR and MH are significantly different. All members of the ALCR, CR, CL and MH groups are genetically different from those of IL, ID and ALID groups. This suggests that more gene flow occurs between individuals within the

group ALCR-CR-CL-MH, than occurs between individuals from the ALCR-CR-CL-MH and IL-ID-ALID groups.

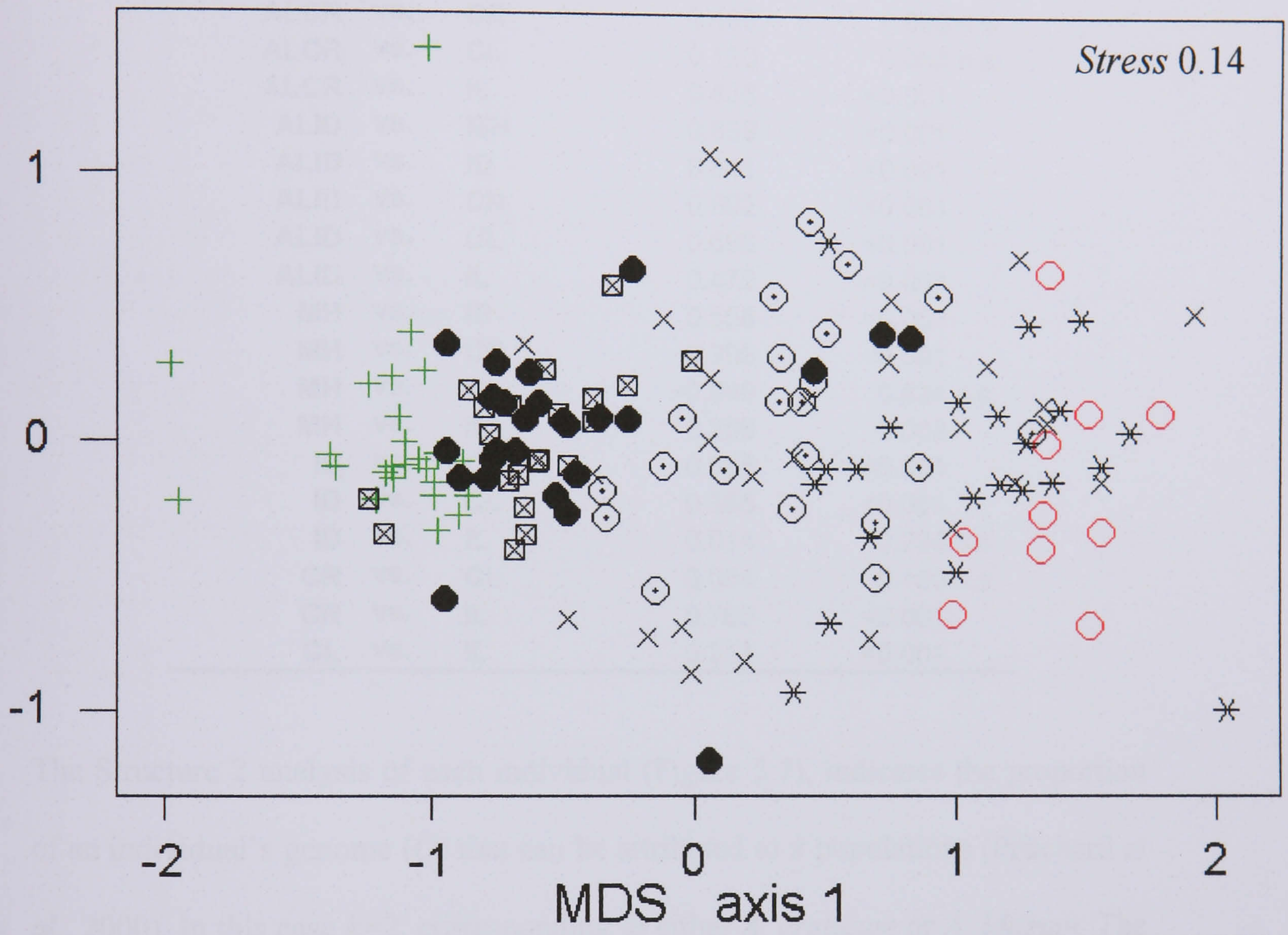


Figure 5.6 MDS ordination of the pairwise genetic distances among all sampled individuals. The individuals are classified into the seven morphological categories: ALCR= allopatric *A. creticum*. CR= *A. creticum* within hybrid zones. CL= *A. creticum* like individuals. MH= Middle hybrid type. IL= *A. idaeum* like individuals. ID= *A. idaeum* within hybrid zones. ALID= allopatric *A. idaeum*.

Table 5.8 The significance of ANOSIM comparisons among all seven morphological groups. *R* indicates the degree of discrimination between the compared groups. *P* values less than 0.05 are significant. Non-significant results are indicated by the annotation n.s.

Morphological category			*R* statistic	*P*-value
ALCR	vs.	ALID	1.000	<0.001
ALCR	vs.	MH	0.517	<0.001
ALCR	vs.	ID	0.999	<0.001
ALCR	vs.	CR	0.103	0.089 n.s.
ALCR	vs.	CL	0.120	0.053 n.s.
ALCR	vs.	IL	0.876	<0.001
ALID	vs.	MH	0.839	<0.001
ALID	vs.	ID	0.511	<0.001
ALID	vs.	CR	0.992	<0.001
ALID	vs.	CL	0.696	<0.001
ALID	vs.	IL	0.475	<0.001
MH	vs.	ID	0.558	<0.001
MH	vs.	CR	0.305	<0.001
MH	vs.	CL	−0.040	0.824 n.s.
MH	vs.	IL	0.256	0.003
ID	vs.	CR	0.955	<0.001
ID	vs.	CL	0.365	<0.001
ID	vs.	IL	0.014	0.238 n.s.
CR	vs.	CL	0.054	0.100 n.s.
CR	vs.	IL	0.755	<0.001
CL	vs.	IL	0.211	<0.001

The Structure 2 analysis of each individual (Figure 5.7), indicates the proportion of an individual's genome (*Q*) that can be attributed to *k* populations (Pritchard *et al.*, 2000). In this case *k*=2, corresponding to either *A. creticum* or *A. idaeum.* The allopatric populations (ALCR and ALID) and the morphological types CR and ID are predominantly a single colour corresponding to either the *A. creticum* genotype or the *A. idaeum* genotype. This indicates little mixing of the two parental genotypes in these morphological categories. In contrast, individuals from the three putative hybrid classes (CL, MH and IL) contain a mixture of both genotypes. The *A. creticum* genotype (red) is strongly represented in the CL and MH groups whilst this genotype is less represented in the IL group.

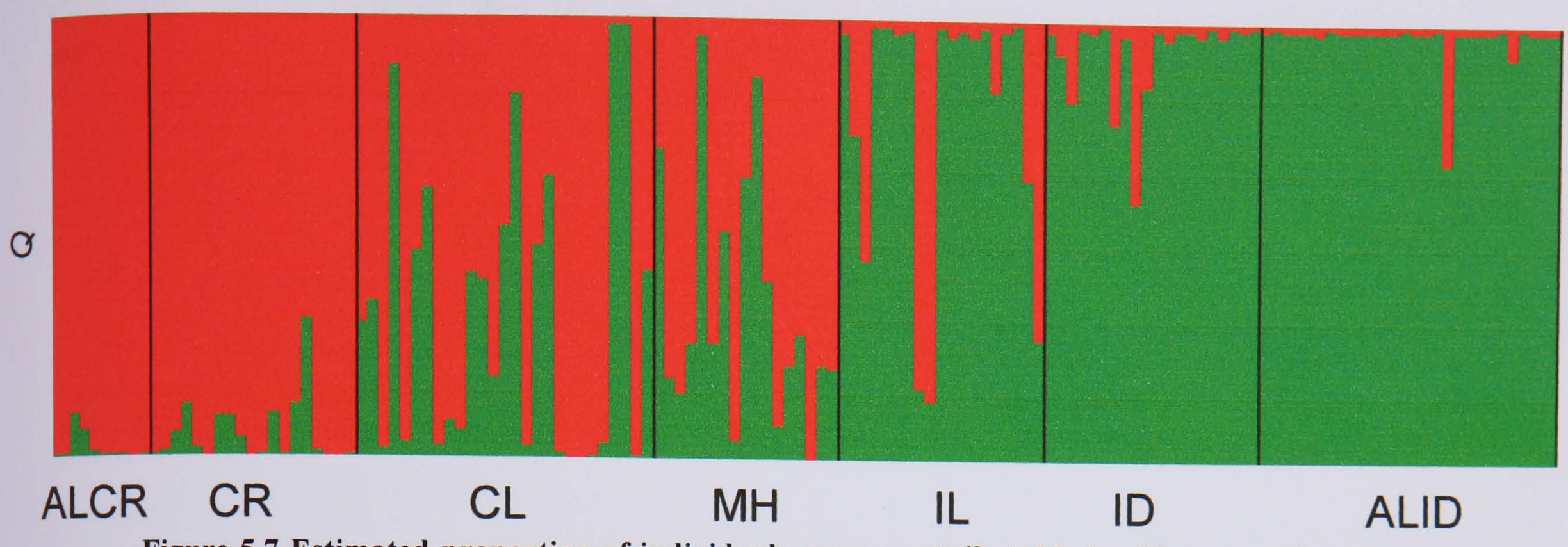


Figure 5.7 Estimated proportion of individuals genome attributable to either *A. creticum* or *A. idaeum*. Each individual is represented by a vertical line, portioned in to *k*=2 coloured segments. Red=*A. creticum* genome and green=*A. idaeum* genome. ALCR and ALID individuals are those from allopatric populations of *A. creticum* and *A. idaeum*, respectively.

The probability of individuals belonging to one of the genotype classes (Table 5.3) was calculated using the New Hybrids program (Anderson and Thompson, 2002) and is shown in Figure 5.8. There are 89 individuals that have a probability above 0.95 of being either pure *A. creticum* or pure *A. idaeum*. Of the remaining individuals, 15 have a probability of between 0.95 and 0.1 of belonging to one of the parental categories. There are 48 individuals that have a probability of 0.1 being a parental type and therefore a probability of 0.9 being a hybrid of some sort. It is not possible to classify with certainty which genotype class these hybrid individuals belong to because the probabilities of them belonging to a specific hybrid group are low. There are twice as many individuals with a probability exceeding 0.8 in the backcross to *A. creticum* group than the backcross to *A. idaeum* group. The smaller number of individuals with a high probability of being a backcross to *A. idaeum* suggests these backcrosses occur less frequently than backcrosses to *A. creticum*. These data show that hybrids exist within the two studied sites and the hybrids have genotypes that are consistent with initial F_1 hybrids, F_2 hybrids and backcrosses to both *A. creticum* and *A. idaeum*. However,

the results indicate that the backcrosses to *A. idaeum* are less frequent than backcrosses to *A. creticum*.

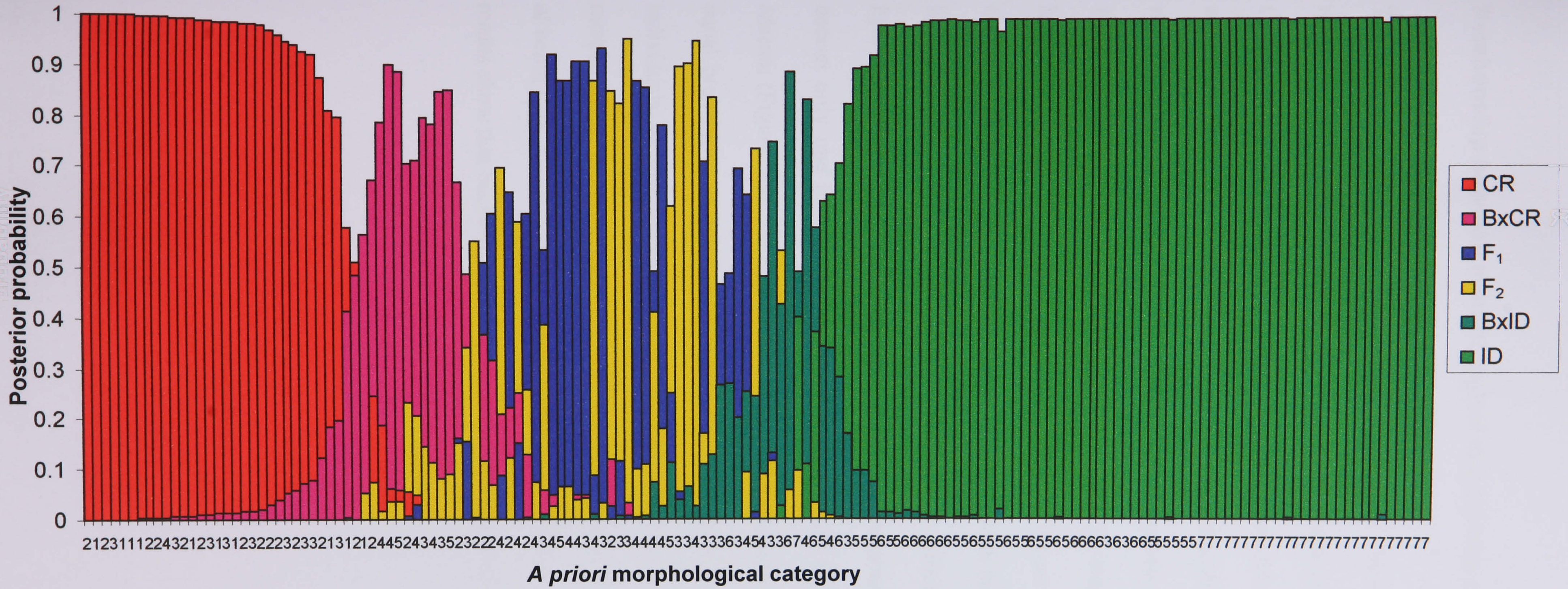


Figure 5.8 The probability calculated in the New Hybrids program of each individual (represented by coloured columns) belonging to the six genotype frequency classes. The height of each coloured bar is proportional to the probability of an individual belonging to the corresponding genotypic class. The genotypic classes are: pure *A. creticum* (CR), pure *A. idaeum* (ID), F_1 and F_2 hybrids and backcrosses to *A. creticum* (B×CR) and *A. idaeum* (B×ID). The individuals are ordered according to the probability that they are pure *A. creticum*, backcrosses to *A. creticum*, F_1 and F_2 hybrids, backcross to *A. idaeum*, and pure *A. idaeum*. The individuals are ordered by probability of being CR, BxCR, F_1, F_2, BxID and ID. To show the correspondence of the seven *a priori* morphological categories to the probabilities the morphological category of each individual is indicated on the X-axis. The *a priori* morphological categories are: 1=ALCR, 2=CR, 3=CL, 4=MH, 5=IL, 6=ID and 7=ALID.

Relationship between morphological and genetic characters

There is a statistically significant relationship between the continuous morphological characters (Table 5.2) and genetic characters based on the ISSRs (Mantel test: r=0.324, P<0.001), indicating that the morphological characters are a reliable indicator of genetic similarity. The relationship between the seven morphological groups (described in Table 5.1) and genetic category is shown in Figure 5.9. The morphological categories of ID and IL comprise individuals that have a high probability of being pure *A. idaeum* (Figure 5.9). The MH group contains individuals with intermediate morphology. With the exclusion of a single individual (pure *A. creticum* genome), individuals within the MH group are hybrids of some sort (F_1, F_2 or backcrosses to *A. creticum*). Within the whole dataset only one individual has a high probability of being a backcross to *A. idaeum* (Figure 5.9). There are two individuals in the ALCR group that have an equal probability of being hybrids or pure *A. creticum* (P=0.5), the rest of the individuals in this group are either pure *A. creticum* or F_2 generation. The morphological groups CR and CL both contain individuals with high probabilities of being pure *A. creticum*, F_1 or F_2 hybrids and backcrosses to *A. creticum*. These results show that backcrosses to *A. idaeum* are comparatively rare.

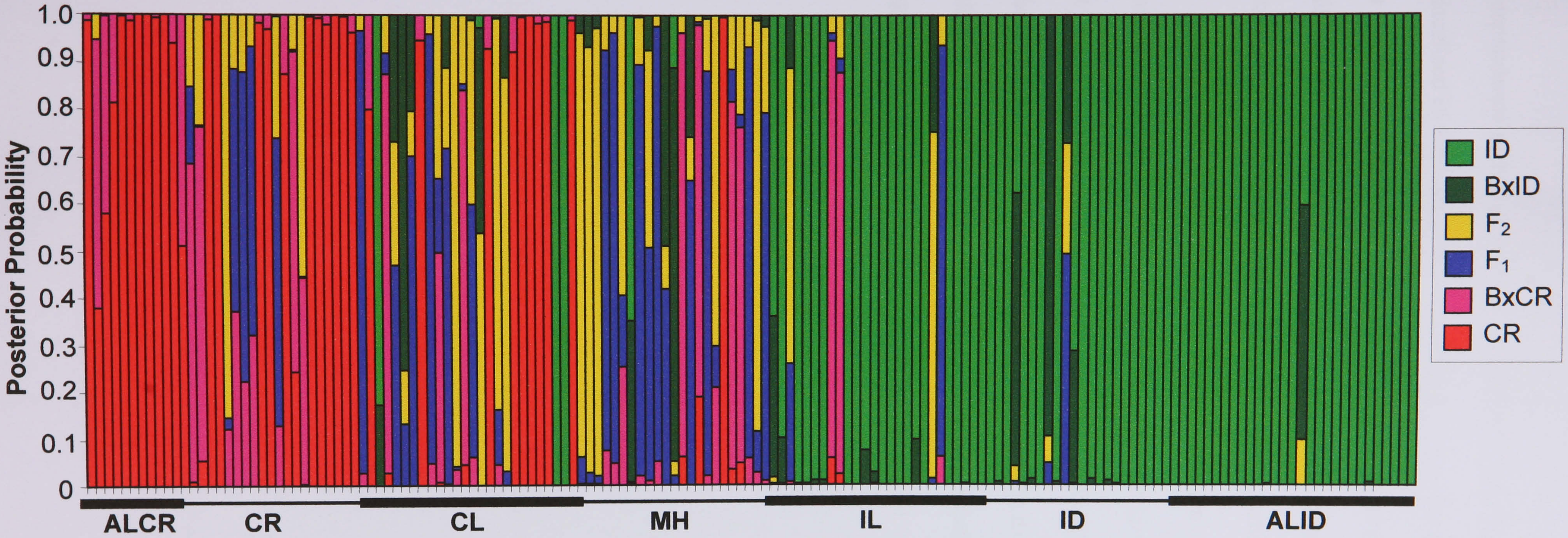


Figure 5.9 The probability calculated in the New Hybrids program of each individual belonging (represented by coloured columns) to the six genotype frequency classes. The individuals are ordered by morphological group, the proportion of each colour in the vertical bars indicating the probability of an individual belonging to each of the genotype classes. The classes are pure *A. creticum* (CR), pure *A. idaeum* (ID), F_1 and F_2 hybrids and backcrosses to *A. creticum* (B×CR) and *A. idaeum* (B×ID). Individuals are grouped by the morphological categories. Allopatric *A. creticum* (ALCR); *A. creticum* (CR); *A. creticum* like (CL); Middle hybrid (MH); *A. idaeum* like (IL); *A. idaeum* (ID); allopatric *A. idaeum* (ALID).

Correspondence of the categorical characters (Table 5.2) with genetic groups can be visualised in three separate N-MDS plots where the categorical character states have been superimposed (Figure 5.10). The categorical characters generally do not follow the same pattern as the continuous characters. Genetically intermediate plants usually display their inflorescence as a flag, which is an *A. creticum* trait. But, both *A. creticum* and *A. idaeum* traits are also seen in genetically intermediate individuals with a reflexed spathe hood (*A. creticum* trait) and an erect spathe hood (*A. idaeum* trait). There are, however, some individuals with a partially reflexed spathe hood (even at full anthesis) that could be assumed to be hybrids. But, several of these individuals are not genetically intermediate instead being genetically similar to allopatric populations of *A. creticum*.

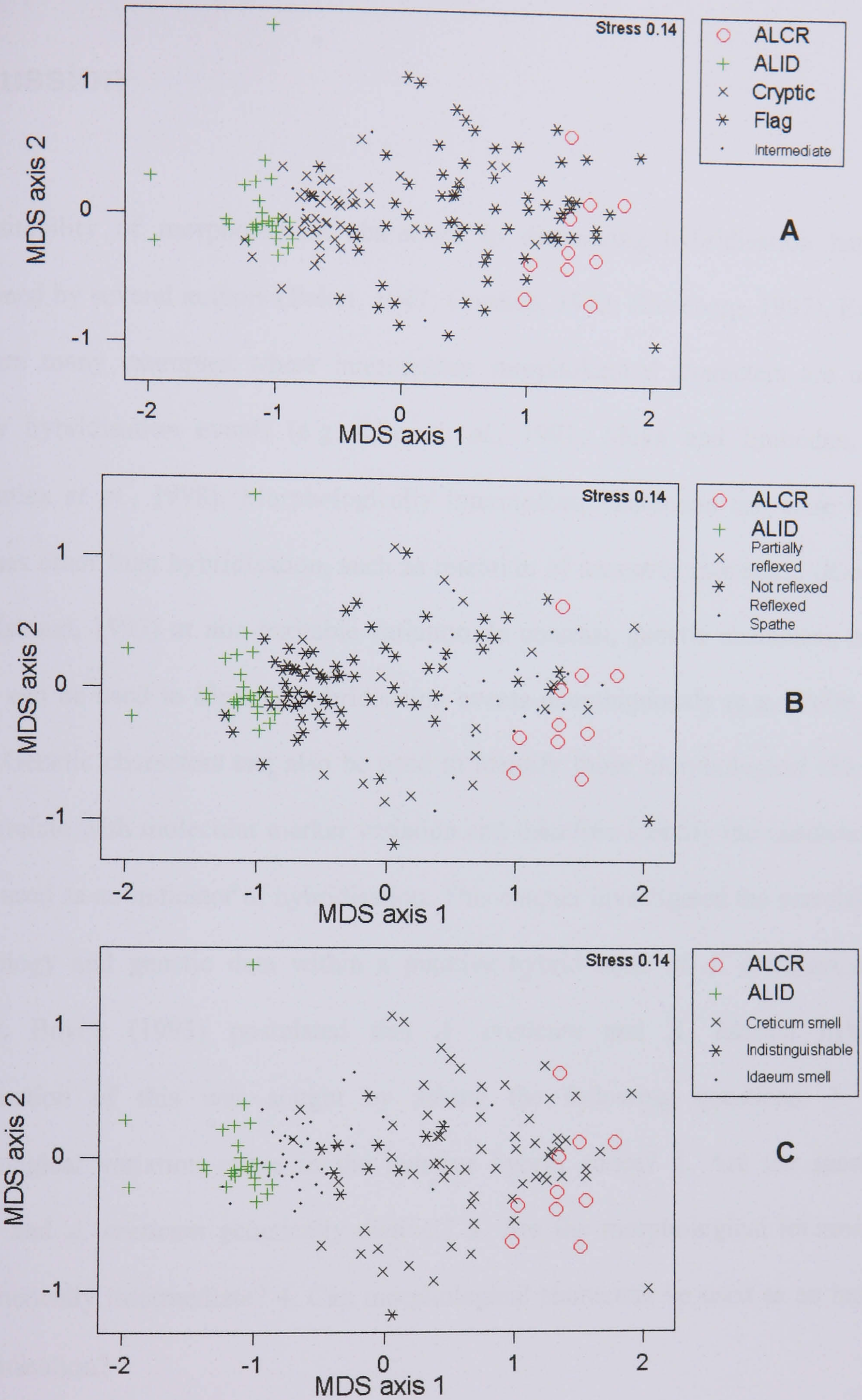


Figure 5.10 MDS plots with three categorical characters superimposed on the graphs. The allopatric samples of *A. creticum* (ALCR) and *A. idaeum* (ALID) are marked for reference in red and green, respectively. A: Inflorescence display mode either flag or cryptic. Flag display is typical of *A. creticum*, cryptic display typical of *A. idaeum*; B: Spathe reflexed. Reflexed spathe hoods are typical of *A. creticum*, non-reflexed spathe hoods typical of *A. idaeum*; C: Smell either like *A. creticum* or *A. idaeum*.

Discussion

The suitability of morphological characters in diagnosing hybridisation has been questioned by several authors (Baker, 1947; Gottlieb, 1972; Rieseberg, 1997). Even so, there are many examples where intermediate morphological characters are used to identify hybridisation events (e.g. Klier *et al.*, 1991; Meyn and Emboden, 1987; Mullenniex *et al.*, 1998). Morphologically intermediate characters can arise through processes other than hybridisation, such as retention of ancestral characters (Rieseberg and Ellstrand, 1993) or non-heritable variation. In contrast, genetic characters, such as ISSRs, can be used to identify hybridisation events unambiguously (e.g. Wolfe *et al.*, 1998). Genetic characters can also be used to identify those morphological characters that correlate with molecular marker variation and therefore identify the characters that can be used as an indicator of hybridisation. This chapter investigated the correlation of morphology and genetic data within a putative hybrid zone of *A. creticum* and *A. idaeum*. Boyce (1993) postulated that *A. creticum* and *A. idaeum* hybridise. Confirmation of this was sought by asking the following questions: 1. What morphological variation exists in the putative hybrid zones? 2. Are the species *A. idaeum* and *A. creticum* genetically distinct? 3. Are the morphological intermediates also genetically intermediate? 4. Can morphological characters be used as an indicator of hybridisation?

Arum idaeum and *A. creticum* were shown to be genetically distinct entities, with individuals of each species forming a monophyletic clade with 90% bootstrap support. These two species are monophyletic, each corresponding to a single evolutionary

lineage, and conform to the evolutionary species concept described by Simpson (1961) and Wiely (1978). The genetic data therefore support the differentiation of *A. creticum* and *A. idaeum* based upon the current taxonomic treatment by Boyce (1993).

Plants with an intermediate morphology were found in this study. Generally, morphologically intermediate plants were also genetically intermediate between *A. creticum* and *A. idaeum*. The intermediate individuals have diagnostic genetic markers from both *A. creticum* and *A. idaeum* and were shown to have a genome that is statistically different from either pure *A. creticum* or pure *A. idaeum.* It is therefore clear that *A. creticum* × *A. idaeum* hybrids exist.

The majority of morphological intermediates have a very low probability of having pure genomes corresponding to either *A. creticum* or *A. idaeum* but they could not be assigned to an individual hybrid class. The morphologically intermediate individuals must therefore be classed as hybrids, but the type of hybrid (F_1, F_2 or backcrosses) cannot be known with certainty. The data suggest that there have been many generations of hybridisation and introgression. This has produced individuals with a highly mixed genome that cannot be distinguished statistically from *A. creticum* or *A. idaeum* or the simple hybrid classes tested.

The genetic pattern observed here is consistent with previously published studies of hybrid zones, where genetically intermediate individuals have also been identified as hybrids (e.g. oaks, Ishida *et al.,* 2003; thistles, O'Hanlon *et al.,* 1999). This study measured both continuous and categorical characters from the inflorescence;

interestingly the results for these two types of characters differed. Variation of the continuous characters was correlated with molecular marker variation, but the categorical characters did not show a relationship with the marker variation. Therefore, continuous characters may be more informative in diagnosing hybridisation.

The results presented here indicate that, in this instance, morphologically intermediate individuals are genetically intermediate and that there is a statistically significant relationship between morphological and genetic similarity. These results are contrary to the studies reviewed by Rieseberg and Ellstrand (1993). The significant relationship between morphological and genetic marker variation within this hybrid zone could be caused by the hybrid phenotype being equally suited to the environment as the parental phenotype, allowing both to coexist. If one parental phenotype is more suited to the environment one morphological type can dominate. For example, in a *Quercus* hybrid zone genetically intermediate individuals were found not to be morphologically intermediate, as the phenotype of one parent was better suited to the environment (Gonzalez-Rodriguez *et al.*, 2004). In *Polystichum* hybrids, individuals with intermediate morphologies were only found in intermediate habitats, not in the parental habitats (Ketner and Mesler, 2000). In the *A. creticum* × *A. idaeum* hybrid zone, parents and hybrids are found in the same habitat, which suggests that they are equally suited to the environment.

The significant relationship between morphological and genetic marker variation could also be caused by additive inheritance of the morphological characters, with no complex interactions between the *A. creticum* and *A. idaeum* genomes. Floral characters are often

the products of several genes that have complex interactions through many transcription factors and transcriptional regulators (reviewed by Buzgo *et al.*, 2004). The significant relationship identified here between morphology and genotype suggests that the genes associated with the studied morphological characters are interacting in a predictable additive manner in the hybrids. This is initially surprising, as the studied morphological characters are likely to be the products of many different genes and such a predictable outcome is unexpected. However, the two species are genetically very similar; they are always united within a clade in the subgenus *Arum* (see Chapter 4) and therefore the genomes are not comparatively divergent. There, is however, a discrepancy between the continuous and categorical characters.

The categorical characters used in the hybrid study did not correlate with genetic marker variation. The patterns displayed by the categorical characters are typified by the mode of flower display, in which intermediate genetic individuals display the category of flag (i.e. inflorescence held above the leaves) associated with *A. creticum.* There are a comparatively small number of genetic intermediates that display the cryptic trait (i.e. inflorescence is displayed below the leaves) associated with *A. idaeum.* The data collected here show an inherent problem associated with the use of categorical characters: they cannot convey intermediacy if there are no unambiguous intermediate states. If flower display was used to assess hybrid status, most individuals would be assumed to have an *A. creticum* genotype as they have an inflorescence with a flag display, but in reality these individuals are genetically intermediate. An intermediate category of inflorescence halfway beneath and halfway above the leaves could easily be misidentified. Similarly, variation in the character flower smell could also be influenced

by bias introduced by the surveyor. Categorical characters should be avoided, instead being converted into continuous characters; for example, flower display can be expressed as the ratio of peduncle to petiole length and smell could be chemically quantified by mass spectroscopy.

In addition to testing the correlation of morphological characters with genetic marker variation, it is also possible to understand biological processes occurring within the hybrid zones. The genetic relatedness among the *a priori* morphological groups gives an insight into the probable direction(s) of gene flow among the plants in the hybrid zones. The groups CR, CL and MH all form a genetic continuum as they are not significantly different; therefore, the majority of gene flow occurs between and within the *a priori* morphological groups CR, CL and MH. These three groups (CR, CL and MH) are distinct from ID and IL, indicating that there are fewer gene flow events between the CR-CL-MH complex and the ID and IL groups.

The conclusion that the CR-CL-MH group has larger amounts of gene flow is supported by estimates of the genotype frequency classes. The majority of the individuals within the groups CR, CL and MH have a probability of less than 0.1 of being genetically pure *A. creticum* or *A. idaeum*, and therefore have a probability of 0.9 of being a mixed genotype. This shows that there is introgression of hybrids to *A. creticum* individuals. There are only three individuals with a substantial probability of being backcrosses to *A. idaeum* (with probabilities of 0.89, 0.84 and 0.75 from the groups MH, ID and ID, respectively). Furthermore, the *a priori* morphological groups ID and IL are significantly genetically different from MH, CL and CR, and therefore gene flow

appears to be insufficient to form a genetic continuum between MH-CL-CR and ID or IL.

The direction of introgression and gene flow between MH-CR-CL can be explained by the differences in the pollination biology of *A. creticum* and *A. idaeum*. *A. creticum* is an obligate outcrosser, whereas *A. idaeum* can self-pollinate as the maturation time of male and female flowers overlap (Diaz and Kite, In press). The direction of crossing in an initial F_1 hybrid is therefore likely to be through transfer of *A. creticum* pollen to *A. idaeum*. Introgression towards *A. creticum* and the genetic continuum among the CR, CL and MH group also suggests that these individuals are outbreeders and not habitual selfers like *A. idaeum* where the ID and IL groups are genetically distinct. A genetic test for the direction of hybridisation has not been carried out, but could be carried out by comparing the uniparentally inherited cytoplasmic genome to the biparentally inherited nuclear genome of hybrids. If the direction of hybridisation is from *A. creticum* (via pollen) to *A. idaeum*, the chloroplast genome of the F_1 hybrids would be expected to be identical to that of *A. idaeum*. In the group MH-CL-CR if crosses occur in all directions it is expected that both chloroplast genomes of *A. creticum* and *A. idaeum* would be present.

It is often noted that introgression has the potential to confer traits from one taxon to another (Arnold *et al.*, 1990; Rieseberg *et al.*, 1990) or to cause a rare taxon to become extinct (Levin *et al.*, 1996). *A. idaeum* is an endemic species to Crete and would be a candidate for extinction through hybridisation. The data indicate that introgression of the *A. creticum* genotype into *A. idaeum* is limited. The limited introgression of *A.*

creticum traits into *A. idaeum* is due to the selfing nature of *A. idaeum*, and the habitual selfing is currently ensuring the survival of *A. idaeum*. There is, however, introgression of the *A. idaeum* genotype into *A. creticum*. The risk of extinction of *A. creticum* is mitigated as *A. creticum* has a distribution beyond Crete, notably in south-west Turkey (Boyce 1993). However, the individuals of *A. creticum* on Crete are geographically isolated by approximately 250km from the main populations on the Turkish mainland. The geographical isolation of *A. creticum* on Crete creates the potential for *A. creticum* there to become distinct from those on the mainland through introgression, as through time, the *A. creticum* genotype on Crete will increasingly contain parts of the *A. idaeum* genotype.

In this hybrid zone morphological variation of continuous characters correlated significantly with genetic marker variation. The patterns of morphological and genetic marker variation have indicated that introgressive hybridisation is occurring towards *A. creticum* at a greater frequency than it is towards *A. idaeum*. In this *A. creticum–A. idaeum* system, the quantitative morphological characters have been shown to be a good indicator of the occurrence of hybridisation and the extent of introgression. In contrast, categorical characters are not a good indicator of the extent of hybridisation.

References

Anderson EC, Thompson EA (2002) A model-based method for identifying species hybrids using multilocus genetic data. *Genetics* **160**, 1217–1229.

Arnold ML, Hamrick JL, Bennett BD (1990) Allozyme variation in louisiana irises - a test for introgression and hybrid speciation. *Heredity* **65**, 297–306.

Baker HG (1947) Criteria of hybridity. *Nature* **159**, 221–223.

Boyce P (1989) A new classification of *Arum* with keys to the infrageneric taxa. *Kew Bulletin* **44**, 383–395.

Boyce P (1993) *The genus Arum. A Kew Magazine Monograph*. HMSO, London.

Buzgo M, Soltis DE, Soltis PS, Ma H (2004) Towards a comprehensive integration of morphological and genetic studies of floral development. *Trends in Plant Science* **9**, 164–173.

Cronberg N, Natcheva R (2002) Hybridization between the peat mosses, *Sphagnum capillifolium* and *S. quinquefarium* (Sphagnaceae, Bryophyta) as inferred by morphological characters and isozyme markers. *Plant Systematics and Evolution* **234**, 53–70.

Diaz A, Kite G (In press) Why be a rewarding trap? The evolution of floral rewards in *Arum* (Araceae) a genus characterised by sapromyophilous and saprocantarophilous pollination systems. *Biological Journal of the Linnean Society*.

Gonzalez-Rodriguez A, Arias DM, Valencia S, Oyama A (2004) Morphological and RAPD analysis of hybridization between *Quercus affinis* and *Q. laurina* (Fagaceae), two Mexican red oaks. *American Journal of Botany* **91**, 401–409.

Gottlieb LD (1972) Levels of confidence in the analysis of hybridisation in plants. *Annals of the Missouri Botanical Garden* **59**, 435–446.

Gottlieb LD (1984) Genetics and morphological evolution in plants. *The American Naturalist* **123**, 681–709.

Ishida TA, Hattori K, Sato H, Kimura MT (2003) Differentiation and hybridization between *Quercus crispula* and *Q. dentata* (Fagaceae): insights from

morphological traits, amplified fragment length polymorphism markers, and leafminer composition. *American Journal of Botany* **90**, 769–776.

Ketner EK, Mesler MR (2000) Evidence for natural selection in a fern hybrid zone. *American Journal of Botany* **87**, 1168-1174.

Klier K, Leoschke MJ, Wendel JF (1991) Hybridization and introgression in white and yellow ladyslipper orchids (*Cypripedium candidum* and *C. pubescens*). *Journal of Heredity* **82**, 305–318.

Levin DA, Francisco-Ortega J, Jansen RK (1996) Hybridization and the extinction of rare plant species. *Conservation Biology* **10**, 10–16.

Meyn O, Emboden WA (1987) Parameters and consequences of introgression in *Salvia apiana* X *S mellifera* (Lamiaceae). *Systematic Botany* **12**, 390–399.

Mullenniex A, Hardig TM, Mesler MR (1998) Molecular confirmation of hybrid swarms in the fern genus *Polystichum* (Dryopteridaceae). *Systematic Botany* **23**, 421–426.

Nei M, Li H (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences USA* **76**, 5269–5273.

O'Hanlon PC, Peakall R, Briese DT (1999) Amplified fragment length polymorphism (AFLP) reveals introgression in weedy *Onopordum* thistles: hybridization and invasion. *Molecular Ecology* **8**, 1239–1246.

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

Rieseberg LH (1997) Hybrid origins of plant species. *Annual Review of Ecology and Systematics* **28**, 359–389.

Rieseberg LH, Beckstromsternberg S, Doan K (1990) *Helianthus annuus* ssp. *texanus* has chloroplast DNA and nuclear ribosomal-RNA genes of *Helianthus debilis* ssp. *cucumerifolius*. *Proceedings of the National Academy of Sciences of USA* **87**, 593–597.

Rieseberg LH, Ellstrand NC (1993) What can molecular and morphological markers tell us about plant hybridisation? *Critical Reviews in Plant Science* **12**, 213–241.

Simpson GG (1961) *Principals of Animal Taxonomy* Columbia University Press, New York.

Swofford DL (1998) PAUP*: phylogenetic analysis using parsimony (*and other methods) version 4. Sinauer, Sunderland, Massachusetts, USA.

Wiley EO (1978) The evolutionary species concept reconsidered. *Systematic Zoology* **41**, 17–26.

Wolfe AD, Xiang QY, Kephart SR (1998) Assessing hybridization in natural populations of *Penstemon* (Scrophulariaceae) using hypervariable inter simple sequence repeat (ISSR) bands. *Molecular Ecology* 7, 1107–1125.

Chapter 6 Correlation of morphological and genetic characters: a discussion

In this thesis, a comparative analysis of morphological and molecular character variation was carried out to further our understanding of the correlation between morphological characters and evolutionary lineages. Comparing variation of morphological and molecular characters can test the use of morphological characters for describing differences within and between species and for understanding processes of evolution. The relationship between morphological character variation and genetic marker variation was investigated at three taxonomic levels in the genus *Arum*: within species (Chapter 3, intraspecific variation), among species (Chapter 4, phylogenetic relationships) and between hybridising species (Chapter 5, interspecific hybridisation).

Intraspecific variation

Species are known to comprise an assemblage of populations that vary in both morphological and genetic characters (Linhart and Grant, 1996). This was true for the tested species, *A. maculatum* and *A. italicum*, as they displayed both morphological and molecular marker variation within and among the sampled populations. The results presented were similar to these of previously published studies where morphological characters correlated with molecular variation, e.g. *Draba* (Scheen *et al.*, 2002) and *Cardamine* (Lihova *et al.*, 2000), and others where the variation did not correlate, e.g. *Begonia* (Matolweni *et al.*, 2000) and *Myrecugenia* (Jensen *et al.*, 2002). It was found

that different morphological characters correlated with molecular marker variation in the two studied species; for example, in *A. maculatum* spotting on the leaves did not correlate with genetic groups, whereas in *A. italicum* similar leaf patterning characters corresponded with two genetic groups. The mixture of results, with some species showing a correlation and others no correlation, indicates that correspondence of a morphological character to an evolutionary group cannot be assumed and must be tested.

In addition to differences between species, the correlation of morphological and molecular marker variation can be dependent on the populations studied, as there were different patterns found in sympatric and allopatric populations of subsp. *neglectum*. The different correlations in sympatric and allopatric populations of subsp. *neglectum* are potentially caused by crossbreeding between the two subspecies or competition for currently unquantified resources. The data presented in Chapter 3 indicate that gene flow is occurring in sympatric populations of the *A. italicum* subspecies. Gene flow between these taxa indicates that they are not isolated lineages. The non-independence of the two *A. italicum* subspecies conflicts with the assumption made at the start of the thesis: all taxa should follow the evolutionary species concept. If described taxa are to correspond with the evolutionary species concept they must match single independent evolutionary lineages (Wiley, 1978). However, at the subspecies level, demonstrated here in *A. italicum*, gene flow may occur, and therefore independence of the lineages depends upon whether populations are sympatric or allopatric. Future studies of interspecific variation should take this into account when investigating interspecific variation of subspecies with overlapping ranges.

Interspecific variation

Phylogenetic analyses of genera are often both concordant and discordant with taxonomies: examples can be found in the genera *Prunus* (Bortiri *et al.*, 2001), *Begonia* (Plana, 2003) and *Pinus* (Gernandt *et al.*, 2003) – the genus *Arum* is no exception. The phylogenetic analysis of the genus *Arum* was used to identify morphological characters used in the taxonomy that do not correspond with monophyletic groups.

The molecular datasets used for the phylogenetic analysis of *Arum* have a low resolving power at and below the sectional level. Other published phylogenetic analyses also have a low resolving power, for example in the genera *Veratrum* (Zomlefer *et al.*, 2003) and *Draba* (Beilstein and Windham, 2003). However, the data collected here could distinguish between the two subgenera in the genus *Arum* (differentiated by flowering time). The sectional and subsectional differentiation, using tuber and floral characteristics, are not supported. Like the evolution of plane and concave leaves in genera of the Fontilaceae (Shaw and Allen, 2000) and floral characters in *Illicium* (Oh *et al.*, 2003) the data suggest that the morphological characters of tuber shape and flower height may have evolved more than once in *Arum*. These characters should therefore not be used in the taxonomy of *Arum* as they cannot be used to define evolutionary groups unambiguously.

Future work is needed to generate genetic data that can be used to distinguish the species in the subgenus *Arum* with a high statistical confidence. To achieve this greater

resolution in the phylogeny, a larger number of plastid and nuclear genetic markers are needed such as the 16 markers used to investigate phylogenetic relationships in cotton (Cronn *et al.*, 2002). An alternative approach to using more sequence data would be to generate ISSR profiles for each species within the subgenus *Arum*. This technique has been successfully used in *Viola* (Yockteng *et al.*, 2003) and in the Asphodelaceae (Treutlein *et al.*, 2003). The ISSR technique offers an alternative to further sequencing as it can reveal differences between and within species as demonstrated by the population and hybridisation sections of this thesis. The major drawback of the ISSR technique is caused by unknown homology of DNA fragments. In the ISSR technique DNA fragments of equal size are assumed to be homologous, but this assumption ignores the potential sequence differences between the primer binding sites. Homology of the fragments could be determined by cloning each fragment from each species and sequencing the clones.

Although the phylogenetic data for *Arum* have a low resolving power, tuber shape (either rhizomatous or discoid) and some inflorescence characters (flag/cryptic and reflexed spathe) may have evolved more than once. The evolution of tuber shape in *Arum* could be due to the two lineages *A. maculatum* and *A. italicum* adapting to the similar habitat of deciduous woodlands. The evolution of flower display (either flag or cryptic) was shown to be convergent (independent evolution of the same character in different lineages) in the related species *A. creticum*, *A. korolkowii* and *A. euxinum*. The convergence may have occurred through the species adapting to exploit similar pollinators; such an occurrence is not uncommon. For example, anther cones that are associated with buzz pollination in the genus *Solanum* have evolved by two mutually

exclusive pathways in the species *Solanum lycopersicum* and *S. dulcamara* (Glover *et al.*, 2004). The anther cone in *S. lycopersicum* is formed by interlocking hairs whereas in *S. dulcamara* the anthers are stuck together by a cellular secretion (Glover *et al.*, 2004). The convergent evolution of floral characters is also widespread in angiosperms; bilateral flowers have evolved more than once throughout angiosperm evolution, coinciding with the evolution of bees 60 million years after the origin of the angiosperms (Dilcher, 2000). If inflorescence characters are likely to be homoplasious, they should be used with caution in taxonomy. Cautious use of inflorescence characters is especially important as it has been demonstrated that closely related taxa with very similar morphological characters can be homoplasious, as exemplified by *Solanum* and *Arum*; character homology cannot be assumed at any taxonomic level.

Interspecific hybridisation

A correlation between morphological and molecular marker variation has been reported as an exception rather than the norm in plant hybrid zones (Rieseberg and Ellstrand, 1993). This thesis has, however, found a significant correlation between morphological and molecular marker variation in a hybrid zone of *A. idaeum* and *A. creticum*. Intermediate morphological characters (between *A. idaeum* and *A. creticum*) were found to be a good indicator of hybridisation. Many reasons are cited for a lack of correlation of morphological and molecular marker variation in hybrid zones, such as: unpredictable interactions between divergent genomes; the dominance of one genotype over another; and selection (e.g. Gonzalez-Rodriguez *et al.*, 2004). The correlation of morphological and molecular marker variation in the *A. creticum* × *A. idaeum* hybrid

zone indicates that continuous morphological characters can be used as indicators of hybrid status.

The correlation between morphological and molecular marker variation in the *A. creticum* × *A. idaeum* hybrid zone exists because the two species are closely related and therefore the genomes are not divergent. Evidence suggesting that *A. creticum* and *A. idaeum* are not highly divergent comes from their phylogeny. The phylogeny of *Arum* (Chapter 4) indicates that *A. creticum* and *A. idaeum* are closely related, as they are always united in a monophyletic clade. Selection acting in the hybrid zones can probably be ruled out in Crete as the hybrid morphs were found growing in the same parental habitats, unlike intermediate populations of *Quercus* or *Polystichum* hybrids (Gonzalez-Rodriguez *et al.*, 2004; Ketner and Mesler, 2000). Variation of continuous floral characters correlated with molecular marker variation in this hybrid zone. Categorical characters were not as useful for predicting molecular marker variation in this hybrid zone. Without testing the relationship between morphological characters with molecular characters, the utility of morphological characters in diagnosing hybrid zones cannot be predicted. Therefore, molecular characters should be used in conjunction with continuous morphological characters for investigating hybrid zones.

Genetic data and taxonomy

Taxonomists often call for the increased use of genetic data, and there have been several papers requesting a formal use of DNA sequences in taxonomies (e.g. Hebert *et al.*, 2003; Tautz *et al.*, 2003). This is a commendable scheme as the inclusion of more data

increases the accuracy of a taxonomy, which is a goal taxonomists strive for (Stace, 1989). The results presented here certainly support the use of DNA data for improving taxonomies; however, there are potential problems associated with this goal, including finding informative DNA sequences and the financial implications of large-scale sequencing projects highlighted by Libscomb *et al.* (2003) and Seberg *et al.* (2003).

It is entirely possible to find DNA regions that are variable in one species but not another, as demonstrated in the *Arum* phylogeny; the *trnL* DNA region could not differentiate among all *Arum* species despite being useful in other genera such as *Geum* (Smedmark and Eriksson, 2002). The problem of finding suitable DNA sections can only be solved by surveying large numbers of loci, which leads to issues including money and time. Sequencing is expensive; more loci would have been surveyed in this study if funds were available. There are also concerns that developing countries will not be able to afford to undertake large sequencing projects (Seberg *et al.*, 2003). However, at the population level, ISSR variation was successful at revealing informative molecular marker variation; screening ISSR variation is cheaper than large scale-sequencing projects. ISSR variation has limited use for phylogenetic systematics because band homology cannot be ensured without laborious cloning and sequencing. Despite this, ISSR variation has been used in phylogenetic investigations, for example in the genus *Viola* (Yockteng *et al.*, 2003), but ISSR variation is used as an addition to sequence variation for which there are fewer doubts of marker homology.

Aside from the technical problems, there is a fundamental reason why DNA data, although useful in understanding the consequences of biological processes, should not

become a primary tool for describing differences among species at the expense of reducing data from other sources. DNA variation cannot currently provide information regarding species biology, for which morphological, behavioural and reproductive data need to be obtained. Once information about species biology is known, it can be used to understand the causes and implications of molecular marker variation. For example, without the knowledge that *A. italicum* subsp. *italicum* and *A. italicum* subsp. *neglectum* share pollinators in Europe (Albre *et al.*, 2003) and Britain (Diaz and Kite, 2002) gene flow between them would not have been suspected. Without this knowledge, separating the populations into sympatric and allopatric would not have been a logical analytical step and patterns of variation would have been hidden by conflicting patterns of data. Thus a combination of data from different sources should be used when investigating population-level variation.

The relationship between phylogenetics and population genetics

By using both phylogenetic and population-level studies, this thesis, has been able to show that the evolutionary closeness of *A. creticum* and *A. idaeum* could account for the strong relationship between morphology and molecular marker variation as the genomes of the two species are not highly divergent. Similarly, the usefulness of leaf patterning corresponding to genetic groups in the species *A. italicum* and the ineffectiveness in *A. maculatum* is only understood by the realisation that these species are not closely related, as indicated by the morphological taxonomy, but are highly divergent.

Phylogenetics shows the evolutionary relationships among species and can be used to ensure the taxonomies reflect these relationships.

In addition to a taxonomic use, phylogenies, such as those in the genus *Lithophragma* (Kuzoff *et al.*, 2001), can show how characters have evolved. For example, ovary evolution in *Lithophragma* is bidirectional: some species have superior and some inferior ovaries (Kuzoff *et al.*, 2001). Phylogenetics alone cannot be used to achieve the main aim of evolutionary biology, i.e. to understanding the processes involved in species change; variation within populations needs to be investigated. Phylogenetic data only allow us to speculate on what has driven morphological change. In *Arum* adaptation to different habitats or a random event could have caused the evolution of different tuber morphologies it is not possible to infer the detailed population processes (e.g. isolation by distance, adaptation to an environmental variable, or selection) involved in species evolution from phylogenetics.

To be able to explain why there are patterns of variation, populations must be studied, as this is where the variation is subjected to selection and random processes. Selection often comes from environmental differences such as elevation (Wilson *et al.*, 2001) or moisture levels (Bruschi *et al.*, 2003). The studies of *A. italicum*, *A. maculatum* and the *A. creticum* × *A. idaeum* hybrid zone show how investigations of population-level variation can uncover potential factors that affect both morphological and molecular marker variation. These factors include isolation by distance observed in *A. maculatum*, and the putative effect of past glaciations and subsequent colonisation in subspecific taxa of *A. italicum*. Investigations of patterns of variation can show how selection is

affecting patterns of variation (e.g. Gonzalez-Rodriguez *et al.*, 2004; Ketner and Mesler, 2000). The *A. creticum* and *A. idaeum* study showed how differences in reproductive behaviour of the two species, in the absence of strong selection, can affect morphological and molecular marker variation within hybrid zones. The outcrossing nature of *A. creticum* is promoting the incorporation of the *A. idaeum* genome into *A. creticum*. In contrast, selfing in *A. idaeum* is limiting introgression from *A. creticum*, which reduces the likelihood that this narrow-range endemic will go extinct through hybridisation. Without a phylogeny we would not understand the relationships between the studied species and without a population-level study we would not be able to show what was causing genetic and morphological variation; both are needed in a study of variation.

Use of molecular markers within this thesis

All molecular markers that are available to modern researchers have advantages and disadvantages; it is up to the researcher to use the most appropriate marker to answer the question. A marker may be useful at one taxonomic level and not another, for example, the slow mutation rate of *rbc*L means that it is a useful marker for distinguishing old evolutionary differences between families (e.g. Chase *et al.*, 1993) but, is not suitable for identifying more recent evolutionary differences at the specific level (Soltis and Soltis, 1998). The type of statistics used, and therefore the type of question that can be answered differs between marker data. Sequences are more often used for phylogenetics than ISSR or AFLP markers as homology can be assigned more

easily to sequence data than AFLP or ISSR marker bands, making these markers less reliable for phylogenetic analysis (Wolfe and Liston, 1998).

The choice of markers for this thesis was heavily constrained by cost and development time, thus relatively inexpensive markers that required only a single stage of optimisation were chosen for Chapters 3 and 4. The most practical markers that could provide data to answer the questions set out in Chapter 1, and meet both budget and time constraints were ISSRs. However, ISSR markers are dominant markers (homo and heterozygotes can not be distinguished) thus, the data could not be used to estimate allele frequencies and hence the amount of heterozygosity. Using a codominant marker (allozyes) Matolweni (*et al.*, 2000) showed that in African *Begonia* populations, there were few heterozygotes, and very limited gene flow among the populations which indicates limited dispersal events. A similar estimate of gene flow is not possible with the ISSR data.

An estimate of gene flow would have been especially useful for the *A. italicum* section. If estimates of gene flow were possible, a better understanding of the frequency of crossing between the evolutionary lineages corresponding to subsp. *italicum* and subsp. *neglectum* could have been gained. Gene flow between taxa can also lead to little differences in heterozygosity among populations, it is possible that crossing between the two *A. italicum* subspecies could lead to high levels of heterozygotes as seen in Italian *Quercus patraea* populations growing near to *Q. pubescens* (Bruschi *et al.*, 2003). An estimate of gene flow between the subspecies would allow an assessment of the

conservation status of subsp. *neglectum,* and is important because subsp. *neglectum* it is a narrow range endemic. Thus, development of microsatellite markers is the most pertinent step to extend the work within Chapter 3.

Unlike the population level work in Chapter 3, the hybrid zone study (Chapter 5) would not benefit from the application of codominant SSR markers. Uni-parentally inherited markers (cp-DNA/mt-DNA) would be the most informative. The ISSR data, as demonstrated by a study in *Penstemon* (Wolfe 1998a, b) was also able to show that *A. idaeum* and *A. creticum* are undergoing introgressive hybridisation. The ISSR marker data cannot be used to identify the direction of crossing in the hybrid zone, as ISSR markers are bi-parentally inherited (Zietkiewicz, 1994). If a cp-DNA marker such as a gene sequence or RFLPs as used by Rieseberg *et al.,* (1990) in sunflowers, were applied to *A. ideaum* X *A. creticum,* it would be possible to test the direction of crossing and identify if either species are in danger of becoming extinct through hybridisation on Crete.

Conclusions

This thesis has presented a comparative analysis of morphological and molecular marker variation at three taxonomic levels within the genus *Arum*. No one morphological character or type of character (e.g. reproductive vs. vegetative) was shown to correlate with molecular marker variation at all the tested taxonomic levels. It has been shown that genetic and morphological variation within species sometimes

correlates depending on the morphological character and the species. Additionally, morphological and genetic character variation has been shown to be affected by geographical distance among populations and by reproductive behaviour. Variation within subspecies can also be affected by cross-breeding of intraspecific lineages that are not completely isolated (e.g. *A. italicum* subspecies).

The phylogenetic data presented have shown that both reproductive and vegetative characters can be homoplasious and misleading to taxonomy. Characters used in taxonomies therefore need to be tested with genetic data to ensure that classifications reflect evolutionary relationships of species. As both vegetative and reproductive characters have been shown to be convergent, it appears that extrinsic factors affecting both vegetative growth and reproductive biology (possibly selection mediated by pollinators) have been important in the diversification of the genus.

Within hybrid zones continuous morphological characters were shown to correlate significantly with molecular marker variation, and could be used as indicators of hybridisation. However, categorical characters were less useful. Continuous characters that show a positive correlation with variation can and should be used in future analyses of hybridisation. The reproductive behaviour of the tested species was found to be having an effect on the direction of introgression, which in future could affect the taxonomic status of at least one of the species.

The importance of testing morphological character variation with molecular marker variation has been shown throughout. By carrying out the investigation at three taxonomic levels within a single genus, it is apparent that even in closely related lineages, character usefulness cannot be assumed from any precedents set by a related lineage. The correlation between morphological and genetic characters is affected by population processes, which may not be uniform throughout a study area. Therefore, without testing, there is always uncertainty when using morphological characters to describe relationships within and between species and when diagnosing hybridisation.

References

Albre J, Quilichini A, Gibernau M (2003) Pollination ecology of *Arum italicum* (Araceae). *Botanical Journal of the Linnean Society* **141**, 205–214.

Beilstein MA, Windham MD (2003) A phylogenetic analysis of Western North American *Draba* (Brassicaceae) based on nuclear ribosomal DNA sequences from the ITS region. *Systematic Botany* **28**, 584–592.

Bortiri E, Oh SH, Jiang JG, Baggett S, Granger A, Weeks C, Buckingham M, Potter D, Parfitt DE (2001) Phylogeny and systematics of *Prunus* (Rosaceae) as determined by sequence analysis of ITS and the chloroplast *trnL-trnF* spacer DNA. *Systematic Botany* **26**, 797–807.

Bruschi P, Vendramin G, Bussotti F, Grossoni P (2003) Morphological and molecular diversity among Italian populations of *Quercus petraea* (Fagaceae). *Annals of Botany* **91**, 707–716.

Chase MW, Soltis DE, Olmstead RG, Morgan D, Les DH, Mishler BD, Duvall MR, Price RA, Hills HG, Qiu YL, Kron KA, Rettig JH, Conti E, Palmer JD, Manhart JR, Sytsma KJ, Michaels HJ, Kress WJ, Karol KG, Clark WD, Hedren M, Gaut BS, Jansen RK, Kim KJ, Wimpee CF, Smith JF, Furnier GR, Strauss SH, Xiang QY, Plunkett GM, Soltis PS, Swensen SM, Williams SE, Gadek PA, Quinn CJ, Eguiarte LE, Golenberg E, Learn GH, Graham SW, Barrett SCH, Dayanandan S, Albert VA (1993) Phylogenetics of seed plants - an analysis of nucleotide sequences from the plastid gene Rbcl. *Annals of the Missouri Botanical Garden* **80**, 528–580.

Cronn RC, Small B, Haselkron T, Wendel JF (2002) Rapid diversification of the cotton genus *Gossipium* (Malvaceae): revealed by sixteen nuclear and chloroplast genes. *American Journal of Botany* **89**, 707–725.

Diaz A, Kite G (2002) A comparison of the pollination ecology of *Arum maculatum* and *Arum italicum* in England. *Watsonia* **24**, 171–181.

Dilcher D (2000) Toward a new synthesis: Major evolutionary trends in the angiosperm fossil record. *Proceedings of the National Acadamy of Sciences USA* **97**, 7030–7036.

Gernandt DS, Liston A, Pinero D (2003) Phylogenetics of *Pinus* subsections Cembroides and Nelsoniae inferred from cpDNA sequences. *Systematic Botany* **28**, 657–673.

Glover GJ, Bunnewell S, Martin C (2004) Convergent evolution in the genus *Solanum*: the specialised anther cone develops through alternative pathways. *Gene* **331**, 1–7.

Gonzalez-Rodriguez A, Arias DM, Valencia S, Oyama A (2004) Morphological and RAPD analysis of hybridization between *Quercus affinis* and *Q. laurina* (Fagaceae), two Mexican red oaks. *American Journal of Botany* **91**, 401–409.

Hebert PDN, Cywinska A, Ball SL, DeWaard JR (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London Series B-Biological Sciences* **270**, 313–321.

Jensen RJ, Schwoyer M, Crawford DJ, Stuessy TF, Anderson GJ, Baeza CM, O MS, Ruiz E (2002) Patterns of morphological and genetic variation among populations of *Myrceugenia fernandeziana* (Myrtaceae) on Masatierra island: Implications for conservation. *Systematic Botany* **27**, 534–547.

Ketner EK, Mesler MR (2000) Evidence for natural selection in a fern hybrid zone. *American Journal of Botany* **87**, 1168–1174.

Kuzoff RK, Hufford L, Soltis DE (2001) Structural homology and developmental transformations associated with ovary diversification in Lithophragma (Saxifragaceae). *American Journal of Botany* **88**, 196–205.

Lihova J, Marhold K, Neuffer B (2000) Taxonomy of *Cardamine amara* (Cruciferae) in the Iberian Peninsula. *Taxon* **49**, 747–763.

Linhart YB, Grant MC (1996) Evolutionary significance of local genetic differentiation. *Annual Review of Ecology and Systematics* **27**, 237–277.

Lipscomb D, Platnick N, Wheeler Q (2003) The intellectual content of taxonomy: a comment on DNA taxonomy. *Trends in Ecology & Evolution* **18**, 65–66.

Matolweni LO, Balkwill K, McLellan T (2000) Genetic diversity and gene flow in the morphologically variable, rare endemics *Begonia dregei* and *Begonia homonyma* (Begoniaceae). *American Journal of Botany* **87**, 431–439.

Oh IC, Denk T, Friis EM (2003) Evolution of *Illicium* (Illiciaceae): Mapping morphological characters on the molecular tree. *Plant Systematics and Evolution* **240**, 175–209.

Plana V (2003) Phylogenetic relationships of the Afro-Malagasy members of the large genus *Begonia* inferred from *trnL* intron sequences. *Systematic Botany* **28**, 693–704.

Rieseberg LH, Beckstromsternberg S, Doan K (1990) *Helianthus annuus* ssp *texanus* has chloroplast DNA and nuclear ribosomal-RNA genes of *Helianthus debilis* ssp. *cucumerifolius*. *Proceedings of the National Academy of Sciences of the United States of America* 87: 593-597

Rieseberg LH, Ellstrand NC (1993) What can molecular and morphological markers tell us about plant hybridisation? *Critical Reviews in Plant Science* **12**, 213–241.

Scheen AC, Elven R, Brochmann C (2002) A molecular-morphological approach solves taxonomic controversy in arctic *Draba* (Brassicaceae). *Canadian Journal of Botany-Revue Canadienne De Botanique* **80**, 59–71.

Seberg O, Humphries CJ, Knapp S, Stevenson DW, Petersen G, Scharff N, Andersen NM (2003) Shortcuts in systematics? A commentary on DNA-based taxonomy. *Trends in Ecology & Evolution* **18**, 63–65.

Shaw AJ, Allen B (2000) Phylogenetic relationships, morphological incongruence, and geographic speciation in the Fontinalaceae (Bryophyta). *Molecular Phylogenetics and Evolution* **16**, 225–237.

Smedmark JEE, Eriksson T (2002) Phylogenetic relationships of *Geum* (Rosaceae) and relatives inferred from the nrITS and trnL-trnF regions. *Systematic Botany* **27**, 303–317.

Soltis DE, Soltis PS (1998) Choosing an Approach and an Appropriate Gene for Phylogenetic Analysis. In: Soltis DE, Soltis, P.S. and Doyle, J.J (ed) Molecular Systematics of Plants II: DNA Sequencing. Kluwer Academic Publishers, Boston, pp 1-42

Stace CA (1989) *Plant Taxonomy and Biosystematics*, 2nd edn. Cambridge University Press, Cambridge.

Tautz D, Arctander P, Minelli A, Thomas RH, Vogler AP (2003) A plea for DNA taxonomy. *Trends in Ecology & Evolution* **18**, 70-74.

Treutlein J, Smith GF, Van Wyk BE, Wink M (2003) Phylogenetic relationships in Asphodelaceae (subfamily Alooideae) inferred from chloroplast DNA sequences (rbcL, matK) and from genomic fingerprinting (ISSR). *Taxon* **52**, 193-207.

Wiley EO (1978) The evolutionary species concept reconsidered. *Systematic Zoology* **41**, 17-26.

Wilson BL, Kitzmiller J, Rolle W, Hipkins VD (2001) Isozyme variation and its environmental correlates in *Elymus glaucus* from the California Floristic Province. *Canadian Journal of Botany* **79**, 139-153.

Wolfe AD, Liston A (1998) Contributions of PCR-Based Methods to Plant Systematics and Evolutionary Biology. In: Soltis DE, Soltis, P.S. and Doyle. J.J. (ed) Molecualr Systematics of Plants II: DNA Sequencing. Kulwer Accademic Press Boston, pp 43-86

Wolfe AD, Xiang QY, Kephart SR (1998a) Assessing hybridization in natural populations of *Penstemon* (Scrophulariaceae) using hypervariable inter simple sequence repeat (ISSR) bands. *Molecular Ecology* **7**: 1107-1125

Wolfe AD, Xiang QY, Kephart SR (1998b) Diploid hybrid speciation in *Penstemon* (Scrophulariaceae). *Proceedings of the National Academy of Sciences USA* **95**: 5112-5115

Yockteng R, Ballard HE, Mansion G, Dajoz I, Nadot S (2003) Relationships among pansies (Viola section Melanium) investigated using ITS and ISSR markers. *Plant Systematics and Evolution* **241**, 153-170.

Zietkiewicz E, Rafalski A, Labuda D (1994) Genome Fingerprinting by Simple Sequence Repeat (SSR)-Anchored Polymerase Chain-Reaction Amplification. *Genomics* **20**: 176-183

Zomlefer WB, Whitten WM, Williams NH, Judd WS (2003) An overview of *Veratrum* s.l. (Liliales : Melanthiaceae) and an infrageneric phylogeny based on ITS sequence data. *Systematic Botany* **28**, 250-269.

Appendix A

Alignment of *trnL* sequences

```
                    ....|....| ....|....| ....|....| ....|....| ....|....|
                            10         20         30         40         50
A.balansan          GACATGTAGA ATGGGACTCT CTCTTTATTC TCGTCCGATT AATCATTTTT
A.euxinum           GACATGTAGA ATGGGACTCT CTCTTTATTC TCGTCCGATT AATCATTTTT
A.alpinum           GACATGTAGA ATGGGACTCT CTCTTTATTC TCGTCCGATT AATCA-TTTT
A.hygrophy          GACATGTAGA ATGGGACTCT CTCTTTATTC TCGTCCGATT AATCC---TT
A.oriental          GACATGTAGA ATGGGACTCT CTCTTTATTC TCGTCCGATT AATCA--TTT
A.cyrenaic          GACATGTAGA ATGGGACTCT CTCTTTATTC TCGTCCGATT AATCATTTTT
A.cyrenaic          GACATGTAGA ATGGGACTCT CTCTT-ATTC TCGTCCGAT- -ATCA-TTTT
A.cyrenaic          GACATGTAGA ATGGGACTCT CTCTTTATTC TCGTCCGATT AATCA-TTTT
A.dioscori          GACATGTAGA ATGGGACTCT CTCTTTATTC TCGTCCGATT AATCA---TT
A.dioscori          GACATGTAGA ATGGGACTCT CTCTT-ATTC TCGTCCGATT -ATCA-----
A.dioscori          GACATGTAGA ATGGGACTCT CTCTT-ATTC TCGTCCGATT AATCA----T
A.elongatu          GACATGTAGA ATGGGACTCT CTCTTTATTC TCGTCCGAT- -ATCA----T
A.jaquemon          GACATGTAGA ATGGGACTCT CTCTTTATTC TCGTCCGATT AATCA---TT
A.korolkow          GACATGTAGA ATGGGACTCT CTCTTTATTC TCGTCCGATT AATCA--TTT
A.korokowi          GACATGTAGA ATGGGACTCT CTCTTTATTC TCGTCCGATT AATCA--TTT
A.pictum            GACATGTAGA ATGGGACTCT CTCTT-ATTC TCGTCCGATT AATCA--TTT
A.byzantin          GACATGTAGA ATGGGACTCT CTCTT-ATTC TCGTCCGATT AATCA----T
A.concinna          GACATGTAGA ATGGGACTCT CTCTTTATTC TCGTCCGATT AATCA---TT
A.palestin          GACATGTAGA ATGGGACTCT CTCTTTATTC TCGTCCGATT AATCA--TTT
A.rupicola          GACATGTAGA ATGGGACTCT CTCTTTATTC TCGTCCGATT AATCA-TTTT
A.idaeum            GACATGTAGA ATGGGACTCT CTCTT-ATTC TCGTCCGATT AATCA--TTT
A.creticum          GACATGTAGA ATGGGACTCT CTCTT-ATTC TCGTCCGATT AATCA----T
A.italicum          GACATGTAGA ATGGGACTCT CTCTT-ATTC TCGTCCGATT AATCA----T
A.maculatu          GACATGTAGA ATGGGACTCT CTCTT-ATTC TCGTCCGATT AATCATTTTT
Biarum_ten          GACATGTAGA ATGGGACTCT CTCTT-ATTC TCGTCCGATT AATCA---TT
B.t.subsp.          GACATGTAGA ATGGGACTCT CTCTT-ATTC TCGTCCGATT AATCA--TTT

                    ....|....| ....|....| ....|....| ....|....| ....|....|
                            60         70         80         90        100
A.balansan          TTTTTTTTTC AAAAGATCAA TCAGACTCTG GAGTGAATGA TTTGATCACT
A.euxinum           TTTTTTTTTC AAAAGATCAA TCAGACTCTG GAGTGAATGA TTTGATCACT
A.alpinum           TTTTTTTTTC AAAAGATCAA TCAGACTCTG GAGTGAATGA TTTGATCACT
A.hygrophy          TTTTTTTTTC AAAAGATCAA TCAGACTCTG GAGTGAATGA TTTGATCACT
A.oriental          TTTTTTTTTC AAAAGATCAA TCAGACTCTG GAGTGAATGA TTTGATCACT
A.cyrenaic          TTTTTTTTTC AAAAGATCAA TCAGACTCTG GAGTGAATGA TTTGATCACT
A.cyrenaic          TTTTTTTTTC AAAAGATCAA TCAGACTCTG GAGTGAATGA TTTGATCACT
A.cyrenaic          TTTTTTTTTC AAAAGATCAA TCAGACTCTG GAGTGAATGA TTTGATCACT
A.dioscori          TTTTTTTTTC AAAAGATCAA TCAGATTCTG GAGTGAATGA TTTGATCACT
A.dioscori          TTTTTTTTTC AAAAGATCAA TCAGACTCTG GAGTGAATGA TTTGATCACT
A.dioscori          TTTTTTTTTC AAAAGATCAA TCAGACTCTG GAGTGAATGA TTTGATCACT
A.elongatu          TTTTTTTTTC AAAAGATCAA TCAGACTCTG GAGTGAATGA TTTGATCACT
A.jaquemon          TTTTTTTTTC AAAAGATCAA TCAGACTCTG GAGTGAATGA TTTGATCACT
A.korolkow          TTTTTTTTTC AAAAGATCAA TCAGACTCTG GAGTGAATGA TTTGATCACT
A.korokowi          TTTTTTTTTC AAAAGATCAA TCAGACTCTG GAGTGAATGA TTTGATCACT
A.pictum            TTTTTTTTTC AAAAGATCAA TCAGACTCTG GAGTGAATGA TTGGATCACT
```

```
A.byzantin    TTTTTTTTTC AAAAGATCAA TCAGACTCTG GAGTGAATGA TTTGATCACT
A.concinna    TTTTTTTTTC AAAAGATCAA TCAGACTCTG GAGTGAATGA TTTGATCACT
A.palestin    TTTTTTTTTC AAAAGATCAA TCAGACTCTG GAGTGAATGA TTTGATCACT
A.rupicola    TTTTTTTTTC AAAAGATCAA TCAGACTCTG GAGTGAATGA TTTGATCACT
A.idaeum      TTTTTTTTTC AAAAGATCAA TCAGACTCTG GAGTGAATGA TTTGATCACT
A.creticum    TTTTTTTTTC AAAAGATCAA TCAGACTCTG GAGTGAATGA TTTGATCACT
A.italicum    TTTTTTTTTC AAAAGATCAA TCAGACTCTG GAGTGAATGA TTTGATCACT
A.maculatu    TTTTTTTTTC AAAAGATAAA TAAAACTCTG AAGTAAATGA TTGGATAACT
Biarum_ten    TTTTTTTTTC AAAAGATCAA TCAAACTCTG GAGTGAATGA TTTGATCACT
B.t.subsp.    TTTTTTTTTC AAAAAATCAA TCAAACTCTG GAGTGAATGA TTTGATCACT

              ....|....| ....|....| ....|....| ....|....| ....|....|
                       110        120        130        140        150
A.balansan    GAATATTCGA CTCTTACTTC AATTTCAATG TTAGAATGGA TTCGCAATAA
A.euxinum     GAATATTCGA CTCTTACTTC AATTTCAATG TTAGAATGGA TTCGCAATAA
A.alpinum     GAATATTCGA CTCTTACTTC AATTTCAATG TTAGAATGGA TTCGCAATAA
A.hygrophy    GAATATTCGA CTCTTACTTC AATTTCAATG TTAGAATGGA TTCGCAATAA
A.oriental    GAATATTCGA CTCTTACTTC AATTTCAATG TTAGAATGGA TTCGCAATAA
A.cyrenaic    GAATATTCGA CTCTTACTTC AATTTCAATG TTAGAATGGA TTCGCAATAA
A.cyrenaic    GAATATTCGA CTCTTACTTC AATTTCAATG TTAGAATGGA TTCGCAATAA
A.cyrenaic    GAATATTCGA CTCTTACTTC AATTTCAATG TTAGAATGGA TTCGCAATAA
A.dioscori    GAATATTCGA CTCTTACTTC AATTTCAATG TTAGAATGGA TTCGCAATAA
A.dioscori    GAATATTCGA CTCTTACTTC AATTTCAATG TTAGAATGGA TTCGCAATAA
A.dioscori    GAATATTCGA CTCTTACTTC AATTTCAATG TTAGAATGGA TTCGCAATAA
A.elongatu    GAATATTCGA CTCTTACTTC AATTTCAATG TTAGAATGGA TTCGCAATAA
A.jaquemon    GAATATTCGA CTCTTACTTC AATTTCAATG TTAGAATGGA TTCGCAATAA
A.korolkow    GAATATTCGA CTCTTACTTC AATTTCAATG TTAGAATGGA TTCGCAATAA
A.korokowi    GAATATTCGA CTCTTACTTC AATTTCAATG TTAGAATGGA TTCGCAATAA
A.pictum      GAATATTCGA CTCTTACTTC AATTTCAATG TTAGAATGGA TTCGCAATAA
A.byzantin    GAATATTCGA CTCTTACTTC AATTTCAATG TTAGAATGGA TTCGCAATAA
A.concinna    GAATATTCGA CTCTTACTTC AATTTCAATG TTAGAATGGA TTCGCAATAA
A.palestin    GAATATTCGA CTCTTACTTC AATTTCAATG TTAGAATGGA TTCGCAATAA
A.rupicola    GAATATTCGA CTCTTACTTC AATTTCAATG TTAGAATGGA TTCGCAATAA
A.idaeum      GAATATTCGA CTCTTACTTC AATTTCAATG TTAGAATGGA TTCGCAATAA
A.creticum    GAATATTCGA CTCTTACTTC AATTTCAATG TTAGAATGGA TTCGCAATAA
A.italicum    GAATATTCGA CTCTTACTTC AATTTCAATG TTAGAATGGA TTCGCAATAA
A.maculatu    GAATATTCGA CTCTTACTTC AATTTCAATG TTAGAATGGA TTCGCAATAA
Biarum_ten    GAATATTCGA CTCTTACTTC AATTTCAATG TTAGAATGGA TTCGCAATAA
B.t.subsp.    GAATATTCGA CTCTTACTTC AATTTCAATG TTAGAATGGA TTCGGAATAA

              ....|....| ....|....| ....|....| ....|....| ....|....|
                       160        170        180        190        200
A.balansan    CTCTTTCATT TTTCATAAAA TTTTTAATTT ATTTTTTATT A-TTTATTT-
A.euxinum     CTCTTTCATT TTTCATAAAA TTTTTAATTT ATTTTTTATT A-TTTATTT-
A.alpinum     CTCTTTCATT TTTCATAAAA TTTTTAATTT ATTTTTTATT A-TTTATTT-
A.hygrophy    CTCTTTCATT TTTCATAAAA TTTTTAATTT ATTTTTTATT A-TTTATTT-
A.oriental    CTCTTTCATT TTTCATAAAA TTTTTAATTT ATTTTTTATT A-TTTATTT-
A.cyrenaic    CTCTTTCATT TTTCATAAAA TTTTTAATTT ATTTTTTATT A-TTTATTT-
A.cyrenaic    CTCTTTCATT TTTCATAAAA TTTTTAATTT ATTTTTTATT A-TTTATTT-
A.cyrenaic    CTCTTTCATT TTTCATAAAA TTTTTAATTT ATTTTTTATT A-TTTATTT-
A.dioscori    CTCTTTCATT TTTCATAAAA TTTTGAATTT ATTTTTTATT A-TTTATTTA
A.dioscori    CTCTTTCATT TTTCATAAAA TTTTGAATTT ATTTTTTATT A-TTTATTTA
A.dioscori    CTCTTTCATT TTTCATAAAA TTTTGAATTT ATTTTTTATT A-TTTATTTA
A.elongatu    CTCTTTCATT TT-CATAAA- -TTTTAA-TT ATTTTTTATT A-TTTATTT-
A.jaquemon    CTCTTTCATT TTTCATAAAA TTTTTAATTT ATTTTTTATT A-TTTATTT-
A.korolkow    CTCTTTCATT TTTCATAAAA TTTTTAATTT ATTTTTTATT A-TTTATTTA
```

```
A.korokowi    CTCTTTCATT TTTCATAAAA TTTTTAATTT ATTTTTTATT A-TTTATTTA
A.pictum      CTCTTTCATT TTTCATAAAA TTTTGAATTT AAATTTTATT A-TTTATTTA
A.byzantin    CTCTTTCATT TTTCATAAAA TTTTGAATTT ATTTTTTATT A-TTTATTTT
A.concinna    CTCTTTCATT TTTCATAAAA TTTTGAATTT ATTTTTTATT A-TTTATTTT
A.palestin    CTCTTTCATT TTTCATAAAA TTTTGAATTT ATTTTTTATT A-TTTATTTA
A.rupicola    CTCTTTCATT TTTCATAAAA TTTTTAATTT ATTTTTTATT A-TTTATTT-
A.idaeum      CTCTTTCATT TTTCATAAAA TTTTTAATTT ATTTTTTATT A-TTTATTT-
A.creticum    CTCTTTCATT TTTCATAAAA TTTTTAATTT ATTTTTTATT A-TTTATTT-
A.italicum    CTCTTTCATT TTTCATAAAA TTTTGAATTT ATTTTTTATT A-TTTATTTT
A.maculatu    CTCTTTCATT TTTCATAAAA TTTTTAATTT ATTTTT-ATT A-TTTATTT-
Biarum_ten    CTCTTTCATT TTTCATAAAA TTTTGAATTT ATTATTTATT AATTTATTT-
B.t.subsp.    CTCTTTCATT TTTCATAAAA TTTTGAATTT ATTATTTATT AATTTATTT-

              ....|....| ....|....| ....|....| ....|....| ....|....|
                       210        220        230        240        250
A.balansan    -----AAATT TAAATAAATA TAAAAATATA ATTTAAA--- ----------
A.euxinum     -----AAATT TAAATAAATA TAAAAATATA ATTTAAA--- ----------
A.alpinum     -----AAATT TAAATAAATA TAAAAATATA ATTTAAA--- ----------
A.hygrophy    -----AAATT TAAATAAATA TAAAAATATA ATTTAAA--- ----------
A.oriental    -----AAATT TAAATAAATA TAAAAATATA ATTTAAA--- ----------
A.cyrenaic    -----AAATT TAAATAAATA TAAAAATATA ATTTAAA--- ----------
A.cyrenaic    ------AATT TAAATAAATA TAAAAATATA ATT-AAA--- ----------
A.cyrenaic    -----AAATT TAAATAAATA TAAAAATATA ATTTAAA--- ----------
A.dioscori    AATTT-AATT TAAATAAATA TAAAAATATA ATTTAAA--- ----------
A.dioscori    AATTT-AATT TAAATAAATA TAAAAATATA ATTTAAA--- ----------
A.dioscori    AATTT-AATT TAAATAAATA TAAAAATATA ATTTAAA--- ----------
A.elongatu    -----AAATT T-AATAAATA TAAAAATATA ATTTAAA--- ------TATA
A.jaquemon    -----AAATT TAAATAAATA TAAAAATATA ATTTAAA--- ------TATA
A.korolkow    AATTTAAATT TAAATAAATA TAAAAATATA ATTTAAA--- -TTAAATATA
A.korokowi    AATTTAAATT TAAATAAATA TAAAAATATA ATTTAAA--- -TTAAATATA
A.pictum      AATTTAAA-- ----TAAATA TAAAAATATA ATTTAAATAA ATAAAATATA
A.byzantin    AATT-AAATT TAAATAAATA TAAAAATATA ATTTAAA--- ----------
A.concinna    AATT-AAATT TAAATAAATA TAAAAATATA ATTTAAA--- ----------
A.palestin    AATTT-AATT TAAATAAATA TAAAAATATA ATTTAAA--- ---------T
A.rupicola    -----AAATT TAAATAAATA TAAAAATATA ATTTAAA--- ------TATA
A.idaeum      -----AAATT TAAATAAATA TAAAAATATA ATTTAAA--- ----------
A.creticum    -----AAATT TAAATAAATA TAAAAATATA ATTTAAA--- ----------
A.italicum    AATT-AAATT TAAATAAATA TAAAAATATA ATTTAAA--- ----------
A.maculatu    -----AAATT TAAATAAATA TAAAAATATA ATTTAAA--- ----------
Biarum_ten    -----AAATT AAAATATAAA TAT---TAAA ATATAAATAA TAATAATAT-
B.t.subsp.    -----AAATT AAAATATATA TAAAA-TAAA ATATAAATAA TAATAATAT-

              ....|....| ....|....| ....|....| ....|....| ....|....|
                       260        270        280        290        300
A.balansan    --TAAATTAA ATATTTAAAT AG------AA ATAATATA-- ---------T
A.euxinum     --TAAATTAA ATATTTAAAT AG------AA ATAATATA-- ---------T
A.alpinum     --TAAATTAA ATATTTAAAT AG------AA ATAATATA-- ---------T
A.hygrophy    --TAAATTAA ATATTTAAAT AGAATTTAAA ATAATATA-- ---------T
A.oriental    --TAAATTAA ATATTTAAAT AG------AA ATAATATA-- ---------T
A.cyrenaic    --TAAATTAA ATATTTAAAT AG------AA ATAATATA-- ---------T
A.cyrenaic    --TAAAT-AA ATATTTAAAT AG------AA ATAATATA-- ---------T
A.cyrenaic    --TAAATTAA ATATTTAAAT AG------AA ATAATATA-- ---------T
A.dioscori    --TAAATTAA ATATTTAAAT AGAATTTAAA ATAATATAAA ---------T
A.dioscori    --TAAATTAA ATATT-AAAT AGAATT--AA ATAATATAAA ----TATATT
A.dioscori    --T-AATTAA ATATT-AAAT AGAATT-AAA ATAATATAAA ---------T
A.elongatu    AATAAAT-AA ATATTTAAAT AGAATTTAAA ATAATATA-- ----GATATT
```

```
A.jaquemon    AATAAATTAA ATATTTAAAT AGAATTTAAA ATAATATA-- ----TATATT
A.korolkow    AATAAATTAA ATATTTAAAT AGAATTTAAA ATAATATA-- ----TATATT
A.korokowi    AATAAATTAA ATATTTAAAT AGAATTTAAA ATAATATA-- ----TATATT
A.pictum      AATAAATTAA AAATATAAAT AAA-TTAAAT TAAAATTAAA ----TATATA
A.byzantin    --TAAATTAA ATATWTAAAT CGAA-TTAAA ATAATATAA- ----TATATT
A.concinna    ---------- -TATTTAAAT CGAATTTAAA ATAAAATAA- ----TATATT
A.palestin    AATAAATTAA AtaTTTAAAT AGAATTTAAA ATAATATAAA ----TATATT
A.rupicola    AATAAATTAA ATATTTAAAT AGAATTTAAA ATAATATA-- ----TATATT
A.idaeum      --TAAATTAA ATATTTAAAT AGAATTTTAA ATAATATA-- ---------T
A.creticum    --TAAATTAA ATATTTAAAT AGAATTTAAA ATAATATA-- ---------T
A.italicum    --TAAATTAA ATATTTAAAT CGAATTTAAA ATAATATA-- ---------T
A.maculatu    --TAAATTAA ATATTTAAAT AG------AA ATAATATA-- ---------T
Biarum_ten    ---------- ---------- ---------- ATAATATAAA TAATAATAAT
B.t.subsp.    ---------- ---------- ----AATAAT AATATATAAT ATAAATAATA

                  ....|....| ....|....| ....|....| ....|....| ....|....|
                          310        320        330        340        350
A.balansan    ATATAATATA AA-------- --TATTATGA TATAATAT-- ----------
A.euxinum     ATATAATATA AA-------- --TATTATGA TATAATAT-- ----------
A.alpinum     ATATAATATA AA-------- --TATTATGA TATAATAT-- ----------
A.hygrophy    ATATAATATA AA-------- --TATTATGA TATAATAT-- ----------
A.oriental    ATATAATATA AA-------- --TATTATGA TATAATAT-- ----------
A.cyrenaic    ATATAATATA AA-------- --TATTATGA TATAATAT-- ----------
A.cyrenaic    ATATAATATA AA-------- --TATTATGA TATAATAT-- ----------
A.cyrenaic    ATATAATATA AA-------- --TATTATGA TATAATAT-- ----------
A.dioscori    ATATAATATA AA-------- --TATTATGA TATAATAT-- ----------
A.dioscori    ATATAATATA AA-------- --TATTATGA TATAATAT-- ----------
A.dioscori    ATATAATATA AA-------- --TATTATGA TATAATAT-- ----------
A.elongatu    ATATAATATA AA-------- --TATTATGA TATAATAT-- ----------
A.jaquemon    ATATAATATA AA-------- --TATTATGA TATAATAT-- ----------
A.korolkow    ATATAATATA AA-------- --TATTATGA TATAATAT-- ----------
A.korokowi    ATATAATATA AA-------- --TATTATGA TATAATAT-- ----------
A.pictum      ATATAA-ATA AA-------- --TATTATGA TATAAATTAT AAATAATAAT
A.byzantin    ATATAATATA AA-------- --TATTATAA TATAATAT-- ----------
A.concinna    ATATAATATA AA-------- --TATTATAA TATAAATT-- ----------
A.palestin    ATATAATATA AA-------- --TATTATAA TATAATAT-- ----------
A.rupicola    ATATAATATA AA-------- --TATTATGA TATAATAT-- ----------
A.idaeum      ATATAATATA AA-------- --TATTATGA TATAATAT-- ----------
A.creticum    ATATAATATA AA-------- --TATTATGA TAT------- ----------
A.italicum    ATATAATATA AA-------- --TATTATAA TATAATAT-- ----------
A.maculatu    ATATAATATA AA-------- --TATTATGA TATAATAT-- ----------
Biarum_ten    ATA-----TA AAATATAATA TATATTATGA TAT------- ----------
B.t.subsp.    ATAATATATA AAATATAATA TATATAATGA TAT------- ----------

                  ....|....| ....|....| ....|....| ....|....| ....|....|
                          360        370        380        390        400
A.balansan    ---------- ---------- ---------- ---AAATTAT AAATAATAAA
A.euxinum     ---------- ---------- ---------- ---AAATTAT AAATAATAAA
A.alpinum     ---------- ---------- ---------- ---AAATTAT AAATAATAA-
A.hygrophy    ---------- ---------- -------AAA TATAAATTAT AAATAATAA-
A.oriental    ---------- ---------- ---------- ---AAATTAT AAATAATAA-
A.cyrenaic    ---------- ---------- ---------- ---AAATTAT AAATAATAA-
A.cyrenaic    ---------- ---------- ---------- ----AATTAT AAATA-TAA-
A.cyrenaic    ---------- ---------- ---------- ---AAATTAT AAATAATAA-
A.dioscori    ---------- ---------- ---------- ---AAATTAT AAATAATAA-
A.dioscori    ---------- ---------- ---------- ----AAATAT AAATAATAA-
```

```
A.dioscori   ---------- ---------- ---------- ---AAATTAT AAATAATAA-
A.elongatu   ---------- ---------- ---------- ----AAT-AT AAATAATAA-
A.jaquemon   ---------- ---------- ---------- ---AAATTAT AAATAATAA-
A.korolkow   ---------- ---------- ---------- ---AAATTAT AAATAATAA-
A.korokowi   ---------- ---------- ---------- ---AAATTAT AAATAATAA-
A.pictum     TATGATATAA ATTATAAATA ATAATTATGA TATAAATTAT AAATAATAA-
A.byzantin   ---------- ---------- ---------- ---AAATTAT AAATAATAA-
A.concinna   ---------- ---------- ------ATAA TATAAATTAT AAATAATAA-
A.palestin   ---------- ---------- ---------- ---AAATTAT AAATAATAAT
A.rupicola   ---------- ---------- ---------- ---AAATTAT AAATAATAAT
A.idaeum     ---------- ---------- ---------- ---AAATAAT AAATAATAA-
A.creticum   ---------- ---------- ---------- ---AAATTAT AAATAATAA-
A.italicum   ---------- ---------- ---------- ---AAATTAT AAATAATAA-
A.maculatu   ---------- ---------- ---------- ---AAATTAT AAATAATAA-
Biarum_ten   ---------- ---------- ---------- ---AAATTAT AAATAATAA-
B.t.subsp.   ---------- ---------- ---------- -------AAA AATAAATAAT

                 ....|....| ....|....| ....|....| ....|....| ....|....|
                         410        420        430        440        450
A.balansan   AATTATGGAT TTGGGTCGTG ATTAATCGTT TGATATGTCA GTATATAATA
A.euxinum    -TATATGGAT TTGGGTCGTG ATTAATCGTT TGATATGTCA GTATATAATA
A.alpinum    --TTATGGAT TTGGGTCGTG ATTAATCGTT TGATATGTCA GTATATAATA
A.hygrophy   --TTATGGAT TTGGGTCGTG ATTAATCGTT TGATATGTCA GTATAT-ATA
A.oriental   --TTATGGAT TTGGGTCGTG ATTAATCGTT TGATATGTCA GTATATAATA
A.cyrenaic   --TTTATGAT TTGGGTCGTG ATTAATCGTT TGATATGTCA GTATATAATA
A.cyrenaic   --TTATGGAT TTGGGTCGTG ATTAATCGTT TGATATGTCA GTATATAATA
A.cyrenaic   --TTATGGAT TTGGGTCGTG ATTAATCGTT TGATATGTCA GTATATAATA
A.dioscori   --TTATGGAT TTGGATCGTG ATTAATCGTT TGATATGTCA GTATAT-ATA
A.dioscori   --TTATGGAT TTGGATCGTG ATTAATCGTT TGATATGTCA GTATAT-ATA
A.dioscori   --TTATGGAT TTGGATCGTG ATTAATCGTT TGATATGTCA GTATAT-ATA
A.elongatu   --TTATGGAT T-GG-TCGTG ATTAATCGTT -GATATGTCA GTATAT-ATA
A.jaquemon   --TTATGGAT TTGGGTCGTG ATTAATCGTT TGATATGTCA GTATAT-ATA
A.korolkow   --TTATGGAT TTGGGTCGTG ATTAATCGTT TGATATGTCA GTATAT-ATA
A.korokowi   --TTATGGAT TTGGGTCGTG ATTAATCGTT TGATATGTCA GTATAT-ATA
A.pictum     --TTATGGAT T-GG-TCGTG AT-AATCGTT TGATATGTCA GTATAT-ATA
A.byzantin   --TTATGGAT TTGGGTCGTG ATTAATCGTT TGATATGTCA GTATAT-ATA
A.concinna   --TTATGGAT TTGGGTCGTG ATTAATCGTT TGATATGTCA GTATAT-ATA
A.palestin   AATTATGGAT TTGGATCGTG ATTAATCGTT TGATATGTCA GTATAT-ATA
A.rupicola   AATTATGGAT TTGG-TCGTG ATTAATCGTT TGATATGTCA GTATAT-ATA
A.idaeum     --TTATGGAT TTGGGTCTTG ATTAATC-TT TGATATGTCA GTATAT-ATA
A.creticum   --TTATGGAT TTGGGTCGTG ATTAATCGTT TGATATGTCA GTATAT-ATA
A.italicum   --TTATGGAT TTGGGTCGTG ATTAATCGTT TGATATGTCA GTATAT-ATA
A.maculatu   --TTATGGAT TTGG-TCGTG ATTAATCGTT TGATATGTCA GTATATAATA
Biarum_ten   --TTATGGAT TTGGGTCGTG ATTAATCGTY TGATATSTCA GTATAT-ATA
B.t.subsp.   AATTATGGAT TTGGGTCGTG ATTAATCGTT TGATATGTCA GTATAT-ATA

                 ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                         460        470        480        490        500
A.balansan   CGTGCGTATT AGGTATATAG GGTTATCTTT TCTGTTCTGT AATTTAGATA
A.euxinum    CGTGCGTATT AGGTATATAG GGTTATCTTT TCTGTTCTGT AATTTAGATA
A.alpinum    CGTGCGTATT AGGTATATAG GGTTATCTTT TCTGTTCTGT AATTTAGATA
A.hygrophy   CGTGCGTATT AGGTATATAG GGTTATCTTT TCTGTTCTGT AATTTAGATA
A.oriental   CGTGCGTATT AGGTATATAG GGTTATCTTT TCTGTTCTGT AATTTAGATA
A.cyrenaic   CGTGCGTATT AGGTATATAG GGTTATCTTT TCTGTTCTGT AATTTAGATA
A.cyrenaic   CGTGCGTATT A-GTATATA- GGTTATCTTT -CTGTTCTGT AATT-AGATA
A.cyrenaic   CGTGCGTATT AGGTATATAG GGTTATCTTT TCTGTTCTGT AATTTAGATA
```

```
A.dioscori    CGTGCGTATT AGGTATATAG GGTTATCTTT TCTGTTCTGT AATTTAGATA
A.dioscori    CGTGCGTATT A-GTATATAG GGTTATCTTT TCTGTTCTGT AATTTAGATA
A.dioscori    CGTGCGTATT AGGTATATAG GGTTATCTTT TCTGTTCTGT AATTTAGATA
A.elongatu    CGTGCGTATT AGGTATATAG GGTTATCTTT TCTGTTCTGT AATTTAGATA
A.jaquemon    CGTGCGTATT AGGTATATAG GGTTATCTTT TCTGTTCTGT AATTTAGATA
A.korolkow    CGTGCGTATT AGGTATATAG GGTTATCTTT TCTGTTCTGT AATTTAGATA
A.korokowi    CGTGCGTATT AGGTATATAG GGTTATCTTT TCTGTTCTGT AATTTAGATA
A.pictum      CGTGCGTATT AGGTATATAG GGTTATCTTT TCTGTTCTGT AATTTAGATA
A.byzantin    CGTGCGTATT A-GTATATAG GGTTATCTTT TCTGTTCTGT AATTTAGATA
A.concinna    CGTGCGTATT AGGTATATAG GGTTATCTTT TCTGTTCTGT AATTTAGATA
A.palestin    CGTGCGTATT AGGTATATAG GGTTATCTTT TCTGTTCTGT AATTTAGATA
A.rupicola    CGTGCGTATT AGGTATATAG GGTTATCTTT TCTGTTCTGT AATTTAGATA
A.idaeum      CGTGCGTATT AGGTATATAG GGTTATCTTT TCTGTTCTGT AATTTAGATA
A.creticum    CGTGCGTATT AGGTATATAG GGTTATCTTT TCTGTTCTGT AATTTAGATA
A.italicum    CGTGCGTATT AGGTATATAG GGTTATCTTT TCTGTTCTGT AATTTAGATA
A.maculatu    CGTGCGTATT AGGTATATAG GGTTATCTTT TCTGTTCTGT AATTTAGATA
Biarum_ten    CGTGCGTATT AGGTATATAG GGTTATCTTT TCTGTTCTGT AATTTAGATA
B.t.subsp.    CGTGCGTATT AGGTATATAG GGTTATCTTT TCTGTTCTGT AATTTAGATA

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        510        520        530        540        550
A.balansan    GAAAGAAGAA TTCCAGCTAC CAACGCAACG CAATCAACTC CATTTGTTAG
A.euxinum     GAAAGAAGAA TTCCAGCTAC CAACGCAACG CAATCAACTC CATTTGTTAG
A.alpinum     GAAAGAAGAA TTCCAGCTAC CAACGCAACG CAATCAACTC CATTTGTTAG
A.hygrophy    GAAAGAAGAA TTCCAGCTAC CAACGCAACG CAATCAACTC CATTTGTTAG
A.oriental    GAAAGAAGAA TTCCAGCTAC CAACGCAACG CAATCAACTC CATTTGTTAG
A.cyrenaic    GAAAGAAGAA TTCCAGCTAC CAACGCAACG CAATCAACTC CATTTGTTAG
A.cyrenaic    GAAAGAAGAA TTCCAGCTAC CAACGCAACG CAATCAACTC CATTTGTTAG
A.cyrenaic    GAAAGAAGAA TTCCAGCTAC CAACGCAACG CAATCAACTC CATTTGTTAG
A.dioscori    GAAAGAAGAA TTCCAGCTAC CAACGCAACG CAATCAACTC CATTTGTTAG
A.dioscori    GAAAGAAGAA TTCCAGCTAC CAACGCAACG CAATCAACTC CATTTGTTAG
A.elongatu    GAAAGAAGAA TTCCAGCTAC CAACGCAACG CAATCAACTC CATTTGTTAG
A.jaquemon    GAAAGAAGAA TTCCAGCTAC CAACGCAACG CAATCAACTC CATTTGTTAG
A.korolkow    GAAAGAAGAA TTCCAGCTAC CAACGCAACG CAATCAACTC CATTTGTTAG
A.korokowi    GAAAGAAGAA TTCCAGCTAC CAACGCAACG CAATCAACTC CATTTGTTAG
A.pictum      GAAAGAAGAA TTCCAGCTAC CAACGCAACG CAATCAACTC CATTTGTTAG
A.byzantin    GAAAGAAGAA TTCCAGCTAC CAACGCAACG CAATCAACTC CATTTGTTAG
A.concinna    GAAAGAAGAA TTCCAGCTAC CAACGCAACG CAATCAACTC CATTTGTTAG
A.palestin    GAAAGAAGAA TTCCAGCTAC CAACGCAACG CAATCAACTC CATTTGTTAG
A.rupicola    GAAAGAAGAA TTCCAGCTAC CAACGCAACG CAATCAACTC CATTTGTTAG
A.idaeum      GAAAGAAGAA TTCCAGCTAC CAACGCAACG CAATCAACTC CATTTGTTAG
A.creticum    GAAAGAAGAA TTCCAGCTAC CAACGCAACG CAATCAACTC CATTTGTTAG
A.italicum    GAAAGAAGAA TTCCAGCTAC CAACGCAACG CAATCAACTC CATTTGTTAG
A.maculatu    GAAAGAAGAA TTCCAGCTAC CAACGCAACG CAATCAACTC CATTTGTTAG
Biarum_ten    GAAAGAAGAA TTCCAACTAC CAACGCAATG CAATCAACTC CATTTGTTAG
B.t.subsp.    GAAAGAAGAA TTCCAACTAC CAACGCAATG CAATCAACTC CATTTGTTAG

                ....|....| ....|....| ....|....| ....|....| ....|....|
                        560        570        580        590        600
A.balansan    AACAGCTTCC ATTGAGTCTC TGCACCTATC CCTTTTTCTC AAAAAAA--C
A.euxinum     AACAGCTTCC ATTGAGTCTC TGCACCTATC CCTTTTTCTC AAAAAAA--C
A.alpinum     AACAGCTTCC ATTGAGTCTC TGCACCTATC CCTTTTTCTC AAAAAAA--C
A.hygrophy    AACAGCTTCC ATTGAGTCTC TGCACCTATC CTTTTT-CTC AAAAAAA--C
A.oriental    AACAGCTTCC ATTGAGTCTC TGCACCTATC CTTTTTTCTC AAAAAAA--C
A.cyrenaic    AACAGCTTCC ATTGAGTCTC TGCACCTATC CCTTTTTCTC AAAAAAA--C
```

```
A.cyrenaic    AACAGCTTCC ATTGAGTCTC TGCACCTATC CCTTTTTCTC AAAAAAA--C
A.cyrenaic    AACAGCTTCC ATTGAGTCTC TGCACCTATC CCTTTTTCTC AAAAAAA--C
A.dioscori    AACAGCTTCC ATTGAGTCTC TGCACCTATC CTTTTTTCTC AAAAAAA--C
A.dioscori    AACAGCTTCC ATTGAGTCTC TGCACCTATC CTTTTTTCTC AAAAAAA--C
A.dioscori    AACAGCTTCC ATTGAGTCTC TGCACCTATC CTTTTTTCTC AAAAAAA--C
A.elongatu    AACAGCTTCC ATTGAGTCTC TGCACCTATC CTTTTTTCTC AAAAAAA--C
A.jaquemon    AACAGCTTCC ATTGAGTCTC TGCACCTATC CTTTTTTCTC AAAAAAA--C
A.korolkow    AACAGCTTCC ATTGAGTCTC TGCACCTATC CTTTTTTCTC AAAAAAA--C
A.korokowi    AACAGCTTCC ATTGAGTCTC TGCACCTATC CTTTTTTCTC AAAAAAA--C
A.pictum      AACAGCTTCC ATTGAGTCTC TGCACCTATC CTTTTTTCTC AAAAAAA--C
A.byzantin    AACAGCTTCC ATTGAGTCTC TGCACCTATC CTTTTTTCTC AAAAAAA--C
A.concinna    AACAGCTTCC ATTGAGTCTC TGCACCTATC CTTTTTTCTC AAAAAAA--C
A.palestin    AACAGCTTCC ATTGAGTCTC TGCACCTATC CTTTTTTCTC AAAAAAA--C
A.rupicola    AACAGCTTCC ATTGAGTCTC TGCACCTATC CTTTTTTCTC AAAAAAA--C
A.idaeum      AACAGCTTCC ATTGAGTCTC TGCACCTATC CTTTTTTCTC AAAAAAA--C
A.creticum    AACAGCTTCC ATTGAGTCTC TGCACCTATC CTTTTTTCTC AAAAAAA--C
A.italicum    AACAGCTTCC ATTGAGTCTC TGCACCTATC CTTTTTTCTC AAAAAAA--C
A.maculatu    AACAGCTTCC ATTGAGTCTC TGCACCTATC CCTTTTTCTC AAAAAAA--C
Biarum_ten    AACAGCTTCC ATTGAGTCTC TGCACCTATC CTTTTTTCTC AAAAAAA--C
B.t.subsp.    AACAGCTTCC ATTGAGTCTC TGCACCTATC CTTTTTTCTC AAAAAAA--C

                 ....|....| ....|....| ....|....| .
                         610        620        630
A.balansan    AAGGATTTGG CTCAGGATTG CCCTTTTTTT A
A.euxinum     AAGGATTTGG CTCAGGATTG CCCTTTTTTT A
A.alpinum     AAGGATTTGG CTCAGGATTG CCCTTTTTTT A
A.hygrophy    AAGGATTTGG CTCAGGATTG CCCTTTTTTT A
A.oriental    AAGGATTTGG CTCAGGATTG CCCTTTTTTT A
A.cyrenaic    -AGGATTTGG CTCAGGATTG CCCTTTTTTT A
A.cyrenaic    AAGGATTGGG CTCAGGA-TG CCCTTTTTTT A
A.cyrenaic    -AGGATTTGG CTCAGGATTG CCCTTTTTTT A
A.dioscori    MAGGATTTGG CTCAGGATTG CCCTTTTTTt A
A.dioscori    -AGGATTTGG CTCAGGATTG CCCTTTTTTT A
A.dioscori    AAGGATTTGG CTCAGGATTG CCCTTTTTTT A
A.elongatu    AAGGATT--G CTCAGGA-TG CCC-TTTTTT A
A.jaquemon    -AGGATTTGG CTCAGGATTG CCCTTTTTTT A
A.korolkow    AAGGATTTGG CTCAGGATTG CCCTTTTTTT A
A.korokowi    AAGGATTTGG CTCAGGATTG CCCTTTTTTT A
A.pictum      AAGGATTTGG CTCAGGATTG CCCTTTTTTT A
A.byzantin    AAGGATTTGG CTCAGGATTG CCCTTTTTTT A
A.concinna    AAGGATTTGG CTCAGGATTG CCCTTTTTTT A
A.palestin    AAGGATTTGG CTCAGGATTG CCCTTTTTTT A
A.rupicola    -AGGATTTGG CTCAGGATTG CCCTTTTTTT A
A.idaeum      -AGGATTTGG CTCAGGATTG CCCTTTTTTT A
A.creticum    -AGGATTTGG CTCAGGA-TG CCC-TTTTTT A
A.italicum    -AGGATTTGG CTCAGGATTG CCCTTTTTTT A
A.maculatu    AAGGATTTGG CTCAGGATTG CCC-TTTTTT A
Biarum_ten    AAGGATTTGG CTCAGGATTG CCCATTTTTT A
B.t.subsp.    AAGGATTTGG CTCAGGATTG CCCATTTTTT A
```

Alignment of ITS1 sequences

```
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                         10         20         30         40         50
A.maculatu      CGCACT-CGC GAACGGTTGA CCCTATCAAA CCCCGGAGGG GGCATGGGTG
A.italcium      CGCACT-CGC GAACGGTTGA -CCTATCAAA CCC-GGAGGG GGCATGGGTG
A.italicum      CGCACT-CGC GAACGGTTGA TCCTATCAAA CCCCGGAGGG GGCATGGGTG
A.oriental      CGCACT-CGC GAACGGTTGA CCCTATCAAA CCCCGGAGGG GGCATGGGTG
A.lucanum       CGCACT-CGC GAACGGTTGA -CCTATCAAA CCC-GGAGGG GGCATGGGTG
A.cyrenaci      CGCACT-CGC GAAYGGTTGA CCCTATCAAA CCCCGGAGGG GGCATGGGTG
A.cyrenaci      CGCACT-CGC GAAYGGTTGA -CCTATCAA- CCCCGGAGGG GGCATGGGTG
A.balansan      CGCACT-CGC GAACGGTTGA CCCTATCAAA CCCCGGAGGG GGCATGGGTG
A.elongatu      CGCACT-CGC GAACGGTTGA CCCTATCAAA CCCCGGAGGG GGCATGGGTG
A.jacquemo      CGCACT-CGC GAACGGTTGA -CCTATCAAA CCC-GGAGGG GGCATGGGTG
A.korolkow      CGCACT-CGC GAACGGTTGA CCCTATCAAA CCCCGGAGGG GGCATGGGTG
A.krorolow      CGCACT-CGC GAACGGTTGA CCCTATCAAA CCCCGGAGGG GGCATGGGTG
A.euxinum       CGCACT-CGC GAATGGTTGA CCCTATCAAA CCCCGGAGGG GGCATGGGTG
A.hygrophi      CGCACT-CGC GAACGGTTGA CCCTATCAAA CCCCGGAGGG GGCATGGGTG
A.dioscori      CGCACT-CGC GAACGGTTGA CCCTATCAAA CCCCGGAGGG GGCATGGGTG
A.dioscori      CGCACT-CGC GAACGGTTGA CCCTATCAAA CCCCGGAGGG GGCATGGGTG
A.palaesti      CGCACT-CGC GAACGGTTGA CCCTATCAAA CCCCGGAGGG GGCATGGGTG
A.creticum      CGCACT-CGC GAACGGTTGA -CCTATCAA- CCCCGGAGGG GGCATGGGTG
A.idaeum        CGCACT-CGC GAACGGTTGA -CCTATCAAA CCTCGGAGGG GGCATGGGTG
A.nigrum        CGCACT-CGC GAACGGTTGA -CCTATCAAA CCTCGGAGGG GGCATGGGTG
A.pictum        CGCACTTCGC GAACGGTTGA -CCTATCAAA CCC-GGAGGG GGCATGGGTG
Biarum1         CGCACT-CGC GAACGGTTGA -CCTACCCA- CCT-GGAGGG GGCAGGGGTG
Biarum2         CGCACT-CGC GAACGGTTGA CCCTACCCA- CCT-GGAGGG GGCAGGGGTG

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                         60         70         80         90         100
A.maculatu      CGGCTTTGGA CGTGTAAATG TTTGCCTTCC CCTGCTATCC CC-RTCCCTC
A.italcium      CGGCTCTGGA CGTGTAAATG TTTGCCTTCC CTTGCTATCC CC-ATCCCTC
A.italicum      CGG-TCTGGA CGTGTAAATG TTTGCCTTCC CCTGCTATCC CC-ATCCCTC
A.oriental      CGGCTTTGGA CGTGTAAATG TTTGCCTTCC CCTGCTATCC CC-ATCCCTC
A.lucanum       CGGCTTTGGA CGTGTAAATG TTTGCCTTCC CCTGCTATCC CC-ATCCCTC
A.cyrenaci      CGGCTTTGGA CGTGTAAATG TTTGCCTTCC CCTGCTATCC CC-ATCCCTC
A.cyrenaci      CGGCTTTGGA CGTGTAAATG TTTGCCTTCC CCTGCTATCC CC-ATCCCTC
A.balansan      CGGCTTTGGA CGTGTAAATG TTTGCCTTCC CCTGCTATCC CC-ATCCCTC
A.elongatu      CGGCTCTGGA CGTGTAAATG TTTGCCTTCC CCTGCTATCC CC-ATCCCTC
A.jacquemo      CGGCTTTGGA CGTGTAAATG TTTGC-TTCC CCTGCTATCC CC-ATCCCTC
A.korolkow      CGGCTTTGGA CGTGTAAATG TTTGCCTTCC CCTGCTATCC CC-ATCCCTC
A.krorolow      CGGCTTTGGA CGTGTAAATG TTTGCCTTCC CCTGCTATCC CC-ATCCCTC
A.euxinum       CGGCTCTGGA CGTGTAAATG TTTGCCTTCC CCTGCTATCC CC-ATCCCTC
A.hygrophi      CGGCTCTGGA CGTGTAAATG TTTGCCTTCC CCTGCTATCC CC-ATCCCTC
A.dioscori      CGGCTCTGGA CGTGTAAATG TTTGCCTTCC CCTGCTATCC CC-ATCCCTC
A.dioscori      CGGCTCTGGA CGTGTAAATG TTTGCCTTCC CCTGCTATCC CC-ATCCCTC
A.palaesti      CGGCTCTGGA CGTGTAAATG TTTGCCTTCC CCTGCTATCC CC-ATCCCTC
A.creticum      CGGCTCTGGA CGTGTAAATG TTTGCCTTCC CCTGCTATCC CC-ATCCCTC
A.idaeum        CGGCTCTGGA CGTGTAAATA TTTGCCTTCC CCTGCTATCC CC-ATCCCTC
A.nigrum        CGGCTCTGGA CGTGTAAATR TTTGCCTTCC CCTGCTATCC CC-ATCCCTC
A.pictum        CGGCTCTGGA CGTGTAAATG TTTGCCTTCC CCTGCTATCC CC-ATCCCTC
Biarum1         TGGCTGAGGA CGTGCAAATG CTCGTCCTCC CCTCCTATCC CCCGTCCCTC
```

```
Biarum2      TGGCTGAGGA CGTGCAAATG CTCGTCCTCC CCTCCTATCC CCCGTCCCTC

                 ....|....| ....|....| ....|....| ....|....| ....|....|
                         110        120        130        140        150
A.maculatu   TCCGGGATAT ATTTCCCTGC CTCCGAYAGC TTCGATCGGT GGGAG----C
A.italcium   TCCGGGATAT ATTTCCCTGC CTCCGACAGC TCCGATCGGT GGGAG----C
A.italicum   TCCGGGATAT ATTTCCCTG- CTCCGACAGC TCCGATCGGT GGGAG----C
A.oriental   TCCGGGATAT ATTTCCCTG- --CCGACAGC TTCGATCGGT GGGAG-----
A.lucanum    TCTGGGATAT ATTTCCCTGC CTCCGATAGC TTCGATCGGT GGGAG----C
A.cyrenaci   TCTGGGATAT ATTTCCCTGC CTCCGATAGC TTCGATCGGT GGGAG----C
A.cyrenaci   TCTGGGATAT ATTTCCCTGC CTCCGATAGC TTCGATCGGT GGGAG----C
A.balansan   TCTGGGATAT ATTTCCCTGC CTCCGATAGC TTCGATCGGT GGGAG----C
A.elongatu   TCCGGGATAT ATTTCCATGC CTTTGACAGC TTCGATCGGT GGGAG----C
A.jacquemo   TCCGGGATAT ATTTCCCTGC CTCCGACAGC TTCGATCGGT GGGAG----T
A.korolkow   TCCGGGATAT ATTTCCCTGC CTCCGACAGC TTCGATCGGT GGGAG----T
A.krorolow   TCCGGGATAT ATTTCCCTGC CTCCGACAGC TTCGATCGGT GGGAG----T
A.euxinum    TCCGGGATAT ATTTCCCTGC CTCCGATAGC TTCGATCGGT GGGAG----C
A.hygrophi   TCCGGGATAT ATTTCCCTGT GTCCGACAGC TCTGATCGGT GGGAG----C
A.dioscori   TCCGGGATAT ATTTCCCTGC CTCCGACAGC TTCGATCGGT GGGAG----C
A.dioscori   TCCGGGATAT ATTTCCCTGT CTCCGACAGC TCCGATCGGT GGGAG----C
A.palaesti   TCCGGGATAT ATTTCCCTGT CTCCGACAGC TCCGATCGGT GGGAG----C
A.creticum   TCCGGGATAT ATTTCCCTGC CTCCGATAGC TTCGATCGGT GGGAG----C
A.idaeum     TCCGGGATAT ATTTCCCTGC CTCCGATAGC TTCGATCTGT GGGAG----C
A.nigrum     TCCGGGATAT ATTTCCCTGC CTCCGATAGC TTCGATCRGT GGGAG----C
A.pictum     TCCGGGATAT ATTTCCCTGC CTCTGAYATC TCCGATCGGT GGGAG----C
Biarum1      TCCGGGATAT TTTTCCCTGC CTCCGACACG TTCGAGCGGT GAGAGAGAGC
Biarum2      TCCGGGATAT TTTTCCCTGC CTCCGACACG TTCGAGCGGT GAGAGAGAGC

                 ....|....| ....|....| ....|....| ....|....| ....|....|
                         160        170        180        190        200
A.maculatu   CTCCCATCGA TCGGGCTGTC GGCCGGCGGG ACGACGAACC TTCCCGGCGC
A.italcium   CTCCCATCGA TCGGGCCGTC GGCCGGCGGG ACGACGAACC TTCCCGGCGC
A.italicum   CTCCCATCGA TCGGGCCGTC GGCCGGCGGG ACGACGAACC TTCCCGGCGC
A.oriental   CTCC-ATCGA TCGGGCTGTC GGCCGGCGGG ACGACGAACC TTCCCGGCGC
A.lucanum    CTCCCATCGA TCGGGCTGTC GGC-GGCGGG ACGACGAACC TTCCCGGCGC
A.cyrenaci   CTCCCATCGA TCRGGCTGTC GGCCGGCGGG ACGACGAACC TTCCCGGCGC
A.cyrenaci   CTCCCATCGA TCAGGCTGTC GGCCGGCGGG ACGACGAACC TTCCCGGCGC
A.balansan   CTCCCATCGA TCGGGCTGTC GGCCGGCGGG ACGACGAACC TTCCCGGCGC
A.elongatu   CTCCCATCGA TCGGGCTGTC GGCCGGCGGG ACGACGAACC TTCCCGGCGC
A.jacquemo   CTCC-ATCGA TCGGGCTGTC GGCCGGCGGG ACGACGAACC TTCCCGGCGC
A.korolkow   CTCCCATCGA TCGGGCTGTC GGCCGGCGGG ACGACGAACC TTCCCGGCGC
A.krorolow   CTCCCATCGA TCGGGCTGTC GGCCGGCGGG ACGACGAACC TTCCCGGCGC
A.euxinum    CTCCCATCGA TCGGGCTGTC GGCCGGCGGG ACGACGAACC TTCCCGGCGC
A.hygrophi   CTCCCATCGA TCGGGTTGTC GGCCGGCGGG ACGACGAACC TTCCCGGCGC
A.dioscori   CTCCCATCGA TCGGGCTGTC GGCCGGCGGG ACGACGAACC TTCCCGGCGC
A.dioscori   CTCCCATCGA TCGGGCTGTC GGCCGGCGGG ACGACGAACC TTCCCGGCGC
A.palaesti   CTCCCATCGA TCGGGCGGTC GGGCGGCGGG ACGACGAACC TTCCCGGCGC
A.creticum   CTCCCATCGA TCGGGCTGT- GGGCGGGGGG ACGACGAAGC TTCCCGGCGC
A.idaeum     CTCCCATCGA TCGGGCTGTC GGGCGGCGGG ACGACGAAGC TTCCCGGCGC
A.nigrum     CTCCCATCGA TCGGGCRGTC GG-CGGGGGG ACGAGGAAGC TTCC-GGGGC
A.pictum     CTCCCATCGA TCGGGCGGTC GG-CGGGGG- ATGACGAAGC TTCCCGGCGC
Biarum1      CTCTTGTGGA TCGGGCCGTC GGGCGGGGGG ATGATGAAGC TTCACGGGT-
Biarum2      CTCTTGTCGA TCGGGGCGTC GGGCGGGGGG ATGATGAAGC TTCACGGGT-
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                        210        220        230        240        250
A.maculatu      GGTATGCGCC AAGGAAAACG GACGTGAAGG CCTGCGTGAT CAGTCCC---
A.italcium      GGTATGCGCC AAGGAAAACG GACGTGAAGG CCCGCGTGAT CAGTCCCC--
A.italicum      GGTATGCGCC AAGGAAAACG GACGTGAAGG CCCGCGTGAT CAGTCCCC--
A.oriental      GGTATGCGCC AAGGAAAACG GACGTGAAGG CCTGCGTGAT CAGTCCC---
A.lucanum       GGTATGCGCC AAGGAAAACG GACGTGAAGG CCTGCGTGAT CAGTCCC---
A.cyrenaci      GGTATGCGCC AAGGAAAATG GACGTGAAGG CCTGCGTGAT CAGTCCC---
A.cyrenaci      GGTATGCGCC AA-GAAAATG GACGTGAAGG GCTGCGTGAT CAGTCCC---
A.balansan      GGTATGCGCC AAGGAAAACG GACGTGAAGG CCTGCGTGAT CAGTCCC---
A.elongatu      GGTATGCGCC AAGGAAAATG GACGTGAAGG CCTGCGTGAT CAGTCCA---
A.jacquemo      GGTATGCGCC A--GAAA-CG GACGTGAAGG -CTGCGTGAT CAGTCCC---
A.korolkow      GGTATGCGCC AAGGAAAACG GACGTGAAGG CCTGCGTGAT CAGTCCC---
A.krorolow      GGTATGCGCC AAGGAAAACG GACGTGAAGG CCTGCGTGAT CAGTCCC---
A.euxinum       GGTATGCGCC AAGGAAAATG GACGTGAAGG CCTGCGTGAT CAGTCCC---
A.hygrophi      GGTATGCGCC AAGGAAAACG GACGTGAAGG CCTGCGTGAT CAGTCCT---
A.dioscori      GGTATGCGCC AAGGAAAACG GACGTGAAGG CCTGCGTGAT CAGTCCC---
A.dioscori      GGTATGCGCC AAGGAAAACG GACGTGAAGG CCCGCGTGAT CAGTCCC---
A.palaesti      GGTATGCGCC AAGGAAAACG GACGTGAAGG CCCGCGTGAT CAGTCCC---
A.creticum      GGTATGCGCC AAGGAAAACG GACGTGAAGG -CTGCGTGAT CAGTCCC---
A.idaeum        GGTATGCGCC AAGGAAAACG GACGTGAAGG -CTGCGTGAT CAGTCCC---
A.nigrum        GGTATGCGCG AAGGAAAACG GACGTGAAGG CCTGCGTGAT CAGTCCC---
A.pictum        GGTATGCGCC AAGGAAAACG GACGTGAAGG CCTGCGTGAT CAGTCCC---
Biarum1         GGTATGCGCC AAGGAAAAGA GAAGTGAAGG CC-GTGTGAT CCGATCCAGC
Biarum2         GGTATGCGCC AA-GAAAAGA GAAGTGAAGG CC-GTGTGAT CCGATCCAGC


                ....|....| ....|....| ....|....| ....|....| ....|....
                        260        270        280        290
A.maculatu      --TGTGGCGG AGGCACGCGG CG-CTCATCA CCGATACAAA ACGAGTCTG
A.italcium      --TGT-GC-G AGGCACGCGG CG-CTCATCA CCGATACAAA ATGAGTMTG
A.italicum      --TGTGGCGG AGGCACGCGG TG-CTCATCA CCGATACAAA ACGAGTCTG
A.oriental      --TGTGGCGG AGGCACGCGG CG-CTCATCA CCGATACAAA ACGAGTCTG
A.lucanum       --TGTGGCGG AGGCACGCGG CG-CTCATCA CCGATACAAA ACGAGTCTG
A.cyrenaci      --TGTGGCGG AGGCACGCGG CGGCTCATCA CCGATACAAA ATGAGTCTG
A.cyrenaci      --TGTGGCGG AGGCACGCGG CG-CTCATCA C-GATACAAA -TGAGTCTG
A.balansan      --TGTGGCGG AGGCACGCGG CG-CTCATCA CCGATACAAA ACGAGTCTG
A.elongatu      --TGTGGCGG AGGCACGCGG CG-CTCATCA CCGATACAAA ATGAGTCTG
A.jacquemo      --TGTGGGCG AG-CACGCGG CG-CTCATCA C-GATACAAA ACGAGTCTG
A.korolkow      --TGTGGCGG AGGCACGCGG CG-CTCATCA CCGATACAAA ACGAGTCTG
A.krorolow      --TGTGGCGG AGGCACGCGG CG-CTCATCA CCGATACAAA ACGAGTCTG
A.euxinum       --TGTGGCGG AGGCACGCGG CG-CTCATCA CCGATACAAA ATGAGTCTG
A.hygrophi      --TGTGGCGG AGGCACGCGG CG-CTCATCA CCGATATAAA ACGAGTCTG
A.dioscori      --TGTGGCGG AGGCACGCGG CG-CTCATCA CCGATACAAA ACGAGTCTG
A.dioscori      --TGTGGCAG AGGCACKCGG CG-CTCATCA CCGATACAAA ACGAGTCTG
A.palaesti      --TGTGGCGG AGGCACGCGG CG-CTCATCA CCGATACAAA ACGAGTCTG
A.creticum      --TGTGGCGG AGGCACGCGG TG-CTCATCA CCGATACAAA ACGAGTCTG
A.idaeum        --TGTGGCGG AGGCACGCGG TG-CTCATCA CCGATACAAA ACGAGTCTG
A.nigrum        --TGTGGCGG AGGCACGCGG TG-CTCATCA GCGATAGAAA ACGAGTCTG
A.pictum        --TGTGACGG AGGCACGCGG GC-GTCATCA CCGATACAAA ACGAGTCTG
Biarum1         CCTGTGGCGG AGGCACGCGG CG-CTCATCA CCGATACGAA ACGAGTCTG
Biarum2         CCTGTGGCGG AGGCACGCGG CG-CTCATCA CCGATACGAA ACGAGTCTG
```