Phylogeography and comparative ecophysiology of *Chrysanthemoides* Turn. Ex Medik. (Tribe Calenduleae)

By
Seranne Howis

Submitted in Fulfilment for the
Degree of Masters of Science

Rhodes University
Botany Department
February 2004

Abstract

Chrysanthemoides is a common Southern African shrub that grows in a variety of habitats, from coastal shrubland and fynbos to mountainous areas as far north as Kenya. The genus has two species and 8 subspecies, the diagnoses and delimitation of which have been based almost exclusively on morphological characteristics. This project aims to investigate, with the use of phylogenetic species concepts, the validity of these subspecies.

Unlike biological species concepts that rely on reproductive isolation as a defining character of a species, phylogenetic species concepts (PSC) are concerned with delimiting evolutionary significant units (ESUs). ESUs are evolutionarily isolated lineages, and under the PSC a species is an aggregation of organisms consistently diagnosable by a fixed character or combination of characters. This project therefore searched for genetic and physiological characters by which to delimit ESUs within the *Chrysanthemoides* genus. DNA sequencing was used to investigate the genetic characters, while gas exchange studies were used to investigate the ecophysiological characters.

DNA sequence analysis indicated that the ESUs can be diagnosed by genetic means and that one species may be of hybrid origin. Field studies of three disparate genetically identifiable ESUs from three disparate climates found that there are noticeable differences in ecophysiological responses of these ESUs in the field. Plants from each ESU were transferred to a greenhouse and grown under identical conditions for several months and compared to determine if these traits are inherent, or elastic in relation to environmental conditions. Under simulated high rainfall conditions, there does not appear to be a significant difference in the photosynthetic traits.

Acknowledgments

Many thanks to the following people:

- Nigel Barker and Brad Ripley for endless enthusiasm, and excellent supervision and thesis critiquing.
- Syd Ramdhani, Craig Peter, Alan Wood, Tony Dold, Pete Phillipson and all those who collected samples for me in their travels.
- Monica Mwale and Syd for advice and counsel.
- Amy Skinner and Paul Barratt for help, support and taxi services, above and beyond.
- A special thank you to Taryn Skikne, the best friend a person could have, for all the bad movies, the good food, the great jokes, and for being Sam to my Frodo. See, I finally made it to Mount Doom. Go me!

This research was supported by the National Research Foundation.

Contents

Chapter 1: Introduction

Chapter 2: Phylogeography of *Chrysanthemoides*

Introduction

Methods

Results

Analyses of the "Complete ITS" data set

Analyses of the "Monilifera ITS" data set

Analyses of the "Limited ITS" data set

Analyses of the "psbA-trnH" data set

Analyses of the "trnL-trnF" data set

Analyses of the combined cpDNA data set

Discussion

Chapter 3: Ecophysiology of selected *Chrysanthemoides* ESUs

Introduction

Methods

Results

Discussion

Appendices

Appendix 3: Final Sequence Alignments

Chapter 1.
Introduction

The central aim of this research was to investigate the use of DNA sequence data for the elucidation of, and subsequent ecophysiological examination of, discrete taxonomic entities within the genus *Chrysanthemoides*, focusing on the species *C. monilifera* Tourn. ex Medik.

Chrysanthemoides as a study model

Chrysanthemoides Tourn. ex Medik. (Compositae) is a genus of southern African shrubs that occur in coastal shrub and fynbos as well as in the mountains of southern and eastern Africa at low to mid altitudes (Griffioen, 1995). Norlindh (1977) mentions that the distributions in tropical Eastern Africa are disjunctive, occurring only in the highland regions. The genus belongs to the tribe Calenduleae and is closely associated with *Osteospermum*. Until a review in 1943 by Norlindh, all the species were placed within *Osteospermum* genus, but the presence of drupes in *Chrysanthemoides* was used as a definitive characteristic to separate the two genera. The evolution of drupes is extremely rare in the Compositae and as far as is known, no other plants of this family in the Old World have drupes (Norlindh, 1977).

The current accepted taxonomy of the *Chrysanthemoides* genus has two species: *Chrysanthemoides monilifera* (L.) T. Norl. and *C. incana* (Burm. f.) T. Norl.

C. monilifera is split into 6 subspecies:

- *C. m. canescens* (DC.) T. Norl.,
- *C. m. monilifera* (L.) T. Norl.,
- *C. m. pisifera* (L.) T. Norl.,
- *C. m. rotundata* (DC.) T. Norl.,
- *C. m. septentrionalis* T. Norl., and
- *C. m. subcanescens* (DC.) T. Norl.

Two subspecies of *C. monilifera*: *C. monilifera rotundata* ("bitou bush") and *C. monilifera monilifera* ("boneseed"); have become naturalized in Australia where they have become invasive and are considered important noxious weeds along their coastline (Scott, 1996).

The unpublished work of Griffioen (1995) examines the variation of *Chrysanthemoides* using both morphology and allozymes. He used the Biological Species Concept to define species level entities (p. 84), and then applied the Morphological Species Concept to subdivide species into subspecies (in combination with their geographical distributions; p. 87). Griffioen (1995) maintained the two species C. *monilifera* and C. *incana*, but he moved the *C. monilifera subcanescens* subspecies to *C. incana subcanescens*, and recognized a new subspecies: *C. monilifera floribunda* R.C. Griffioen (which he further divided into two forms, based on ecological, electrophoretic and morphological data).

See Table 1.1. for the taxonomic history of the *Chrysanthemoides* entities recognized by Griffioen (1995) and their previous designations. Figure 1.1 shows the phenogram for *Chrysanthemoides* derived from morphological data, as presented in Griffioen's thesis (Figure 1, pp 23); Figure 1.2 shows the phylogram based on a parsimony analysis of the characters used in Griffioen's phylogenetic analysis that are presented in Appendix 3 of Griffioen's thesis, reanalyzed for the current investigation. In the first tree, taken directly from Griffioen's thesis, *C. incana* is basal to the *C. monilifera* clade. In the second tree, based on a reanalysis of raw data presented in Griffioen's thesis, *C. incana* is derived from within the *C. monilifera* clade.

Each of the subspecies of *Chrysanthemoides* have different preferred habitats and ranges, for example: *C. m. rotundata* occurring along the coastlines on the eastern side of the coast of South Africa (up to 150m above sea level), whilst *C. m. canescens* occurs 1000- 2500m above sea level. They differ quite strongly in their leaf morphology as well: the former subspecies has large thick obovate leaves with margins either entire or scalloped, whilst the later subspecies has smaller thin elliptical leaves with toothed margins.

As can be seen from the maps (Figures 1.3 to 1.10 below), there are several different distribution patterns amongst the subspecies: *C. monilifera septentrionalis* stretches far north up Africa, broken in several places by the disjunction in its high altitude forest habitat; whereas several other subspecies and varieties overlap and intermingle in the south east regions of the Western Cape. The former distribution pattern provides a sufficiently diverged subspecies for comparison purposes in terms of level of genetic divergence of the more closely associated subspecies. The latter pattern allows for an investigation into the possible levels of hybridization and introgression between subspecies, as well as a study of the biogeographical patterns of population structure and cohesion within the subspecies.

Table 1.1: History of taxa recognized by Griffioen (1995) in the top row, and their previous designations in the columns below that. "*O.*" is an abbreviation for *Osteospermum*. (Adapted from Griffioen, 1995).

Author (Date)	*C. monilifera monilifera*	*C. monilifera floribunda*	*C. monilifera pisifera*	*C. monilifera canescens*	*C. monilifera septentrionalis*	*C. monilifera rotundata*	*C. incana incana*	*C. incana rangei*	*C. incana subcanescens*
Linnaeus (1753)	*O.moniliferum*		*O. pisiferum*						
Bergius (1676)	*O.moniliferum*		*O. pisiferum* *O. cilitum*						
Burman (1768)							*O. incanum*		
Jacquin (1798)							*O. spinosum*		
Cassini (1818)							*Eriocline obovata*		
Thunberg (1836)	*O.moniliferum*		*O. piliferum*						
De Candolle (1836)	*O.moniliferum*		*O. pisiferum* *O. cilitum*	*O. pisiferum* var *canescens*		*O. moniliferum* var. *rotundatum* *O. macrocarpum*	*O. moniliferum* var. *lanosum* *O. spinescens*		*O. subcanescens* var. *virescens* & var. *angustifolium*
Drege (1843)				*O. pisiferum* var *canescens*			*O. moniliferum* var. *lanosum* *O. spinescens*		
Harvey & Sonder (1836)	*O.moniliferum* var. *verum*			*O.moniliferum* var. *pisiferum*		*O.moniliferum* var. *rotundatum*	*O. moniliferum* var. *lanosum* *O. spinescens*		
Engler (1892)					*O.moniliferum*				
Muschler (1910)								*O. rangei*	
Compton & Pillans (1931)							*O. lanosum*		
Norlindh (1943)	C. monilifera	C. monilifera	C. monilifera ssp. pisifera	C. monilifera ssp. canescens	C. monilifera ssp. septentrionalis	C. monilifera ssp. rotundata	C. incana	C. incana	C. monilifera ssp. subcanescens

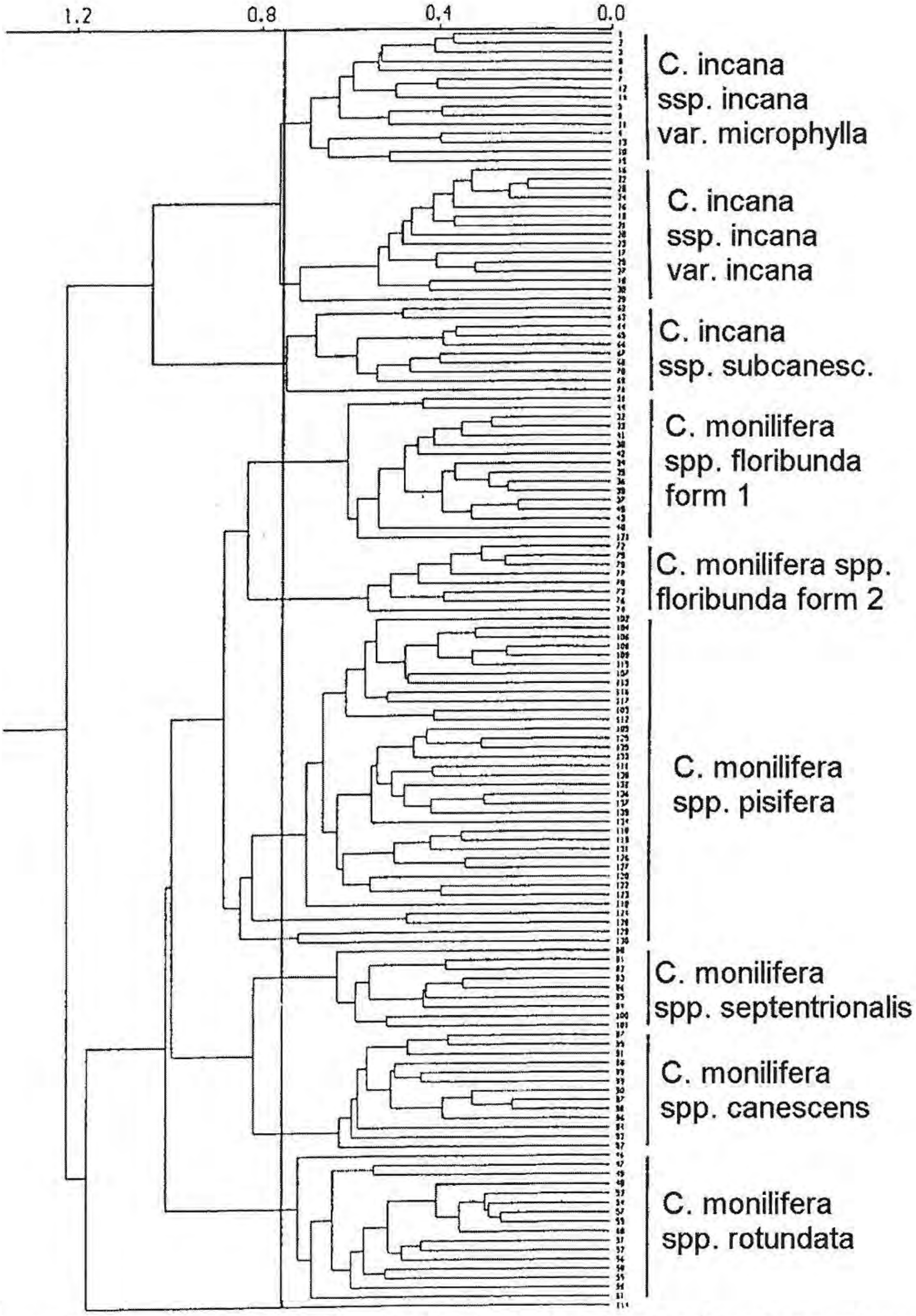

Figure 1.1: Phenogram for *Chrysanthemoides* (showing the 0.75 phenon line), based on morphological data. (Adapted from Griffioen, 1995)

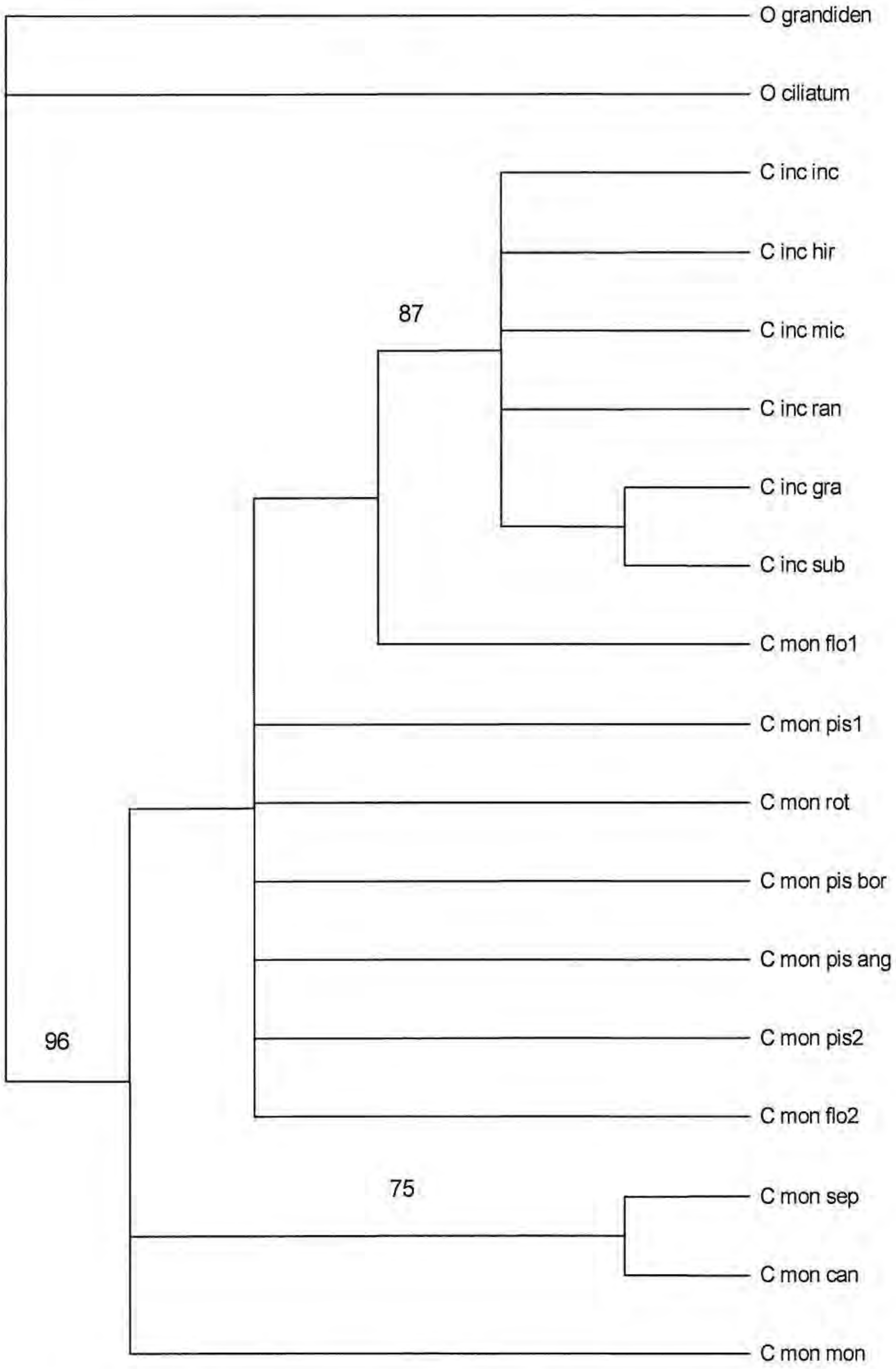

Figure 1.2: Morphological data recoded from Griffioen (1995): Strict Consensus tree of 32 Most Parsimonious trees. (CI=0.491; RI=0.695 Length = 57; Bootstrap support above branches)

Abbreviations: O grandiden = Osteospermum grandiden; O ciliatum = Osteospermum ciliatum; C inc inc = Chrysanthemoides incana spp incana var incana; C inc hir = C. i. spp incana var hirsuta; C inc mic = C. i. spp incana var microphylla; C inc ran = C. i. spp incana var rangei; C inc. gra. = C. i. spp incana var gracilis; C inc sub = C. i. spp subcanescens; C mon flo1 = C. monilifera spp floribunda form 1; C mon pis1 = C. m. spp pisifera pisifera form 1; C mon rot. = C. m. spp rotundata.; C mon pis bor = C. m. spp pisifera var borealis; C mon pis ang = C. m. spp pisifera var angustifolia; C mon pis2 = C. m. spp pisifera pisifera form 2; C mon flo2 = C. monilifera spp floribunda form 2; C mon sep = C. m. spp septentrionalis; C mon can = C. m. spp canescens; C mon mon = C. m. spp monilifera.

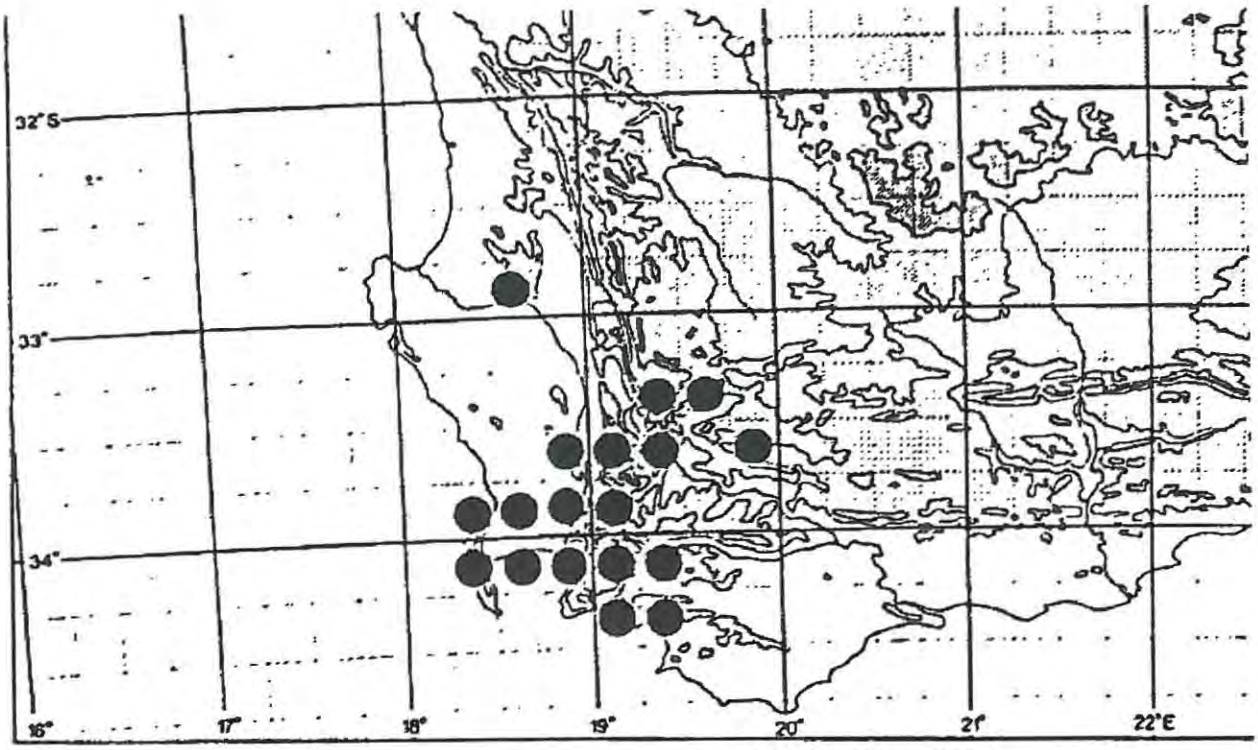

Figure 1.3: Distribution map of *Chrysanthemoides monilifera ssp. monilifera* (Griffioen, 1995)

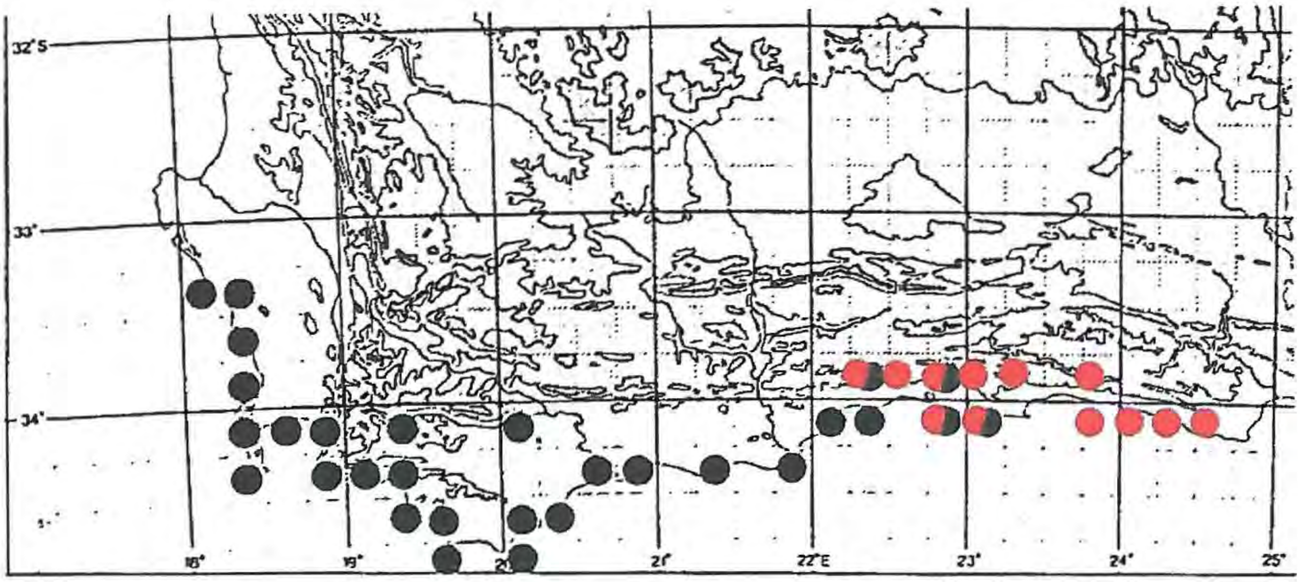

Figure 1.4: Distribution map of *Chrysanthemoides monilifera* ssp. *floribunda* [form 1 = black; form 1 = red] (Griffioen, 1995)

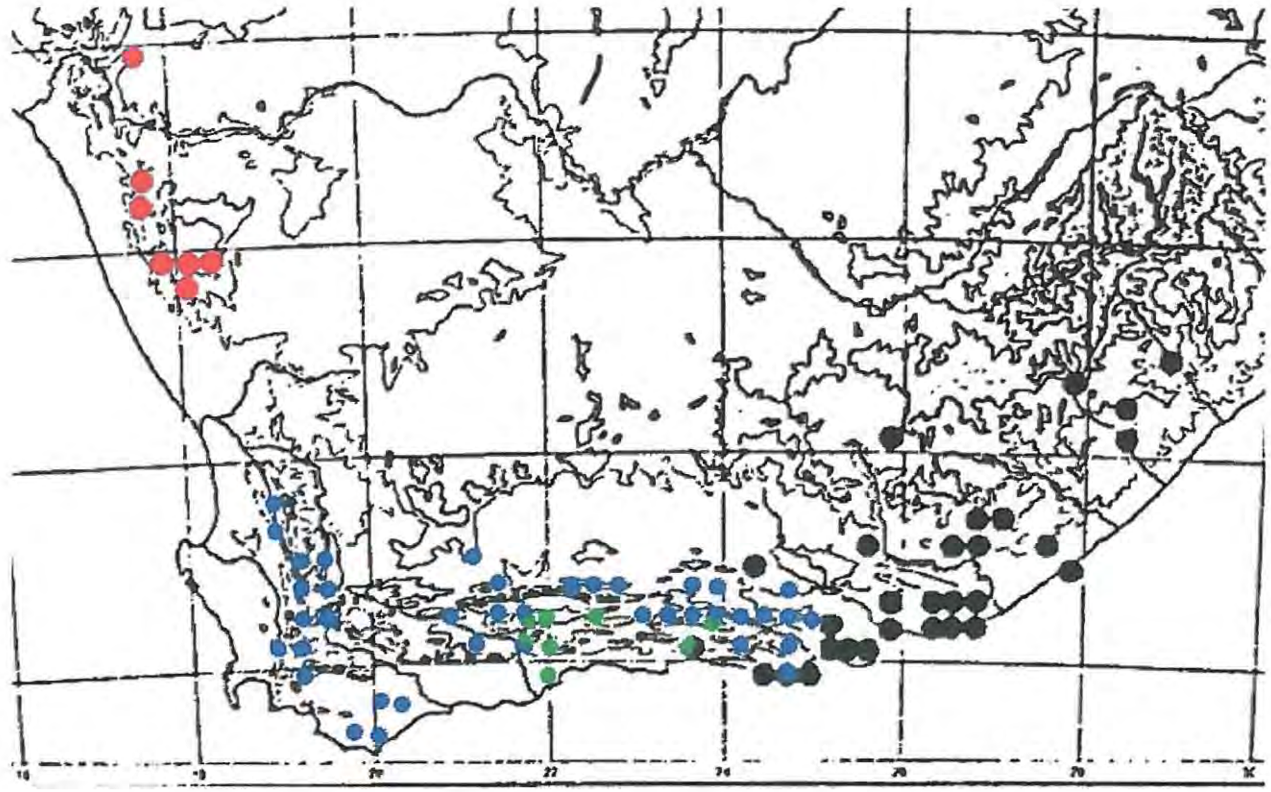

Figure 1.5: Distribution map of *Chrysanthemoides monilifera ssp. pisifera* [black circle = form 1; blue = form 2; red = *borealis*; green = *angustifolia*] (Griffioen, 1995)

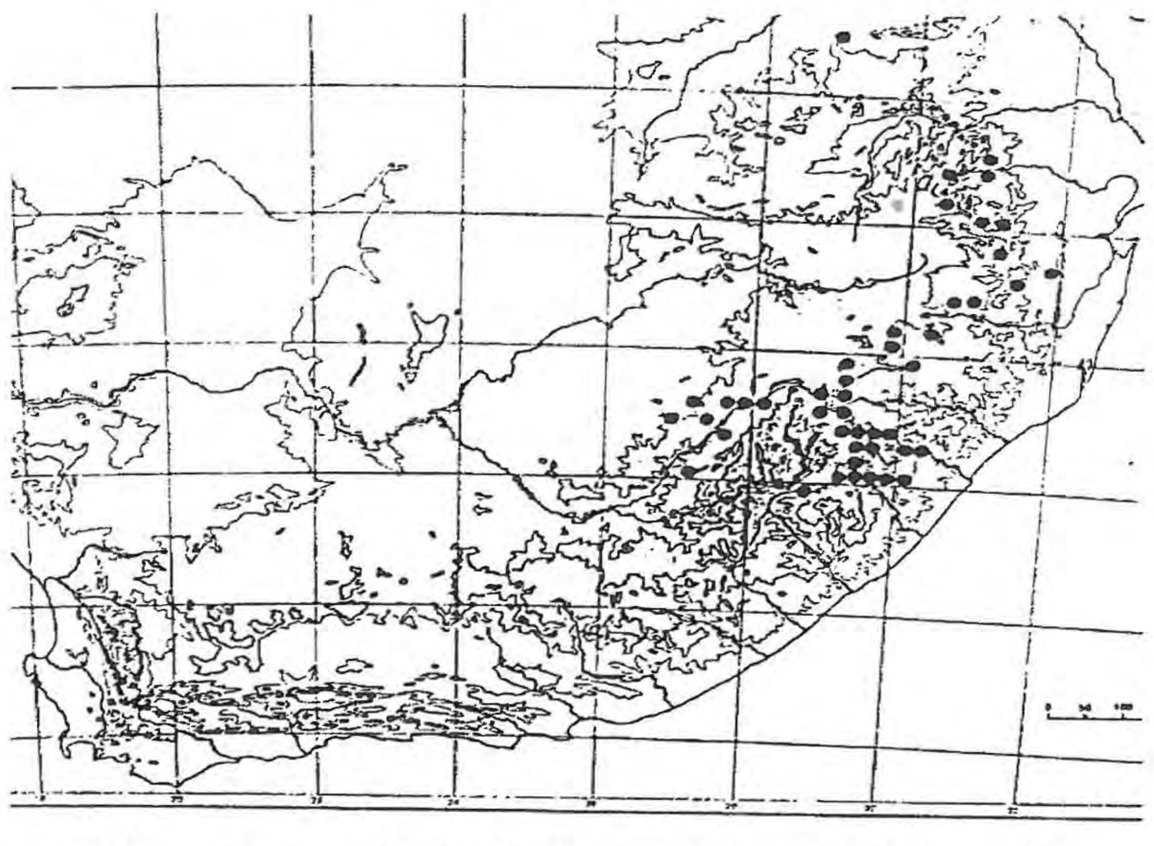

Figure 1.6: Distribution map of *Chrysanthemoides monilifera ssp. canescens* (Griffioen, 1995)

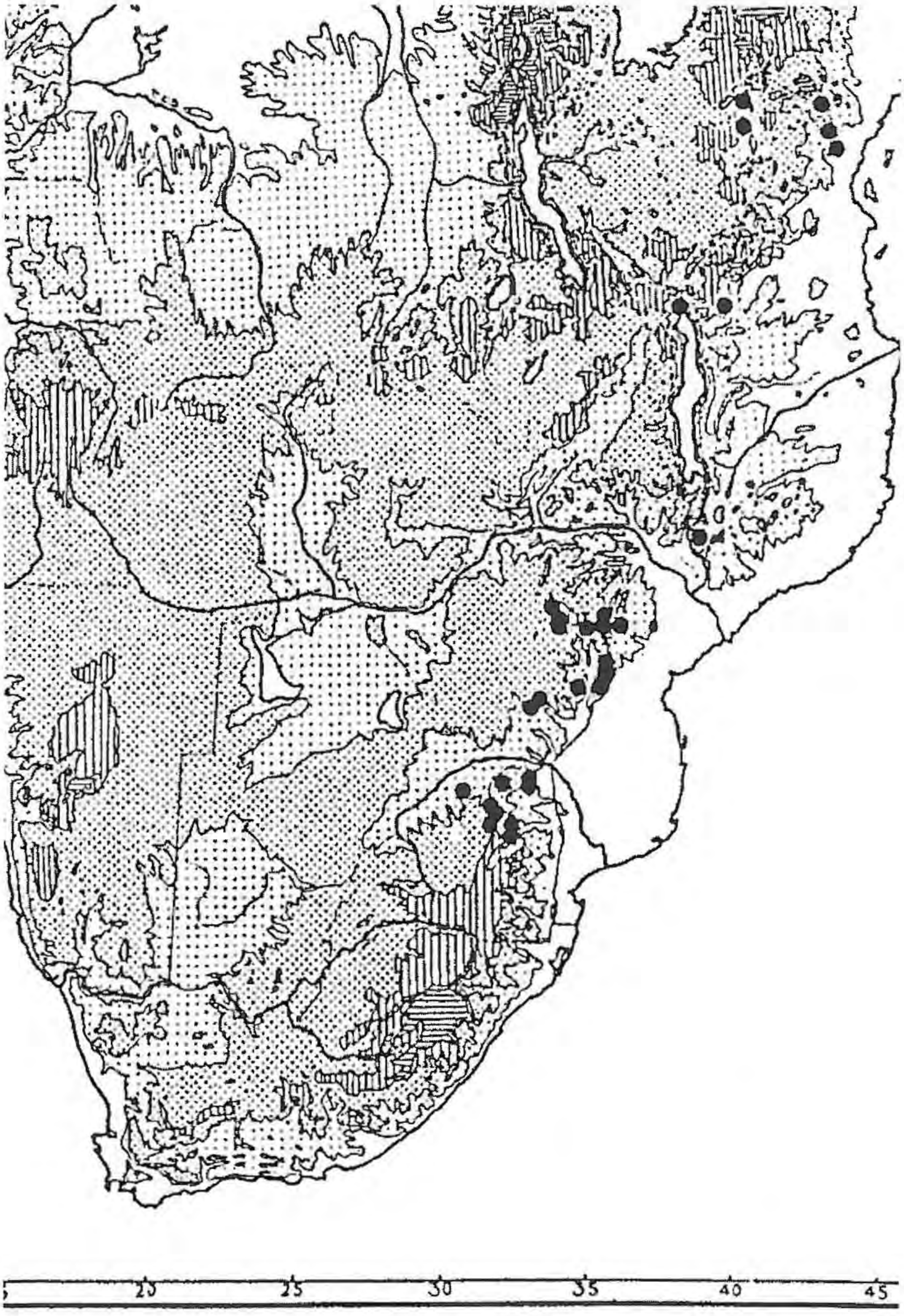

Figure 1.7: Distribution map of *Chrysanthemoides monilifera ssp. septentrionalis* (Griffioen, 1995)

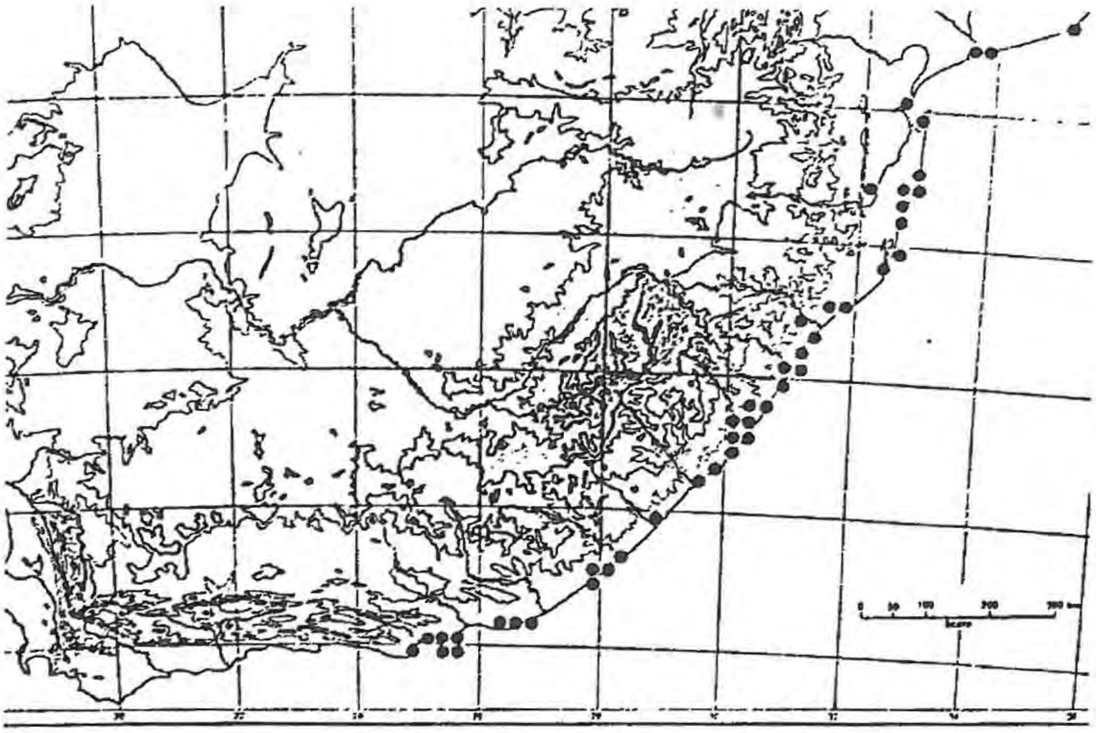

Figure 1.8: Distribution map of *Chrysanthemoides monilifera ssp. rotundata* (Griffioen, 1995)

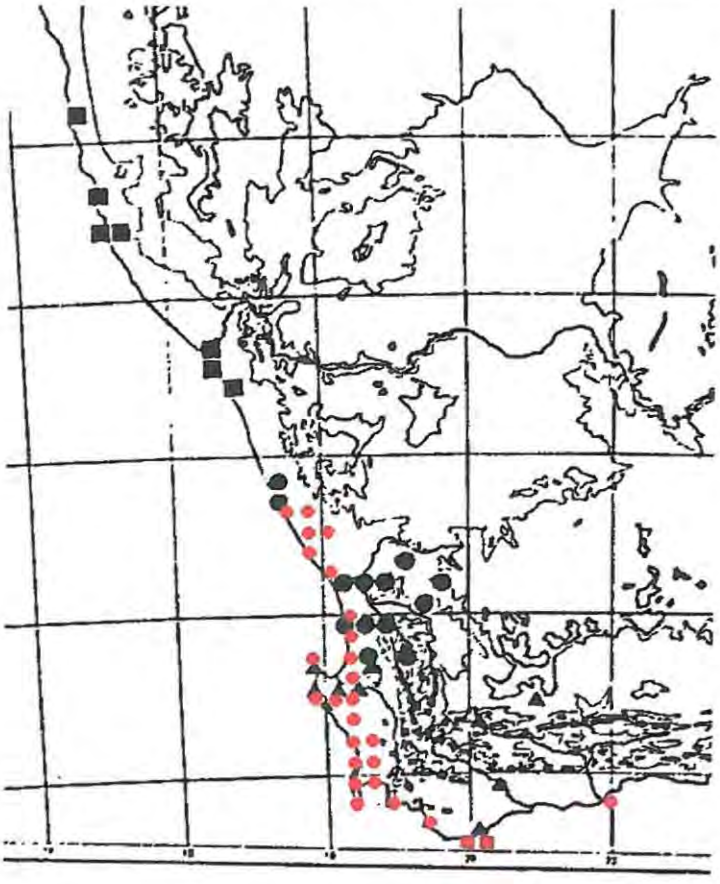

Figure 1.9: Distribution map of *Chrysanthemoides incana* ssp. *incana* [black triangle = *microfolia*; black square = *rangei*; black circle = *gracilis*; red circle = *incana*; red square = *hirsuta*) (Griffioen, 1995)

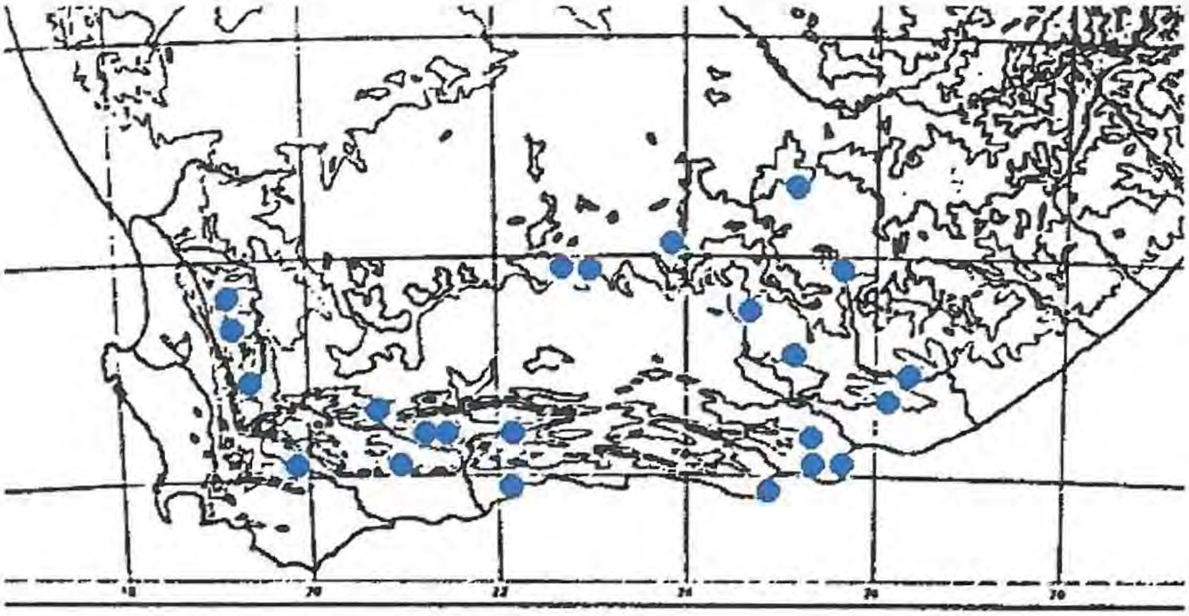

Figure 1.10: Distribution map of *Chrysanthemoides incana* ssp. *subcanescens* (Griffioen, 1995)

As this introduction serves to demonstrate, the taxonomy of the species and infraspecific taxa of *Chrysanthemoides* has been difficult, if not controversial. This, in part, has been a result of the application of different species concepts and taxonomic tools to the study of this genus. Central to the resolution of this problem is the recognition and delimitation of species and infraspecific taxa. It is thus pertinent and relevant that an overview of species concepts as related to this study is undertaken.

Species concepts

While species are regarded as the basic units of taxonomy, taxonomists have differing opinions of what these units should be, and what criteria should be used to delimit them. Species concepts remain controversial in plant systematics, and the delimitation of species boundaries remains a widely varying practise (Davis *et al*., 1991). There are many species concepts (22 at last count; Mayden, 1997).

The different ways species arise and are described is crucially linked to the comprehension of biodiversity and evolution (de Meeus, 2003). For systematists, the question of "by what criteria shall species taxa be identified" lies at the heart of the species-concept debate (Hey *et al*. 2003). Hey *et al*. (2003) note that

> "...when a taxon is to be a tool for the study of evolutionary entities, then the question becomes the following: what criterion will aid best in the discovery of the locations, boundaries and properties of evolutionary entities? Importantly, the answer might not be the same for all kinds of organisms" (p. 600).

Historically, and until the recent development of molecular tools, taxonomic groups have been identified using morphological criteria alone ("morphospecies"). This has left uncertainty in some cases as to the validity, and support, of groups delineated using these criteria (Hendry *et al.*, 2000). The last few decades have seen the development of molecular techniques that can serve as independent means to evaluate the validity of taxonomic groupings and designations. Hendry *et al.* (2000) point out that many genetic studies of taxonomies that were formerly based exclusively on morphology have uncovered paraphyletic or polyphyletic groupings, which have precipitated some taxonomic shuffling and reassignments. Hey (2001) notes that the idea of species as evolutionary groups is in stark contrast to the categorical tradition that is imbedded in most minds; the tradition of thinking of species as evolutionary groups is only 140-yrs old.

This re-examination process has also been applied to species designations in particular groups, and has lead to both some confirmations and some refutations of species status in these studies (Hendry *et al.*, 2000). They remind readers that while the recognition of a group of organisms as a distinct species (or not) comes down to a simple dichotomous choice ("yes" or "no"); the choice of a cut-off point or a level of differentiation (or isolation) that defines that species is nowhere near as simple. Evolutionary groups can also be difficult to study, as they are often very indistinct, with fuzzy boundaries between groups, and the any conjoining forces can be subtle (Hey, 2001).

A complete philosophy of science must contain both a theory of what is underlying reality (ontology) and a set of operations by which knowledge can be gained about that reality (epistemology) (Mishler & Theriot, 2000). In simple terms, species concepts have both grouping and ranking criteria: the grouping criterion indicates how organisms are to be put together into a group, whereas the ranking criterion indicates how the particular group thus formed is to be recognized as a species rather than some other lower or higher rank (Mishler & Theriot, 2000). Paul (2002) notes that there has been a tendency to confuse species concepts (simply the ideas on the kind of entity that is designated a species category) and species criteria (the actual standards for judging whether an entity qualifies as a member of a species category). The species criteria provide a list on which to judge if groups of individuals are distinct species (e.g. there are intrinsic barriers to genetic exchange, species are diagnosable by fixed character states) whilst the species concept should be concerned with defining a species according to some biological universal (Paul, 2002).

There are also multiple attempts by various authors to group the many species concepts into general categories (taxonomies of species concepts as it were):

Luckow (1995) groups species concepts into two general categories: (a) those based on causal theories or mechanisms, where species are units undergoing evolution; and (b) historically based concepts, where species are the end products of evolution. The BSC and others such as the Ecological Species Concepts (Van Valen, 1976) are grouped into the category of Mechanistic Species Concepts. These begin with a theory of how speciation (evolution) works, and then base species recognition on the most important causal factor or factors leading to speciation. The PSC is grouped with the historically based concepts.

This pattern of groupings was echoed by Harrison (1998) when he grouped seven main species concepts into (i) those clearly motivated by an interest in the process of speciation (including the BSC) and (ii) those that are clearly retrospective and emphasise pattern rather than process (including the PSC).

Mayden (1997) however, grouped the 22 species concepts in his review under two main groups: Those applicable to sexual organisms only (including the BSC here), and those that could also be used on asexual organisms as well (the PSC was included here).

The two main species concepts that will be under discussion here are the Biological Species Concept and the Phylogenetic Species Concept. This is because the BSC has been used to determine the current ranking of certain *Chrysanthemoides* taxa as subspecies. Griffioen (1995) uses the BSC to divide *Chrysanthemoides* into two species. He also cites the occurrence of morphologically intermediate populations as evidence of inter-breeding and thus assigns subspecies status, rather than species status. This study will attempt to apply the PSC to re-evaluate the taxonomic identity and ranking of several "subspecies" of *Chrysanthemoides*).

The Morphological Species Concept

The published classifications of *Chrysanthemoides* subspecies to date have been based on a Morphological Species (subspecies) Concept, which has several weaknesses. One of these weaknesses, mentioned by Taylor *et al.* (2000), is that a species diagnosed by morphology may often comprise more than one species when diagnosed by other species concepts. Another possible weakness is that morphological characters used to diagnose the species or subspecies may be dependant on environmental factors, rather than being inherent characteristics that are retained when the plants are grown under different environmental conditions. This is especially true of leaves, which have not often been regarded as reliable characters in classifications, due to their high variability and plasticity in response to environmental characters. For example, in a study on *Dodonaea* by West *et al.* (1984) the use of leaf morphology to separate *D. viscosa* from *D. angustissima* and *D. cuneata* was found to be inappropriate, as an apparently continuous variation in the leaf morphology exists among the three species, with no other significant morphological variation in other organs to assist in distinction. Although the three species do occur as well defined taxa in some populations, many morphologically intermediate individuals and populations also exist (West *et al.*, 1984).

The Biological Species Concept

One of the oldest and most commonly used species concepts is the Biological Species Concept of Mayr (1940). The BSC states that species are

> "… groups of actually or potentially interbreeding natural populations which are reproductively isolated from other such groups" (Mayr, 1940; 1963).

In the BSC, the inability to interbreed is the most important causal factor in speciation, and the basic idea is that all the individuals within a species have a similar morphology and general biology because they exchange genetic material (Linder *et al.*, 1994). It should be noted that this species concept was designed for birds and animals and is problematic in practical application to other organism classes. It is almost impossible to apply to asexual organisms, where each organism would be its own species (because asexual organisms are, by definition, reproductively isolated from all other organisms; Mishler & Theriot, 2000; de Meeus, 2003). Another minor problem frequently brought up is simply that it is practically impossible to make all the necessary crosses to test genetic compatibility (Mallet, 1995).

There have been many more major problems associated with the BSC, and one of the main ones is the absolute nature of total reproductive isolation. The identification of a BSC species can clearly be defined by 100% reproductive isolation. Hendry *et al.* (2000) point out that severe problems would occur

> "… if this criterion was universally adopted, as many current taxonomic species would no longer be recognised, owing to rampant hybridisation and introgression in the wild".

Under the BSC, gene flow is expected among populations but not between species. This view has made intraspecific patterns the primary focus of phylogeography; however, species boundaries are poorly known in most taxa, are blurry during speciation and can be semi-permeable to gene flow after speciation (Hare, 2001).

There is a problem in defining species under the BSC if there is any gene flow, and according to a strict isolation concept, any gene flow between species negates the BSC (Noor, 2002). Several taxa that biologists consider as good species exchange genes in some parts of their genomes and a strict version of the BSC might assume that any gene exchange demonstrates that reproductive isolation is incomplete and that species status has not been attained (Noor, 2002). Such a strict definition would not match the intuition of most biologists. If limited gene flow is permitted between BSC defined species, then the cut-off point necessary for species designation becomes subjective and arbitrary (How much reproductive isolation defines a species, ask Hendry *et al.* (2000): 90% isolation? 10% isolation?)

Despite Harrison (1998) claiming that the BSC is

> "… clearly motivated by an interest in the process of speciation",

there are also issues with the process of speciation and how the BSC accommodates this process. de Meeus (2003) has several points of contention about the BSC and indicates that species identification and speciation cannot be disentangled as long as the BSC will be used to describe life. This is because the BSC exactly describes how members of a group must behave to be considered as members of the same species (reproductive isolation) and, at the same time, how speciation occurs (evolution of reproductive isolation). A strict application of the BSC makes the speciation process difficult because it requires the evolution of a character (reproductive isolation) that is hardly adaptive by itself (de Meeus, 2003). In allopatry (where the two groups of interest are geographically isolated), it is ineffective and in sympatry (where there is no geographical barrier), it is deleterious, because it requires organisms to be sexually selective in comparison with those who are not selective (de Meeus, 2003). The major advantage of sexual reproduction lies in recombination; with the BSC, we see that speciation, because it requires that breeding only occurs between the alike, significantly weakens the advantage of sex for those species fitting in the BSC (de Meeus, 2003). Mallet (1995) also points out that since theories of speciation involve a reduction in ability or tendency to interbreed, species cannot themselves be defined by interbreeding without confusing cause and effect.

Speciation is usually driven by natural or sexual selection, which causes functional divergence. However, the BSC does not delineate the evolutionary forces that cause divergence, and, as such, the BSC considers reproductively isolated units to be species, even if differential adaptation is not apparent (Noor, 2002). Phrased another way, Avise (2000) pointed out this long-recognised drawback of the BSC: its difficulty in ranking allopatric populations.

Under an interbreeding concept, species cohesion is due to interbreeding or gene flow within species and an absence of gene flow between species, by definition (Mallet, 1995). This is related to the Cohesion Species Concept (Templeton, 1989) where species are

> "... the most inclusive population of individuals having the potential for phenotypic cohesion through mechanisms".

Because gene flow between species is conceptually impossible under interbreeding concepts, it is extremely hard to imagine how speciation, which must often involve a gradual cessation of gene flow, can occur (Mallet, 1995).

Griffioen (1995) has used the BSC in his study on *Chrysanthemoides* to divide the genus into two species, and to confer the status of "subspecies" (rather than "species") on the other taxonomic entities within the genus (based on the presence of multiple hybridization events).

The Phylogenetic Species Concept

Several alternative species concepts have been proposed in recent years (Avise, 2000) and several attempts have been made at forging a species concept compatible with phylogenetic systematics or cladistics (Mishler & Theriot, 2000). The history of the PSC progressed from efforts to fix the BSC, to applying cladistic analysis to the problem, to formulating a concept fully compatible with phylogenetic theory but not dependant on prior cladistic analysis (Wheeler & Platnick, 2000). Several such synthesis concepts have been called *the* Phylogenetic Species Concept, leading to confusion in the literature (Mishler & Theriot, 2000) and while various formulations of the PSC have been advanced, all agree that species recognition should emphasise criteria of phylogenetic relationship (descent) and not reproductive relationships (Avise, 2000). Diagnosability criteria issues have plagued previously formulated versions of the PSC, and the resolving power already available in molecular assays of rapidly evolving genes means that recently derived mutations often can be found that distinguish local populations, family units and even individuals (Avise, 2000).

This raises the question: what then is the level (cut-off point) of genetic differentiation sufficient to define a species?

A species definition must have both an ontology and an epistemology (Cracraft, 2000). That is, if the definition is heavy on the theory, but light on how species are to be identified in the real world, then one must look for the methodology that allows the definition to be applied (Cracraft, 2000). If the definition is very operational in its approach, then one must look to see whether the entities identified by that approach make theoretical sense with respect to some process-level phenomenon (Cracraft, 2000). The PSC *sensu* Wheeler and Platnick (PSC-WP; Wheeler & Platnick, 2000) is more epistemological, whilst the PSC *sensu* Mishler and Theriot (PSC-MT; Mishler & Theriot, 2000) is more ontological (Cracraft, 2000). Mayden (1997) divides the two PSCs into one based on Diagnosability (PSC-WP) and the other based on Monophyly (PSC-MT).

The PSC-WP has a definition of species as

> "... the smallest aggregation of populations (sexual) or lineages (asexual) diagnosable by a unique combination of character states in comparable individuals" (Nixon & Wheeler, 1990).

The PSC-MT has a definition of species as

> "... an irreducible cluster of organisms, diagnosably distinct from other such clusters, and within which there is a parental pattern of ancestry and descent" (Cracraft, 1989).

There are other differences between the two PSCs. According to Mishler & Theriot (2000), the PSC-WP is explicitly not based on synapomorphy, but on a shared combination of characters; whilst the PSC-MT holds that organisms should be grouped into species on the basis of evidence for monophyly. Differences in underlying philosophy remain still: PSC-WP has emphasised epistemology in its central focus on character evidence, whereas PSC-MT has emphasised ontology in its central focus on monophyly (Mishler & Theriot, 2000).

The other major point of contention, according to Mishler & Theriot, (2000), is that of reticulation, as the two versions of the PSC differ strongly in how they view reticulate relationships and characters in cladistic analysis. The PSC-WP states that units having reticulate relationships are inappropriate for phylogenetic analysis (because this is inherently a study of branching relationships; Mishler & Theriot, 2000). The PSC-MT states that reticulation can occur throughout the hierarchy of life and so is not a special species problem, but rather one of more general difficulty;

> "As barriers to reticulation are often not complete, reticulation is not a complete barrier to cladistic analysis" (Mishler & Theriot, 2000).

Platnick & Wheeler (2000) rebut this accusation by pointing out that

> "… the existence of hybrids is not unexpected based on what population geneticists have learned and does not negate the existence of either parent species".

Hey *et al.* (2003) ask the very important question of

> "When does one decide that there is one, or more than one, evolving entity?"

They suggest two fairly simple answers: (i) Don't decide whether or where to draw lines of demarcation, but instead present the full picture that research has revealed, in its full complexity rather than trying to impose an artificial reduction of complexity; (ii) Make a decision regarding demarcations, while also recognising that the decision is an oversimplification demanded by the practical concerns of the research.

The question of the demarcation used to define a species or a subspecies (i.e. how much genetic divergence defines a species), based on the PSC is a difficult one to answer. Under the PSC, many well-substantiated BSC sub-species will no doubt be elevated to species status (Wheeler & Platnick, 2000).

While morphological data has traditionally been used to delimit species and continues to be widely used today, recent studies have used DNA sequence data to test these morphology-based taxonomies (Wiens *et al.*, 2002). Barraclough *et al.* (1999) note that one recent approach for testing such ideas is to examine geographical and ecological patterns within closely related groups of species and correlate this to their phylogeny.

Weins & Penkrto (2002) point out that each population is a phylogenetic species if the haplotypes of all its members are joined in a contiguous section of an unrooted network. It is unclear, however, how exactly one divides up these networks into species. The PSC (*sensu* Wheeler and Platnick) holds the view that phylogenetically distinct populations ("species") are the smallest aggregation of populations that are distinguishable from each other (Nixon & Wheeler, 1990; Crozier, 1997) and must possess at least one unique character (or unique set of characters) that consistently distinguishes them from all other populations (Nixon & Wheeler, 1990; Vogler *et al.*, 1993; Luckow, 1995). As mentioned above, various formulations of PSC have been advanced, but all versions agree that species recognition should emphasise criteria of phylogenetic relationships (descent), not reproductive relationships (Avise, 2000). For current purposes of this thesis, the definition being used is that of Wheeler and Platnick (2000), based on Nixon and Wheeler (1990).

Evolutionarily Significant Units

The concept of ESUs was first proposed by Ryder (1986). ESUs are groups of populations derived from different common ancestors (Crozier, 1997), i.e. lineages that are evolutionarily isolated (Schwartz, 1999). The ESU has been proposed as a unit of conservation and is now widely applied for policy and management purposes, and the concept is frequently used for prioritising genetic diversity management activities (Cavers *et al.*, 2003). In terms of conventional taxonomy, an ESU most often corresponds to species or subspecies boundaries, but in some circumstances, can extend to isolated populations (Karl *et al.*, 1999). Riddle and Hafner (1999) suggest phylogeographic ESUs as an improvement over "species" as measures of biological diversity as an alternative concept to the idea of “species”. Ryder (1986) indicated that recognition of ESUs was not easy, and required the use of natural history information, morphometrics, distribution data, cytogenetic analysis and genetic information as well. Ryder (1986) also indicated that concordance between the sets of data derived by different techniques is a criterion for identifying ESUs.

As stated above, the concept of ESUs does not rely on BSC (which relies on reproductive isolation as a defining character), but rather on the PSC. Avise, (2000) points out that it is possible to reconcile the two species concepts to a certain degree:

> “… that within a suspected biological species, taxonomic subspecies should be demarcated by any pronounced and concordant phylogeographic partitions observed across multiple genetic traits. Thus, subspecies should conform to ESUs”.

The means by which ESUs were distinguished in the past have generally been based on morphological data, with some investigation being made into allozyme and isozyme analysis (Davis *et al.*, 1991; Griffioen, 1995; Rajakaruna *et al.*, 1999). Morphological characters, unlike molecular characters, have a phenotype that may be many developmental steps from the genotype that controls them. The nature of some polygenic morphological characters means that morphological character states may not be as discrete as DNA sequence base changes (Freudenstein *et al.*, 1994). Molecular techniques (such as DNA sequencing or fingerprinting) have more recently been applied in a wide range of studies (Avise *et al.*, 1987; Vogler *et al.*, 1993; Freudenstein *et al.*, 1994) in an effort to distinguish ESUs.

ESUs identification can be defined (by Moritz, 1994) as

> "… reciprocal monophyly for organelle haplotypes **and** significant divergence of allele frequencies at nuclear loci" (bold my emphasis),

clearly indicating that two or more genomic sources of sequence data is preferred, if not downright necessary. Hey *et al.* (2003) comment that an ESU should show evidence of being genetically separate from other populations and contributes substantially to the ecological or genetic diversity found within the species as a whole.

Avise (2000) adds yet another element to ESUs, commenting that the greater the intraspecific differentiation the better. Within a BSC defined species, taxonomic subspecies should be demarcated by any pronounced (and concordant) phylogeographic partitions observed across multiple genetic traits (Avise, 2000). Thus, theoretically, subspecies and ESU designations should be concordant. Stebbins (1950) describes subspecies and varieties as

> "… series of populations having certain morphological and physiological characteristics in common inhabiting a geographical subdivision of the range of the species, or a series of similar ecological habitats, and differing in several characteristics from typical members of other subspecies, although connected with one or more of them by a series of intergrading forms."

He suggests that using several degrees of rank may produce more confusion than order, however, the nature of the categories to be recognised depend on which number of ranks is most convenient and succeeds in providing a clear picture of the variation in the genus. ESUs provide an alternative to this potentially confusing hierarchical rank-based approach, and do not require subsequent ranking decisions to be made after ESUs have been identified.

Avise (2000) asks

> "How much difference is enough for ESU qualification?",

and suggests the clearer the differentiation, the better. He also asks

> "How deep must the branch separation in a gene tree be for ESU recognition?"

and again suggests that the deeper the separation of the branches, the better. He also points out that any universal definition that demarcates ESU status from non-ESU status is arbitrary to some extent and fails to concede that some situations truly are intermediate. Moritz (1994) suggested a specific genetic cut-off (based on mitochondrial DNA monophyly and nuclear gene differences) for conferring ESU status in animals (Hey *et al.*, 2003). Given that mtDNA diversity will often be a poor indicator of demographic boundaries (Hudson, 2003), this particular proposal may not be ideal (Hey *et al.*, 2003).

In accordance with Ryder's (1986) suggestion that multiple sources of data be utilised to designate ESU status, other researchers have pointed out that basing ESUs on genetic data only (particularly neutral data) may not fully address the problems of conservation and further factors, especially ecological ones, must also be taken into account (Cavers *et al*., 2003). Cavers *et al*., (2003) present the idea that concordance between neutral genetic and functional adaptive information is a good strategy for the identification of ESUs. This approach is fundamental to this project. A further suggestion in their paper, that as a starting point, identification of molecular ESUs provides a valuable practical framework for further study, is the focus and aim of chapter 2 of this thesis.

Hey *et al*. (2003) suggest that the current ESU criteria should be replaced by

> "... a single, better criterion that would, inherently, by its nature, dispel uncertainty".

Hey *et al*. (2003) state that the principal claim of this sort is that ESUs should be groups of individuals that share a unique character, or a suite of characters that distinguish them from individuals of other ESUs (the same criterion applied to "species" in the PSC-WT). Hey *et al*. (2003) claim that this definition is not for reasons of efficiency, but because such criteria are inherently unambiguous indicators of real evolutionary entities. This suggestion comes full circle back to the PSC-WT, and simply replaces the word "species" with "ESU". Hey *et al*. (2003) do, however, note that

> "... we should recognise that there is not a single species concept, nor a research protocol, that can remove the inherent difficulty and uncertainty that accompanies research on evolving populations".

Aims

1) To identify taxonomic units (ESUs) within Chrysanthemoides using (DNA sequence) data.

In order to test the morphological and biological species and infraspecific taxa, data from two genetic sources (nuclear and chloroplast) were utilised in an effort to identify ESU's with a genetic basis. The Internal Transcribed Spacer region of the nuclear ribosomal DNA and the *psbA-trnH* and *trnL-trnF* spacer regions of the chloroplast genome were used in this component of the study.

2) To correlate the ESUs identified above with currently recognised taxa based on morphology.

Generally there is a close correspondence between molecular phylogenies to those based on morphological data, indicating that molecular differences can reflect differences in other aspects of the organism (Crozier, 1997). Herbarium specimens were made of each plant that DNA samples were taken from. Using these, the morphological taxonomic designation of each plant included here was determined. This designation was then used to determine if the genetically identified ESUs correlated to previously identified morphological entities.

3) To correlate the ESUs identified above to geographical areas and/or climatic factors

The localities of each sample's origin were noted on the phylogenetic trees to determine if any correlation between the genetically delimited ESUs and geographical distribution or climatological variables exists.

4) To determine whether some of the ESUs identified above have unique ecophysiological traits, and if these traits correlate to climatic or environmental factors.

In order to investigate the processes that may limit or allow possible expansion of the geographic distributions of the genealogical lineages, the functional ecophysiology of three selected ESUs was investigated across their environmentally disjunct habitats in a study to determine if the ESU's distribution could be explain by possible ecological or environmental limitations. Three field study sites were chosen and the ecophysiological functioning of the plants that occurred at the three sites was compared to determine if there are any physiologically unique traits inherent to each group. Further more, plants taken from each ecophysiological study site were grown in a greenhouse in an attempt to discover if their physiological traits remain when they are subjected to identical conditions, or if they are physiologically plastic enough that they change to suit their local environment.

The first three aims are addressed in the following chapter, while the last aim is the subject of Chapter 3.

Chapter 2.

Molecular Phylogeography of *Chrysanthemoides*

Introduction

Phylogeography

Phylogeography, as defined by Bermingham & Moritz (1998), is an investigation of the fundamental links between population processes and regional patterns of diversity and biogeography. Avise (1998) further defines phylogeography as being concerned with principles and processes governing the geographical distribution of genealogical lineages, especially those at the intraspecific level, but notes that there are relatively few instances of molecular phylogeographic analyses at the intraspecific level in the botanical literature (Avise, 2000).

Schaal *et al.* (1998) note that within species, it is genetic exchange that has traditionally been emphasised as the determinant of genetic structure, rather than historical relationship. They further point out that phylogeography, which uses the historical information inherent in gene trees, is not merely an extension of phylogenetic principles to the intraspecific level. Rather, it characterises population subdivision by recognising geographical patterns of genealogical structure across the range of a species. Phylogeographic analysis relies on interpreting patterns of congruence (or lack of congruence) between the geographical distribution of haplotypes and their genealogical relationships (Schaal *et al.*, 1998). A pattern of congruence is seen if clades of closely related haplotypes are geographically restricted and occur in proximity to each other (Schaal *et al.*, 1998). Congruence indicates a long standing pattern of highly restricted gene flow, with this pattern arising because novel mutations remain localised within the geographical context of their origins (Schaal *et al.*, 1998). Phylogeography can provide a theoretical framework that can be used to test ideas of genetic isolation without the restriction of taxonomic preconceptions based solely on morphological divergence (Schaal *et al.*, 1998).

In any phylogeographic investigation, the accuracy of the spatio-temporal reconstruction will be affected by the resolution that is possible in terms of geographical structure and phylogeny (Burban & Petit, 2003). The geographic resolution will be greater if the genomes targeted are subject to limited gene flow, whereas the phylogenetic reconstructions will be facilitated if they are uniparentally inherited (Burban & Petit, 2003).

Bermingham & Moritz (1998) concede that phylogeography has been subject to criticism for being overly reliant on using a single gene system alone as a marker of evolutionary descent. Any single gene tree represents only a small part of the full genetic history of a species and cautions that care must be exercised in drawing population-level conclusions from gene-tree data at a few select loci only (Avise 2000; See Figure 2.1). Hare (2001) notes that a gene tree for a single cytoplasmic or nuclear locus

> "provides a slim and sometimes misleading representation of the population histories through which alleles were transmitted".

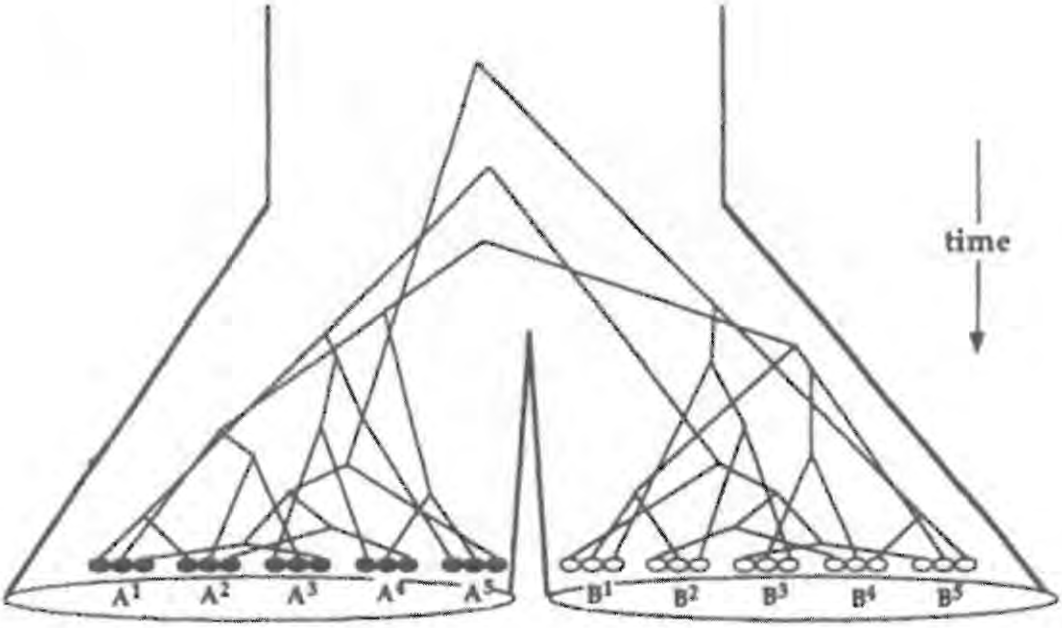

Figure 2.1: Illustration of the difference between patterns of phylogeny from a single gene tree and three gene trees (genealogies) that create a species tree. A and B (1-5) represent individuals in two different populations, each of which have had 3 independent genes sampled. (from Shaw, 1998, p.49).

The possible pitfalls inherent in this limited approach (such as interspecific hybridisation or the inadvertent amplification of pseudogenes) can be overcome by testing for phylogenetic congruence across nuclear and chloroplast genes (Bermingham & Moritz, 1998). The question of

> "how many genealogical pathways are needed to estimate any major features in a phylogeny?"

is answered by Avise (2000) as

> "relatively few is the answer from computer simulations".

The two primary sources of molecular variation used for phylogenetic purposes in plants have been the chloroplast genome (Palmer, 1987; Palmer et al., 1988; Olmstead et al., 1990) and the nuclear ribosomal DNA repeat region (Baldwin, 1992; Baldwin, 1993; Suh *et al.*, 1993; Grube *et al.*, 2000). The ability to work with two sources of data (uniparentally inherited chloroplast DNA and biparental nuclear ITS DNA) also offers its own advantages: more phylogenetically informative characters, possible complementarity of phylogenetic signal and the ability to track the different histories of the two sets of DNA (McDade *et al.*, 2000). The biparental pattern of inheritance of ITS can make hybridisation events easier to detect, and a short cpDNA intergenic spacer alone may not provide enough synapomorphic mutations to resolve close interspecific relationships. These relationships may be assessed by sequencing multiple rapidly evolving noncoding regions in the chloroplast genome (Sang *et al.*, 1997).

There are some concerns with using nuclear genes in a phylogeographic context. Hare (2001) presents two practical concerns with using nuclear haplotypes for phylogeography: (i) phylogenetic methods of gene tree reconstruction could be impeded by recombination and (ii) poor resolution could result from low mutation rates, leading to too few informative polymorphisms. There is also a low expectation of nuclear monophyly among recent population isolates (Hare, 2001).

There are also concerns with using organellar genomes. Because cytoplasmic loci are usually inherited uniparentally (maternally in almost all angiosperms; Palmer, 1987; Reboud & Zeyl, 1994), they will not have genealogical patterns that are representative of the entire population history (Hare, 2001) but rather trace matriarchal lineages (Palmer, 1987). However, when these two separate (and often discordant) sources of genetic data are used concurrently, they can highlight historical hybridisation and introgression events or provide evidence of ancestral polymorphisms when compared with discordant nrDNA data (Palmer, 1987). There is also a definite necessity for more than one source of genetic sequence data, especially when one is investigating or assigning taxonomic status (Tautz, 2003).

There is, however, going to be some inherent degree of phylogenetic discordance across gene trees, and this is simply a consequence of Mendelian inheritance in combination with the random patterns of lineage sorting at unlinked loci through a sexual pedigree (Avise, 2000). Gene trees may also conflict with other biogeographic information for several reasons, including chance historical events leading to phylogeographic patterns that at face value appear inconsistent with historical geography (Avise, 2000).

Another potential source of disagreement relevant to phylogeography in plants is highlighted by Schaal *et al.* (1998) and that is the relationship between morphological and genetic differentiation: Morphological divergence may have little relationship to the degree of genetic differentiation between lineages, and it can thus be difficult to predict the genetic cohesiveness of a group based on its morphological differentiation (or taxonomic classification) alone. Schaal *et al.* (1998) also note that, as a result of a lack of genetic differentiation (and the potential for hybridisation) genetic exchange may in fact be occurring between what appear to be independent lineages. They note that some studies have shown that chloroplast haplotypes can be more strongly correlated with their geographical position than with any morphologically based subspecies designation (Terauchi *et al.*, 1991; Mayer & Soltis, 1994).

Hare (2001) states that the marker of first choice for phylogeography is and will continue to be cytoplasmic DNA for the initial characterisation of population structure, testing population monophyly or inferring maternal gene flow. He also notes that given the potential for complications (both technological and biological) when analysing nuclear haplotype data at the intraspecific level, it has been uncertain whether a phylogeographical study design could generally be extended to noncytoplasmic markers. He notes that in theory, phylogeographical structure is likely to be less pronounced at diploid nuclear loci than with cytoplasmic loci because of their different effective population size. The effective population size of a locus is related to the number of breeding adults; but it might be larger or smaller, depending on factors such as ploidy and mode of inheritance (Hare, 2001). In dioecious populations with equal ratios of both males and females and each sex having equal variance in reproductive success,

> "autosomal nuclear loci have an effective population size four times larger than that of uniparentally inherited cytoplasmic markers" (Hare, 2001).

It should be noted that there is some disagreement in the literature on the use of cpDNA. Hare's (2001) assertion that the first choice of markers for phylogeographical study are those of chloroplast genes is strongly contradicted by Burban & Petit (2003) who maintain that in studies based on cpDNA, phylogeographic insights have been reduced because of low levels of geographical structure and the limited phylogenetic resolution. Burban & Petit (2003) do however agree that using multiple markers with contrasted inheritance is a sound concept.

Regions of DNA suitable for molecular phylogeography studies in plants

The suitability of a genome for phylogenetic study is dependant on the mode and tempo of evolution of the genome: the tempo of change is a major determinant of the taxonomic level at which a particular genome (or DNA sequence thereof) is most informative (Palmer, 1987). Within that genome, several criteria should be met in the selection of a specific sequence for phylogenetic analysis: the most important being sufficient length of sequence to provide enough phylogenetically informative nucleotide positions (Olmstead and Palmer, 1994). Concurrent with this is the necessary rate of sequence divergence and that this divergence of sequence (not just the tempo of change in the genome) be appropriate to the phylogenetic question being addressed (Olmstead and Palmer, 1994).

In the short history of molecular systematics in plants, much has relied on the chloroplast genome and more recently on the nuclear genome (Soltis & Soltis, 1998). Both chloroplast DNA (cpDNA) and nuclear ribosomal DNA (nrDNA) have been used to infer phylogenies at different taxonomic levels (Avise 2000). Coding regions have been widely used at higher taxonomic levels (family, tribe etc), but they are less informative at lower taxonomic levels since they are highly conservative. Therefore, there has been a growing interest in the use of noncoding sequences at lower taxonomic levels such as genus and/or species (Gielly *et al.*, 1996). Non-coding sequences are generally expected to reveal more information than coding regions at lower taxonomic levels (Gielly *et al.*, 1996), although mutation rates differ across different lineages (Wolf, 1987; Gaut, 1993), and rates of substitution can vary by as much as a factor of five (Gaut, 1993). Also, although the rate of mutation of coding regions may be similar, rates of mutation of non-coding regions can be more variable (Gaut, 1993), a factor that was encountered when initial screening for suitable chloroplast noncoding regions for this project was undertaken.

There are many genes in several different regions of the plant's genome that can be used, but for the purposes of this project, the nuclear ITS region and the chloroplast *trnL-trnF* and *psbA-trnH* regions were selected following preliminary screening trials.

The Internal Transcribed Spacer region

The Internal Transcribed Spacer (ITS) conventionally includes the entire ITS-1, 5.8S gene and ITS-2 portion of the nuclear rRNA cistron. Within the ITS, the highly conserved 5.8S gene sequence is useful for designing primers (Coleman, 2003). The ITS regions are part of the nuclear rDNA transcript but are not incorporated into ribosomes. They appear to play a role in the maturation of nuclear rDNAs, bringing the large and small subunits into close proximity within a processing domain. This function suggests that ITS-1 and ITS-2 are under some evolutionary constraints in structure and sequence (Soltis & Soltis, 1998).

There were some initial misgivings about the suitability of the ITS for phylogenetic studies, and these arose from the observation that it is a multigene family with the potential for variation among repeats (there are hundreds of tandem copies of the rRNA cistrons at the nucleolar organiser locus in a typical eukaryotic genome; Coleman, 2003). Accumulating evidence now suggests that significant variation among ITS sequences of an organism is found only within organisms that are hybrids, either diploid or polyploids, of disparate parents. In other non-hybrid organisms, a process called concerted evolution rapidly homogenizes the many copies of this multigene family, such that ITS can be treated as a single gene (Coleman, 2003). The spacers do have a nucleotide substitution rate high enough to generate intra- and interspecific variability (Fuertes Aguilar *et al.* 1999a). Fuertes Aguilar *et al.* (1999a) also note, however, that the accumulation of such variability (and therefore the observed divergence between evolving lineages) is commonly affected by concerted evolution. This mechanism (which causes the homogenization of sequences of the ribosomal DNA tandem repeats) is relevant to explain the observation that ITS intraspecific variability is lower than variation among species.

The use of the ITS genes (located in the 18S - 26S nuclear ribosomal DNA) is well documented and has proven to be a useful source of characters for phylogenetic studies of many families, due to (i) its small size (useful for PCR work), (ii) highly conserved flanking regions (allowing for primers to be easily coded), (iii) the higher amount of sequence divergence compared to their flanking coding regions, (iv) their rapid concerted evolution and (v) their high copy number (which, in combination with its small size, makes them suitable for direct sequencing of PCR products, instead of requiring the use of cloning techniques; Gielly *et al.*, 1996; Grube *et al.* 2000). There does exist, however, the potential for reticulate evolution due to recombination which can make lineage delimitation complicated (Bermingham & Moritz, 1998).

The use of the ITS region is not without its critics. Alvarez and Wendel (2003) surveyed the plant phylogenetic studies over five years, and found that of 244 papers, 66% of papers that involved comparisons at the genus level or below used ITS sequence data. Alvarez and Wendel (2003) then compared the CI and RI of multiple data sets from studies that used both ITS and other loci for the same set of samples, to aid in the estimation of homoplasy in the various data sets. They found that in only one of 34 comparisons was the CI for ITS judged to be higher than that of the other molecular marker used. Alvarez and Wendel (2003) firmly stated that they

> "... recommend that ITS no longer be routinely utilized for phylogenetic analyses, opting instead for using several or more different single-copy nuclear loci."

The more rapidly evolving regions of the nrDNA (of which ITS is an example) have been used to investigate relationships at various taxonomic levels, from family (Suh *et al.*, 1993) to genus (Baldwin, 1993; Francisco-Ortega *et al.*, 1997; Dubouzet & Shinoda, 1998; Conti *et al.*, 1999; Grube *et al.*, 2000) to the intrageneric and species level (Chatterton *et al.*, 1992; Wojciechowski *et al.*, 1993; Sang *et al.*, 1994; Kollipara *et al.*, 1997; Lashermes *et al.*, 1997; Conti *et al.*, 2000; Prentice *et al.*, 2003; Yang *et al.*, 2003), with Baldwin (1993) having observed ITS sequence divergence among populations of a species of *Calycadenia* (Asteraceae). Bermingham & Moritz (1998) do however indicate that the relatively slower coalescent time for nuclear genes (relative to organellar DNA) can limit their use in population-level studies. The ITS region is also taxonomically useful for interspecific studies and has become a major focus of comparative sequencing at the specific and generic level, and also has been found to have a high information content at lower taxonomic levels (Soltis & Soltis, 1998; Grube *et al.*, 2000). The variation in the ITS region can be sufficient to allow for the construction of phylogenetic relationships (Baldwin *et al.*, 1995).

Grube *et al.* (2000) note that data from any single locus are sufficient to support a phylogenetic hypothesis but they are, however, not sufficient to reject it. This is because a concordance of a single gene genealogy with the phylogenetic hypothesis, based on phenotypic or biogeographical characters, suggests divergence between phylogenetic species. But when these discrete clades of alleles do not correlate with phenotypic or biogeographical data, the two conflicting data sets are insufficient to determine species boundaries (Grube *et al.*, 2000). The correlation between sequence data and phenotypic characters or biogeographical distribution will be true only when sufficient time has passed for genetic isolation to result in the fixation of different character states among the sibling species under investigation. In the case of recently diverged species, however, a single locus may not accurately separate species (Grube *et al.*, 2000). As Ferguson *et al.* (2002) point out, organellar DNA is an obvious source of data for development of a phylogeny independent of the nuclear-based ITS phylogeny.

The Chloroplast genome (cpDNA)

Organelle genomes are ideal markers for phylogeny reconstruction because their sequences record the history of a lineage uncomplicated by recombination (Harrison, 1991). The inheritance of cpDNA is clonal, and the chloroplast genome is inherited most commonly from one parent: through the maternal line in flowering plants, and the paternal parent in certain gymnosperms (Palmer, 1987; Olmstead & Palmer, 1994; Reboud & Zeyl, 1994). Between 20 and 200 copies of this plastid genome are found in each mature chloroplast and these multiple genomes are clustered into nucleoids, which are scattered throughout the stroma (Palmer, 1987).

The chloroplast genome is a relictual molecule of about 150kbp (135kb to 160kb) (Olmstead & Palmer, 1994; Clegg *et al.*, 1995) although it can reach up to 220 kb in size (Gaut *et al.*, 1993). Between only 50 and 100 of the plastid polypeptides (i.e. genes) are encoded by cpDNA; the rest are encoded by nuclear DNA, synthesised cytoplasmically and then transported across the chloroplast envelope (Palmer, 1987; Gaut *et al.*, 1993; Olmstead & Palmer, 1994).

cpDNA has three functional categories of DNA in the genome: (i)non-coding regions that do not code for transfer RNA, ribosomal RNA or proteins; (ii) coding genes (rRNA and proteins); and (iii) chloroplast introns (Clegg *et al.*, 1995). The cpDNA genome is extremely condensed compared to nrDNA and most of the noncoding DNA in the chloroplast genome is found in very short segments separating functional genes (Clegg *et al.*, 1995). It is a relatively stable genome, varying little in size, with marked conservation of gene content and substantial conservation of a primitive gene order and structural organisation (Palmer, 1987; Olmstead & Palmer, 1994). It also accumulates nucleotide substitutions relatively slowly (Palmer, 1987) and has a lower rate of intraspecific mutation than nuclear DNA (Palmer & Zamir, 1982; Clegg *et al.*, 1984a; Clegg *et al.*, 1984b; Perl-Treves & Galun, 1985) which has made the chloroplast genome an ideal focus for studies of plant evolutionary history (Clegg *et al.*, 1995).

There are three main regions in a typical chloroplast genome: the Inverted Repeat (IR), the Small Single Copy region (SSC) and Large Single Copy regions (LSC). The IR regions are actually two identical sequences of usually 25 kb arranged as an inverted duplication that separates the remainder of the 150 kb genome into a SSC and LSC (Sigiura, 1989). The three regions vary in the rates of nucleotide substitution, with the IR having three- to four-fold lower rates of silent nucleotide substitution than LSC or SSC (Wolf *et al.*, 1987; Gaut *et al.*, 1993).

Much of the size variation in the chloroplast genome is associated with that of the large IR regions which are present in all land-plants, with the exception of one group of legumes, which lack one entire copy of the duplicated element (Palmer 1987). By contrast, the 217 KB genome of Geranium, the largest land-plant cpDNA, also features the largest known inverted repeat of length 76kb (Palmer, 1987; Palmer et al. 1987; Olmstead & Palmer, 1994). Most of the known cpDNA size variation occurs by simple changes in the length of the 0-76kb IR, unaccompanied by very few changes in sequence complexity of this region (Palmer 1987). Changes in genome complexity occur primarily by length mutation: the addition of new sequences; or deletion of existing ones, rather than by a gradual drift of repeated elements until such time as they become unmatched single copies (Palmer, 1987). Most of these length mutations are extremely short (involving 1-10 base pair sequence tracts located in non-coding regions; Palmer, 1987).

In a similar fashion to animal mtDNA, the chloroplast genome may be considered a single, non-recombining unit of inheritance; the mutation rate of which varies for different regions of the genome, with the most variation occurring within the large single-copy regions and not in the inverted repeats (Schaal *et al.*, 1998). Further comparisons of cpDNA to the mitochondrial genomes (mtDNA) of both plant and animals show that cpDNA has a blend of the most conservative mutational tendencies of each mitochondrial genome (Palmer, 1987). Animal mtDNA, whilst extremely conservative in size and structure, changes extraordinarily rapidly in primary sequence; and whilst gene order is identical, the sequences of these positionally conserved genes evolve at a fast rate (Palmer, 1987). Plant mtDNA evolution is the opposite to that of animals: plant mtDNA changes rapidly in size and structure, but slowly in primary sequence (Palmer, 1987). Plant mtDNA's can be "bewilderingly large and variable in size, ranging from 200kb to 2000kb" and their gene order is also highly variable; a result of both a high rate of inversion, and of the frequent loss and gain of recombination repeats (Palmer, 1987).

Olmstead & Palmer (1994) noted that cpDNA has provided useful intraspecific variation in some, but not all, species studied (See Soltis *et al.*, 1992 and Barker *et al.*, 2003 for comprehensive lists). A later paper by Jordan *et al.* (1996) noted that due to low evolutionary rates, cpDNA has been traditionally been used to study plant systematics above the species level but that intraspecific variation in cpDNA has been observed in the majority of species examined and thus has proved useful in gaining insights into evolutionary processes within and among populations (Wagner et al., 1987; Soltis et al., 1989; Lavin et al., 1991; Milligan, 1991; Dong & Wagner, 1994; Gielly & Taberlet, 1994; Jordan *et al*., 1996). The usefulness of cpDNA seems to have advanced quite rapidly, such that Schaal *et al.* (1998) note that, as of their 1998 paper, virtually all published plant phylogeography studies had relied on the chloroplast genome alone as their source of genetic information. This sentiment later was echoed again by Hare (2001).

In principle, cpDNA sequencing should involve the selection of a sequence whose substitution rate is appropriate to the phylogenetic problem at hand (Olmstead & Palmer, 1994). To this end, a number of regions (the *trnL-trnF*, *psbA-trnH* and *trnS-trnfM* spacer regions and the *rps*16 intron) were screened to determine if they could provide the necessary intraspecific variability. Non-coding regions of cpDNA, (presumably under less functional constraint and thus evolving more rapidly) may also provide more useful phylogenetic information at lower taxonomic levels (Olmstead & Palmer, 1992; Sang *et al*., 1997; Gielly *et al*., 1996; Klechner, 2000), with 50% of all chloroplast variation attributable to small insertion/deletion mutations (Schaal *et al*., 1998).

trnS-trnfM

The *trnS-trnfM* region has not been as widely used as some other cpDNA non-coding regions but it has been used at species level studies by Stehlik *et al.* (2002) in a study of *Eritrichium nanum*, where they found sufficient variation at the species level to enable the differentiation of populations, suggesting that it would be suitably variable for phylogeographic studies in *Chrysanthemoides*.

rps16

The *rps16* intron is also not as widely used as other cpDNA regions. Asmussen & Chase (2001) refer to rps16 as one of the faster evolving regions of the chloroplast genome. It displays levels of sequence divergence between two and three times lower than that of the ITS region, and therefore been suggested as a useful tool above genus level, but below family level (Liden et al., 1997; Oxelman et al., 1997; Baker *et al*., 2000) It has proven useful at the genus level, in *Gunnera* (Wanntorp *et al*., 2001; Wanntorp & Wanntorp, 2003), *Cymopterus* (Downie et al., 2002) and *Alectryon* (Edwards & Gadek, 2001) and has also been used successfully at species level, in *Silene aegaea* (Popp & Oxelman, 2001), also suggesting that it would be a useful tool in the study of *Chrysanthemoides*.

psbA-trnH

The *psbA-trnH* intergenic spacer region lies in the inverted repeat region of the chloroplast genome, adjacent to the *matK* gene and near to the boundary with the evolutionarily plastic single-copy region (Sang *et al.*, 1997). The *psbA* chloroplast gene belongs to the Photosystem II (PSII) protein complex and codes for the PSII D1-protein (Chandler *et al.*, 2001). The $trnH^{HIS}$ (GUG) gene belongs to the transfer RNS gene system and transfers for the amino acid histidine (Chandler *et al.*, 2001).

At higher taxonomic levels, the swiftly mutating region may be too variable to be phylogenetically informative, with Miller *et al.* (2003) finding 47% of all sites in the *psbA-trnH* region to be variable at the tribal level, with many large deletion events that could not be confidently scored as characters. These large indels (most commonly deletion events, rather than insertions) occur very frequently in the *psbA-trnH* (Sang *et al.*, 1997; Mort *et al.*, 2002). These gaps have, however, been found to be uninformative at the genus level in some studies (Chandler *et al.*, 2001).

The *psbA-trnH* region has been used in phylogenetic studies at the family level (Klak *et al.*, 2003), the tribal level (Miller *et al.*, 2003), the intergeneric level (Aldrich et al., 1988; Asmussen & Liston, 1998) as well as at the intrageneric level (Olmstead & Palmer, 1992; Gielly *et al.*, 1996; Sang *et al.*, 1997; Kim et al, 1999; Chandler *et al.*, 2001). Higher rates of mutations have been detected in the *psbA-trnH* intergenic spacer among species of peonies than in the *trnL-trnF* intergenic spacer, suggesting that the former might be more useful than the latter for phylogenetic studies at the intrageneric level or below (Sang *et al.*, 1997). Chandler *et al.* (2001) found that at the generic level in *Gastrolobum*, 54% of the bases in the *psbA-trnH* region were variable (and 32% informative), suggesting that there was enough sequence diversity in this noncoding region to be useful at or below species level for *Chrysanthemoides*.

trnL-trnF

The *trnL-trnF* intergenic spacer region is located in the large single copy region (Mummenhoff *et al.*, 2001) and is one of the most commonly used non-coding regions of cpDNA in phylogenetic studies (Sang *et al.*, 1997; and papers listed therein). It has often been used in combination with other chloroplast regions (such as *psbA-trnH*, *matK* and *trnK*; Miller *et al.*, 2003). Like the *psbA-trnH* region, the *trnL-trnF* region has been found to have many variances in length and also large indel (mostly deletion) events (Miller *et al.*, 2003; Mort *et al.*, 2003).

The *trnL-trnF* region has been used at many levels of phylogenetic study, from the tribal and sub-tribal level (Bayer and Starr, 1998; Kim et al, 1999; Miller *et al.*, 2003) to the intrageneric and species level (Olmstead & Palmer, 1992; Mes and t'Hart, 1994; Gielly *et al.*, 1996; Sang *et al.*, 1997; Chandler *et al.*, 2001). The large number of studies using *trnL-trnF* at the species level suggested that is was another potentially useful source of data for *Chrysanthemoides*.

Methods

Sampling strategy and sample collection

Samples from each of the eight subspecies (of *C. incana* and *C. monilifera*) were collected from a diverse range of places, including some sites common to Griffioen's (1995) collection vouchers. The maps on the following pages show the sample sites for DNA voucher collection points (listed in Table 2.1 below). For each sample leaf material was collected into silica gel according to the method of Chase and Hills (1991).

Table 2.1: List DNA specimens used in analysis, as well as collection sites and morphological designation. Collectors initials: SH-Seranne Howis; CP - Craig Peter; SR - Syd Ramdhani; NPB - Nigel Barker; BSR - Brad Ripley; TD - Tony Dold; MP – A. Massawe & P. Phillipson; AW- Alan Wood.

Voucher	Locality	South	East	Genus	Species	Subspecies
PiBotDpt	Grahamstown: Botany Dept, Rhodes University	33° 19' 00"	26° 31' 00	Chrys.	monilif.	pisifera pisifera f2
PiMatt	Grahamstown: Memorial grove, Grahamstown	33° 21' 30"	26° 33' 30"	Chrys.	monilif.	pisifera pisifera f2
PiThmbB	Salem: Thomas Baines Nature Reserve.	33° 23' 30"	26° 29' 30"	Chrys.	monilif.	pisifera pisifera f2
pi fin	Grahamstown: Featherstone Kloof, Grahamstown	33° 20' 30"	26° 31' 30"	Chrys.	monilif.	pisifera pisifera f2
pi CSTf	Joubertina: Joubertina	33° 49' 30"	23° 51' 30"	Chrys.	monilif.	pisifera pisifera f2
Pe2	Port Elizabeth: Port Elizabeth	33° 59' 00"	25° 40' 30"	Chrys.	monilif.	rotundata
RKOS	Boesmansriviermond: Kenton-on-Sea, beach	33° 41' 00"	26° 40' 30"	Chrys.	monilif.	rotundata
SH 40	Knysna: Langkloof junction, N2/R62, roadside	34° 06' 30"	23° 04' 00"	Chrys.	monilif.	pisifera pisifera f2
SH 42	Kareedouw: Assegai Hotel, just before Kareedouw	33° 56' 30"	24° 19' 00"	Chrys.	monilif.	pisifera pisifera f2
SH 43	Kareedouw: same Loc as SH 42	33° 56' 30"	24° 19' 00"	Chrys.	monilif.	pisifera pisifera f2
SH 44	Kareedouw: Mid-Kareedouw pass to N2, steep slope	33° 58' 30"	24° 15' 00"	Ost.	junceum	
SH 49	Joubertina: 5km west of Joubertinia	33° 38' 00"	23° 47' 30"	Chrys.	monilif.	pisifera pisifera f2
SH 50	Kruisvallei: pass to Knysna, R339, top of hill	33° 51' 30"	23° 11' 30"	Chrys.	monilif.	floribunda f2
SH 51	Karatara: 90km from Oudtshoorn, 2km from Kaykoe turnoff	33° 46' 30"	22° 55' 00"	Chrys.	monilif.	pisifera pisifera f2
SH 52	George: Outeniqua pass layby.	33° 53' 30"	22° 23' 30"	Chrys.	monilif.	floribunda f2
SH 53	Mosselbay: layby above Brak River	34° 03' 00"	22° 14' 00"	Chrys.	monilif.	floribunda f2
SH 54	Mosselbay: after turnoff to MossellBay N2	54° 10' 34"	22° 07' 30"	Chrys.	monilif.	floribunda f1
SH 55	Herbertsdale: 30km East of Albertinia	34° 10' 30"	21° 57' 30"	Chrys.	monilif.	floribunda f1
SH 57	Riversdale: 19.5 km west of Albertinia	34° 10' 30"	21° 19' 00"	Chrys.	monilif.	floribunda f1
SH 58	Riversdale: 54km west of Albertinia btn Riversdale & Heidelberg	34° 05' 30"	21° 00' 00"	Chrys.	monilif.	floribunda f1
SH 59	Stormsvlei: 10km from Heidelberg, going to Barrydale	34° 03' 30"	20° 50' 30"	Chrys.	monilif.	pisifera angustifolia
SH 60	Warmwaterberg: Tradouw's pass, 4km into pass	33° 58' 30"	20° 42' 30"	Chrys.	monilif.	pisifera pisifera f2
SH 61	Warmwaterberg: Tradouw's pass, 9km in, past highpoint sign	33° 57' 00"	20° 42' 30"	Chrys.	monilif.	pisifera pisifera f2
SH 62	Swellendam: road to Capetown, Riviersonderend (184)	34° 06' 30"	20° 20' 30"	Chrys.	monilif.	monilifera
SH 65	Swellendam: 42km from Bredasdorp	34° 10' 30"	20° 19' 00"	Chrys.	monilif.	floribunda f1
SH 66	Bredasdorp: between Bredasdorp and Struisbaai	34° 37'	20° 04'	Chrys.	monilif.	floribunda f1
SH 68	Bredasdorp: Cape Agulhas (almost) Struisbaai, beach front	34° 49' 00"	20° 01' 30"	Chrys.	monilif.	floribunda f1
SH 70	Baardskeerdersbos: turn to Gansbaai, before cement road starts	34° 39' 30"	19° 31' 30"	Chrys.	monilif.	floribunda f1
SH 71	Baardskeerdersbos:13km east of Gansbaai, Pearly Beach turnoff	34° 37' 30"	19° 30' 00"	Chrys.	monilif.	floribunda f1
SH 72	Stanford: 10km east of Hermanus, just before Yaughtclub	34° 25' 00"	19° 24' 30"	Chrys.	monilif.	pisifera angustifolia
SH 73	Hermanus: Scenic drive, by cellphone mast in fynbos	34° 20'	19° 09'	Chrys.	monilif.	floribunda f1
SH 75	Kirstenbosch Gardens, outside seed office (origin unknown)			Chrys.	incana	incana
SH 77	Capetown: road out of CT, North of BigBay, Bloubergstrand	33° 47' 30"	18° 29' 30"	Chrys.	incana	incana
SH 79	Darling: Turn to Grotto Bay	33° 30' 00"	18° 19' 30"	Chrys.	incana	incana
SH 80	Darling :Turn to Grotto Bay	33° 30' 00"	18° 19' 30"	Chrys.	incana	incana

Voucher	Locality	South	East	Genus	Species	Subspecies
SH 86	Moravia: Piketberg, high up road, in fynbos	32° 50' 30"	18° 43' 30"	Chrys.	monilif.	monilifera
SH 87	Moravia: Piket-bo-berg pass, jnctn NewCaledon + MoutonsValley rds	32° 48' 30"	18° 43' 00"	Chrys.	monilif.	monilifera
SH 88	Gydo pass: Gydo pass to Ceres from Citrusdal (last 20km to Ceres)	33° 14' 30"	19° 20' 00"	Chrys.	monilif.	floribunda f1
SH 90	Barrydale: Tradouw's pass (2km in)	33° 55' 30"	20° 42' 30"	Chrys.	monilif.	pisifera angustifolia
SH 91	Ladismith: Seweweekspoort	33° 24' 30"	21° 24' 00"	Chrys.	monilif.	floribunda f1
SH 94	Kangogrotto: Swartberg pass, 12km into it, just before tar and dirt junction	33° 23' 30"	22° 6' 30"	Chrys.	monilif.	floribunda f1
SH 97	George: Montagu pass	33° 50' 30"	22° 20' 30"	Chrys.	monilif.	pisifera angustifolia
SH 100	Nature's Valley: Riverside, Nature's valley	33° 59' 00"	23° 33' 30"	Chrys.	monilif.	pisifera pisifera f2
SH 102	Nebo: Lekkerwater, De Hoop (1)	24° 58' 00"	29° 56' 30"	Chrys.	monilif.	floribunda f1
SH 103	Nebo: Lekkerwater, De Hoop (2)	24° 58' 00"	29° 56' 30"	Chrys.	monilif.	floribunda f1
SH 104	Nebo: Lekkerwater, De Hoop (3)	24° 58' 00"	29° 56' 30"	Chrys.	monilif.	floribunda f1
SH 105	Nebo: Lekkerwater, De Hoop (4)	24° 58' 00"	29° 56' 30"	Chrys.	monilif.	floribunda f1
SH 111	Joubertina: Joubertina	33° 49' 30"	23° 51' 30"	Chrys.	monilif.	pisifera
SH 112	Nature's Valley: Tsitsikama National Park, timberlot plantation	33° 58' 30"	23° 38' 30"	Chrys.	monilif.	floribunda f2
SH 120	Humansdorp: St Francis Bay, beachside	34° 12' 30"	24° 50' 00"	Chrys.	monilif.	rotundata
SH 122	Seymour: Seymour	32° 33' 30"	26° 46' 00"	Chrys.	monilif.	pisifera
SR 173	Humansdorp: Jeffries bay	34° 03' 00"	24° 55' 00"	Chrys.	monilif.	pisifera
SR 178	Joubertina: Storm's river	33° 58' 00"	23° 51' 30"	Chrys.	monilif.	pisifera pisifera f2
SR 179	Sedgefield: Sedgefield	34° 01' 00"	22° 48' 30"	Chrys.	monilif.	floribunda f2
SR 180	Wilderness: Wilderness	33° 59' 30"	22° 34' 30"	Chrys.	monilif.	floribunda f1
SR 181	Botterkraal: Noetsie	34° 00' 30"	22° 21' 30"	Chrys.	monilif.	floribunda f2
SR 184	Plettenberg Bay: Plettenberg Bay	34° 03' 00"	23° 21' 30"	Chrys.	monilif.	floribunda f2
SR 188	Witelsbos: Witelsbos	33° 59' 30"	24° 07' 00"	Chrys.	monilif.	pisifera pisifera f2
SR 189	Kareedouw: Kareedouw	33° 57' 00"	24° 17' 30"	Chrys.	monilif.	pisifera pisifera f2
CP 480	Kozi mouth	26° 00' 00"	32° 45' 00"	Chrys.	monilif.	rotundata
CP 485	Mhlosinga: Mabibi camp	27° 19' 00"	32° 43' 30"	Chrys.	monilif.	rotundata
CP 491	Umkomaas: Umkomaas	30° 11' 30"	30° 48' 00"	Chrys.	monilif.	rotundata
AW 20	Darling: Swartwater	33° 16' 30"	18° 16' 00"	Chrys.	incana	subcanescens
BSR1	Joubertina: Tsitsikama National Park, top of Storm's river	33° 58' 00"	23° 51' 30"	Chrys.	monilif.	pisifera pisifera f2
BSR2	Joubertina: Tsitsikama National Park	33° 59' 00"	23° 51' 00"	Chrys.	monilif.	pisifera pisifera f2
NPB 1815	Underberge: Drakensburg (low altitude)	29° 52' 30"	29° 10' 30"	Chrys.	monilif.	canescens
NPB 1818	Underberge: Drakensburg (high altitude)	29° 51' 30"	29° 11' 00"	Chrys.	monilif.	canescens
NPB 1820	Underberge: Drakensburg (Sonie pass)	29° 35' 00"	29° 17' 30"	Chrys.	monilif.	canescens
NPB 1830	Grahamstown: Featherstone Kloof	33° 20' 30"	26° 31' 30"	Osteosp.	junceum	
NPB 1831	Strydomsberg: Van Stadens	33° 35' 30"	25° 12' 30"	Chrys.	monilif.	pisifera
NPB 1832	Skoenmakerskop: Sardinia Bay	34° 02' 00"	25° 34' 00"	Chrys.	monilif.	rotundata
NPB 1833	Skoenmakerskop: SchoenmakersKop	34° 02' 30"	25° 31' 30"	Chrys.	monilif.	floribunda f1
NPB 1836	Great Fish Point: FishRiver	33° 30' 00"	27° 07' 30"	Chrys.	monilif.	rotundata
NPB 1877	Groot Kragga: Garcia's Pass	34° 01' 00"	21° 15' 00"	Chrys.	monilif.	floribunda f1
NPB 1879	Stanford: Shaw's pass between Caledon and Hermanus	34° 19' 00"	19° 25' 00"	Chrys.	monilif.	floribunda f1
NPB 1880	Capetown: Table Mountain (Cable way station)	33° 57' 00"	18° 27' 30"	Chrys.	monilif.	floribunda f1
NPB 1881	Capetown: Table Mountain (along contour road)	33° 56' 30"	18° 27' 30"	Chrys.	monilif.	monilifera
NPB 1882	Grahamstown: Signal Hill	33° 20' 30"	26° 31' 30"	Chrys.	monilif.	floribunda f1
NPB 1887	Franschoek: Top of Franschoek pass	33° 55' 30"	19° 08' 30"	Chrys.	monilif.	monilifera
NPB 1888	Villiersdorp: Theewaterskloof	33° 59' 30"	19° 17' 30"	Chrys.	monilif.	floribunda f1
MP 452	Tanzania: Kilimanjaro	04° 21' 00"	37° 55' 00"	Chrys.	monilif.	septentrionalis
TD 4442	Willowvale: Willowvale	32° 16' 00"	28° 31' 00"	Chrys.	monilif.	rotundata

Figure 2.2.1: Distribution map of samples sites. Numbers without initials were collected by SH.

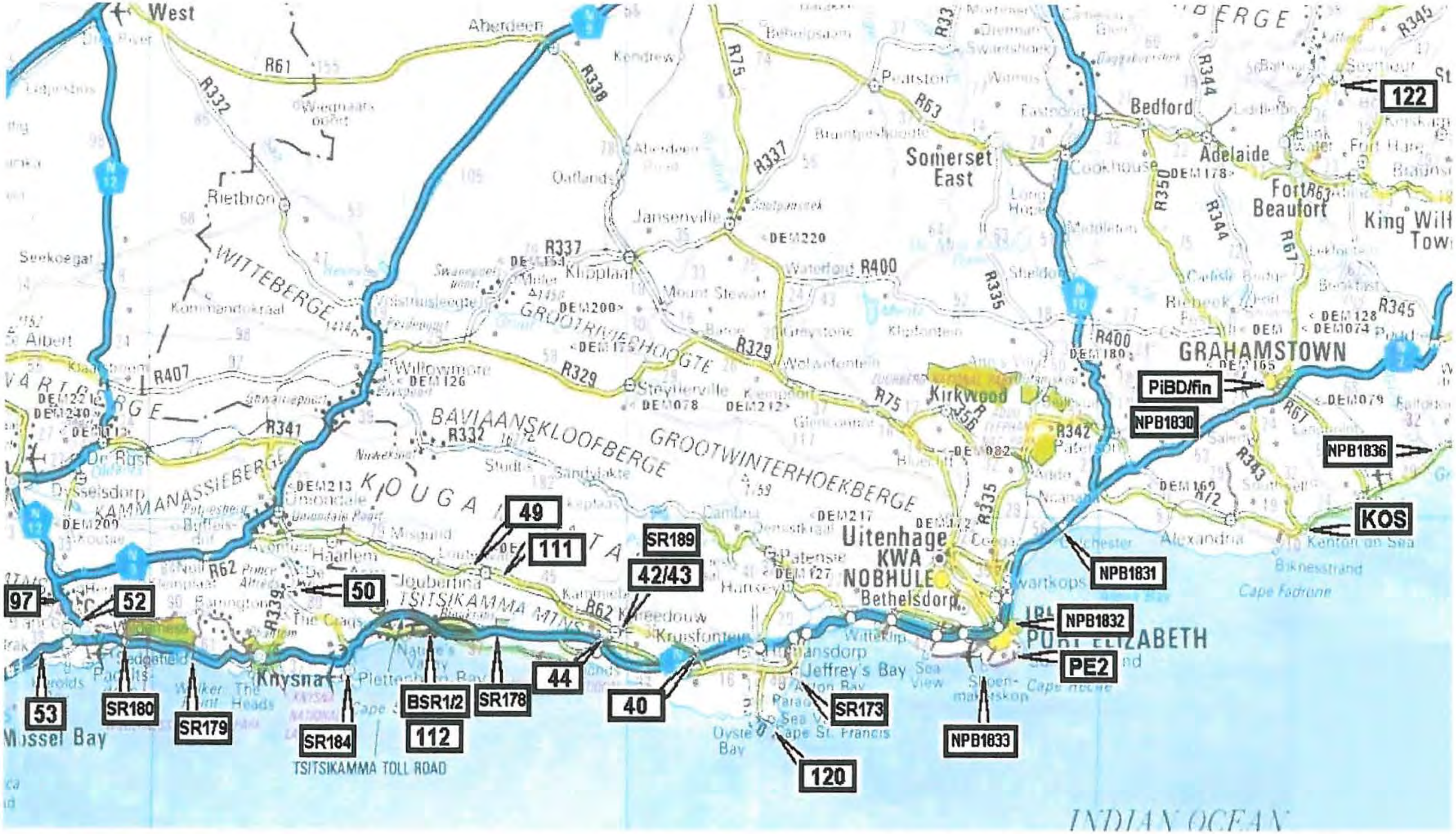

Figure 2.2.2: Distribution map of samples sites. Numbers without initials were collected by SH.

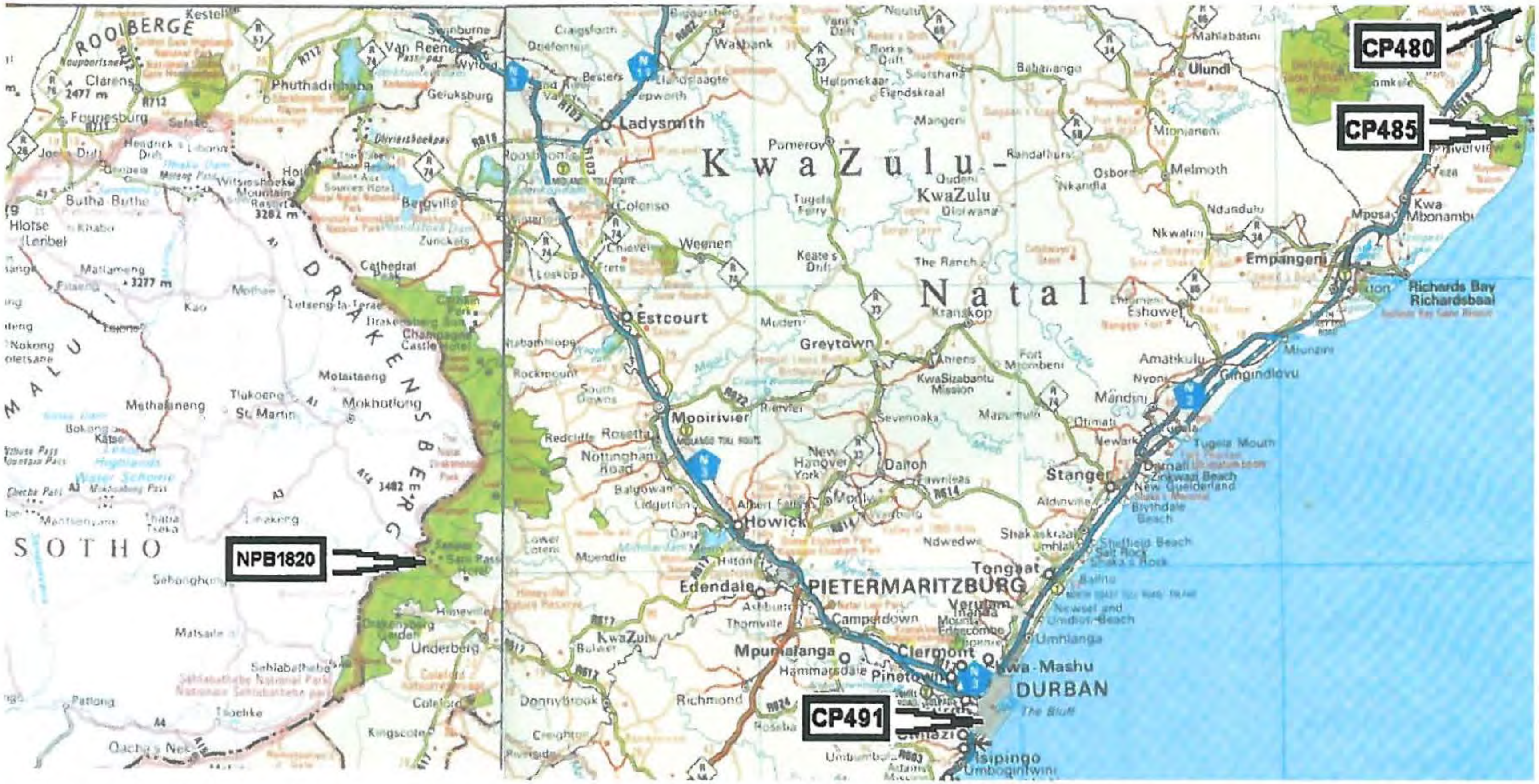

Figure 2.2.3: Distribution map of samples sites. Numbers without initials were collected by SH.

DNA extraction

All samples were extracted using a modified CTAB DNA extraction protocol (Doyle & Doyle 1987; Appendix 1).

PCR amplification and sequencing of ITS

PCR amplifications were conducted either on a ThermoHybaid PCRSprint Temperature Cycling System or a Corbett Research PC-960G Microplate Gradient Thermal Cycler using the following conditions: 95°C for 1 min, 52°C for 1 min and 72°C for 3 minutes repeated between 30 and 40 cycles (depending on the necessary amount of product needed to ensure clean product) with a 10 minute 72°C extension period at the end of the PCR program. The PCR reagents and their volumes are presented in Appendix 2.

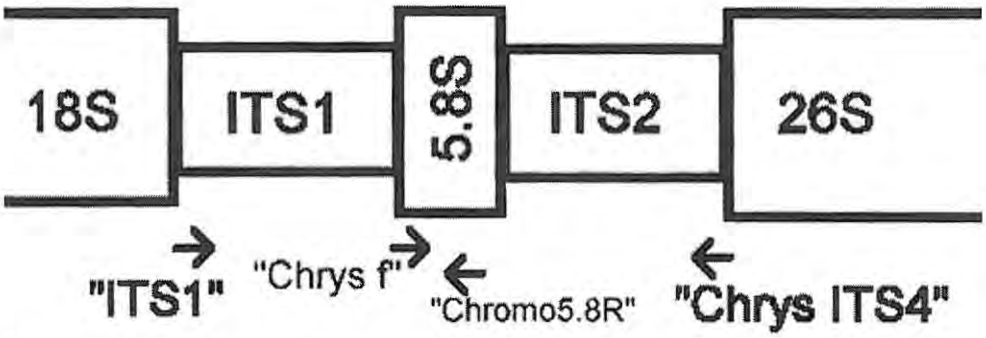

Figure 2.2.4: Diagrammatic representation of the ITS region, showing ITS1 and 2 and the bracketing spacer regions. Arrows indicate starting point and direction of primers (names in inverted commas; details of primers in Methods section and also in Table 2.2)

Table 2.2: Table of ITS primers (and their details: Reference, Length and Sequence) used in PCR and sequencing .

Name	Reference	Length	Sequence
ITS Primers			
ITS1	White *et al.* 1990	19	TCC GTA GGT GAA CCT GCG G
ITS4	White *et al.* 1990	20	TCC TCC GCT TAT TGA TAT GC
ITS5	White *et al.* 1990	22	GGA AGT AAA AGT CGT AAC AAG G
ITS-Chrys-4		20	TCC TCC GCT TAT GGA TAT GC
Chromo-5.8-R	Barker *et al.*, 2003	18	GAT TCT GCA ATT CAC ACC
Danth-5.8-F	Barker *et al.*, 2003	15	GAC TCT CGG CAA CGG
Chrysan-5.8-F		20	GAC TCT CGG CAA CGG ATA TC
Chrysan-i(1)		15	ACC A(A/G)A CAC GCA CAT
Chrysan-i(2)		15	ACC A(A/G)A CAC GCA CAC

ITS primers 4 and 5 (White *et al.*, 1990; Table 2.2) were initially used to amplify the ITS region of the nuclear genome. A primer was subsequently designed from these initial sequences that was specific for the *Chrysanthemoides* ITS 4 primer site (ITS-Chrys-4; see Table 2.2).

The PCR product was run on 1% agarose gels, which consisted of 0.5g agarose in 50ml TBE buffer (10.8g Tris(hydroxymethyl)aminomethane, 5.5g Boris acid and 0.93g EDTA made up to 1L with distilled water). Each gel contained 30µl ethidium bromide and the PCR products bands were visualised using a UV transilluminator.

A clean clear bright band was taken as a positive result. Any smearing indicated an unsatisfactory PCR result, and PCR conditions were altered as necessary to reduce smearing (e.g. a reduction in number of cycles; a reduction in quantity of primers used in the PCR reaction).

The PCR product was cleaned using either the QIAGEN QIAquick PCR purification kit or the PROMEGA Wizard SV Gel and PCR purification kit and resuspended in 30µl of dH_2O. The final product was checked for brightness (i.e. product concentration) by running 1µl of the product, with 5µl water and 5µl of a loading buffer (Bromophenol blue and xylene cyanol in glycerol) on a 1% Agarose gel that contained 15µl of ethidium bromide and visualised by means of a UV transilluminator.

Cleaned PCR product was sequenced using ABI prism BigDye Terminator v3.0 and v3.1 Ready Reaction Cycle sequencing kit (Applied Biosystematics) according to manufacturers instructions with the primers: "ITS1", "ITS-Chrys-4", "Chromo-5.8-R", "Chrysanth-5.8-F" (See Figure 2.2.4 and Table 2.2 for details).

The initial internal primer for sequencing ITS-2 was a primer designed for *Danthonia* ("Danth-5.8-F"; Barker *et al.*, 2003; Table 2.2), but this short primer did not produce clean sequences. For this reason a new longer primer was necessary and was thus designed for sequencing ITS-2 starting from the 5.8S region ("Chrys-5.8-F"; Table 2.2). The ITS-1 region was sequenced using a primer designed for *Chromolaena* ("Chromo-5.8-R"; Barker *et al.*, 2003; Table 2.2) and this primer rendered clean sequences and was thus used throughout the study.

Problems with multiple and different ITS paralogues in samples of C. incana

The ITS-1 region of the *C. incana* samples presented a problem for gathering sequence data. Once initial PCR products had been sequenced, it was discovered that the region has multiple and different copies (paralogues) of ITS, present in approximately equal proportions. This was evidenced by the fact that parts of the sequence trace files consisted of multiple peaks.

When there are two alleles present in a heterozygous individual, both alleles usually are amplified from a target locus, such that subsequent assays fail to distinguish between alternative genetic configurations possible for the two haplotypes when they differ at multiple nucleotide positions (Avise, 2000).

Schaal *et al.* (1998) maintain that for heterozygous individuals, the two alternative alleles must be analysed individually as Operational Taxonomic Units. Zhang & Hewitt (2003) suggest several alternative methods for haplotype determination in heterozygous individuals, including (i) cloning of PCR product; (ii) signal-intensity dependant inference (taking the less intense peaks as the less prevalent sequence) and (iii) allele-specific amplification. Initial attempts at cloning the problematic samples were made, but were unsuccessful. A subtractive method (described below) was initially used, but two allele specific internal primers were finally designed and used (See Appendix 4). Avise (2000) notes that despite the recent availability of these and other physical isolation procedures for individual haplotypes from diploid nuclear genes, few attempts as yet have been made to capitalise upon these approaches as a starting point for estimating intraspecific nuclear gene trees.

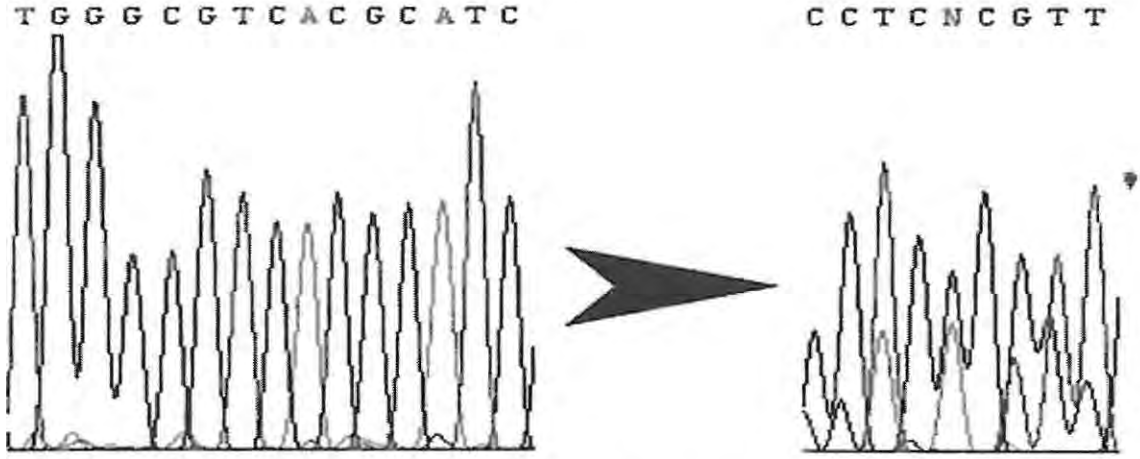

Figure 2.2.5: Examples from *C. incana* sequence trace file, sequenced with "Chrys-5.8-R" primer. Left = clean sequence trace, approximately position 100; Right = double signal trace, just after position 240.

In accordance with Schaal *et al.*'s (1998) statement concerning the analysis of each of the sequences separately, the separate paralogues were initially obtained by a process of comparing the *C. incana* trace files against a consensus sequence from other subspecies, where all the copies were homogenous. By this means, it was possible to deduce (by a process of subtraction) what the *C. incana* paralogues were. In Figure 2.2.5, an example of this technique could be used on the right hand trace file. If the consensus sequence from the other subspecies for this segment of data were CCACACGTT, then the complementary sequence would be GGTCCCTGG

```
GCACTTGTCGTCCTGTCC 80_2
GCACTT---GTCCTGTCC 80_1
GCACCAGTCGTCCTGTCC 78_2
GCACTT---GTCCTGTCC 78_1
GCACTTGTCGTCCTGTCC 75_2
GCACTT---GTCCTGTCC 75_1
GCACTT---GTCCTGTCC Cons. Seq.
```

Figure 2.2.6: Segment of ITS2 data, showing sequences 1 and 2 of three *C. incana* samples as well as a comparative consensus sequence with their three base insert.

Figure 2.2.6 shows three samples of *C. incana* where the duplicate sequence (sequence 2) has a three base insert that differentiates it from sequence 1 (identical to the consensus sequence for the rest of the subspecies).

An attempt at cloning the two paralogues was made (outsourced to the Rhodes University Microbiology, Biotechnology and Biochemistry department), using the Promega pGEM-T Easy Vector cloning kit, as per manufacturer's instructions.

Longer and clearer sequence data were obtained by means of two allele-specific internal primers (Chrys-I(1) and Chrys-I(2); Table 2.2). These primers were designed using the cleanly sequenced areas of ITS-1 (prior to the three base insert) in the *C. incana* samples. The two primers differed in their final base, utilizing the presence of nucleotide additivity at the final base position as a means to separate the two paralogues.

Three C. *incana* sequences (one paralogue per sample) were included in the "Complete" ITS data set (See *Construction of data sets* below) to investigate where in the ITS trees the C. *incana* samples would place. Further partial data for another five *C. incana* samples were obtained, but not included in the analyses.

PCR amplification and sequencing of cpDNA

Primers were screened across six samples (one from each of *C. m. pisifera, C. m. canescens, C. m. rotundata, C. m. floribunda*, C. *incana* and *Osteospermum junceum*) to determine if any intergeneric, inter- and intraspecific variability could be distinguished in each region.

Table 2.3: Table of cpDNA primers (and their details: Reference, Length and Sequence) used in PCR and sequencing for each region

Name	**Reference**	**Length**	**Sequence**
trnL-trnF			
Tab C	Taberlet *et al.*, 1991	20	CGA AAT CGG TAG ACG CTA CG
Tab F	Taberlet *et al.*, 1991	20	ATT TGA ACT GGT GAC ACG AG
Tab E	Taberlet *et al.*, 1991	20	GGT TCA AGT CCC TCT ATC CC
psbA-trnH			
psbA	Sang *et al.*, 1997	22	GTT ATG CAT GAA CGT AAT GCT C
trnH	Sang *et al.*, 1997	23	CGC GCA TGG TGG ATT CAC AAA TC
trnS-trnfM			
trnS	Demesure *et al.*, 1995	20	GAG AGA GAG GGA TTC GAA CC
trnfM	Demesure *et al.*, 1995	20	CAT AAC CTT GAG GTC ACG GG
rps16			
rps16F	Oxelman *et al.*, 1997	24	GTG GTA GAA AGC AAC GTG CGA CTT
rps16R2	Oxelman *et al.*, 1997	24	TCG GGA TCG AAC ATC AAT TGC AAC

trnS-trnfM

PCR temperature conditions were as for ITS, but only 30 cycles were used. The primers used to amplify this region of the chloroplast were "trnS" and "trnfM" (Demesure *et al.*, 1995; Table 2.3). PCR product was checked and cleaned as described in the methods for ITS above.

Cleaned PCR products were sequenced using ABI prism BigDye Terminator v3.0 and v3.1 Ready Reaction Cycle sequencing kit (Applied Biosystematics) according to manufacturers instructions with the primers "trnS" and "trnfM" (See Table 2.2 for details).

rps16

PCR temperature conditions were as for ITS, but only 30 cycles were used. The primers used to amplify this region of the chloroplast were "rps16F" and "rps16R2" (Oxelman *et al.*, 1997; Table 2.3). PCR product was checked and cleaned as described in the methods for ITS above.

Cleaned PCR products were sequenced using ABI prism BigDye Terminator v3.0 and v3.1 Ready Reaction Cycle sequencing kit (Applied Biosystematics) according to manufacturers instructions with the primers "rps16F" and "rps16R2" (See Table 2.2 for details).

psbA-trnH

PCR temperature conditions were as for ITS, but only 30 cycles were used. The primers used to amplify this region of the chloroplast were "psbA" and "trnH" (Sang *et al.*, 1997; Table 2.3). PCR product was checked and cleaned as described in the methods for ITS above.

Cleaned PCR products were sequenced using ABI prism BigDye Terminator v3.0 and v3.1 Ready Reaction Cycle sequencing kit (Applied Biosystematics) according to manufacturers instructions with the primers "psbA" and "trnH" (See Table 2.3 for details).

trnL-trnF

PCR temperature conditions were as for ITS, but only 30 cycles were used. The primers used to amplify this region of the chloroplast were "tab c" and "tab f" (Taberlet *et al.,* 1991; Table 2.3). PCR product was checked and cleaned as described in the methods for ITS above.

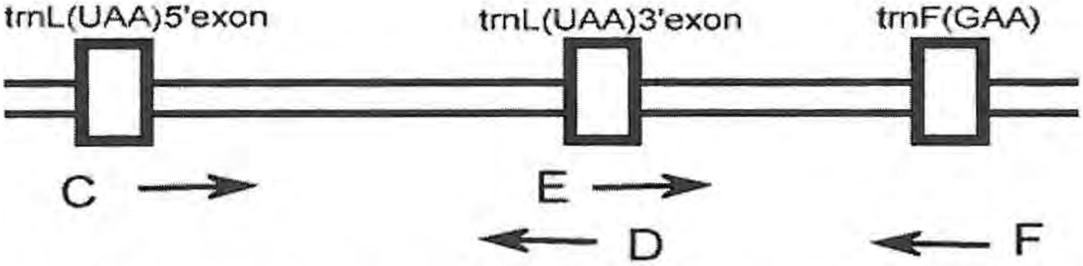

Figure 2.2.7: Diagrammatic representation of the trnL-trnF region, showing arrows indicating starting point and direction of primers (details of primers in Methods section and also in Appendix 4) (After Taberlet *et al.,* 1991).

Cleaned PCR products were sequenced using ABI prism BigDye Terminator v3.0 and v3.1 Ready Reaction Cycle sequencing kit (Applied Biosystematics) according to manufacturers instructions with the primers "tab c", "tab f" and occasionally "tab e", when "tab f" did not render a clean sequence (See Figure 2.2.7 and Table 2.3);

Sequence checking and alignment

Sequence data was checked and edited using SEQUENCHER (Version 3.1.1; Gene Code Corporation). Assembled sequences were exported from Sequencher, and imported into the alignment software DAPSA (version 4.0, 1996, E.H.Harley, Dept of Chemical Pathology, University of Cape Town) and aligned manually by eye.

Construction and analysis of data sets

The initial primer screening revealed extremely low rates of sequence divergence and insufficient variability in the *rps16* and the *trnS-trnfM* data sets to provide any useful data. It was thus decided that only the *psbA-trnH* and *trnL-trnF* regions would be studied.

There are a number of different data sets and combinations thereof. These include three plastid data sets (the *psbA-trnH* data, the *trnL-trnF* data, and the combination of these into a "cpDNA" data set) and three ITS data sets: (i) "Complete ITS", the full 78 sample data set including 75 *C. monilifera* and three *C. incana* sequences, (ii) "Monilifera ITS", the 75 samples *C. monilifera* data set, and (iii) "Limited ITS", the 18 taxa smaller data set for comparative purposes with the cpDNA data.

Table 2.4: Summary table of molecular data sets, the number of taxa in each and the analyses methods used on each. (Pars. = Parsimony analysis; NJ = Neighbor Joining; ML = Maximum Likelihood; Stat. Pars. = Statistical parsimony).

		Analyses			
Data set name	**No. Taxa**	**Pars.**	**NJ**	**ML**	**Stat. Pars.**
Complete ITS	78	YES	YES	YES	NO
Monilifera ITS	75	YES	YES	YES	NO
Limited ITS	18	YES	YES	YES	NO
psbA-trnH	18	YES	YES	YES	YES
trnL-trnF	18	YES	YES	YES	YES
cpDNA	18	YES	YES	YES	YES

Phylogenetic analyses

Several approaches have been developed for the purpose of analysing molecular sequence data as an aid to phylogeny reconstruction and these generally fall into three classes: parsimony, distance methods and maximum likelihood methods (Olmstead and Palmer 1994). A fourth method, statistical parsimony, produces node-and-line haplotype networks, instead of trees, and considers sequence data from the perspective of haplotypes, rather than individual characters or sequences (i.e. two samples can have a shared haplotype, in which case the haplotype represents both of them together).

Parsimony analysis

Parsimony has become the method of choice for most molecular phylogenetic studies in plants (Olmstead and Palmer, 1994). Parsimony is based on the assumption that the most likely tree is the one that requires the fewest number of changes to explain the data in the alignment (Hall, 2001). The basic premise of parsimony is that taxa sharing a common characteristic do so because they inherited that characteristic from a common ancestor. Any conflict that occurs with that assumption is explained by mean of homoplasy ("extra steps" or hypotheses that are required to explain the data). Homoplasic characters are said to arise in several different ways: (i) reversal (a characteristic that changed and then reverted back to its original state), (ii) convergence (unrelated taxa evolving the same characteristic independently) or (iii) parallelism (different taxa may have similar embryological mechanisms that predispose a character to develop in a certain way; Hall, 2001).

Parsimony (or "minimum change") is the criterion for selecting that tree or trees that minimise the number of evolutionary changes (steps), including homoplasies, required to explain the data. In the case of nucleotide sequences, the data are the aligned sequences. Each site in the alignment is a character, and each character can have four different possible nucleotide states (A/T/C/G) in different taxa (there can also be a "-" indicating a gap, corresponding to a deletion in that sample, or an insertion in other the samples in the alignment). Invariant characters (those that are the same in all the taxa) and autapomorphic characters (those that occur in only one taxon) are ignored in a parsimony analysis. Parsimony analysis may result in several trees of equal length differing only slightly in topology. Parsimony analysis may result in several trees, typically differing only slightly, that are consistent with the same number of events and are therefore equally parsimonious. This result offers the advantage of finding equally (and nearly equally) parsimonious trees, and thus enabling the examination of other basically equivalent reconstructions, and also the testing of testing alternative (or forced) topologies to evaluate the cost of other phylogenetic hypotheses (Olmstead and Palmer, 1994).

Because ITS is a biparentally inherited multicopy gene, there may sometimes be two or more base-calls for bases occurring in each of the different paralogues at the same position in the sequence within a single specimen (i.e. at one position, the chromatograph may show more than one peak). PAUP* allows for the use of ambiguous base-calls in sequence analysis and this facility was used (e.g. if there are two equally tall peaks at one point, one indicating "T" and the other "C", then the ambiguous base coded as "Y").

As noted by Maddison (1991) in a set of most-parsimonious trees for a data matrix, there may be several distinct classes (islands) of trees. An island consists of interconnected parsimonious trees (all less than a certain length), with each tree within an island differing from each other by only a single rearrangement of branches (Maddison, 1991). Trees in different islands may have different implications for character evolution and each run of a random input heuristic search will typically find only one island (Maddison, 1991). Data sets with low retention indices are ideal candidates for multiple islands, especially if they contain more than 21 terminal taxa (Maddison, 1991). While the largest ITS data set ("Complete ITS") certainly has more than 21 terminal taxa (78 to be precise), it does not display low retention indices. For completeness sake, the largest ITS data set (Complete) was used in a random input analysis, performed to investigate if there were multiple islands of equally most parsimonious trees. A HEURISTIC search was conducted on the trees found by this method. For the analyses of the larger datasets, MAXTREES was set to 5000. Strict consensus trees were produced from the set of equally most parsimonious trees obtained. Bootstrap support values were calculated for 100 replicates with (MAXTREES = 1000).

Treatment of gaps

Different methods of treating gaps in analyses have been shown to influence the resulting phylogenetic hypothesis (Simmons & Ochoterena, 2000, and papers listed therein). Treatment of gaps can vary widely, from secondarily mapping gaps onto the tree inferred from base characters alone, to treating all gaps as separate characters or character states (Simmons & Ochoterena, 2000). Putatively homologous gaps (those with identical 5' and 3' termini) are treated differently to those with different 5' and/or 3' termini, as at least one indel event must be postulated to transform one gap into another (Simmons & Ochoterena, 2000). Gaps are often not included in phylogenetic analyses, because presumed identical gaps may in fact have multiple origins in unrelated taxa (Johnson & Soltis, 1995; Simmons & Ochoterena, 2000). Others, however, consider gap characters better than substitution characters, as it is unlikely that indels would be repeated in the exact same position, with the same length (and especially sequence, in the case insertions) unlike base substitutions, which can be changed to a specific base multiple independent times (Lloyd & Calder, 1991). Gaps have been found to be good characters in coding and noncoding regions (Simmons & Ochoterena, 2000) and have been considered by some researchers to be

> "excellent markers to identify monophyletic groups" (van Dijk *et al.*, 1999).

Gaps can be coded either as 5th character states for nucleotides or as separate presence/absence characters. These two treatments are not identical to each other in practise, and treating gaps longer than one base as 5th states for each position treats each adjacent position as though they were independent, despite being putatively caused by a single indel event (Eernisse & Kluge, 1993). To code gaps as separate characters, an extra presence/absence character for every gap is added to the data matrix. The corresponding aligned positions in the sequence where the gap occurs are then coded as inapplicable. Problems with this approach have been noted by Maddison (1993), who makes mention of potential problems in the placement of terminals on trees with missing values. For sequence-based analysis, this optimization artefact occurs only when a paraphyletic group (for which the gap is present) separates two groups that share one or more bases, and when two or more bases are present in at least one of the groups (Maddison, 1993).

For each data set, three analyses were performed: one where gaps were considered as a 5th character state, one where gaps were coded manually as binary characters, and an analysis based exclusively on the substitution data.

Distance methods: Neighbor-Joining

Neighbor Joining (NJ) is a distance method (whereas Parsimony and Maximum Likelihood (ML) are character-based methods) and distance based methods convert the aligned sequences into a distance matrix of pairwise differences (distances) between the sequences. Hall (2001) summarises Neighbor Joining thus: NJ manipulates a distance matrix, reducing its size in each step, and then reconstructs the tree from that series of matrices. From the original matrix, it first calculates the net divergence of one taxon from all other taxa as the sum of the individual distances from the taxon. It then uses that net divergence to calculate a corrected distance matrix. NJ then finds the pair of taxa with the lowest corrected distance and calculates the distance from each of those taxa to the node that joins them. A new matrix is then created in which the new node is substituted for those two taxa. NJ does not assume that all taxa are equidistant from the root (unlike UPGMA). NJ, like Parsimony, is a minimum-change method, but it does not guarantee finding the tree with the smallest overall distance.

It should be noted that distance matrix methods can handle larger data sets in less computer time than can parsimony methods, suggesting that distance matrix methods will continue to be valuable for analysing very large data sets. Studies have shown that distance matrix methods that do not assume uniform rates (i.e. not UPGMA) and parsimony methods both perform well at reconstructing phylogenies using DNA sequences (Olmstead and Palmer, 1994).

For the purposes of the Neighbor Joining tree construction, the Jukes-Cantor (Jukes and Cantor, 1969) model of DNA sequence evolution was used. This model is the classic null model for DNA sequence evolution and assumes that all base substitutions are equally likely and that all positions in a sequence have equal expected rates of substitution (Huelsenbeck & Kirkpatrick, 1996). Investigations into the accuracy of the Neighbor Joining method for large data sets have found that the similarity between the true tree and reconstructed trees were extremely high, even when large numbers of taxa were involved (up to 150; Strimmer & von Haeseler, 1996).

In this study, PAUP was used to conduct Neighbor Joining analyses, with distances calculated using the Jukes-Cantor method. Missing data was ignored for pairwise comparisons, and negative branch lengths were set to zero. For each of the analyses, a Jukes-Cantor Neighbor-Joining distance tree was plotted, and Bootstrap support values were calculated from 100 replicates, and then transferred onto the NJ tree.

Maximum likelihood analysis

Maximum likelihood is a statistical procedure that estimates the likelihood of a hypothesis given the data and a deterministic model. As before, the data are the aligned nucleotides. The number of possible trees for larger number of taxa grows at a rapid rate, and as is becomes impossible to evaluate all trees, a heuristic search method must be employed to seek the most likely tree. Maximum likelihood almost always produces a single tree. The advantages of the method are that it allows users to specify the evolutionary model they want to use, and that the likelihood of the resulting tree is known. A disadvantage is that the process is considerably slower than either Parsimony or Neighbor Joining, and can be limited by the computing capacity of the computer used to run the search (Hall, 2001).

Because ML is based on explicit models of DNA evolution, the ModelTest (version 3.04, Posada & Crandall, 1998) software was used to identify the model of DNA substitution that best fit the data. For each data set, a Maximum Likelihood tree was plotted, and a Bootstrap support tree (100 replicates) was also obtained.

Statistical Parsimony

Statistical parsimony estimates haplotype relationships with accompanying confidence bounds on the pairwise connections (Crandall & Templeton, 1996). The method has its greatest statistical power when there are few differences and many similarities between haplotypes (Crandall & Templeton, 1996), a situation that can lead to a lack of resolution in other phylogenetic reconstruction methods (such as Parsimony or ML). Crandall (1994) has shown that the method outperform parsimony when few characters are available to differentiate haplotypes, leading to a more accurate estimation of phylogenetic relationships for data with low levels of divergence. This approach also allows for uncertainty in the cladogram estimation and, therefore, it does not rely on a single estimate of phylogenetic relationships, but is robust over a set of plausible alternative phylogenies (Templeton & Sing, 1993).

The probability of parsimony (as defined in Templeton *et al,* 1992) is calculated for DNA pairwise differences until the probability exceeds 0.95. The number of mutational differences associated with the probability just before this 95% cut-off is then the maximum number of mutational connections between pairs of sequences justified by the "parsimony" criterion (beyond this cut-off point, networks become disconnected). The gene trees are displayed as linked networks of ovals (haplotypes) or rectangles (putative basal or primitive haplotypes), linked by lines with small dots (representing missing haplotypes, i.e. absent sequences that differ by one substitution).

Due to the multicopy nature of ITS, and the presence of ambiguous base-calls in the ITS data, the TCS (Version 1.13; Clement *et al.*, 2000) software could not be used to plot gene trees for ITS. It could however be used to analyse the three cpDNA data sets.

Separate versus combined analysis

Character sampling for phylogenetic analysis involves two important elements: whether characters are independent, and whether there are enough of them (Olmstead and Palmer, 1994). Sufficient character sampling is necessary not only to achieve resolution, but to have enough characters for synapomorphies to outweigh the inevitable randomly occurring homoplasies (Olmstead and Palmer, 1994). Increasing the number of characters can reveal a consistent phylogenetic signal that maybe hidden by the noise in each individual data set (Santiago-Valentin & Olmstead, 2003).

Concern about non-independence of characters within data sets is a primary reason to gather data from two or more independent sources (e.g. cpDNA vs nuclear data) and to analyse them separately (Olmstead and Palmer, 1994). (Obtaining characters from independent sources also is valuable for increasing the amount of evidence brought to bear on a particular question). Although heterogeneity present in the data sets may influence phylogenetic reconstruction and combining data might be misleading (Santiago-Valentin & Olmstead, 2003). This is a common view and Morgan (1997) notes that the different inheritance patterns of the nuclear and cytoplasmic DNA sources of the two data sets do not support combining them, especially if there is evidence of hybridization.

Other researchers note that many recent studies have indicated that combined molecular data sets using regions with different levels of variation has provided resolution at different areas of cladograms, and phylogenetic resolution and levels of support are improved by directly combining independent molecular data sets (Chase & Cox, 1998; Soltis *et al.*, 1998; Whitten *et al.*, 2000). The combined use of nuclear and cpDNA sequences also increases the likelihood of inferring a species tree instead of just a gene tree, which would allow any possible assessment of the role of hybridization in the evolution of a species or genus (Soliva *et al.*, 2001).

There are two opposing approaches to the utilization of multiple data sets in phylogenetic analysis: (i) Total evidence: which requires that all the data be pooled into one large data set and analyses as one; and (ii) Congruence: trees derived from each separate data set should be compared to determine congruence. 'Total evidence' has been defined as the analysis of an unpartitioned body of evidence, ideally all the data available for a group of terminal taxa, that is characterised as seeking a single, best-fitting hypothesis, which in cladistics involves maximising character congruence (Williams, 1994). 'Taxonomic congruence' (also referred to as 'consensus'), the competing approach, seeks a consensus of hypotheses obtained from different data sets (Williams, 1994).

The advantages and disadvantages of each approach can depend on the heterogeneity of the data across multiple data sets. If data heterogeneity is rare, then the total evidence approach can lead to the best results, and if data heterogeneity is common, then combining all the data sets can lead to an erroneous answer that can obscure phylogenetic signal (Huelsenbeck, *et al.*, 1996). Proponents of total evidence hold that all of the independent characters available to the systematists should be combined and then analysed using parsimony. An extension of that argument says that all of the available taxa should be combined (both living and fossil) in a phylogenetic analysis (Huelsenbeck, *et al.*, 1996).

Total evidence is often promoted because it is said to maximise the informativeness and explanatory power of the character data used in the analysis (Huelsenbeck, 1996). Simulation studies have shown that a greater number of characters translate into greater accuracy under a wide variety of circumstances (de Queiroz *et al.*, 1995). Enhancing the detection of this signal may result in restructuring such that the combined tree contains groups not found in any of the trees generates from separate analyses of the different data sets and the resolution of the character conflict embodied in the combined tree may represent a better estimate of the phylogeny than would even a fully resolved consensus tree (de Queiroz, 1993).

Consensus is usually employed when an investigator either has both molecular and morphological data sets, or has different gene-, rRNA-, or protein-sequence data sets available (Eernisse and Kluge, 1993). de Queiroz *et al.* (1995) note that choosing to analyze data sets separately does not necessitate the use of consensus trees, and instead separate analysis may be seen as a means of exploring possible disagreements between data sets. One argument in favor of separate analysis concerns the ability to use areas of disagreement that may highlight conflicts caused by natural selection, hybridization or lineage sorting (De Queiroz *et al.*, 1995) as the use of both nrDNA and cpDNA could in this case. Mason-Gamer *et al.* (1996) echo this sentiment when they note that the decision to keep sets separate generally reflects a hypothesis that either (1) different evolutionary processes are acting on different data sets or portions of data sets, or (2) different data sets reflect different phylogenetic histories. Other advantages attributed to consensus methods include the equal weighting they give to each data set, thus reducing the potential for data sets with relatively large numbers of characters to swamp data sets with fewer characters, and the conservative estimate of phylogenies they are thought to give (de Queiroz, 1993). Arguments against consensus argue that giving data sets equal weight results in an arbitrary differential weighting of characters; that consensus trees do not necessarily indicate the most-parsimonious pattern of character change; and that consensus trees can actually contradict combined trees (de Queiroz, 1993). Separate analysis also does not distinguish between cases where combining partitions can aid in phylogenetic signal amplification, or where it can hinder phylogenetic analysis (Huelsenbeck, *et al.*, 1996). Separate analysis can also lead to more variable phylogenies from each data set, and it can be more difficult to find the same taxonomic groupings from partitioned data sets (Huelsenbeck, *et al.*, 1996). Another problem associated with consensus approaches is the inability of consensus to evaluate relative support for conflicting groups among the trees summarised (Nixon *et al.*, 1996).

Instead of choosing between the two approaches, Huelsenbeck, *et al.* (1996) advocate the idea of conditional combination, which they note sits midway between the two extremes, as it prevents combination of heterogeneous data sets when partitions are clear, as well as allowing for interpretation of evolutionary processes when data sets are determined to be inappropriate for combination.

Testing for incongruence

The incongruence length difference test (ILD; Mickevich & Farris, 1981) with multiple randomizations as described by Farris *et al.* (1994) was used to test for data set incongruence. If two data sets are congruent (i.e. they track the same underlying history) then it should not matter which characters are assigned to which data set, the resulting trees should remain the same (Kellogg *et al.*, 1994). The ILD test first calculates the sum of the lengths of the two original trees, then all the characters are randomly partitioned into two new data sets (of the same size at the original two), and the two most parsimonious trees are computed and their lengths summed (Kellogg *et al.*, 1994). This is repeated to generate a distribution of the sums of tree lengths, and if the total length of the observed data falls within the distribution of the random data sets, then the division of the two data sets is inferred to be arbitrary, and there is no significant difference (incongruence) between them (Kellogg *et al.*, 1994). Otherwise, if the total length of the observed data sets falls outside of 95% of the randomly partitioned data, the data sets are considered to be significantly incongruent. The ILD test is implemented in PAUP* and was used to investigate possible incongruence between the *trnL-trnF*, *psbA-trnH* and "Limited ITS" data sets.

Results

Initial plastid primer screening results

Initial primer screening of the samples from *C. m. pisifera, C. m. canescens, C. m. rotundata, C. m. floribunda*, C. *incana* and *Osteospermum junceum* showed 2/523 variable bases for *rps16* and 7/509 variable bases for *trnS-trnfM*; and of these variable bases, none were phylogenetically informative (see Appendix 3 for sequence alignments). Initial attempts at the sequencing of the *C .m. canescens* and *C. incana* samples with the *rps16* primers, and the *C. m. floribunda* and the *O. junceum* samples with the *trnS-trnfM* primers, were unsuccessful and the extremely low level of variability observed in the initial sequences from both regions did not warrant further attempts to utilize these sequences. Thus these two chloroplast regions were not pursued for further investigation.

The sequences from the *psbA-trnH* and the *trnL-trnF* regions showed sufficient variable characters to warrant further investigation (17/540 and 7/660 respectively, of which more than 50% were phylogenetically informative, including several phylogenetically informative multibase gaps in the *psbA-trnH* region). The higher percentage of informative characters in the *trnL-trnF* regions (relative to *rps16* and *trnS-trnfM*, both of which had no informative characters), combined with the ease of sequencing the region, and its universal usage prompted continued usage of this region.

Random input analysis

The random input analysis conducted on the "Complete ITS" data set did not discover multiple islands of most parsimonious trees. The tree that resulted from the analysis was identical in topology to the tree in Figure 2.3.1.

Sequence diversity

Table 2.3.1 provides the number of total characters, variable characters and potentially parsimony-informative characters for each region, when gaps are considered as missing data, as binary characters or as a 5th state. Sequences for the "Complete ITS", "psbA-trnH" and "trnL-trnF" data sets can be found in Appendix 3. The ITS-1 region was 257-264 bp long, the ITS-2 region was 224-230 bp long and the 5.8s rRNA gene was 165 bp long (identified by comparison to the Asteraceae ITS sequences of Goertzen *et al.*, 2003).

Table 2.3.1: Summary statistics of all data sets analysed using parsimony, and the resulting tree statistics.

DNA region	No. taxa	No. chars	Gaps as:	Var Chrs	Inf chrs	Length	Ci	Ri	No. trees
Complete ITS	78	754	missing	137	51	144	0.924	0.962	5000
			binary	144	57	154	0.903	0.957	5000
			5th	171	67	180	0.900	0.970	5000
Monilifera ITS	75	754	missing	101	41	124	0.927	0.967	5000
			binary	88	45	132	0.902	0.959	5000
			5th	121	51	149	0.893	0.971	5000
Limited ITS	18	754	missing	98	20	31	0.839	0.898	6
			binary	87	26	37	0.784	0.862	6
			5th	117	34	53	0.849	0.909	6
psbA-trnH	18	568	missing	21	8	8	1	1	1
			binary	26	13	15	0.933	0.985	5
			5th	59	44	46	0.957	0.991	10
trnL-trnF	18	684	missing	15	5	9	0.667	0.870	42
			binary	19	7	11	0.727	0.900	94
			5th	19	7	11	0.727	0.900	94
cpDNA	18	1252	missing	36	13	19	0.737	0.922	168
			binary	43	20	28	0.796	0.937	1927
			5th	78	51	59	0.881	0.972	1927

Table 2.3.2: Details on Insertions (number x length), deletions (number x length), transitions (A ↔ T; C ↔ G) and transversions (all other single base changes) for each of the small data sets.

DNA region	Insertions	Deletions	Transitions	Transversions
Limited ITS	6 x 1-base 1 x 3-base	5 x 1-base 1 x 3-base 1 x 8-base	17	45
psbA-trnH	2 x 1-base 1 x 14-base	2 x 1-base 1 x 4-base 1 x 6-base 1 x 10-base	9	12
trnL-trnF	2 x 1-base	2 x 1-base	10	7
cpDNA	4 x 1-base 1 x 14-base	4 x 1-base 1 x 4-base 1 x 6-base 1 x 10-base	19	19

Table 2.3.2 provides the numbers of indels and base changes in the "Limited ITS", "psbA-trnH", "trnL-trnF" and "cpDNA" data sets for comparative purposes (the samples are from the same plant in data set each, allowing for a comparison between each sequenced region). In the cpDNA, the numbers of transitions and transversions are nearly equal, while the number of transversions in the ITS data set is much higher than the number of transitions. There is also a higher incidence of multi-base deletions in the cpDNA than in the ITS data.

Table 2.3.3: Numbers and percentages of variable and potentially parsimony-informative characters per each data set.

DNA region	Total No. chars	Gap coding	No. variable Characters	% variable	No. pars. Inform. characters	% informative
Complete ITS	754	missing	137	18.2%	51	6.7%
		binary	144	19%	57	7.5%
		5th	171	23%	67	8.8%
Monilifera ITS	754	missing	101	13.3%	41	5.4%
		binary	88	11.7%	45	6%
		5th	121	16%	51	6.7%
Limited ITS	754	missing	98	13%	20	2.6%
		binary	87	11.5%	26	3.5%
		5th	117	15.4%	34	4.5%
psbA-trnH	568	missing	21	3.7%	8	1.4%
		binary	26	4.6%	13	2.3%
		5th	59	10.4%	44	7.7%
trnL-trnF	684	missing	15	2.2%	5	0.7%
		binary	19	2.8%	7	1%
		5th	19	2.8%	7	1%
cpDNA	1252	missing	36	2.9%	13	1%
		binary	43	3.4%	20	1.6%
		5th	78	6.2%	51	4.1%

From Table 2.3.3 above, it is evident that the ITS region as a faster rate of mutation than the chloroplast genes (as evidenced by the much higher number of variable characters for the same set of taxa, "Limited ITS" set versus the cpDNA sets). When the gaps are considered as a 5th state, the combined chloroplast data has almost the same percentage of potentially parsimony-informative characters at the equivalent set of "Limited ITS" data (4.1% vs. 4.5% respectively). However, when gaps are considered as missing data (i.e. are ignored) then the "Limited ITS" data has more than twice as many informative characters than the combined cpDNA data (2.6% vs. 1% respectively). When gaps are considered as binary characters, the number of informative characters in the "Limited ITS" data set is intermediate between the number of informative characters from the other two gap coding methods. This is due to the lengths of the ITS indels, with nearly equal numbers of both single base indels and longer multibase indels (refer to the "Limited ITS" data set in Table 2.3.2).

In the cpDNA data, there are twice as many variable characters when gaps are included (rather than ignored), and four times as many informative characters. When gaps are considered as a 5th state, the *psbA-trnH* region has more than twice as many variable characters (10.4% vs. 3.7%) and nearly six times more parsimony-informative (7.7% vs. 1.4%) than when those gaps are considered as missing data. The *trnL-trnF* region however, has far less informative gap characters (2.8% vs.2.2%) and the number of potential parsimony-informative character differ by only two between the two coding methods. When gaps are considered as binary characters, the *psbA-trnH* region shows far fewer informative characters; this is due to the length of the gaps in the *psbA-trnH* region, as coding gaps as 5th state characters gives each gap a much higher weighting than is desirable, as each base in the gap is considered as a separate mutational event, when in fact, a single deletion event can cause a large deletion. The *trnL-trnF* region, however, shows no difference in the number of informative characters when gaps are considered as 5th state characters or binary characters, and this is because the gaps in the *trnL-trnF* region are all single base indels, and the different coding techniques result in identical weightings.

In the "Limited ITS" data, the difference between the various gap handling methods is much less: 15.4% vs. 13% ("5th state" vs. "missing" gap coding respectively) for variable characters and 4.5% vs. 2.6% ("5th state" vs. "missing" gap coding respectively) for informative characters. The numbers of variable characters in the ITS region when gaps are considered as binary characters is closer to the number when gaps are considered as missing data, than when gaps are considered as 5th state characters (11.5% and 13% for "binary" and "missing" gap coding respectively vs. 15% for "5th state" gap coding).

Despite the fact that the *trnL-trnF* region is 116 bases longer, the *psb*A-*trn*H region has more variable characters than the *trnL-trnF* region (21 vs. 15 characters), especially when the gaps are included, considered either as a 5th state (59 vs. 19 characters) or considered as binary characters (26 vs. 19). The excessive weighting that 5th state gap coding gives to longer gaps can be clearly illustrated by the percentage of variable characters in each region that were informative: 75% of the *psbA-trnH* variable characters were informative when the longer gaps in this region were considered as 5th state characters, and when gaps were ignored only 38% of the variable characters were informative. In contrast, 37% of the variable characters in the *trnL-trnF* region were informative when the single-base gaps in that region were included, and 33% were informative when gaps were ignored.

As noted above, the higher percentage of informative characters when gaps are considered as 5th state characters (as opposed to missing or binary coding) is an artefact of the length of the gaps in combination with the pitfalls of that particular gap coding method. Never the less, there is still a slightly higher percentage of potentially parsimony-informative characters when gaps are included in the analyses.

Phylogenetic trees and haplotype networks

The trees plotted on the following pages have the sample number and the subspecies morphological classification of the plant that the samples were collected from. Any sample that lacks collector's initials before it (i.e. CP, NBP, MP, AW, BSR or SR) was collected by myself (SH). The bootstrap values for the Neighbor Joining trees have been enclosed in boxes to make them easier to see amongst the branches. Grey blocks highlight clades that are named on the tree and referred to in the text.

The sample numbers for the haplotype networks have had the initials of the collector (i.e. CP, SR or NBP) removed, due to space constraints of fitting labels within the nodes of the networks. The samples are the same samples used for the cpDNA analyses and the sample numbers in their entirety can be found in those trees.

Analyses of the "Complete ITS" data set

Parsimony analysis

Figure 2.3.1 shows the strict consensus tree following analysis of the "Complete ITS" data set, with gaps considered as missing data, as binary characters or as a 5th character state (analyses were terminated after 5000 trees were obtained, and then these trees were searched to completion). There is little resolution to the tree at the subspecies level, and a rather comb-like topology. A few things should be noted though: (i) the *C. m. septentrionalis* sample comes out as basal to the rest of the genus and (ii) the *C. m. canescens* samples are well supported as a group (100%) and also come out basal to the rest of the *C. monilfera* clade, but above the *C. m. septentrionalis* sample, (iii) the *C. incana* samples form a monophyletic clade placed amongst the *C. m. rotundata* and *C. m. pisifera* samples. This placement renders *C. monilifera* paraphyletic with regards to species definitions.

The bootstrap support at the basal areas of the tree are higher when gaps are considered as a 5th state, than if they are considered as missing data, indicating that the gap data are phylogenetically informative. The drop in bootstrap support for the "FPM" branch (equivalent in composition to the "West" branch in the Neighbour Joining tree, Figure 2.3.2) when gaps are considered as missing data or considered as binary characters is most likely due to a eight-base deletion that is present in the samples of the "FPM" clade. The difference in bootstrap support between the two gap coding methods ("5th character state" and "binary character")

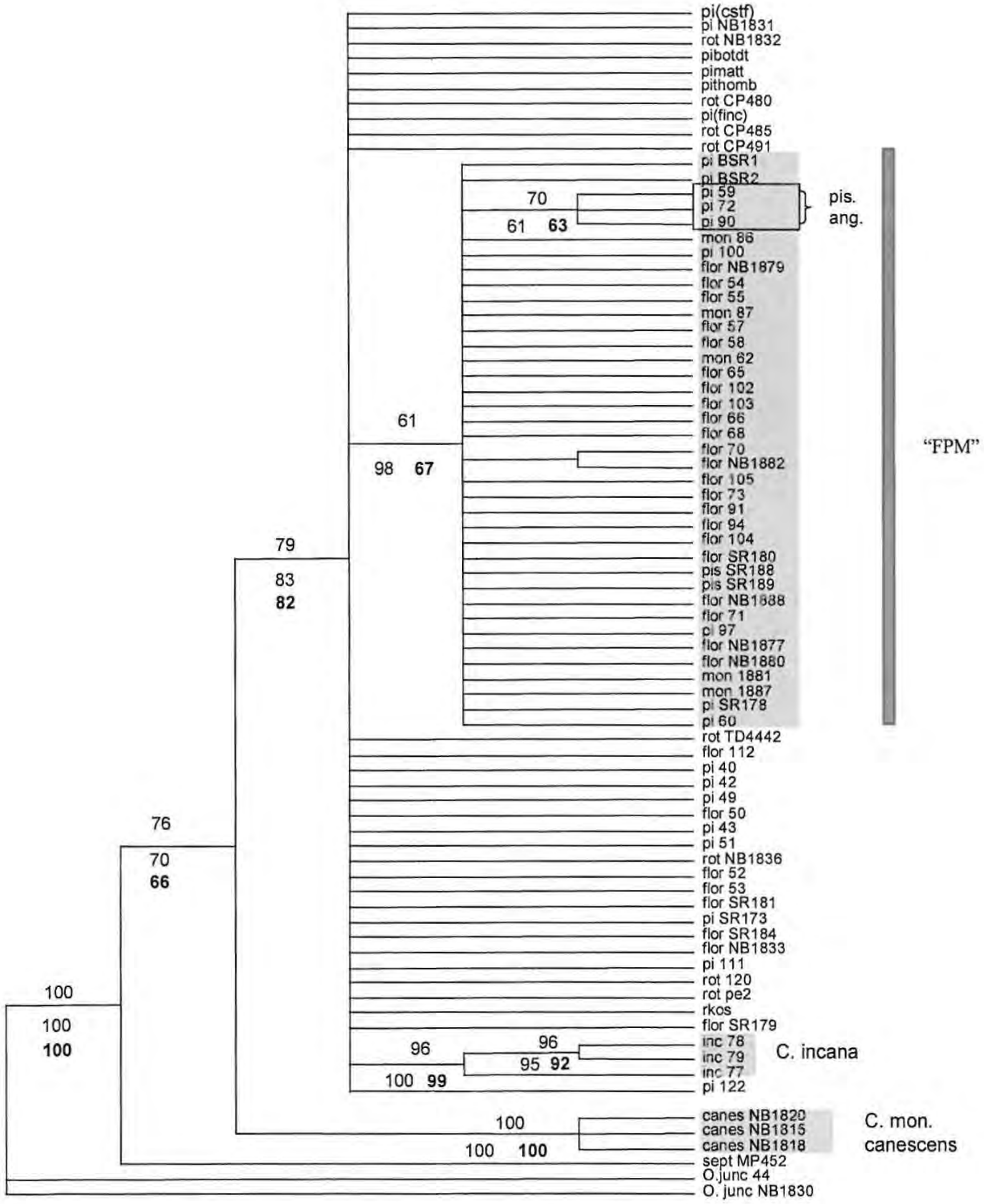

Figure 2.3.1: Complete ITS data set: Strict consensus tree of 5000 most parsimonious trees. Gaps considered as a 5th state (CI=0.890; RI=0.968 Length = 172; Bootstrap support below branches), as binary characters (CI=0.903; RI=0.957; Length = 154; Bootstrap support below branches in bold) or as missing data (CI=0.922; RI=0.962 Length = 141; Bootstrap support above branches). "pis. ang." indicates the C. m. pisifera angustifolia clade. "FPM" indicates the clade containing *C. m. floribunda* form 1, *C. m. monilifera* and *C. m. pisifera pisifera* form 2 samples.

illustrates the difference in weighting that gap coding methods can give to the larger gaps. (The support for the "FPM" branch drops from 98% support when gaps are considered as a 5th state character, to 67% when gaps are coded as binary characters; due to the eight-base deletion that defines the "FPM" branch, being reduced to a single character in "binary" coding, rather than existing as eight independent characters in "5th state" coding).

There are *C. m. pisifera* and *C. m. floribunda* samples both inside and outside of the "FPM" branch, and thus this branch does not reflect any existing subspecies delimitations. However, the *C. m. floribunda* samples that do occur inside the "FPM" branch are form 1 (*sensu* Griffioen) and those that occur outside the branch are form 2. A similar pattern is found for *C. m. pisifera*: those that occur outside the branch are *C. m. pisifera pisifera* form 1, whilst those inside are not. *C. m. pisifera* var *angustifolia* receives some moderate support within the "FPM" branch.

Neighbor Joining analysis

Figure 2.3.2 shows the Neighbor Joining tree obtained from the "Complete ITS" data set using the Jukes Cantor correction. There are many similarities to the Parsimony tree (Figure 2.3.1), in terms of the strong support for *Chrysanthemoides* as a genus, the basal placement of *C. m. septentrionalis* and the placement and support of *C. m. canescens*. The difference lies in a greater structuring of the tree's terminal branches.

The two major branches are designated "East" (which lacks any significant bootstrap support) and "West" (which has 55% bootstrap support) as the samples in the three clades that occur in the "East" branch all occur more than 22° East (as do *C. m. canescens* and *C. m. septentrionalis*). The *C. incana* sequences form a nested clade (with 97% bootstrap support), internal to the "East" branch of the tree (basal to the *C. m. rotundata* clade), rendering *C. monilifera* paraphyletic. The long branch lengths in the *C. incana* samples indicate that, despite their placement within the "East" branch of *C. monilifera*, they do have a relatively large number of sequence differences from the *C. monilifera* samples.

Maximum Likelihood analysis

For the ITS data sets, ModelTest identified the best model as HKY + G (Hasegawa *et al.*, 1985 model with variable sites assumed to follow a discrete gamma distribution) and the

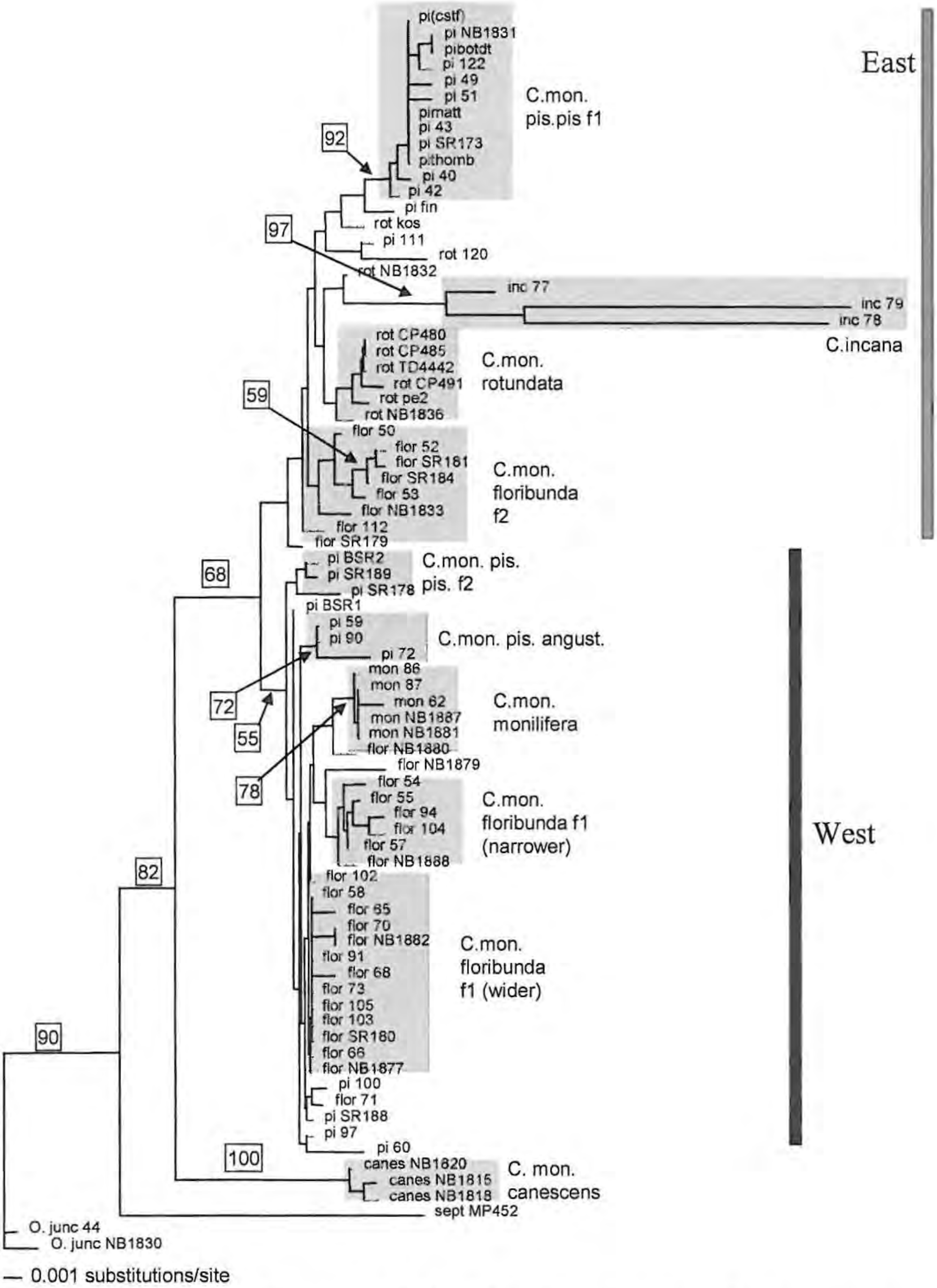

Figure 2.3.2: Complete ITS data set: Neighbor Joining Jukes-Cantor tree. (Bootstrap support above branches).

specific settings were: Lset; Base=(0.2213; 0.2801; 0.2898); Nst=2; TRatio=1.3847; Rate=gamma; Shape=0.1455; Pinvar=0.

Figure 2.3.3 shows the Maximum Likelihood tree following analysis of the "Complete ITS" data set. As with the other trees resulting from the other two analytical methods, the *C. m. septentrionalis* and *C. m. canescens* clades are basal to the rest of the *C. monilifera* subspecies. *C. m. canescens* retains its 100% support and *C. m. monilifera* also has a relatively high support value (73%). The "East"/"West" branch split remains, as does some of the internal structure to these major branches: C. *m. pisifera angustifolia* remains (with bootstrap support of 66%), as does C. *m. floribunda* form 2 and C. *m. pisifera pisifera* form 1. Only the "West" branch has bootstrap support, and that only of 53%. Also important is the placement of the well-supported *C. incana* clade (79%) in the "East" branch.

Analyses of the "Monilifera ITS" data set

Parsimony analysis

Figure 2.3.4 shows the strict consensus Parsimony tree following analysis of the "Monilifera ITS" data set, gaps considered as a 5th state, as binary characters or as missing data (analyses were terminated after 5000 trees were obtained, and then these trees were searched to completion). The tree topology is very similar the same as the tree obtained from the "Complete ITS" data set (although not identical).

Compared to the tree obtained from the "Complete ITS" data set, this tree has lower bootstrap support at the basal branches of the tree, and higher bootstrap support towards the tips of branches. It also has slightly higher RI and CI values. *C. m. pisifera angustifolia* again receives some support within the "FPM" branch (ranging from 57% to 62% depending on the gap handling method).

Neighbor Joining analysis

Figure 2.3.5 shows the Neighbor Joining tree obtained from the "Monilifera ITS" data set using the Jukes Cantor correction. The tree structure remains much the same as that of the "Complete ITS" data set. *C. m. pisifera angustifolia* now has moderate support (66%) that was absent in the "Complete ITS" data set tree. The support for *C. monilifera* as a species remains strong (97%) and *C. m. rotundata* has acquired bootstrap support (63%) and the E/W split noted above is retained.

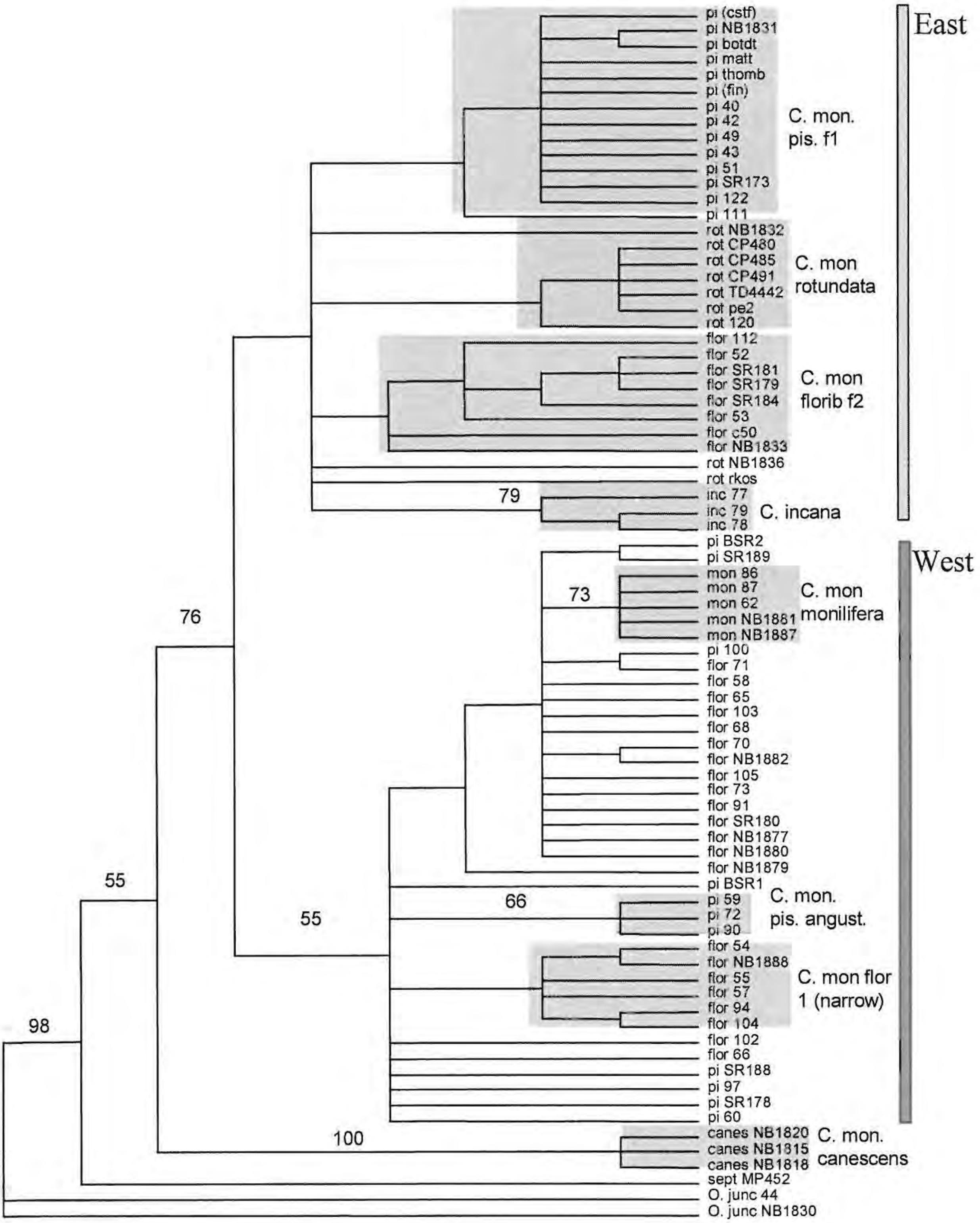

Figure 2.3.3: Complete ITS data set: Maximum Likelihood tree. (Bootstrap support above branches).

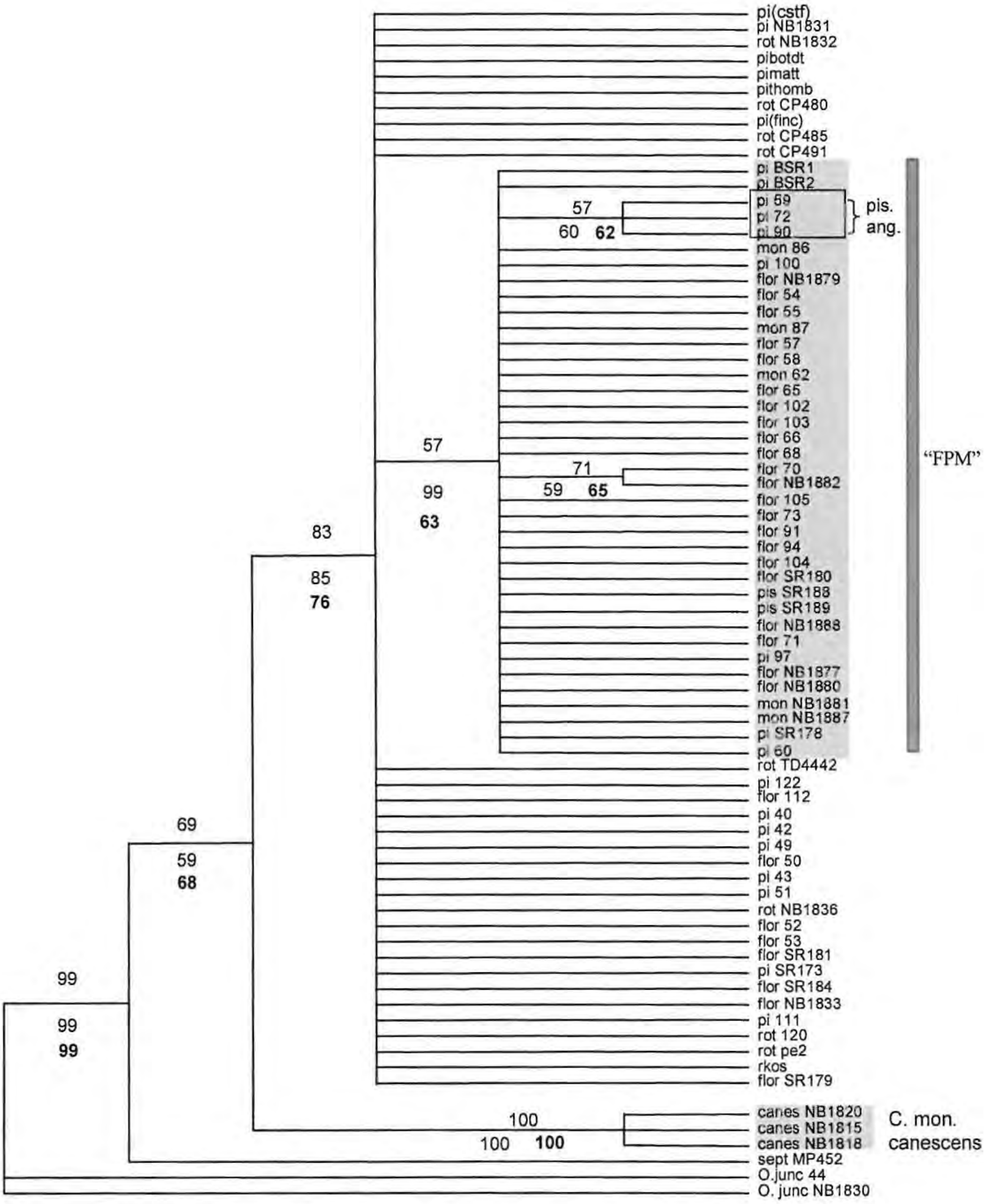

Figure 2.3.4: Monilifera ITS data set: Strict consensus tree of 5000 equally most parsimonious trees. Gaps considered as a 5th state (CI=0.893; RI=0.971 Length = 149; Bootstrap support below branches); as binary characters (CI=0.902; RI=0.959 Length = 132; Bootstrap support below branches in bold) or as missing data (CI=0.927; RI=0.967 Length = 124; Bootstrap support above branches).). "pis. ang." indicates the C. m. pisifera angustifolia clade. "FPM" indicates the clade containing *C. m. floribunda* form 1, *C. m. monilifera* and *C. m. pisifera pisifera* form 2 samples.

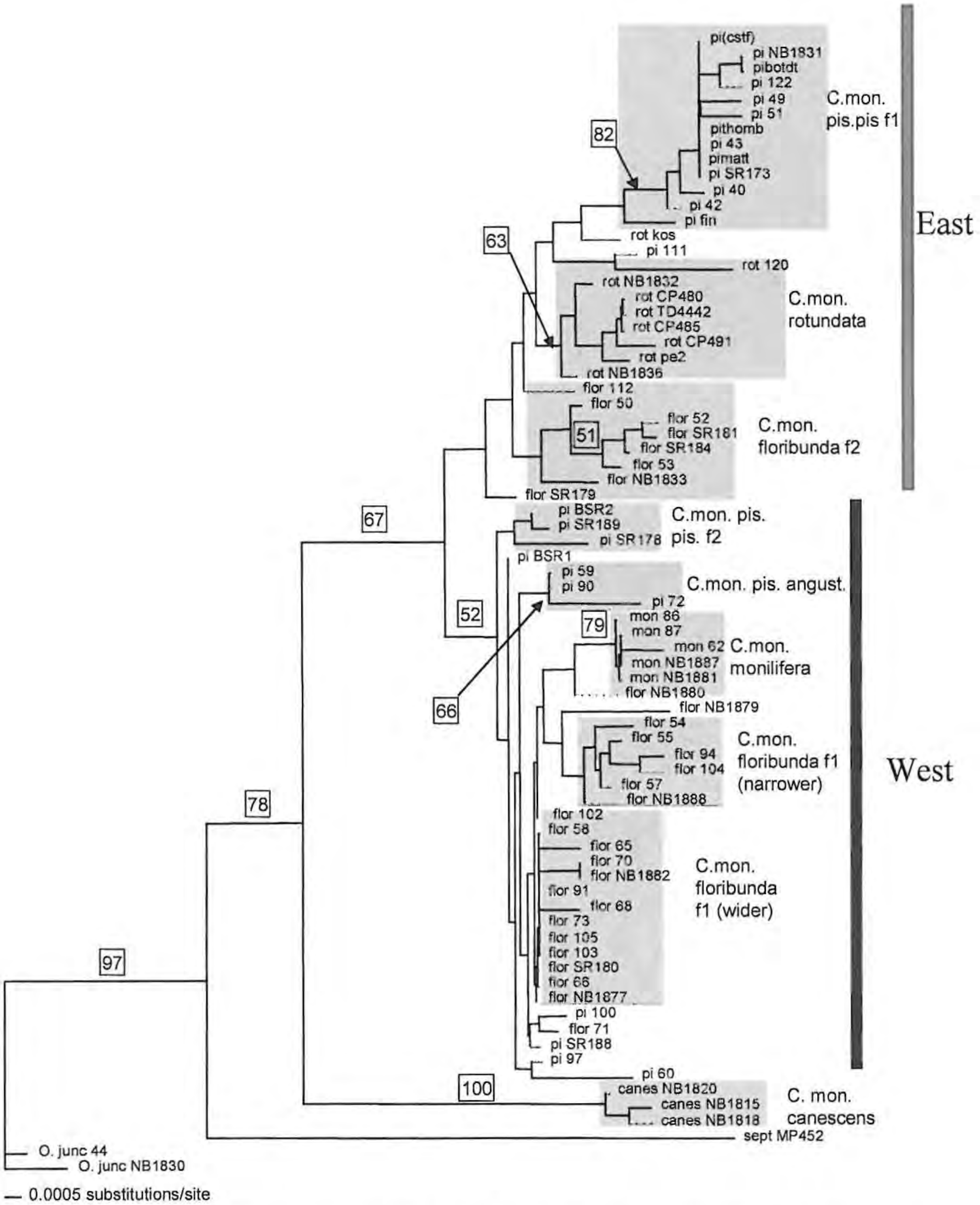

Figure 2.3.5: Monilifera ITS data set: Neighbor Joining Jukes-Cantor tree. (Bootstrap support above branches).

Maximum Likelihood analysis

Figure 2.3.6 shows the Maximum Likelihood tree following analysis of the "Monilifera ITS" data set. The structure is very similar to Figure 2.3.3, except for the excluded *C. incana* clade in the "East" branch. Much like the tree obtained from the "Complete ITS" data set , there is little definitive resolution to the terminal branches. The bootstrap support that is present is much the same at the basal levels of the tree, except that the "West" branch loses bootstrap support.

Analyses of the "Limited ITS" data set

Parsimony analysis

Figure 2.3.7 shows the strict consensus Parsimony tree of six most parsimonious trees found following analysis of the "Limited ITS" data set, with gaps either considered as missing data, as binary characters or as a 5th character state. At the more basal levels, the bootstrap support is higher when gaps are considered as missing characters, whereas at the tips of the branches, support is often higher when gaps are codes as 5th state characters.

This smaller data set has more structure in the branches of the tree, although comb-like structuring stills exists in the terminal branches of the tree. The tree phylogeny agrees with the structure seen in the "Complete ITS" Neighbor Joining tree (Figure 2.3.2), even though the "Complete ITS" data set parsimony-derived tree in Figure 2.3.1 has little definitive structure to it.

Neighbor Joining analysis

Figure 2.3.8 shows the Neighbor Joining tree obtained from the "Monilifera ITS" data set using the Jukes Cantor correction, and again, *C. monilifera septentrionalis* and *C. monilifera canescens* are basal to the rest of the tree. There are good bootstrap support values at the base of the tree (84%), and the *C. monilifera pisifera pisifera* form 1 clade has 57% support but also has good support values for internal nodes (95% and 77%). The *C. incana* samples, however, are split, with the Kirstenbosch sample appearing in the "East" branch and the other two appearing further down the tree (above the *C. m. canescens* and *C. m. septentrionalis* branches, as well as the "57", "58" and "178" *C. m. floribunda* form 1 branch), rendering the species paraphyletic.

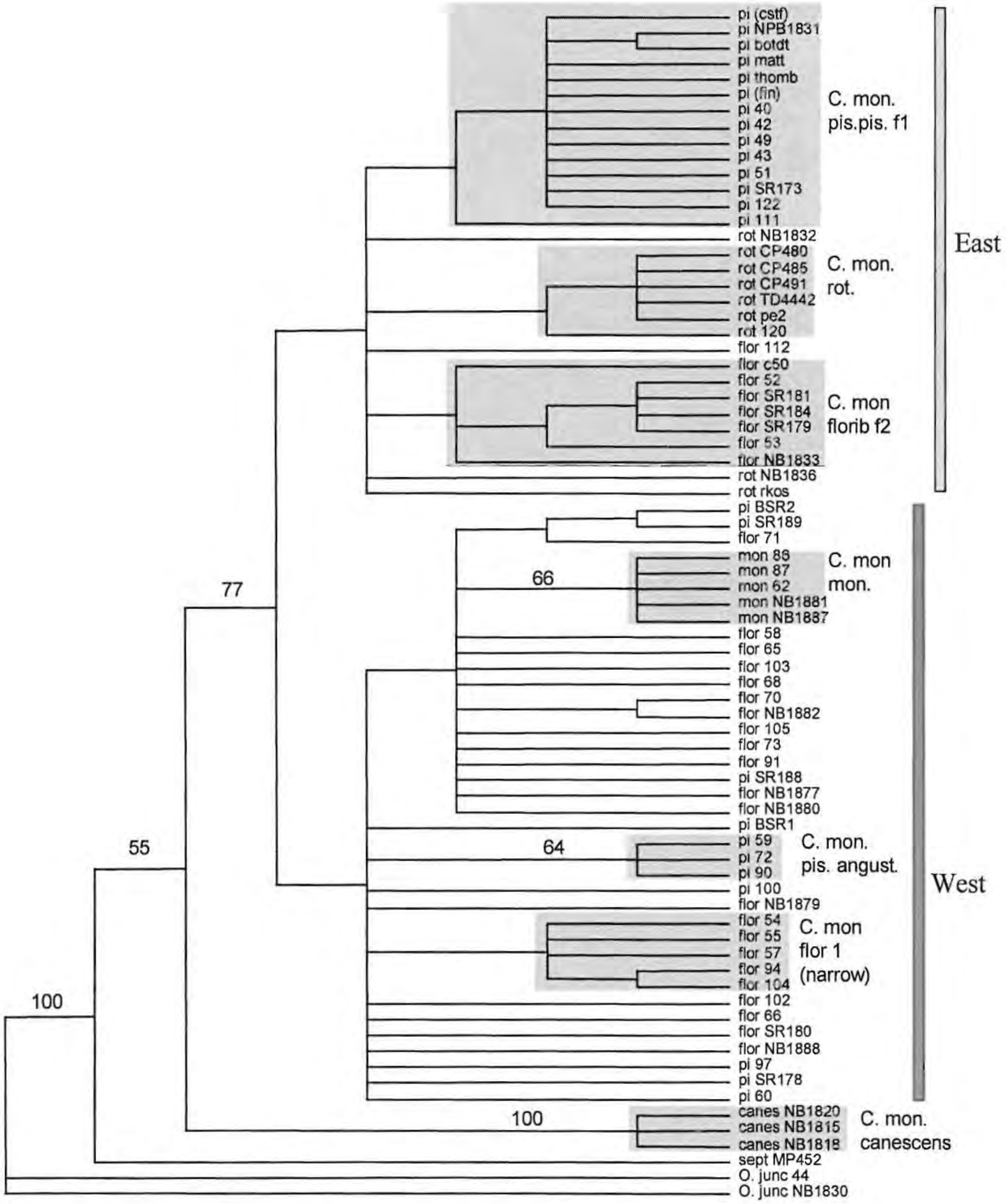

Figure 2.3.6: Monilifera ITS data set: Maximum Likelihood tree. (Bootstrap support above branches).

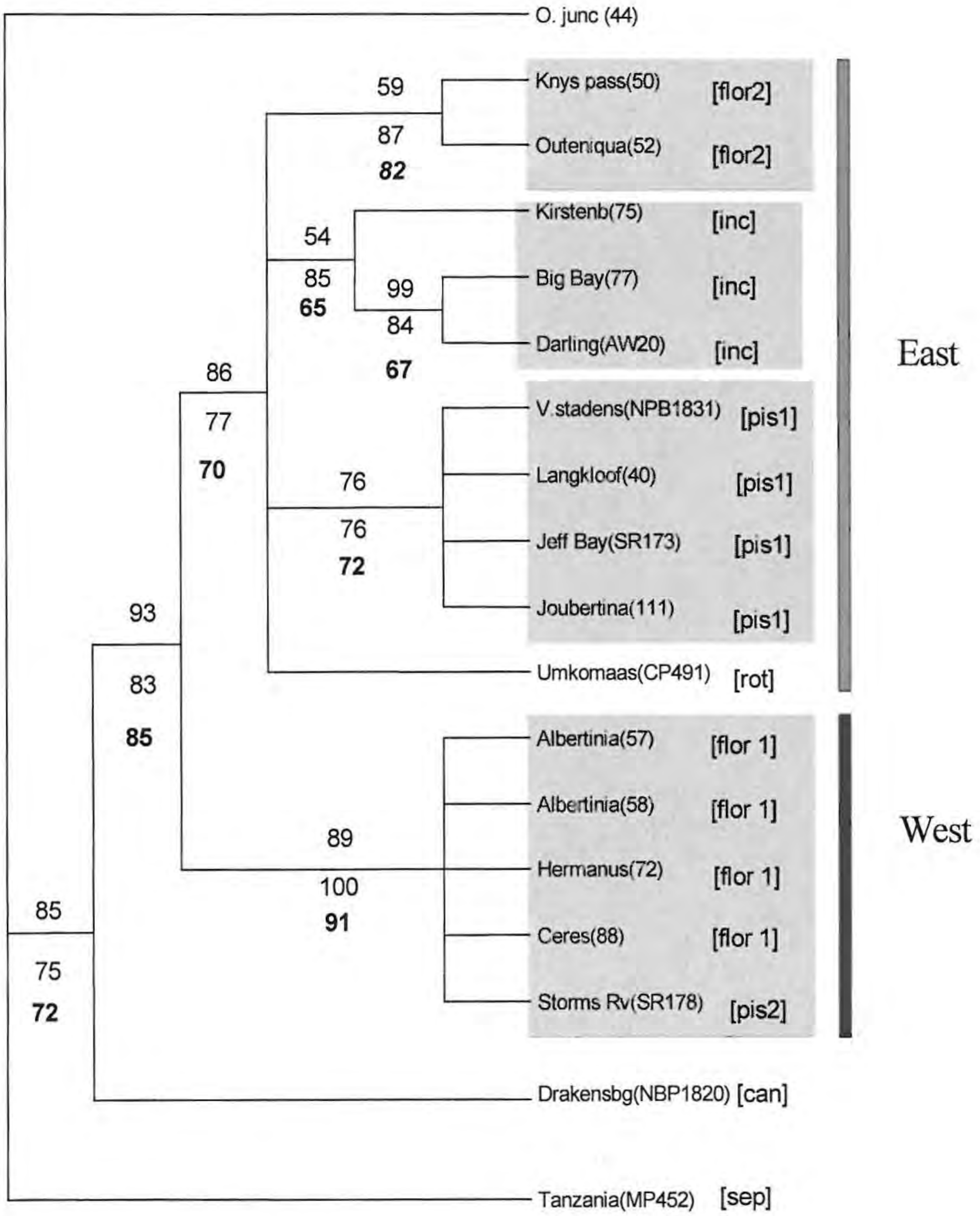

Figure 2.3.7: Limited ITS data set: Strict consensus tree of 6 equally most parsimonious trees. Gaps considered either as a 5th state (CI=0.849; RI=0.909 Length = 53; Bootstrap support below branches), as binary characters (CI=0.784; RI=0.862; Length = 37; Bootstrap support below branches in bold) or as missing data (CI=0.839; RI=0.967 Length = 31; Bootstrap support above branches). Subspecies designations in square brackets after sample number.

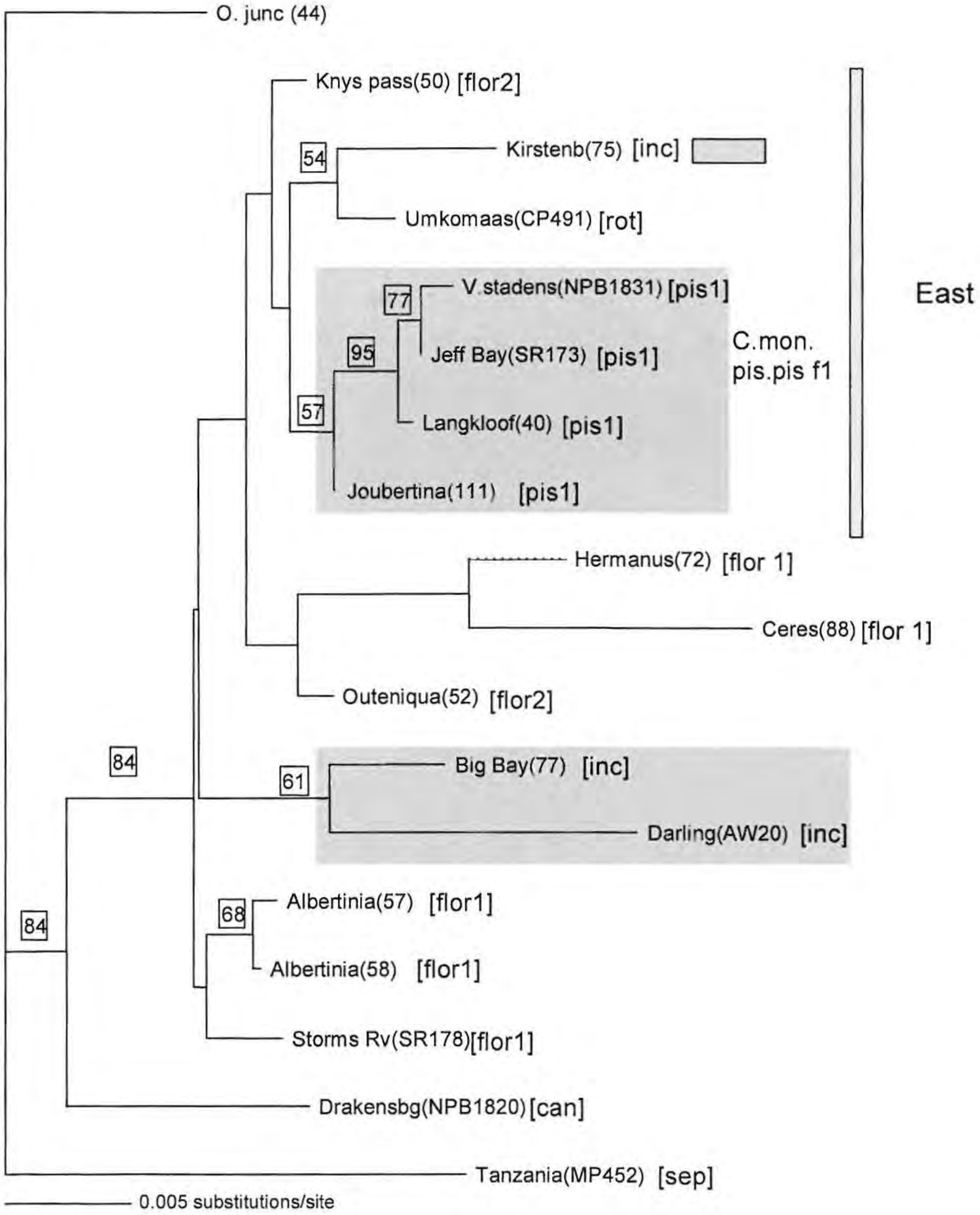

Figure 2.3.8: Limited ITS data set: Neighbor Joining Jukes-Cantor tree. (Bootstrap support above branches). Subspecies designations in square brackets after sample number.

Maximum Likelihood analysis

Figure 2.3.9 shows the Maximum Likelihood tree following analysis of the "Limited ITS" data set. The tree structure is much the same as the Parsimony tree for the same set of data (Figure 2.3.7) except with slightly more resolution at the terminal branches. The "West" branch has 75% support, and the *C. m. pisifera pisifera* form 1 clade now has 54% support. In this tree however, *C. incana* is retrieved as monophyletic.

Analyses of the "psbA-trnH" data set

Parsimony analysis

Figures 2.3.10 and 2.3.11 are the Strict consensus trees following analysis of the "psbA-trnH" data set, with gaps considered either as missing data (Fig 2.3.10), or as a 5th state or a binary character (Fig 2.3.11). Analysis with gaps considered as missing data results in a single most parsimonious tree and resulting CI and RI of 1, whereas analysis with gaps considered as a 5th state results in 10 trees, with CI = 0.957, RI = 0.991 and a length of 46, and gaps considered as binary characters resulted in five trees, with CI = 0.933; RI = 0.985 and a length of 15.

Despite the high CI and RI values, there is some lack of resolution with six samples not showing any branching structure ("178", "1831", "40", "491", "173" and "111"), most likely due to identical sequences. Both trees do, however, have the *C. incana* samples grouping together, with the analysis with gaps considered as a 5th state providing higher bootstrap support values than ignoring gaps altogether (100% vs. 70%; 80% vs. 68%), but the analysis with gaps considered as missing data provides slightly more resolution in the upper "floribunda" branch of the trees. The "floribunda" branch has 100% support in both trees and when the raw sequence data is compared, some of the major defining characteristics of the "floribunda" clade are three multibase deletions. These are four-, six- and 10-base deletion gaps that are present in these nine samples. It is a deletion (not an insertion), as the *Osteospermum* outgroup still possesses these four, six and10-base sequences, and the odds of these three exact sequence arising by chance again in the other *Chrysanthemoides* branch is extremely unlikely (and unparsimonious). Considering gaps as binary characters does not lower the bootstrap support for this branch, despite the length of the gaps (which would receive higher weightings when gaps are considered as 5th state character, due to their length). There is also a 14-base insert that is present only in the *C. incana* samples (here, considering

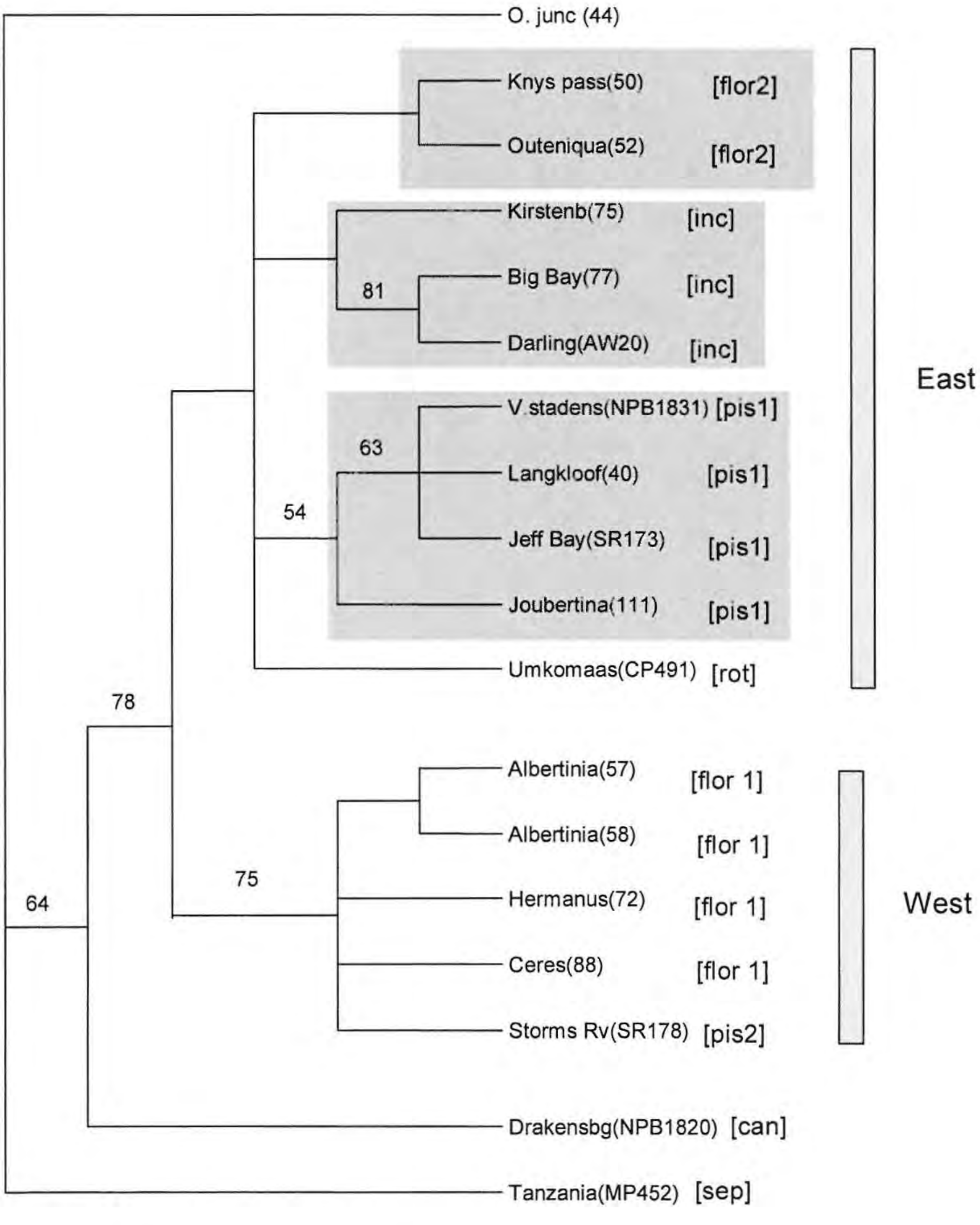

Figure 2.3.9: Limited ITS data set: Maximum Likelihood tree. (Bootstrap support above branches). Subspecies designations in square brackets after sample number.

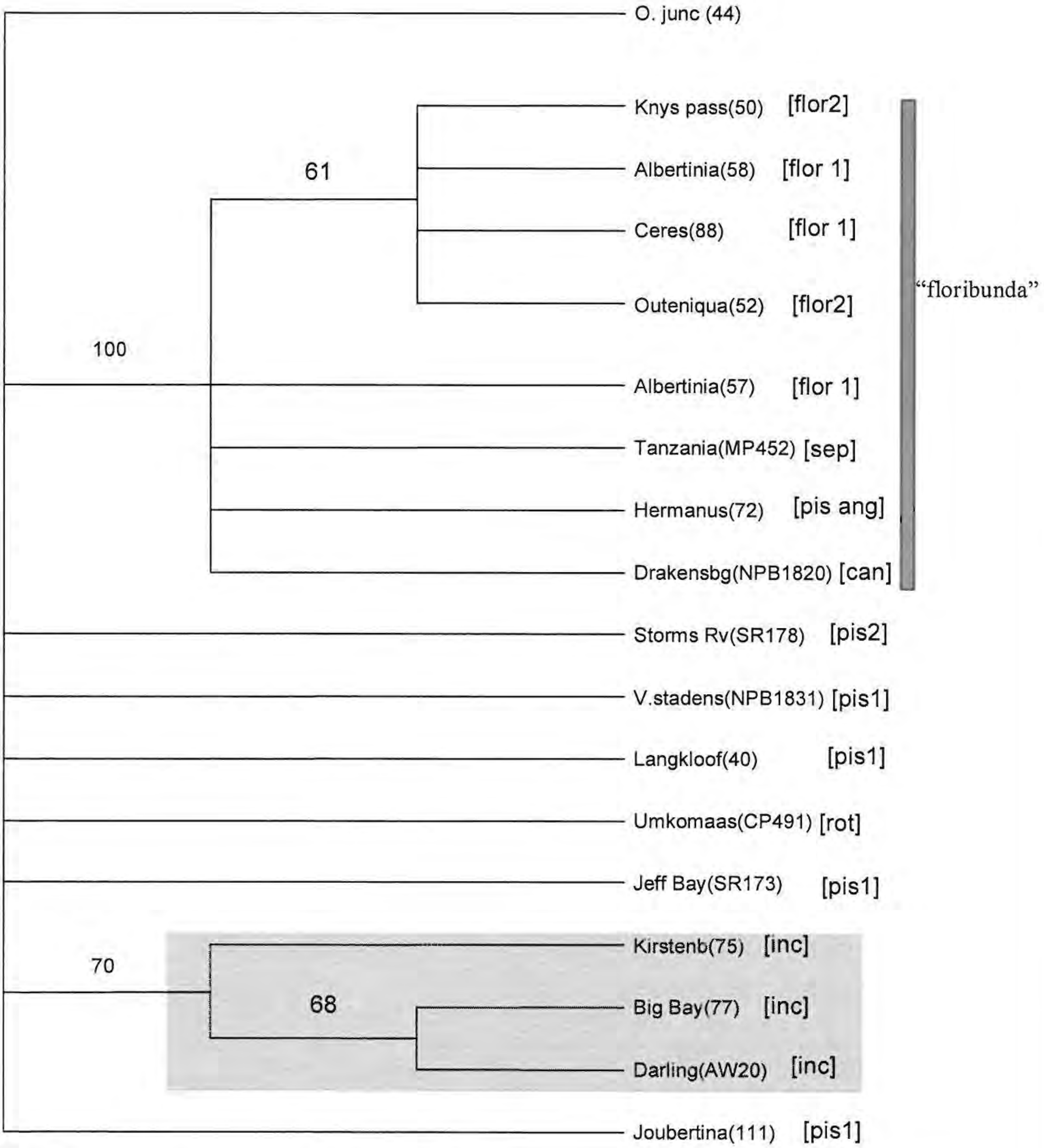

Figure 2.3.10: *psbA-trnH* data set: Single most parsimonious tree (length = 8). Gaps considered as missing data. (CI = 1; RI = 1; bootstrap support above branches).

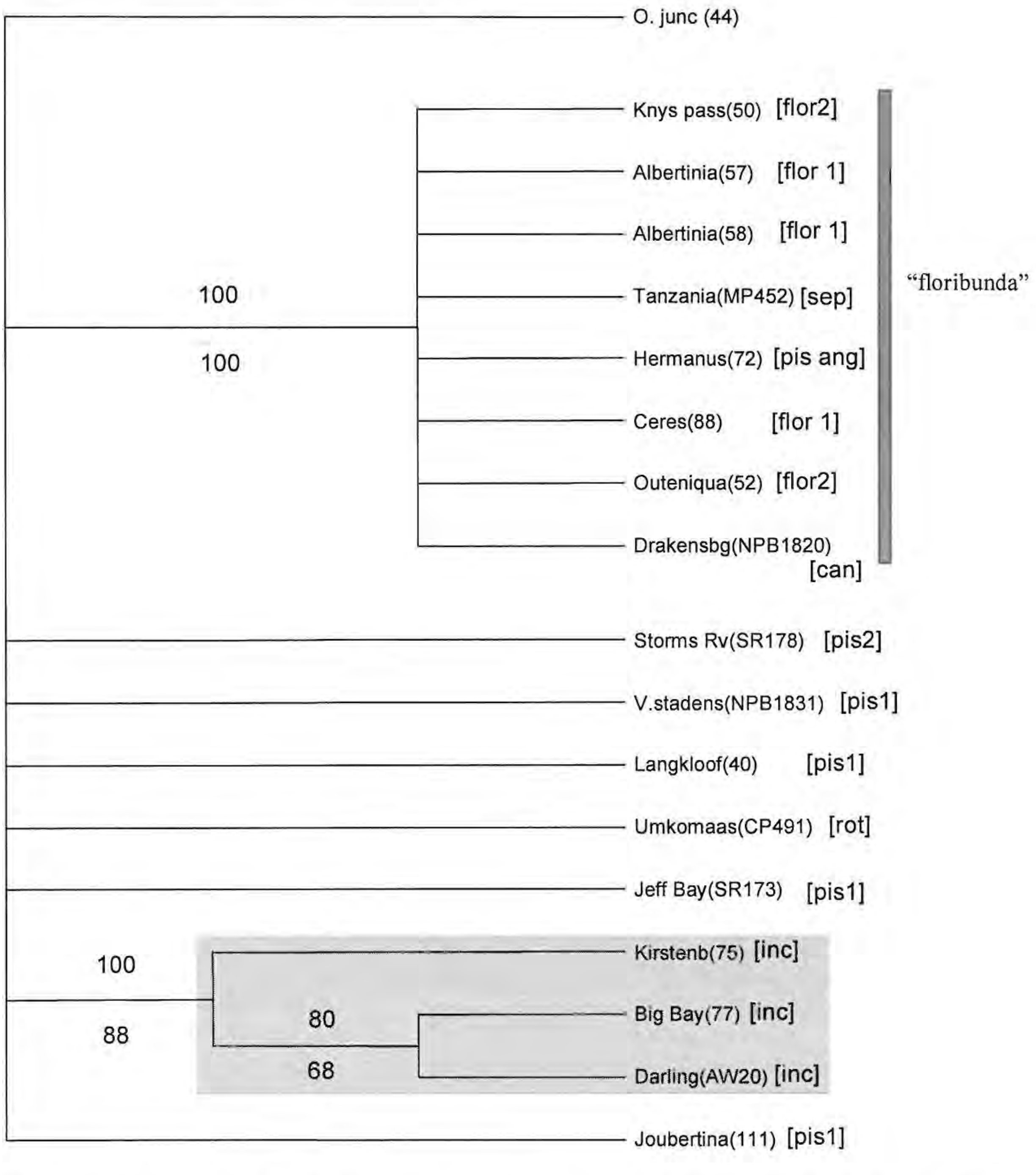

Figure 2.3.11: *psbA-trnH* data set: Strict consensus parsimony tree of 10 trees for the 5th state gap coding and five trees for the binary gap coding. Gaps considered as a 5th state (CI = 0.957; RI = 0.991; Length = 46; Bootstrap support above branches) or coded as binary characters (CI = 0.933; RI = 0.985; Length = 15; Bootstrap support below branch).

gaps as binary characters does lower the bootstrap support for this branch).

There does appear to be some correlation between genetic phylogeny and morphology, with the *"floribunda"* branch including all of the *C. m. floribunda* form 1 and form 2 samples, as well as the *C. m. canescens* (1820) and *C. m. septentrionalis* (452) samples and one *C. m. pisifera pisifera* form 2 sample and a *C. m. pisifera angustifolia* sample.

Unlike the ITS tree, which seems to more closely follow geographical distribution for the *C. m. floribunda* samples, the cpDNA seems to more closely follow the morphological subspecies designations for the floribunda samples, grouping them together. Unlike the ITS tree, however, the *C. m. canescens* and *C. m. septentrionalis* samples come out as derived, not basal to the tree, being placed in the "floribunda" clade, suggesting that they were derived from a *C. m. floribunda* ancestor.

The *C. incana* sample "75" is taken from a plant growing in the Kirstenbosch Botanical Gardens, but it should be noted that the plant was deliberately relocated to the Gardens and the original locality of the plant is unknown. Based on morphology, the plant is most likely *C. incana incana gracilis*, which occurs from Clanwilliam to Calvinia and westwards to Hondeklip Bay from 20-700m above sea level (Griffioen, 1995). The *C. incana* sample "77" is *C. incana incana* var *incana* and the other sample "AW20" is from the other subspecies of *C. incana*, *C. incana subcanescens*. "AW20" is placed as derived within the *C. incana* clade.

Neighbor Joining analysis

Figure 2.3.12 shows the Neighbor Joining distance tree obtained from the "psbA-trnH" data set using the Jukes Cantor correction. There are two strongly supported branches in the tree. The "floribunda" clade (including the *C. m. septentrionalis*, *C. m.canescens* and two *C. m. pisifera* samples) has 98% support. The two *C. m. pisifera* samples ("1831" and "111") are basal to the "floribunda" clade, but the branch length that separates the floribunda clade from those two basal samples is very long. The *C. m. septentrionalis* sample is basal to another branch inside the "floribunda" clade and has a relatively long branch length (of a similar length to the branch that supports the "floribunda" clade as a whole). The *C. m. septentrionalis* sample is basal to a minor branch within the "floribunda" clade that contains a 59% supported clade that includes samples from *C. m. floribunda* form 1 and form 2. The other branch in the "floribunda" clade includes the *C. m. canescens sample*, a *C. m. floribunda* (form 1) sample and a *C. m .pisifera angustifolia* sample. The other major branch

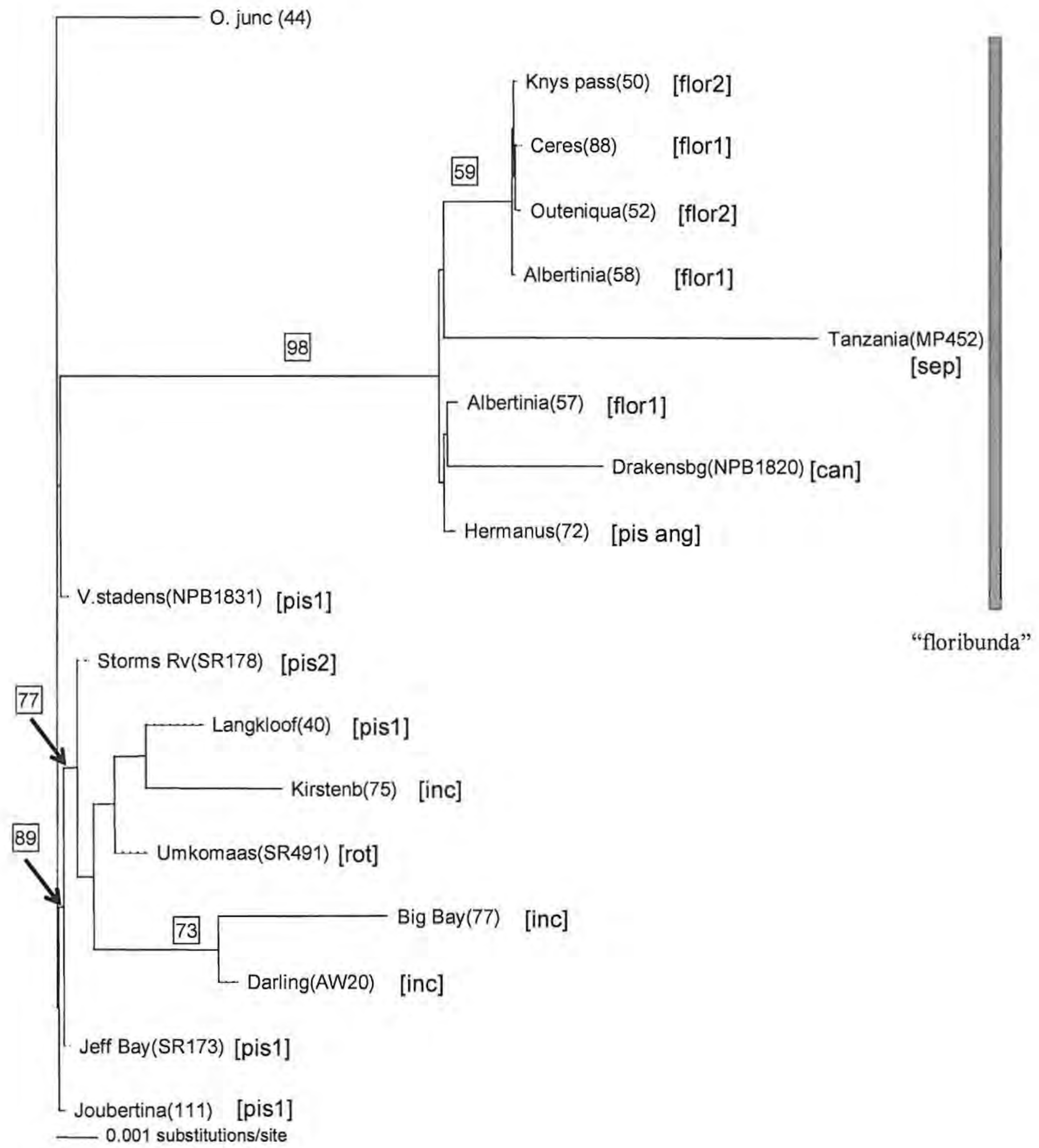

Figure 2.3.12: *psbA-trnH* data set: Jukes-Cantor Neighbor Joining distance tree (bootstrap support above branches).

in the tree (with 89% support) includes the *C. incana* samples, the *C. m. rotundata* samples and four *C. m. pisifera* samples

Maximum Likelihood analysis

For the "psbA-trnH" data set, ModelTest identified the best model as the transversion model (TVM; Posada & Crandall, 1998). Specific settings were: Lset; Base=(0.3228 0.1237 0.1628); Nst=6; Rmat=(0.5845 0.7017 0.0599 0.7795 0.7017); Rates=equal; Pinvar=0.

Figure 2.3.13 shows the Maximum likelihood tree following analysis of the "psbA-trnH" data set. The topology of the tree is identical to that of the strict consensus parsimony tree with gaps considered as missing data (Figure 2.3.10) but with slightly different bootstrap support for the major branches.

Statistical Parsimony analysis

Figure 2.3.14 shows the TCS gene network for the "psbA-trnH" data set with gaps considered as a 5th character state. Figure 2.3.15 shows the TCS gene network for the "psbA-trnH" data set with gaps considered as missing data. Several haplotypes collapse into one node (indicated by several names within one node) due to identical sequences.

The second analysis method (with gaps considered as missing data) produces a complete articulated gene network, whilst considering gaps as a 5th state renders three separate gene networks. This is most likely due to the large insertions and deletions that occur in the *psbA-trnH* region. The three separate networks in Figure 2.3.14 are (i) the group of *C. incana* samples, with a large 14-base insert defining these samples; (ii) the "*Osteospermum-pisifera-rotundata*" grouping; and (iii) the "*floribunda-septentrionalis-canescens*" (F-S-C) group (including a *C. m. pisifera angustifolia* sample), all of which share a four-, six- and 10-base deletion.

When gaps are ignored (or considered as binary characters), as in Figure 2.3.17, the networks in the previous figure become articulated. The "*Osteospermum-pisifera-rotundata*" group is joined to the "*incana*" group, and then this larger group is articulated to the F-S-C group. The network suggests that *C. incana* is derived from a *C. m. pisifera* ancestor, and that the *C. m. floribunda* ESUs are derived from *pisifera* as well, and then from the *C. m. floribunda* clade, *C. m. canescens* and *C. m. septentrionalis* are derived.

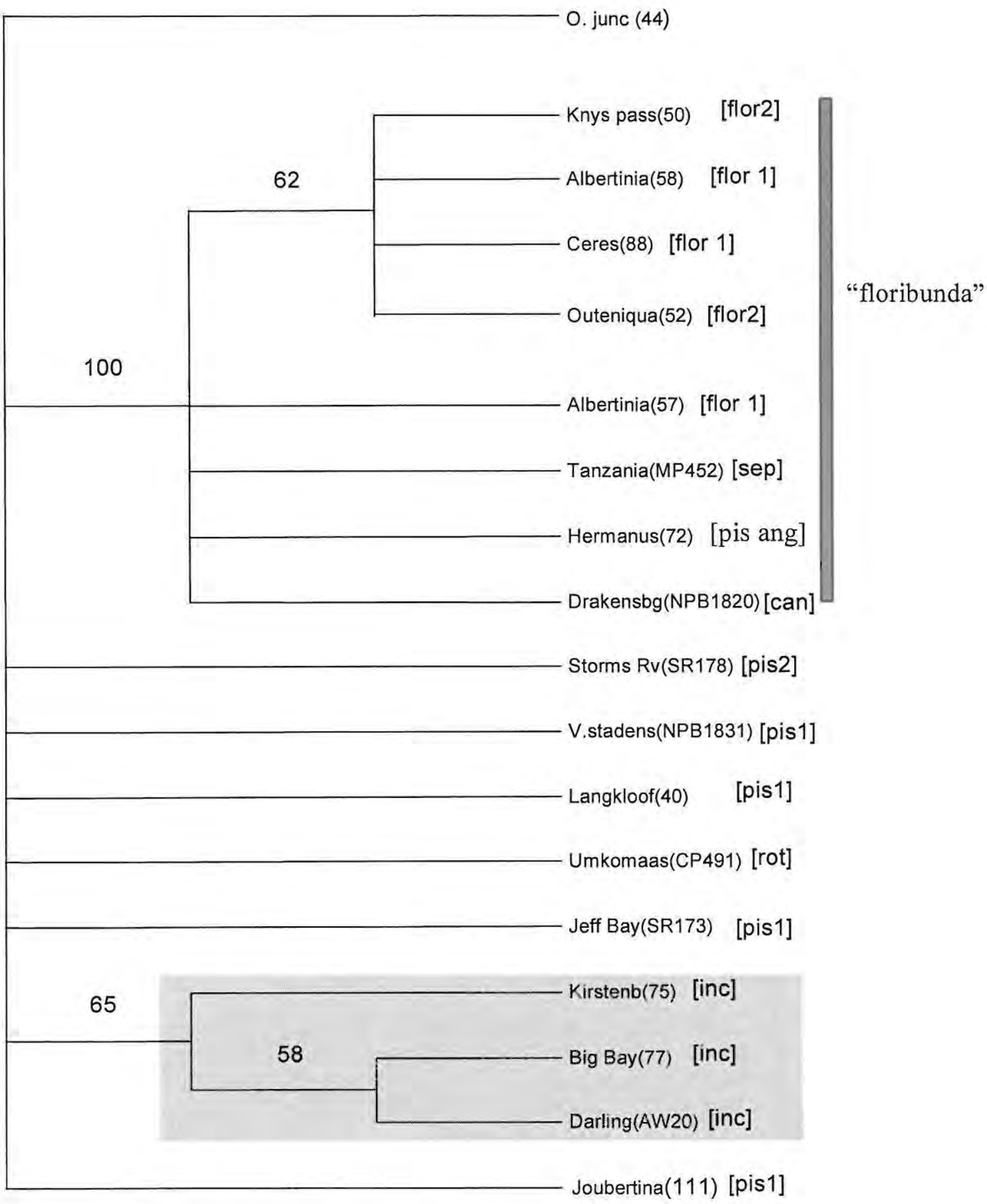

Figure 2.3.13: *psbA-trnH* data set: Maximum Likelihood tree. (Bootstrap support above branches).

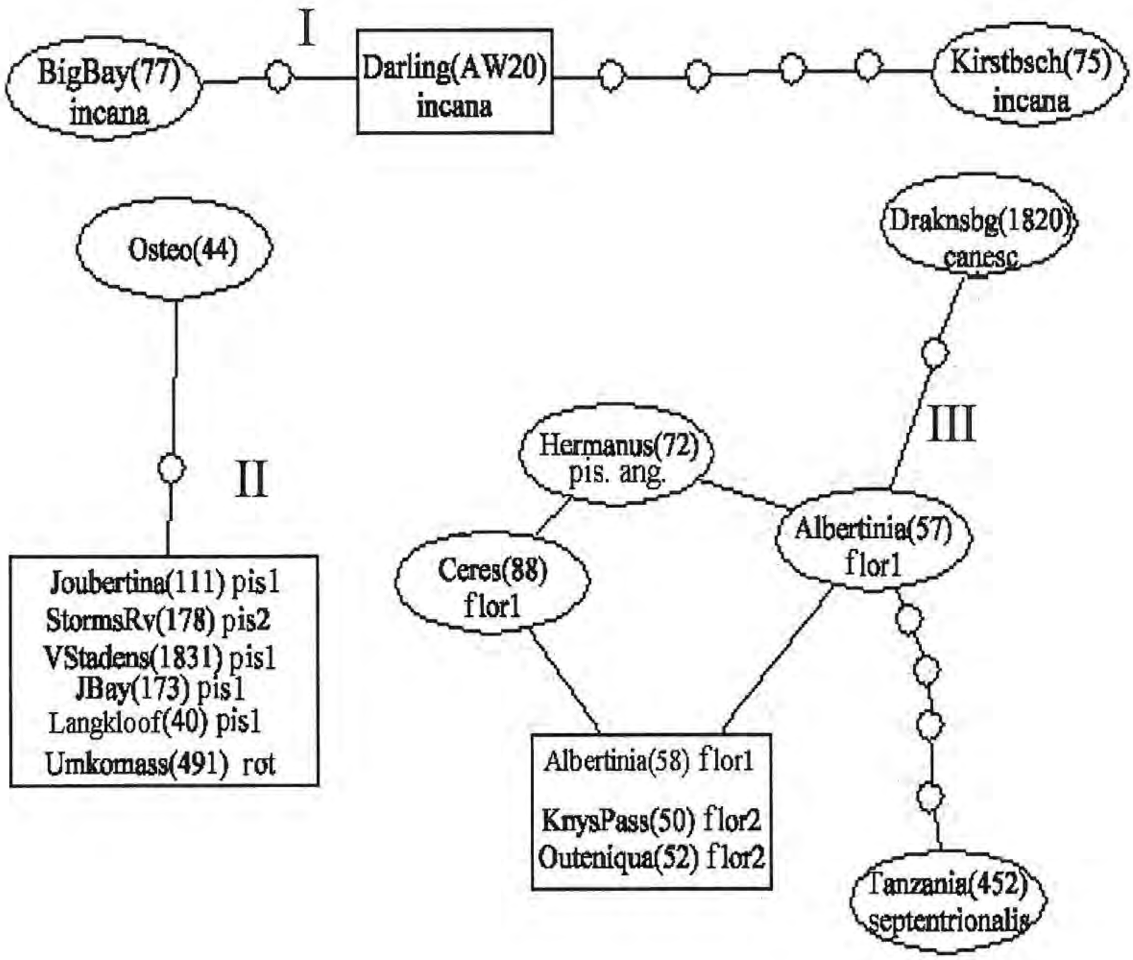

Figure 2.3.14: *psbA-trnH* data set: TCS gene network. Gaps considered as a 5th state.

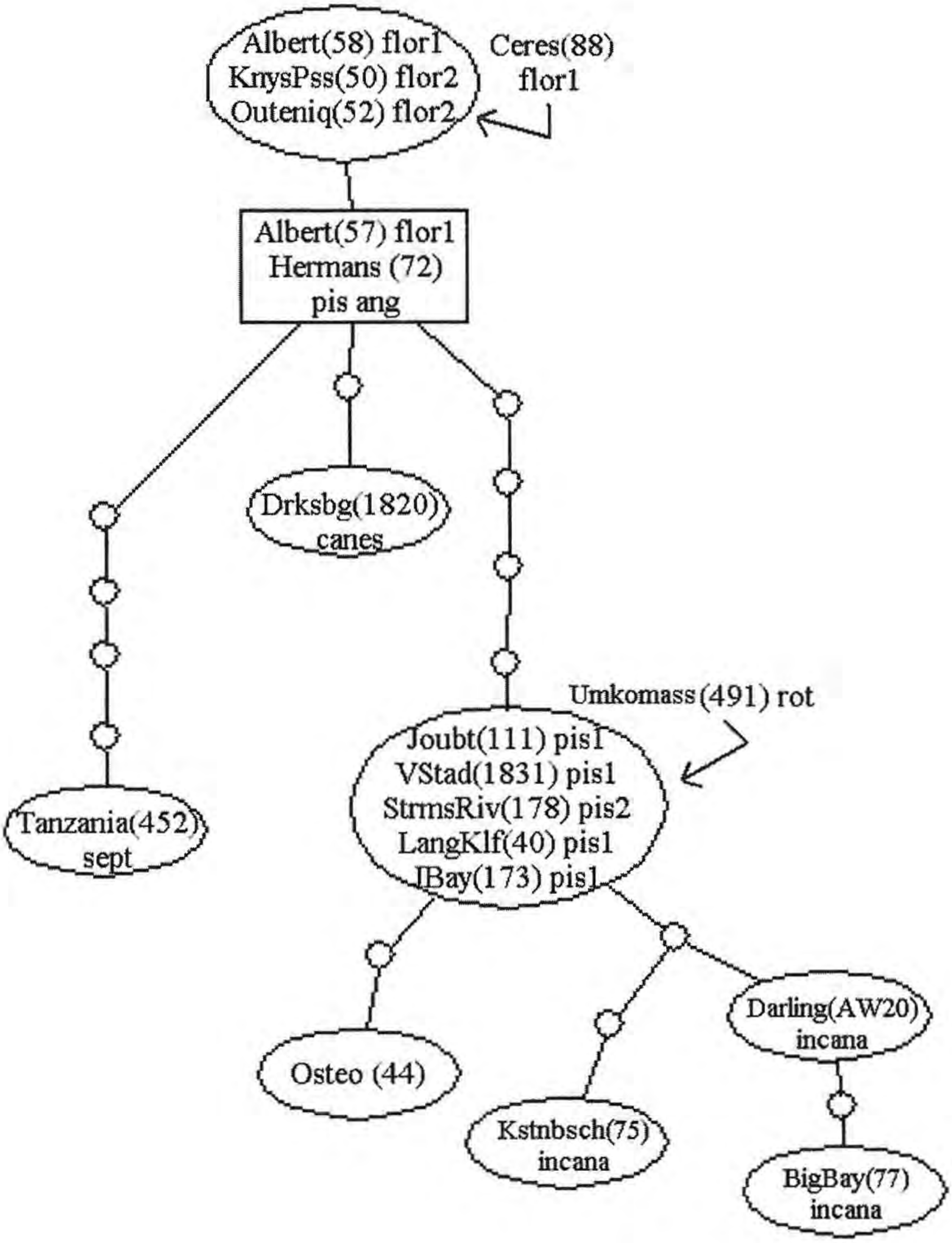

Figure 2.3.15: *psbA-trnH* data set: TCS gene network. Gaps considered as missing data.

Analyses of the "trnL-trnF" data set

Parsimony analysis

Figures 2.3.16 and 2.3.17 are the strict consensus trees for the *trnL-trnF* data, with gaps considered either as missing data (Fig 2.3.16), or as a 5th state or binary character (Fig 2.3.17). Analysis considering gaps as missing data results in 42 most parsimonious trees (Length = 9; CI = 0.667; RI = 0.870), whereas the analysis considering gaps as a 5th state or as binary characters results in 94 trees (Length = 11; CI = 0.727; RI = 0.900 for both coding methods). There is very little bootstrap support for trees where gaps were considered as missing data, whereas considering gaps as a 5th state provided some bootstrap support for the branches, greater topological resolution and higher CI and RI values. All these factors indicate that gaps are important as phylogenetically informative characters in the *trnL-trnF* data. The branch bootstrap support in the basal regions of the tree are lower when gaps are considered as binary characters, rather than a 5th state.

The "floribunda" branch grouping that was present in the trees from the "psbA-trnH" data set is also present in the trees from this data set, although with very little phylogenetic resolution, especially when gaps were considered as missing data (Figure 2.3.16). There is little phylogenetic resolution to the *C. incana* samples, with the *C. incana* samples as part of a basal polytomy in the "floribunda" branch in the analysis with gaps considered as a 5th state (Figure 2.3.17). The relatively well supported grouping of "Knysna pass (50)" and "Hermanus (72)" (59% and 60% support in the two trees) seems unusual, as the two samples are not only different subspecies, but also were collected from sites about 500km apart.

Neighbor Joining analysis

Figure 2.3.18 shows the Neighbor Joining distance tree following analysis of the "trnL-trnF" data set using the Jukes Cantor correction. This tree has very few high bootstrap support values, suggesting that the phylogeny provided by the tree may not be very accurate.

Two main branches are present, (i) the "pisifera" clade, again including the *C. m. rotundata* sample; and (ii) the "floribunda" clade, with two of the *C. incana* samples as basal to the rest of the clade. The finer phylogenetic structure in the two branches does not appear to make much logical sense, and in combination with the lack of high bootstrap support, the phylogeny that the tree suggests must be viewed with caution.

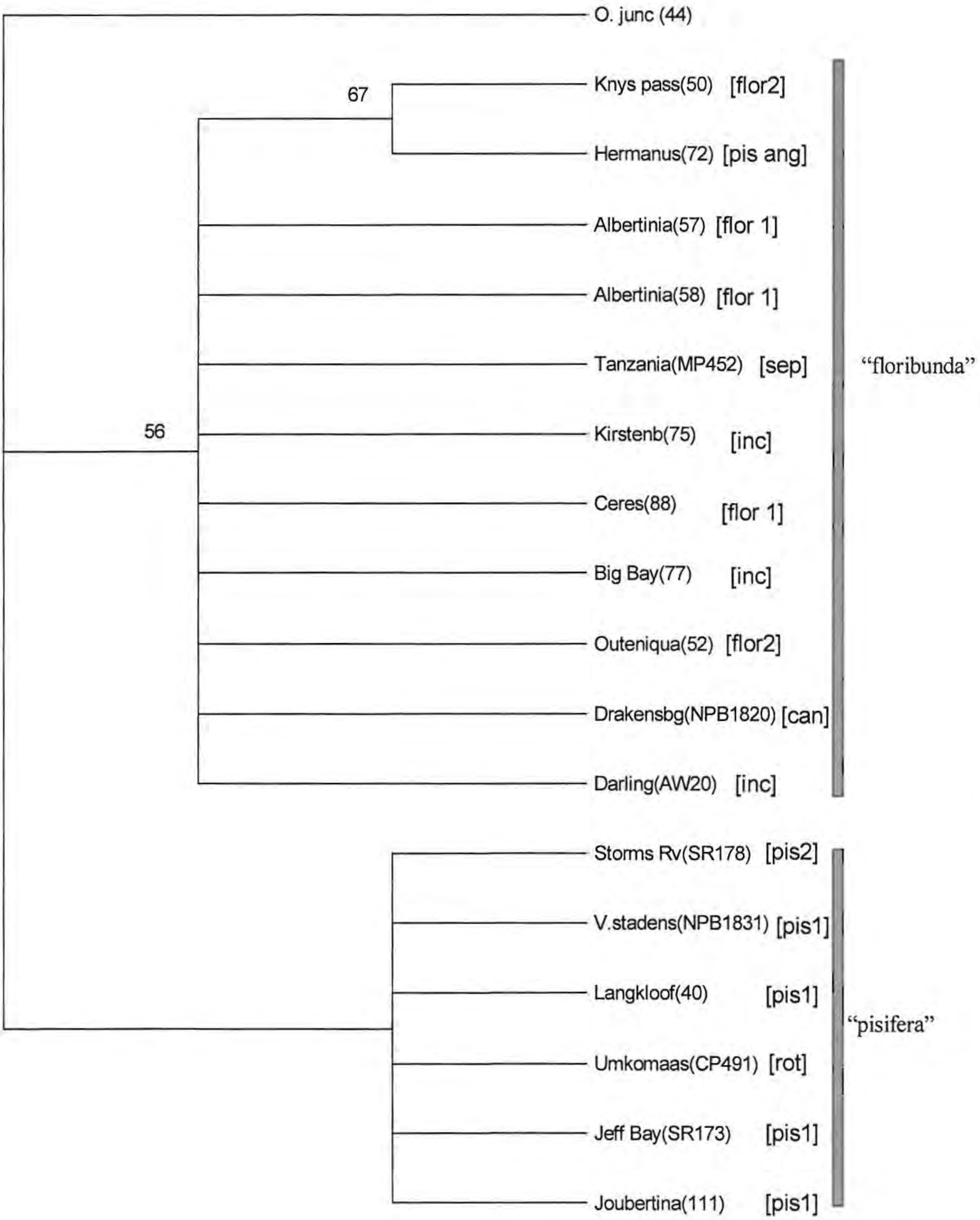

Figure 2.3.16: *trnL-trnF* data set: Strict consensus tree of 42 most parsimonious trees. Gaps considered as missing data. CI = 0.667; RI = 870; Length = 9. (Bootstrap support above branches).

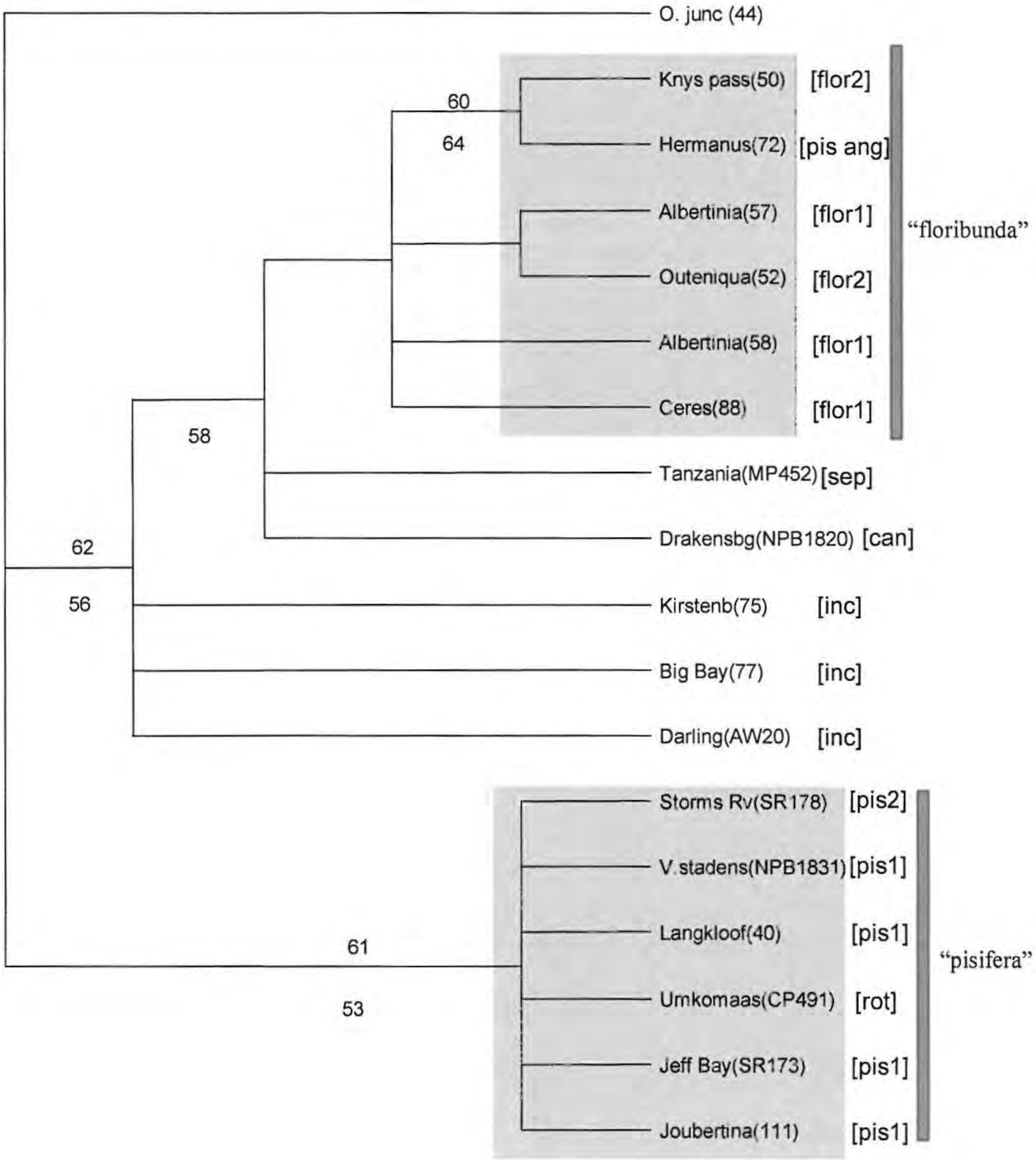

Figure 2.3.17: *trnL-trnF* data set: Strict consensus tree of 94 Most Parsimonious trees. Gaps considered as a 5th state (CI = 0.727; RI = 0.900; Length = 11; Bootstrap support above branches) or as binary characters (CI = 0.727; RI = 0.900; Length = 11; Bootstrap support below branches).

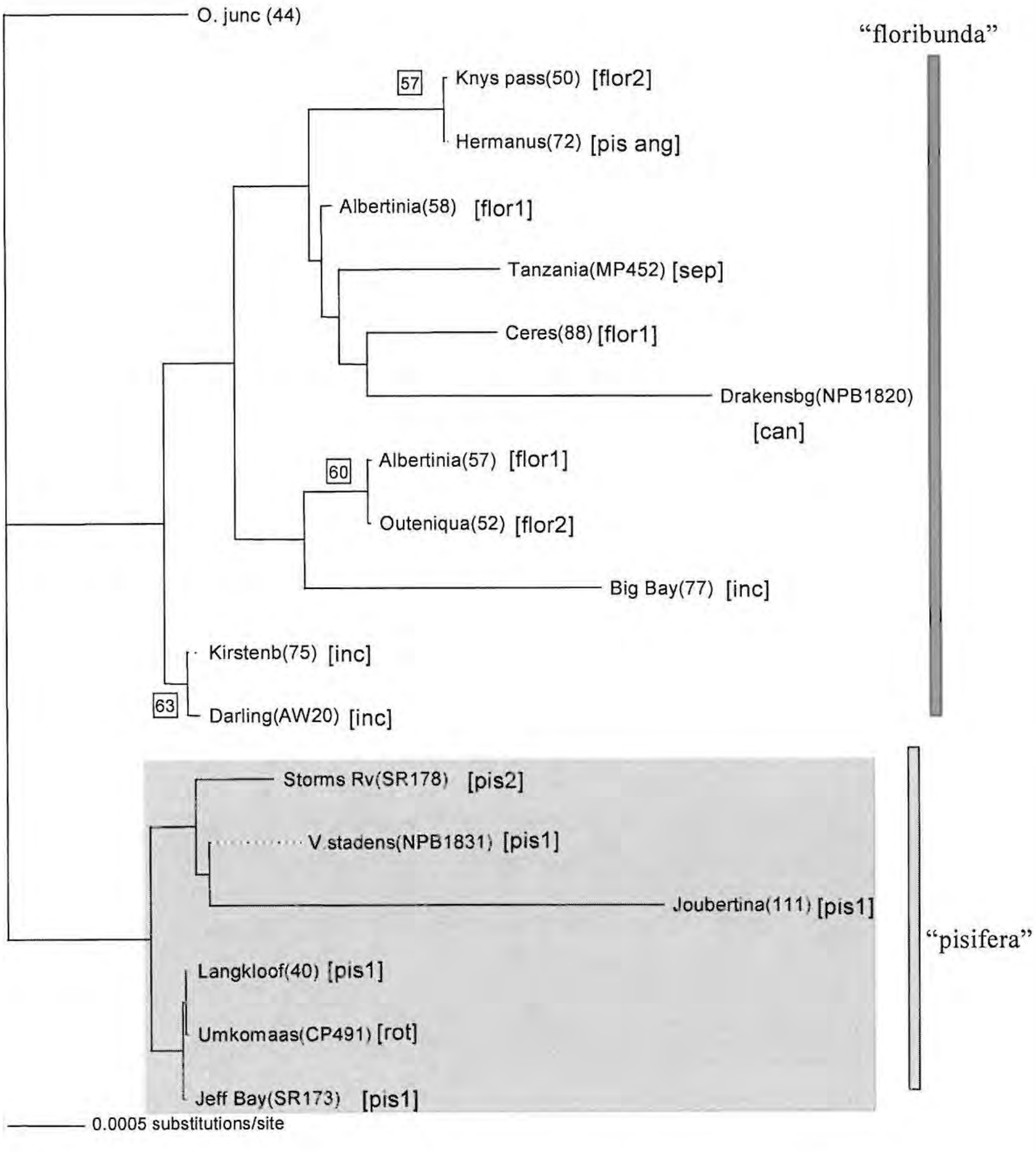

Figure 2.3.18: *trnL-trnF* data set: Jukes-Cantor Neighbor Joining distance tree (Bootstrap support above branches).

Maximum Likelihood analysis

For the "trnL-trnF" data set, ModelTest identified the best model as the Tamura & Nei (1990) model, with some sites assumed to be invariable (TrN+I). Specific settings were: Lset; Base=(0.3369 0.1891 0.1816); Nst=6; Rmat=(1.0000 4.1739 1.0000 1.0000 1.8208); Rates=equal; Pinvar=0.9371.

Figure 2.3.19 shows the Maximum likelihood tree for the "trnL-trnF" data set. The topology is very similar to the strict consensus tree obtained when gaps were considered as missing data (Figure 2.3.16) but with greater resolution in the "pisifera" clade (which is paraphyletic), and with the unusual placement of one *C. incana* sample ("Big Bay(77)") in a more terminal clade while the other two *C. incana* samples are basal to the "floribunda" clade. It should be noted that there is no significant bootstrap support for any of the branches and thus, any conclusion drawn from the tree's topology should be made cautiously.

Statistical Parsimony analysis

Figure 2.3.20 shows the TCS gene network for the "trnL-trnF" data set with gaps considered as a 5th state. Figure 2.3.21 shows the TCS gene network for the "trnL-trnF" data set with gaps considered as missing data. Several haplotypes collapse into one node (indicated by several names within one node) due to identical sequences. While both networks have some reticulation in their structure, considering gaps as missing data results in much more reticulation in the network's topology, and some unusual placements of haplotypes in the network (e.g. the Tanzania sample "452", placed close to "58" but only a single mutation away), suggesting that the *trnL-trnF* region by itself may not be all that phylogenetically informative.

Analyses of the combined cpDNA data set

Parsimony analysis

Figures 2.3.22 and 2.3.23 are the Strict consensus parsimony trees following analysis of the combined cpDNA data set, with gaps considered as missing data (Fig 2.3.22), or as a 5th character state or binary characters (Fig 2.3.23). Analysis of data with gaps considered as missing data results in 168 most parsimonious trees (Length = 19; CI = 0.737; RI = 0.922), whereas the analysis with gaps included results in 1927 trees. When gaps considered as a 5th

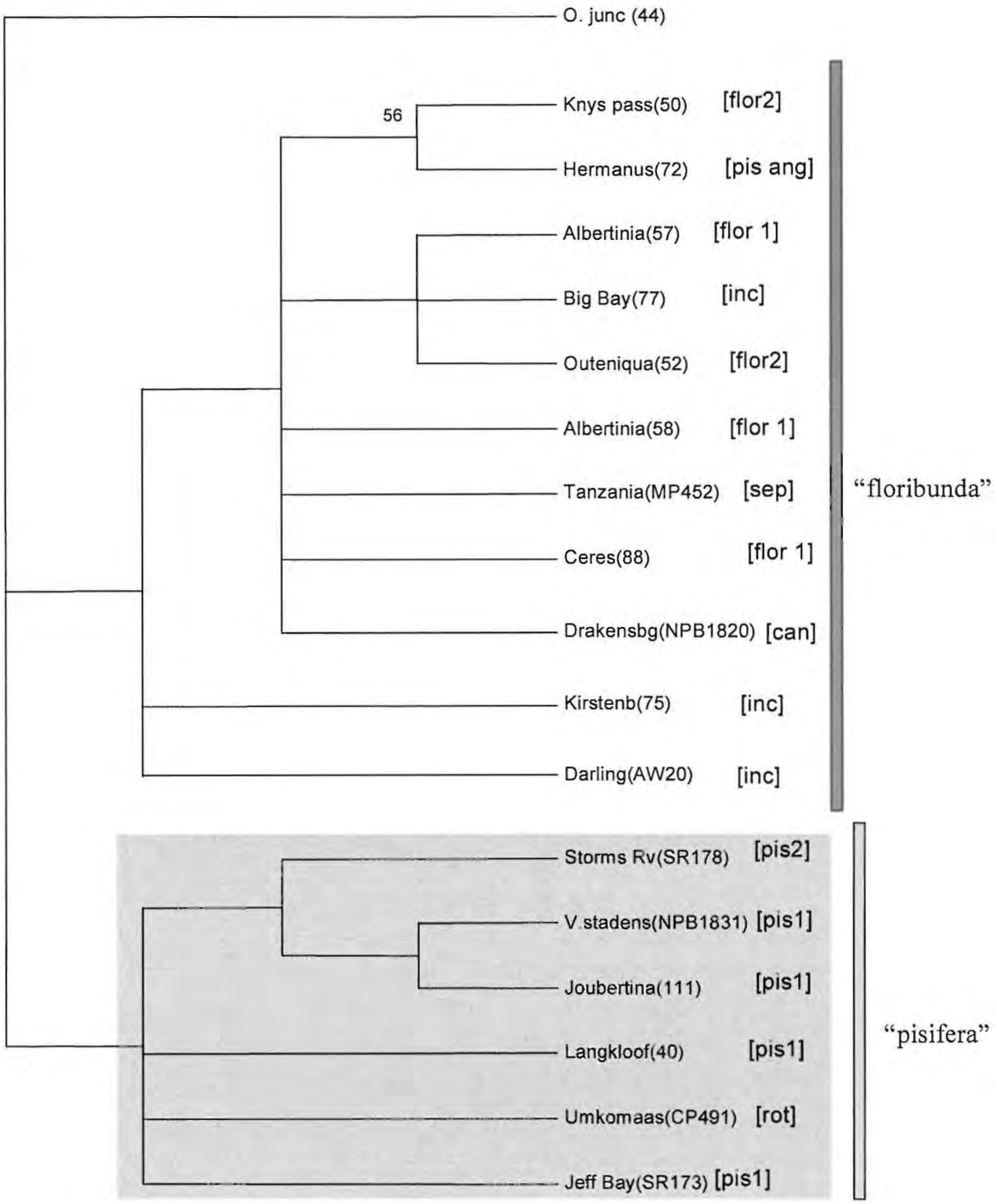

Figure 2.3.19: *trnL-trnF* data set: Maximum Likelihood tree. (No significant bootstrap support).

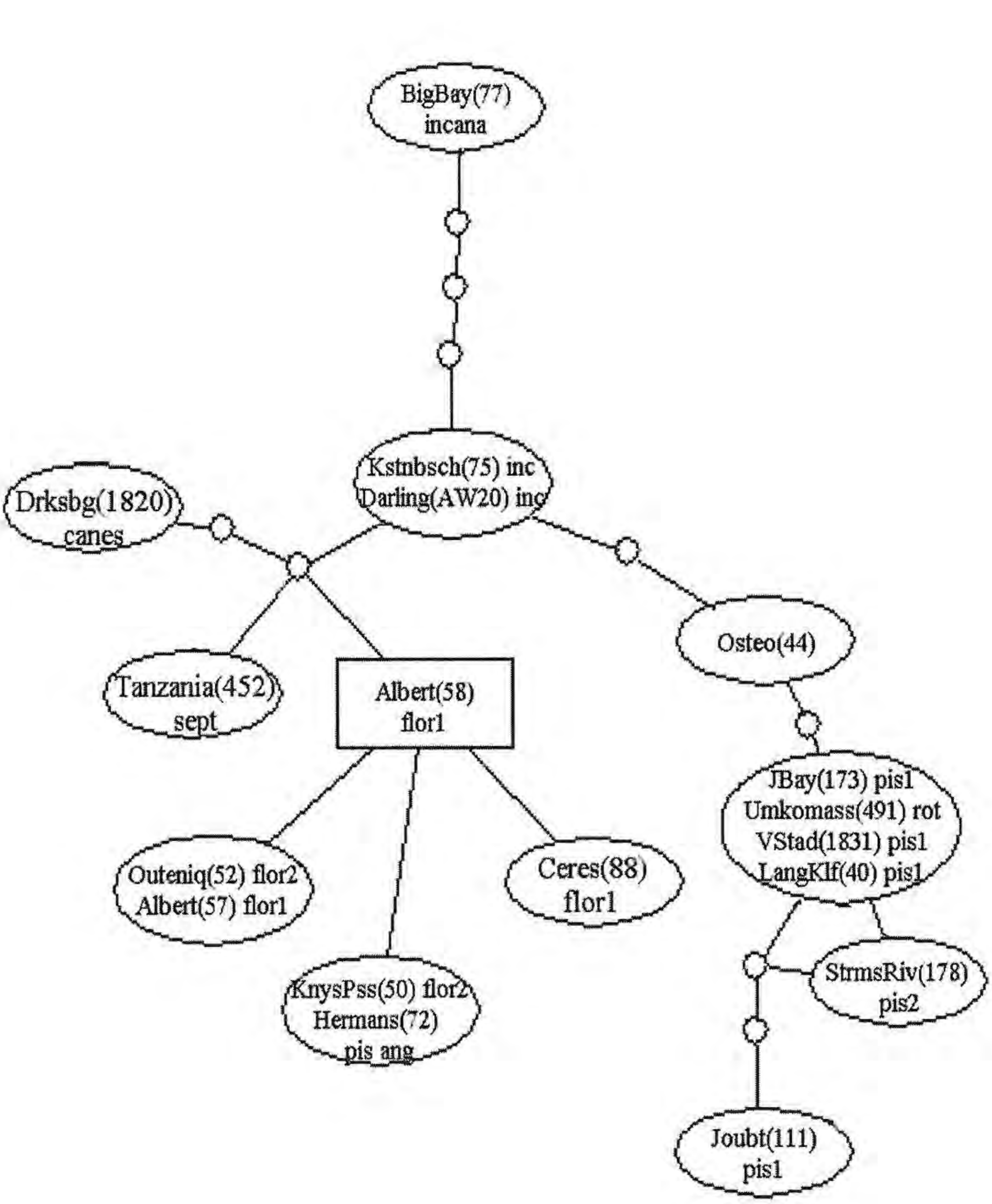

Figure 2.3.20 : *trnL-trnH* data set: TCS gene network. Gaps considered as a 5th state.

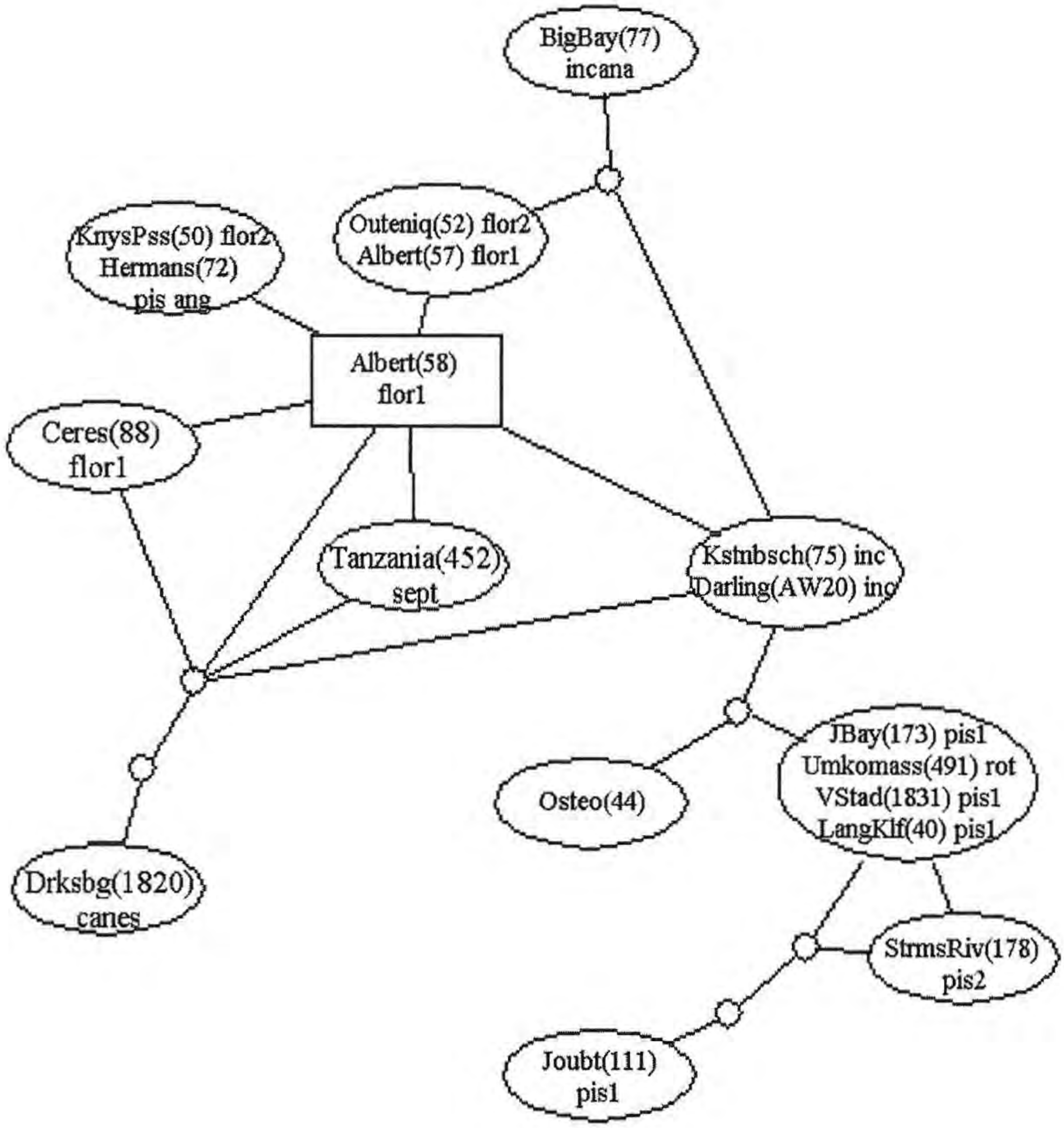

Figure 2.3.21: *trnL-trnF* data set: TCS gene network. Gaps considered as missing data.

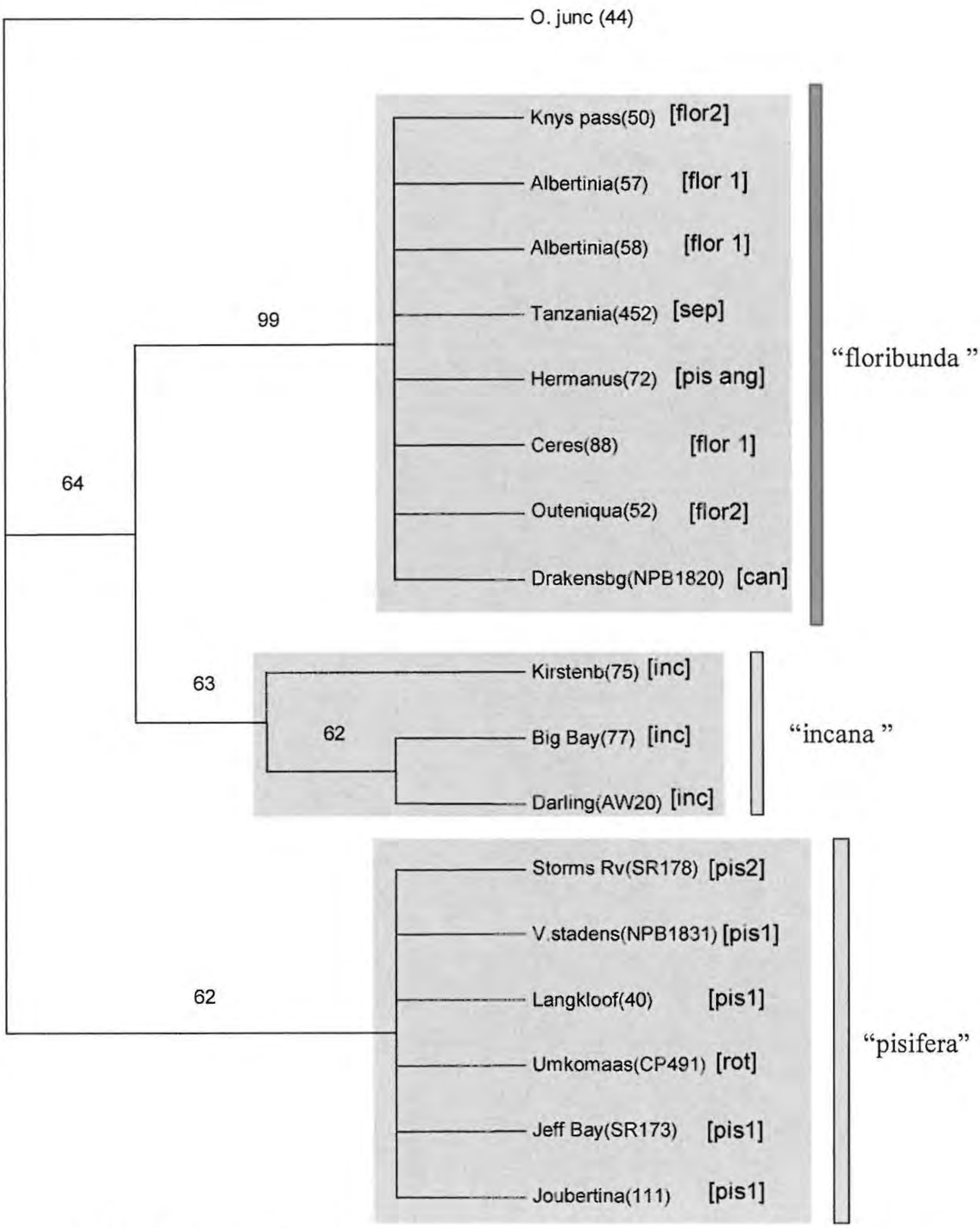

Figure 2.3.22: Combined cpDNA data set: Strict consensus parsimony tree of 168 most parsimonious trees. Gaps considered as missing data CI = 0.737; RI = 0.922; Length = 19. (Bootstrap support above branches).

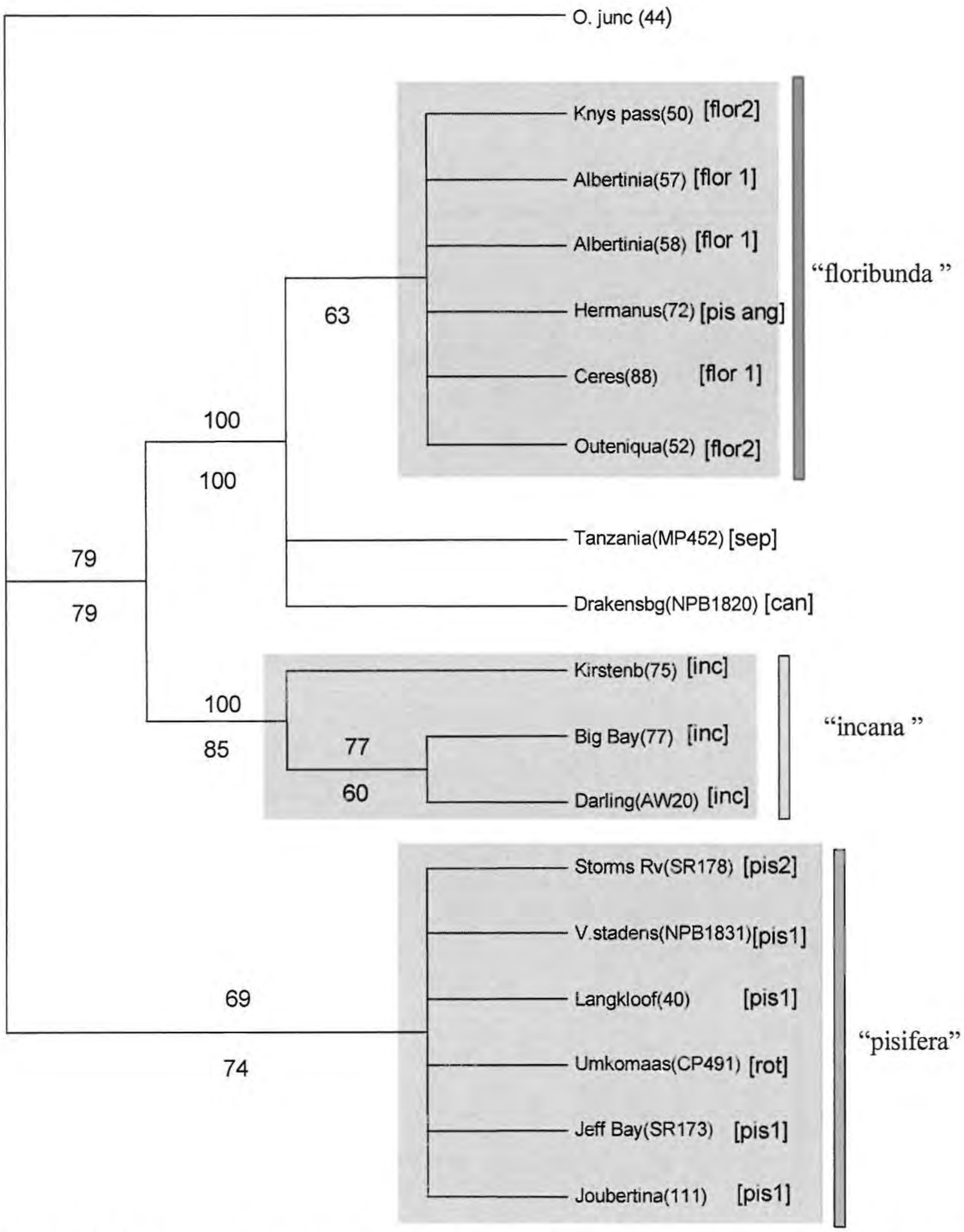

Figure 2.3.23: Combined cpDNA data set: Strict consensus tree of 1927 most parsimonious trees. Gaps considered as a 5th state (CI = 0.881; RI = 0.972; Length = 59; Bootstrap support above branches) or as binary characters (CI = 0.786; RI = 0.937; Length = 28; Bootstrap support below branches).

state, the trees have a CI = 0.881, RI = 0.972 and a length of 59, where as gaps considered as binary characters results in trees with CI = 0.786; RI = 0.937 and a length of 28.

Combining the two sets of data provides more resolution to the phylogenetic tree, and some higher bootstrap support values, especially for the tree with the gaps considered as a 5th state, indicating not only that combining the data provided a stronger phylogenetic signal, but also that the gaps are important phylogenetically informative characters in the chloroplast data.

The combined data set rendered three main branches in the tree. The two "floribunda" and "incana" branches both have 100% support (in the analysis with gaps considered as a 5th state or as binary characters, but less in the tree with gaps considered as missing data) indicating that the gap characters were important in defining the support for the branches, along with the higher RI and CI values for the tree with gaps included in the analysis. The "floribunda" clade branch has all of the *C. m. floribunda* samples (and one *C. m. pisifera* sample), with the *C. m. septentrionalis* and *C. m. canescens* samples as basal to the "floribunda" clade.

The *C. incana* clade is strongly supported (100%) in the tree with gaps considered as a 5th state, (85% supported when gaps are considered as binary characters) but much less supported (only 53%) in the tree with gaps considered as missing data, indicating again that the *psbA-trnH* 14-base insert characters are important in defining the *C. incana* clade. Again, the lower bootstrap support for the *C. incana* branch when gaps are considered s binary characters is a result of the higher weighting that 5th state coding gives to large gaps.

The second major branch (69% supported in the tree with gaps considered as a 5th state, 74% when gaps are considered as binary characters and 62% in the tree with gaps considered as missing data) has most of the *C. m. pisifera* samples and the *C. m. rotundata* sample.

Neighbor Joining analysis

Figure 2.3.24 shows the Neighbor Joining distance tree following analysis of the combined cpDNA data set using the Jukes Cantor correction. Again, three branches are present, (i) the "*pisifera*" clade (with 67% support), (ii) the "*floribunda*" clade (with 99% support) and (iii) the "incana" clade, with 61% support). The "*floribunda*" clade was two branches within it, neither with any significant bootstrap support.

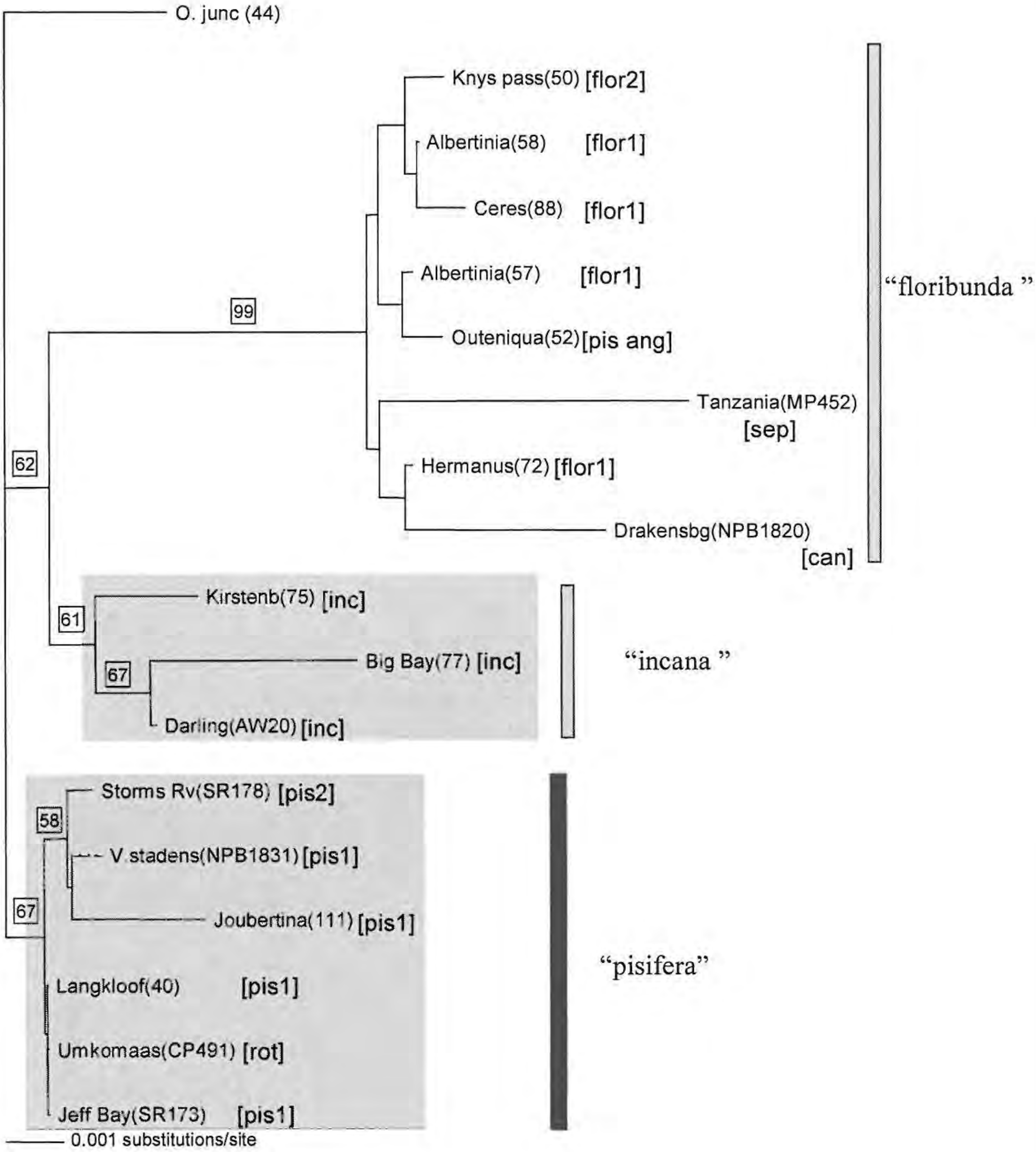

Figure 2.3.24: Combined cpDNA data set: Jukes-Cantor Neighbor Joining Distance tree (Bootstrap support above branches).

Maximum Likelihood analysis

For the combined cpDNA data set, ModelTest identified the best model as General Time Reversible model with some sites assumed to be invariable (GTR+I; Rodriguez *et al*, 1990). Specific settings were: Lset; Base=(0.3300 0.1593 0.1726); Nst=6; Rmat=(0.7352 1.6520 0.0958 1.1867 0.7575); Rates=equal; Pinvar=0.8900.

Figure 2.3.25 shows the Maximum likelihood tree following analysis of the combined cpDNA data set. The topology is one again very similar to the strict consensus parsimony tree with gaps considered as missing data (Figure 2.3.23) but also with greater resolution for some of the branches. This resolution does not come with very much bootstrap support and thus may not be all that informative.

Statistical Parsimony Analysis

Figure 2.3.26 shows the TCS gene network for the complete cpDNA data set with gaps considered as a 5th state. Figure 2.3.27 shows the TCS gene network for the cpDNA data set with gaps considered as missing data. The network where the gaps were considered as a 5th state results once again in three separate networks (similar to that in Figure 2.3.14, due to the large insertions and deletions in the *psbA-trnH* region). The inclusion of the "trnL-trnF" data set has however provided more resolution to the "*Osteospermum-pisifera-rotundata*" clade.

The network with gaps considered as missing data is a single network that makes good evolutionary sense, especially if one considers the *Osteospermum* sample as a rooting point for the network. From this rooting point (outgroup) the network bifurcates into a *C. incana* branch and a *C. monilifera* branch. Even if the gaps characters have been excluded, they can still be mapped back onto the network. The *C. incana* branch has a 14-base insertion that differentiates it. The *Osteospermum* node, the *incana* branch and the "pisifera-rotundata" nodes of the branch all lack the four-, six- and 10-base deletions that characterise the derived F-S-C grouping. This derivative nature of the F-S-C clade is clearly shown in the network (despite the lack of gap characters in the analysis), with the most geographically isolated and genetically disparate *C. m. septentrionalis* and *C. m. canescens* nodes branching off from a *C. m. floribunda* node. The characters that do separate the *C. m. canescens* and *C. m. septentrionalis* samples are all substitutions, not indels

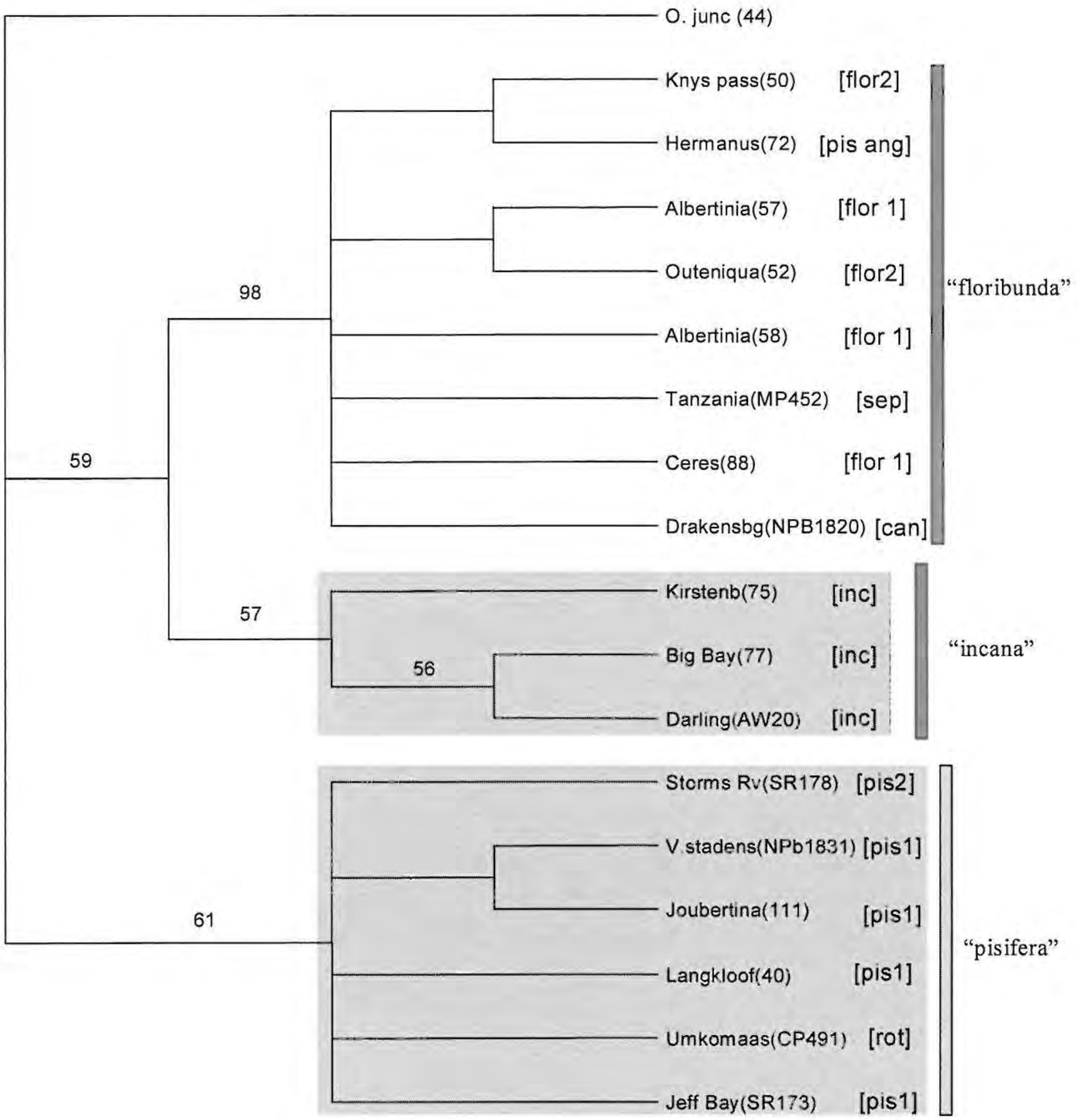

Figure 2.3.25: Combined cpDNA data set: Maximum Likelihood tree. (Bootstrap support above branches).

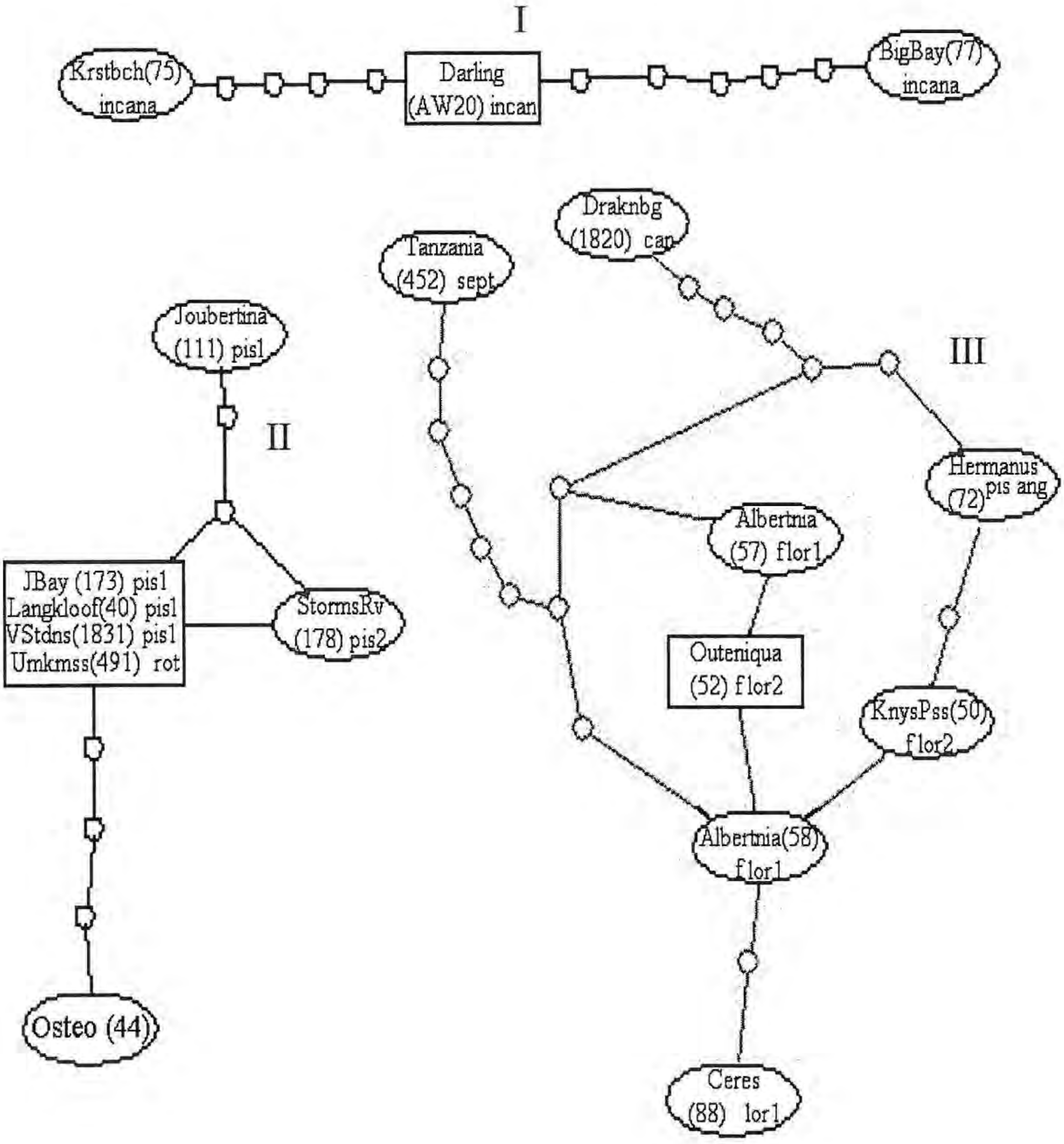

Figure 2.3.26: Combined cpDNA data set: TCS gene network. Gaps considered as a 5th state.

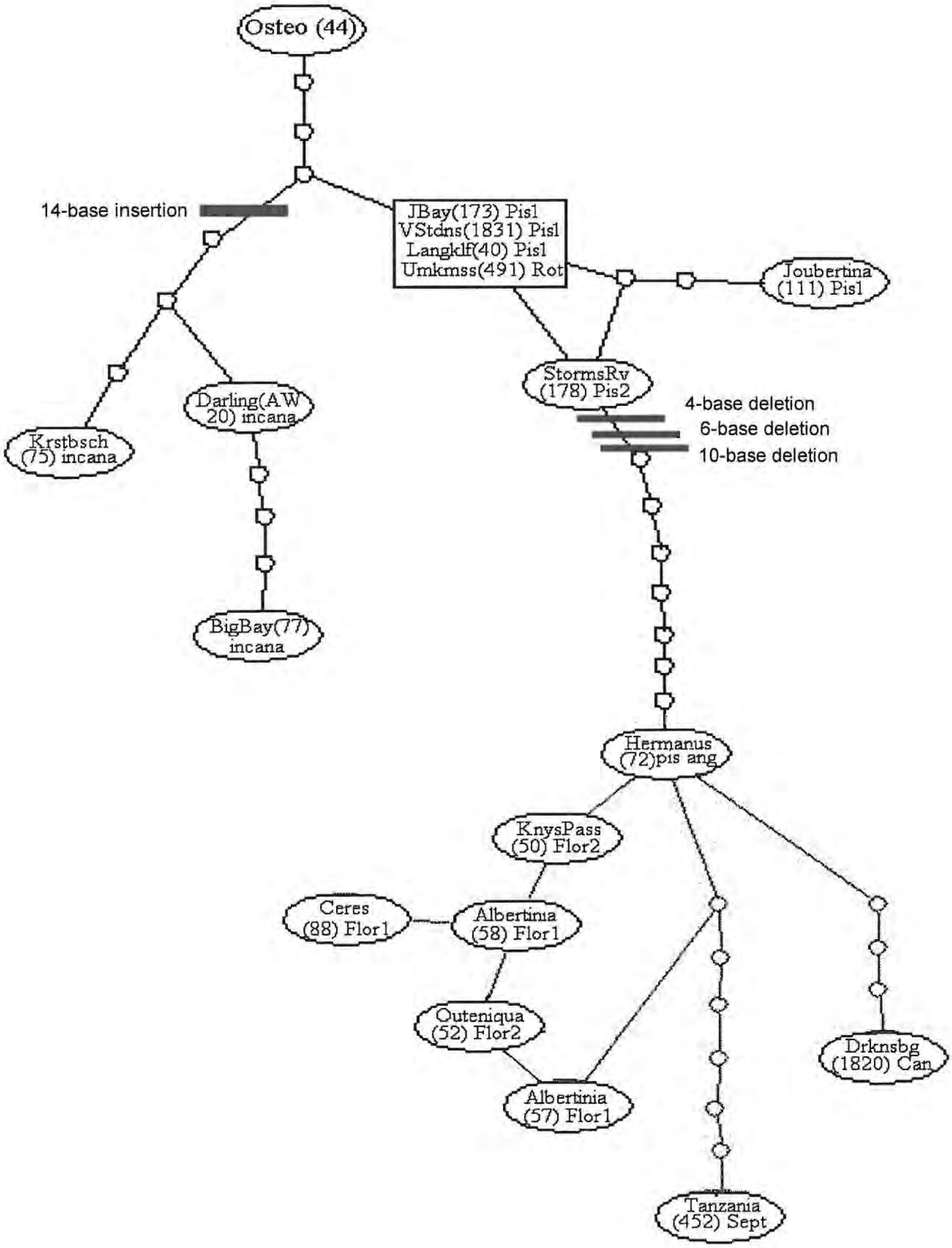

Figure 2.3.27: Combined cpDNA data set: TCS gene network. Gaps considered as missing data.

Discussion

The introductory chapter outlined four aims of this thesis. Three of these are addressed in this chapter. Each of these aims is now discussed, with additional issues arising out of this research addressed later in this section.

The identification of ESUs within *Chrysanthemoides*

Genetic data are able to differentiate a number of clades within the species *C. monilifera* (Figure 2.3.2). These clades correlate with "species" (e.g. *C. incana*); "subspecies" (e.g. *C. monilifera rotundata*) and "forms" (e.g. *C. monilifera pisifera pisifera* form 1) as well as entities below form level (e.g. the split in *C. monilifera floribunda* form 1 into two different clades). This multi-level correlation suggests that the use of the term "ESU" is warranted in the context of this discussion.

ESU identification requires that each ESU have unique molecular characteristics. Table 2.4.1 below lists the defining characteristics of each ESU in each DNA region from Figure 2.3.2 (as well as for the "West" branch in the ITS data noted in Figure 2.3.2, 2.3.5 and 2.3.7). *psbA-trnH* provides defining characters for the *C. monilifera pisifera* "subspecies", and the *trnL-trnF* regions provide characters that define *C. monilifera pisifera* and *C. monilifera floribunda*, but neither region provides separation of "subspecies" into "forms". ITS data provides the highest number of defining molecular characters for each ESU (except for C. m. *floribunda* form 1 ("Wide") which has no defining characters in any data set. The characters that do define *C. monilifera floribunda* in the *psbA-trnH* region define the "subspecies" as whole).

Table 2.4.1: Defining molecular characteristics (insertions, deletions and substitutions) of each ESU from each data set and the West branch.

	psbA-trnH			trnL-trnF			ITS		
ESU	Ins.	Del.	Subst.	Ins.	Del.	Subst.	Ins.	Del.	Subst.
C. incana	1x14		1				1x3	1	9
C. m. septentrionalis			5			1		1x1, 1x2	9
C. m. canescens			2			2			8
C. m. rotundata									2
C. m. pis. pis. fm 1				1x1		2			4
C. m. pis. pis. fm 2									1
C. m. floribunda fm 1	1x1	1x10	5		1x1	1			
C. m. floribunda fm 1 (Nar)	1x4								2
C. m. floribunda fm 2	1x6								2
C. m. monilifera									2
C. m. pis. angustifolia									2
"West" Branch								1x14	4

The NJ analysis is the only analytical method to resolve these ESUs, and so much of the discussion that follows is based on the results from that analysis (Figure 2.3.2).

Correlation of ESUs with morphological entities

When the clades of the tree in Figure 2.3.2 are compared to morphological designation, there is clear correlation for some branches: C. *m. monilifera* is a well-supported group (78%) as is *C. m. pisifera pisifera* form 1 (92%). Others are not so well supported (e.g. the *C. m. rotundata* clade, the *C. m. floribunda* form 1 "narrow" clade).

The branch lengths within the "West" branch are much shorter and there is a rather confusing structure within this branch. Figure 2.4.1 below is a simplified illustration of the basic structure of the tree with regard to where each clade is placed in the tree: the most basal clade of the West branch is *C. m. pisifera pisifera* form 2, then *C. m. pisifera angustifolia*, then *C. m. floribunda* form 1 (wide), then *C. m. floribunda* form 1 (narrow) and finally, *C. m. monilifera,* the most derived clade. The designations of "narrow" and "wide" refer to the widths of the leaves (and are not noted by Griffioen, but were noted by myself in examination of the samples), being either a slightly narrower-leafed form or a wider-leafed form of *C. m. floribunda* form 1.

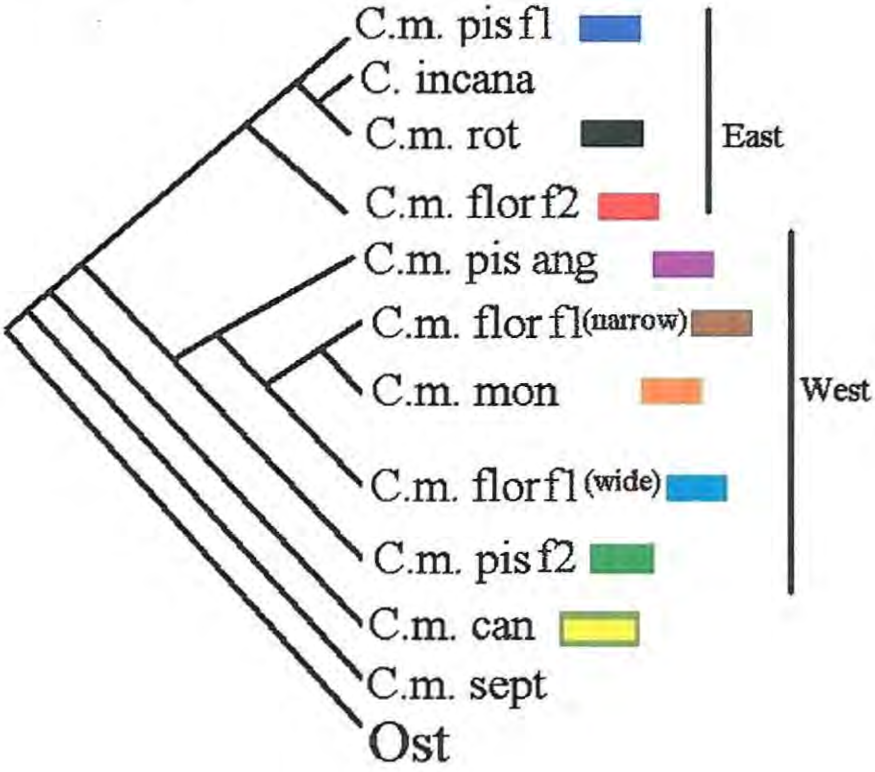

Figure 2.4.1: Simplified structure of Neighbor Joining tree from Figure 2.3.4. Colour blocks for maps sample sites in Figure 2.3.6.

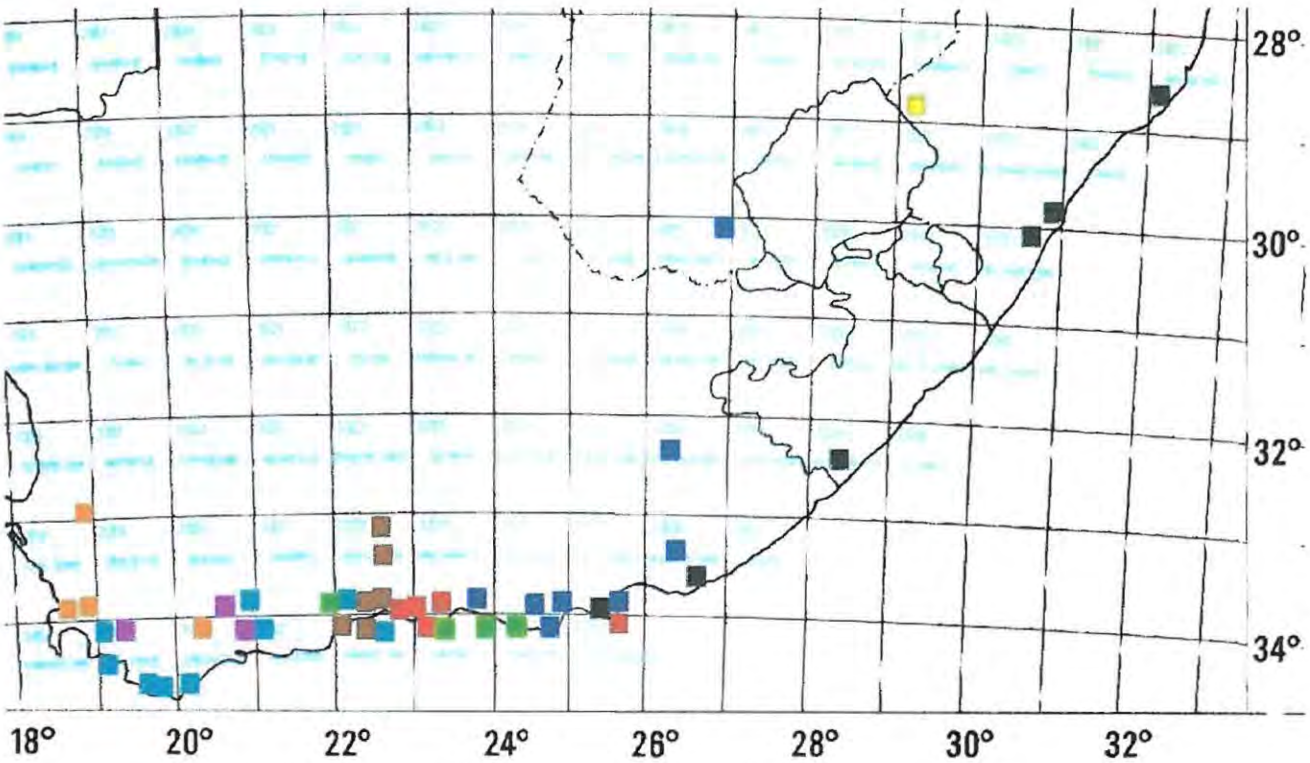

Figure 2.4.2: Map illustrating distributions of sample sites, colour coding taken from Figure 2.3.5.

C. m. floribunda form 1 "narrow" (coded brown on the map in Figure 2.4.2) occurs only within a narrow band that stretches from Prince Albert in the north areas to Mossel Bay in the south (see Figure 2.4.2a). There is some overlap with the other "wide" form (coded purple), especially in the De Hoop area, where, of four samples collected ("SH 102", "103", "104" and "105"), one sample ("104") was the narrow-leafed form and the other three were wide-leafed forms. The narrow-leafed form's localities stretch over the Outeniqua Mountains (though it should be noted that a series of passes through the mountains do link the two sides, and the samples were taken from roadsides). This clade shows slightly longer phylogenetic tree branch lengths in the NJ tree, and this, in combination with some minor morphological differentiation and its limited geographical distribution suggests that it is subject to some limits in gene flow that may illustrate the initial stages of divergence as a result of isolation.

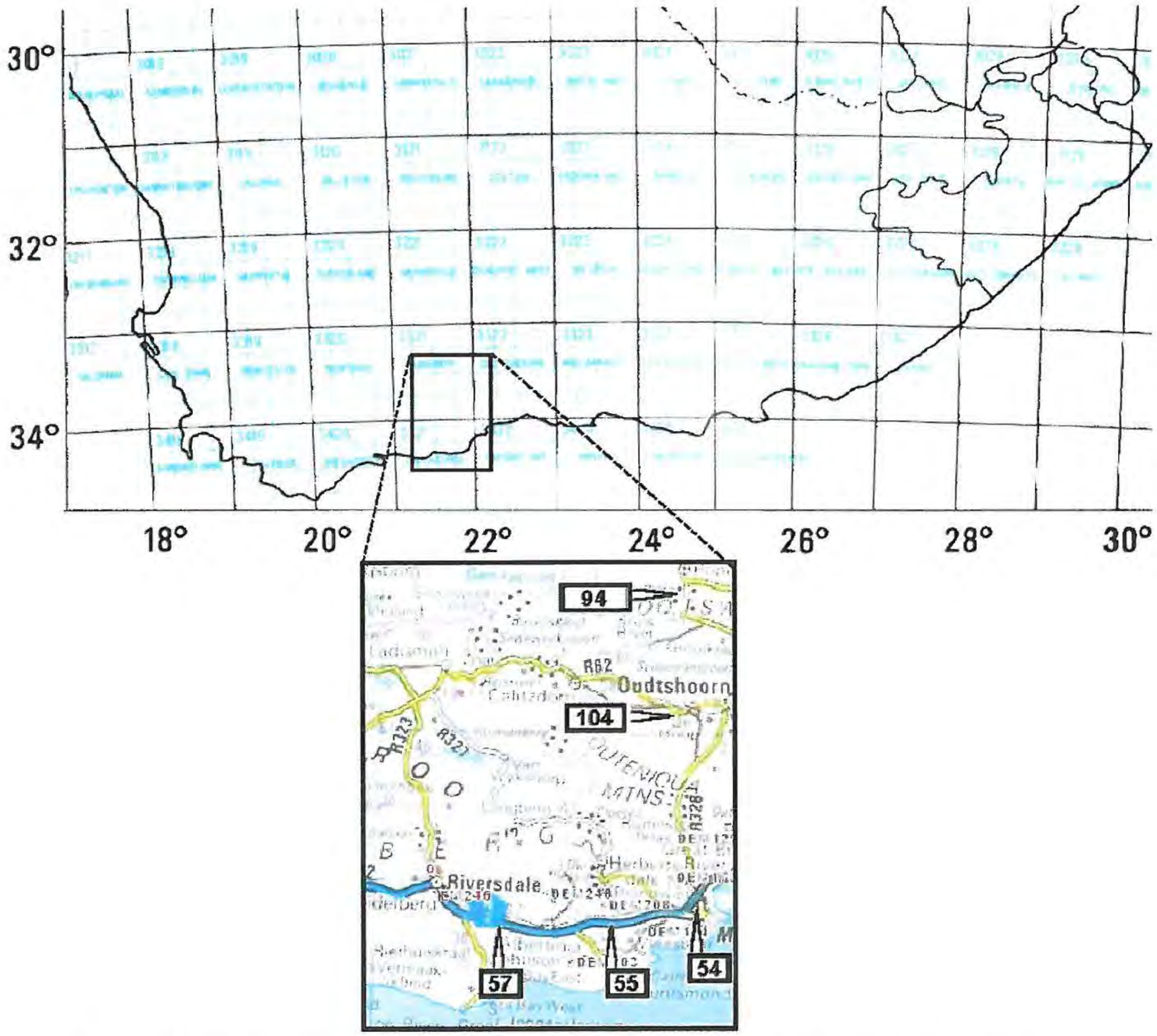

Figure 2.4.2a: Map illustrating distributions of *C. monilifera floribunda* form 1 "narrow" sample sites (coded brown in Figure 2.3.6).

Multiple correlations between entities derived from the genetic data and Griffioen's morphological "subspecies" and "forms" were found, and there are one or two instances of greater definition within subspecies than recognised by Griffioen's morphological designations. *C. monilifera septentrionalis* and *C. monilifera canescens* both had strong support (100%) as distinct clades in all the "Complete ITS" and "*Monilifera* ITS" data set trees. Most of the rest of the definition at the subspecies level is seen in the Neighbor Joining trees of the "Complete ITS" and "*Monilifera* ITS" data sets: (i) *C. monilifera monilifera* forms a distinct clade with good support (75-78%), (ii) *C. monilifera rotundata* forms a distinct clade with only moderate support in the "Monilifera" data set (67%), (iii) *C.*

monilifera pisifera pisifera form 1 *sensu* Griffioen forms a distinct clade with strong support (82-88%), (iv) *C. monilifera pisifera angustifolia* has moderate support (70%) in the "Monilifera" data set, (v) *C. monilifera floribunda* form 2 is a distinct, but unsupported clade, (vi) *C. monilifera floribunda* form 1 is split into two clades: a narrow-leafed (discussed in detail above, also Figures 2.4.1 and 2.4.2) and a wider-leafed form that had a wider geographical spread (this split was not observed in Griffioen's morphological designations). No samples were collected from the *C. monilifera pisifera borealis* variety.

The *C. incana* samples formed distinct clades in most of the trees, except for the "Limited" ITS data set NJ tree (Figure 2.3.8), where the split is between "75" (*C. incana incana gracilis*(?)) which is located in the "East" branch and "77" (*C. incana incana* var. *incana*) and "AW20" (*C. incana subcanescens*) which group together in a more basal position on the tree. The relationships of C. incana are discussed in detail later.

Correlation of ESUs with geographical areas and/or climatic factors

The maps in Figures 2.4.1 to 2.4.8 clearly illustrate a correlation between the genetically defined clades, and their geographical distributions. These genetically defined clades not only correlate with morphologically defined entities, but also with the distribution patterns observed for the morphologically defined entities in Griffioen, 1995).

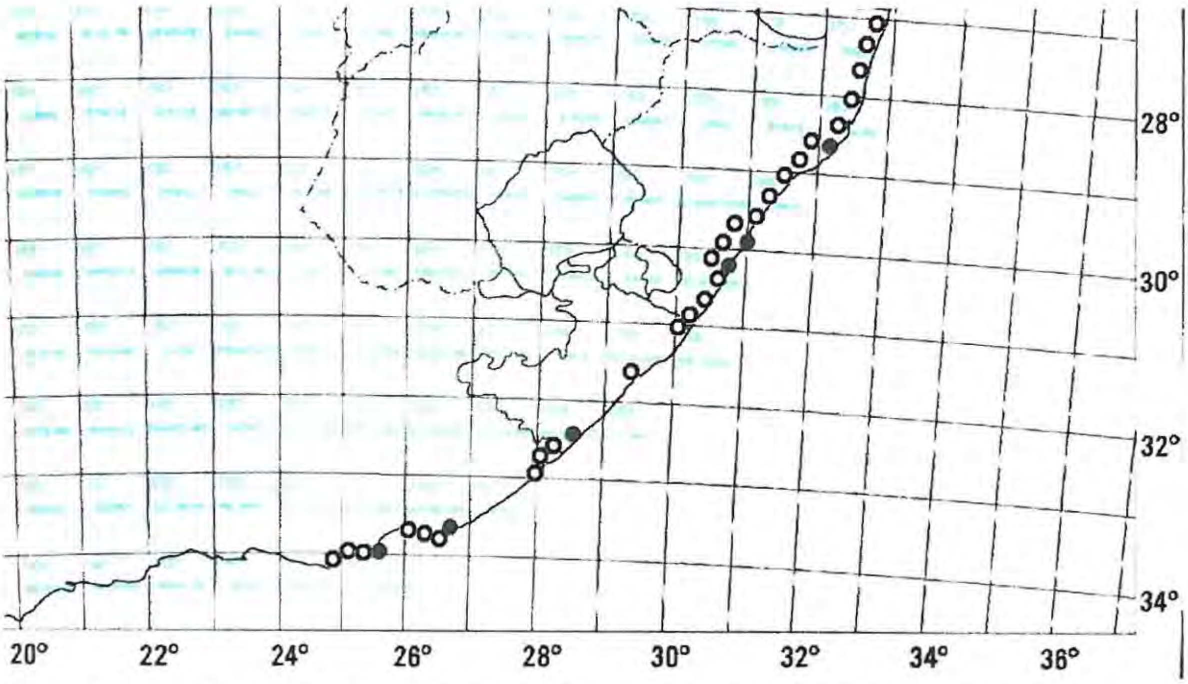

Figure 2.4.1: Distribution map for sample sites for *C. m. rotundata* ESU (black filled circles indicate samples collected for this project, empty circles indicate sample sites from Griffioen, 1995).

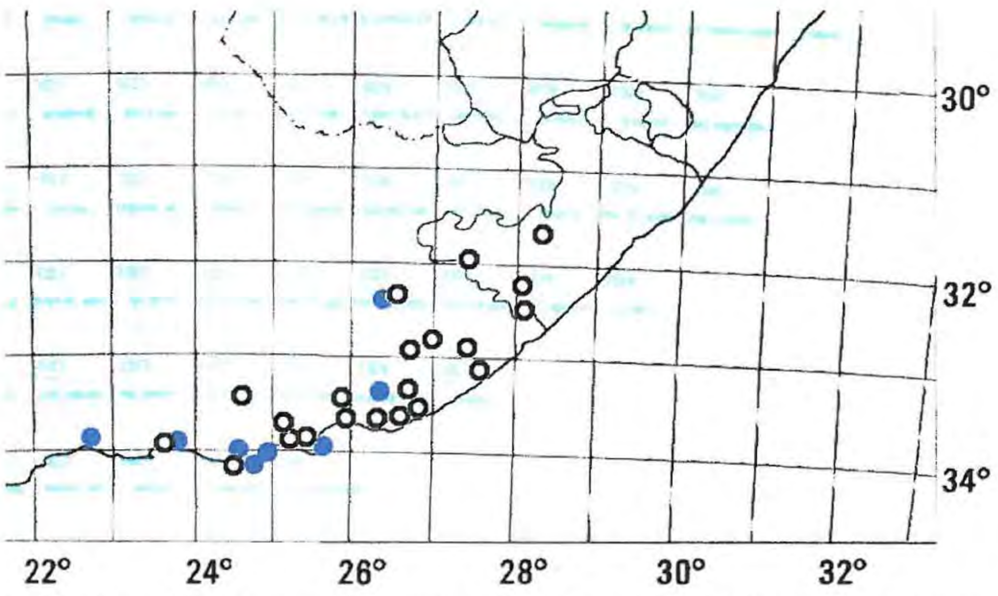

Figure 2.4.2: Distribution map for sample sites for *C. m. pisifera pisifera* form 1 ESU (blue filled circles indicate samples collected for this project, empty circles indicate sample sites from Griffioen, 1995).

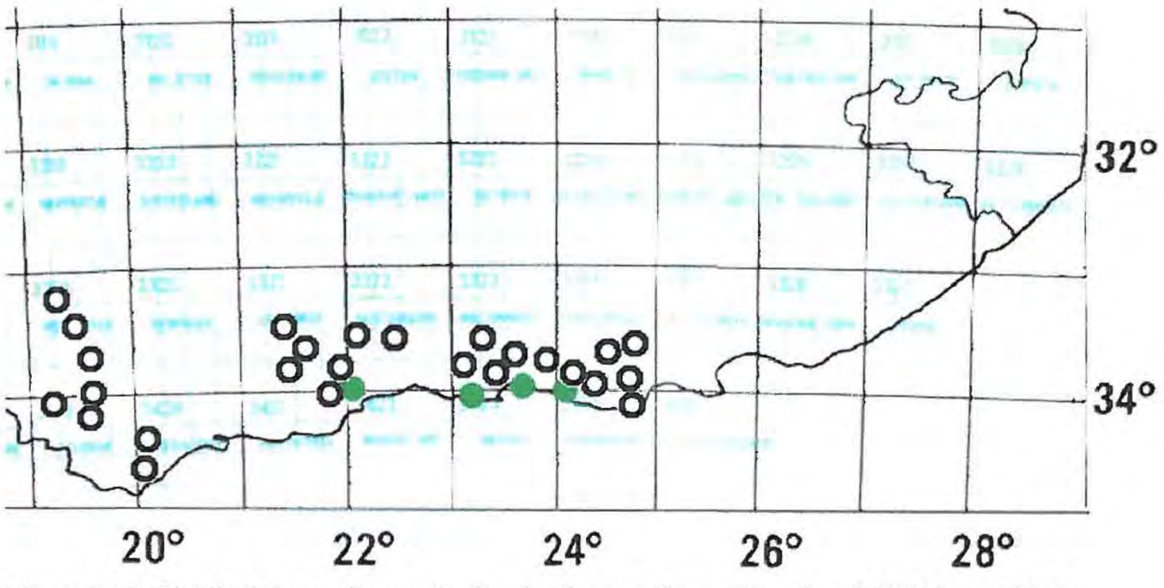

Figure 2.4.3: Distribution map for sample sites for *C. m. pisifera pisifera* form 2 ESU (green filled circles indicate samples collected for this project, empty circles indicate sample sites from Griffioen, 1995).

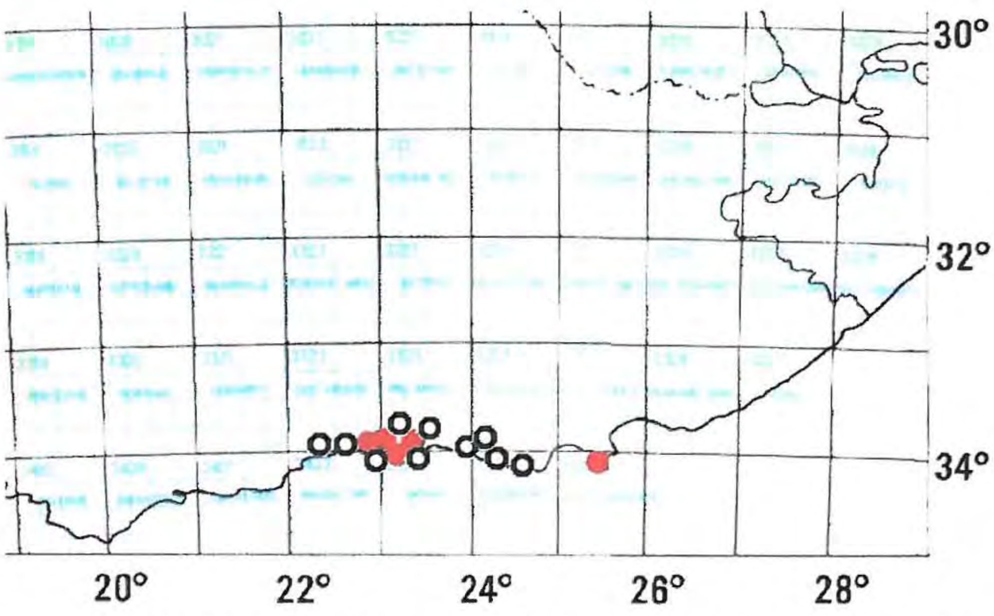

Figure 2.4.4: Distribution map for sample sites for *C. m. floribunda* form 2 ESU ESU (red filled circles indicate samples collected for this project, empty circles indicate sample sites from Griffioen, 1995).

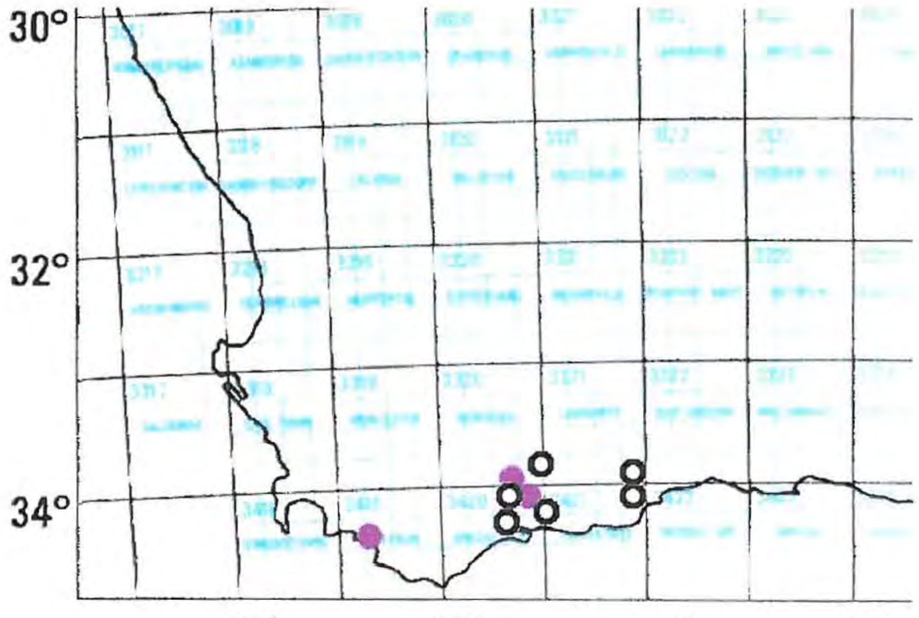

Figure 2.4.5: Distribution map for sample sites for *C. m. pisifera angustifolia* ESU (purple filled circles indicate samples collected for this project, empty circles indicate sample sites from Griffioen, 1995).

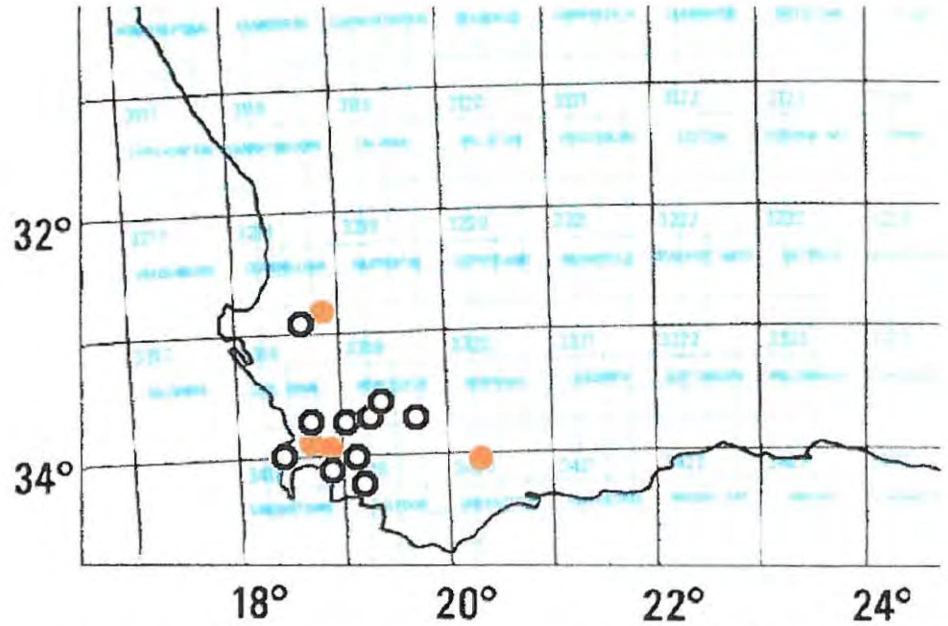

Figure 2.4.6: Distribution map for sample sites for *C. m. monilifera* ESU (orange filled circles indicate samples collected for this project, empty circles indicate sample sites from Griffioen, 1995).

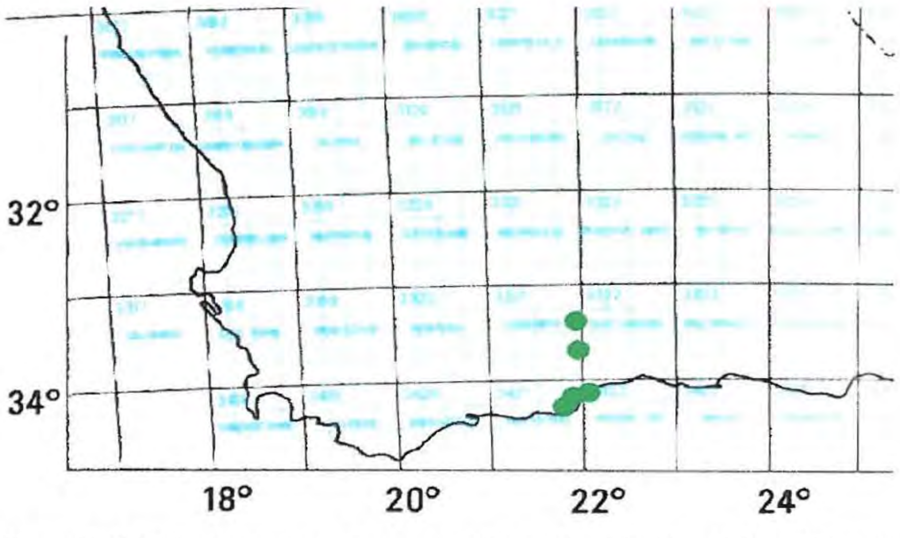

Figure 2.4.7: Distribution map for sample sites for *C. m. floribunda form 1* "narrow" ESU (green circles indicate samples collected for this project, samples from Griffioen presented in Figure 2.4.8).

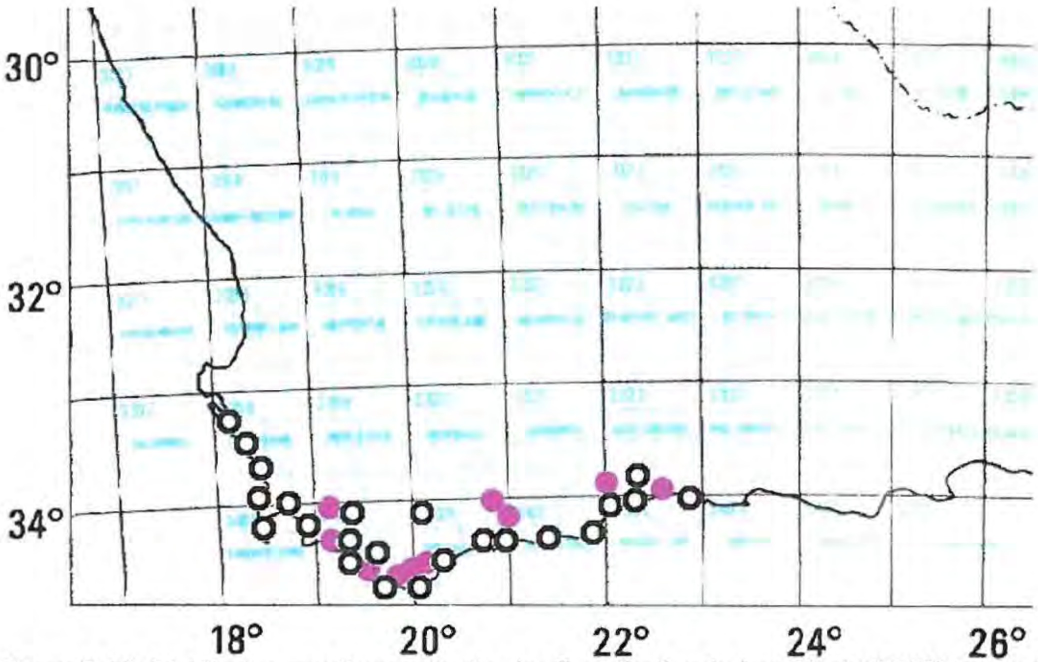

Figure 2.4.8: Distribution map for sample sites for *C. m. floribunda form 1* "wide" ESU (purple filled circles indicate samples collected for this project, empty circles indicate sample sites from Griffioen, 1995).

One of the most interesting findings from this work is the existence of a major dichotomy into an East and West clade (bootstrap support exist for the West branch only, Figure 2.3.2) The East-West geographical split seen in the ITS data (See Figure 2.3.2, illustrated on Figure 2.4.9 below) is also observed in *Euphorbia* subsect. *Meleuphorbia* (See Figure 2.4.10 below; Ritz *et al*., 2003). Ritz *et al.*'s (2003) study of Euphorbia species sampled plants in both the Western and Eastern Cape provinces (from the Little and Great Karoo respectively), and a split was observed in the phylogeny based on ITS data of species of section Meleuphorbia. This split was found to correlate to the geographical origin of the samples (Figure 2.4.10).

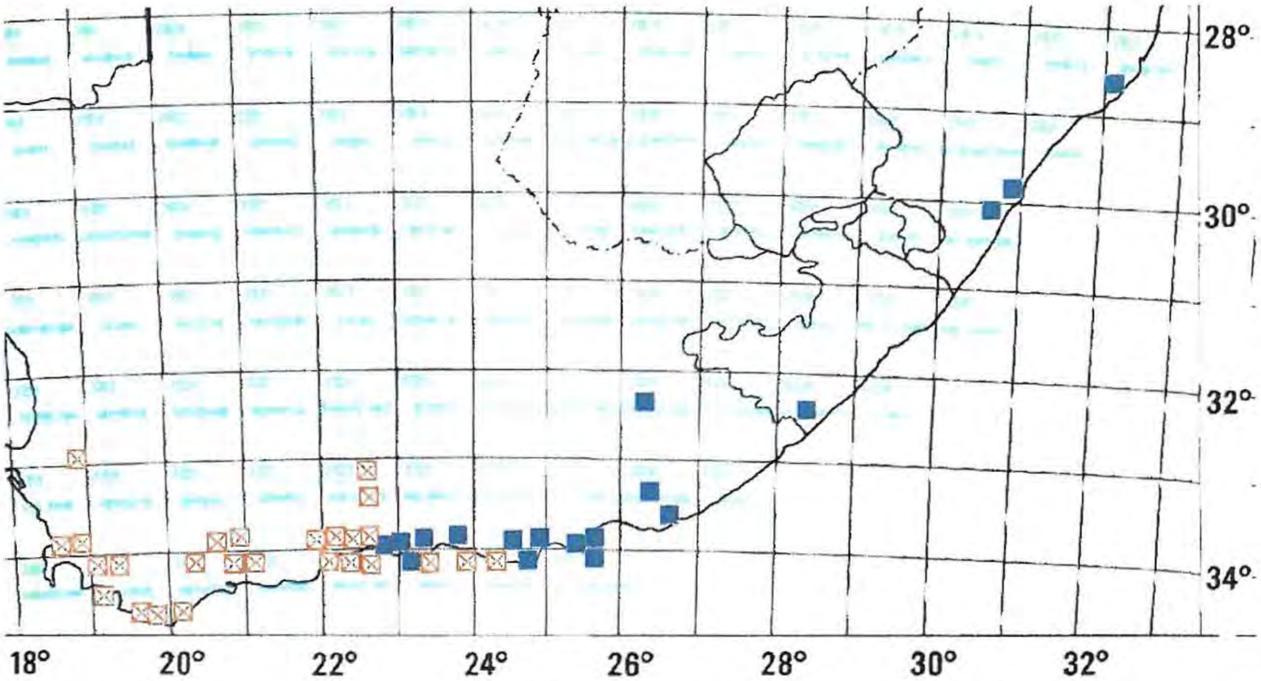

Figure 2.4.9: Distribution map for East/West split in *Chrysanthemoides* ITS sequence data. Solid blue squares show data sample collection points from "East" branch of tree (including *C. m. rotundata, C. m. pisifera pisifera* form 1 and *C. m. floribunda* form 2 ESUs). Orange squares with cross show data sample collection points from "West" branch of tree (including *C. m. monilifera, C. m. pisifera pisifera form 2, C. m. pisifera angustifolia,* and *C. m. floribunda* form 1 ESUs)

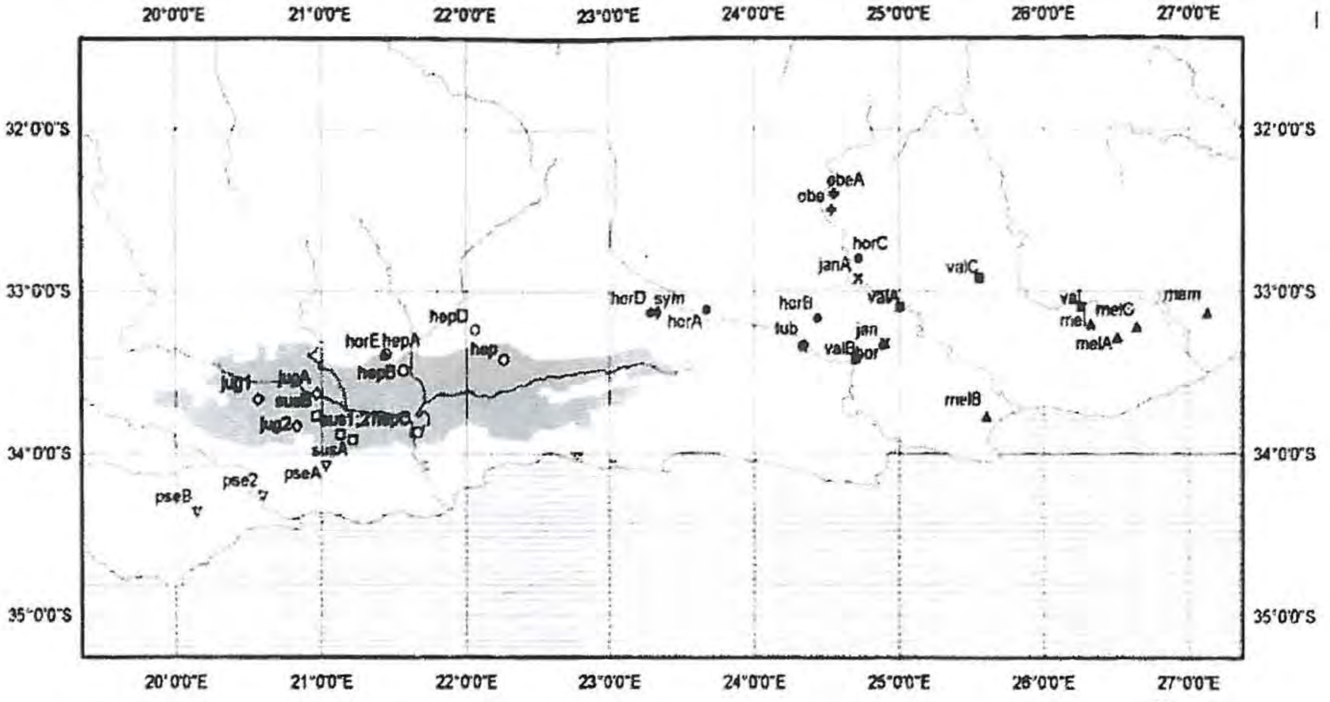

Figure 2.4.10: Distribution of data from Ritz *et al.*'s (2003) study of Euphorbia taxa. Filled symbols indicate taxa belonging to the Eastern clade, empty symbols indicate taxa belonging to the Western clade. The shaded area indicates the extent of the Little Karoo (exact species designations and full names of taxa abbreviations are to be found in Ritz *et al.*'s paper).

Ritz *et al.* (2003) attribute the geographical split in *Euphorbia* to differing rainfall patterns, with the western areas having winter rainfall and the eastern areas having summer rainfall. Although *Chrysanthemoides* flowers nearly all year round, the peak flowering time is during the highest rainfall period. This difference in peak flowering time could result in reproductive isolation between winter and summer rainfall zones. This is a possible explanation for the main split observed in the *Chrysanthemoides* data set, but at the more fine scale of distribution disjunctions, other factors (such as soil characteristics) may also play a role (Griffioen, 1995). Patterns of speciation in the Cape Flora (Linder & Vlok, 1991; Cowling & Holmes, 1992) revealed distribution patterns localised to soil types and water availability and Griffioen (1995) does note that certain subspecies of *Chrysanthemoides* are usually found on certain types of soils. Alternatively, rainfall quantities may also affect ESUs. An example of this instance may be the split between *C. monilifera pisifera pisifera* form 1 and *C. monilifera floribunda* form 2 (both part of the East branch). The two ESUs occur on opposite sides of the Tsitsikama mountain ranges; both have a similar yearly seasonal rainfall pattern, but differ greatly in the amount of rain received (see Chapter 3 for detailed climate diagrams).

Mutations in each of the two isolated clades (E/W) could only accumulate once the current pattern of rainfall seasonality had been established. It can thus be expected that the *Euphorbia* "species" and *the Chrysanthemoides* "subspecies" have had roughly the same period of time in isolation (and thus have diverged fractionally morphologically, as well as genetically, although rates of sequence evolution may differ in the two groups).

Comparing the data from Ritz *et al.*'s (2003) study to the *Chrysanthemoides* data, ITS data of *Euphorbia* were found to have 53/546 phylogenetically informative bases for 20 taxa (9.7% of all the characters were informative), as opposed to 20/754 bases for 18 samples (from 8 taxa) in *Chrysanthemoides* (2.7% of all the characters were informative). The species of *Euphorbia* that were investigated all showed strong morphological similarity, mirroring their low level of interspecies sequence divergence, and Ritz *et al.* (2003) note that the taxonomic status of many of the species are controversial, especially with regard to the assignment of "species" or "subspecies" status to several taxa. Barker *et al.* (2003) note that the species concepts applied in one group can differ to that in another, giving rise to different phylogeographic interpretations depending on how much variation a particular taxonomists is prepared to accept within the "species".

The succulent *Euphorbia* specimens studied by Ritz *et al.* (2003) have a much lower effective distance for seed dispersal than *Chrysanthemoides*, and Ritz *et al.* (2003) suggest that limited seed dispersal (combined with small range size) would prompt rapid parapatric speciation, driven by environmental changes. Goldblatt and Manning (2002) note that long-distance seed dispersal by birds (as shown by *Chrysanthemoides*) is one of the least common dispersal methods in the Cape Flora, and is especially rare in plants on nutrient-poor substrates. They also note that there is an assumption that plants on such soils cannot afford to allocate resources to protein-rich drupaceous fruits. *Chrysanthemoides* tends to grow in disturbed areas (and by roadsides), mostly (but not always) in fairly fertile soils (Griffioen, 1995) and may not entirely fit this model.

Goldblatt and Manning (2002) also highlight the correlation between seed dispersal mechanism and genus size: genera with small dry wind-dispersed seeds (such as *Osteospermum* and *Muraltia*) have a high number of species in the genus, with each species having a limited distribution; whilst genera with fleshy, widely-dispersed seeds tend to have wide ranges and few species in the genus (e.g. *Chrysanthemoides* (a close relative to *Osteospermum*) and *Nylandtia* (a close relative to *Muraltia*), both of which have only two species per genus, each with wide distributions). Thus, the *Euphorbia* species with limited seed dispersal and narrow ranges would have built up genetic divergence faster than the widely distributed and widely seed-dispersed *Chrysanthemoides*, which experiences a higher rate of gene flow between populations, (thus increasing the occurrence of genetic interchange and slowing the rate of genetic divergence between taxa). This may explain the greater amount of sequence variation in *Euphorbia* than in *Chrysanthemoides* noted in the comparison above.

The "*C. incana* problem" and evidence for hybridization

C. incana is a problematic taxon, for a number of reasons. Firstly, the presence of multiple polymorphic copies of the ITS that make DNA sequence reading from trace files difficult, secondly, the effect that the inclusion of C. *incana* sequences have on the phylogeny of *C. monilifera* (rendering that species paraphyletic), and thirdly, the incongruence between the cpDNA and nrDNA phylogenies.

Multiple paralogues

Initial efforts at sequencing the ITS-1 region of the C. *incana* samples proved problematic. Closer inspection of the sequencing trace files revealed two disparate sequences, one overlaying the other, both present in approximately equal quantities. The PCR product was run on an agarose gel in an attempt to separate the two paralogues. Two bands were seen on the gel (Figure 2.4.11), one longer product band, brighter in intensity, and a shorter product band, much fainter. The larger product band was cut out of the gel and cleaned using the Promega Wizard SV Gel and PCR purification kit.

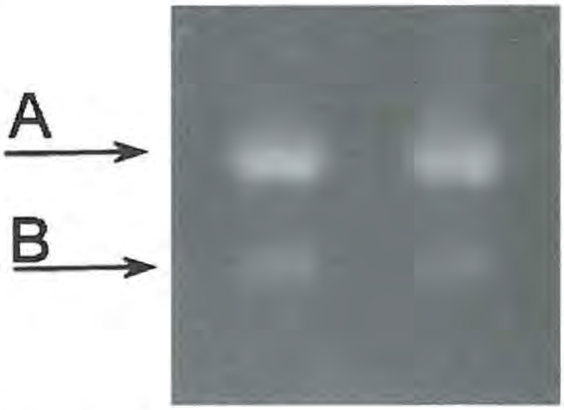

Figure 2.4.11: Agarose gel of *C. incana* samples showing larger brighter PCR product band (A) and shorter fainter PCR product band (B).

This cleaned product was sequenced again, with slightly cleaner sequences, but still showing two peaks at every base beyond a certain point in the sequence (Figure 2.4.12).

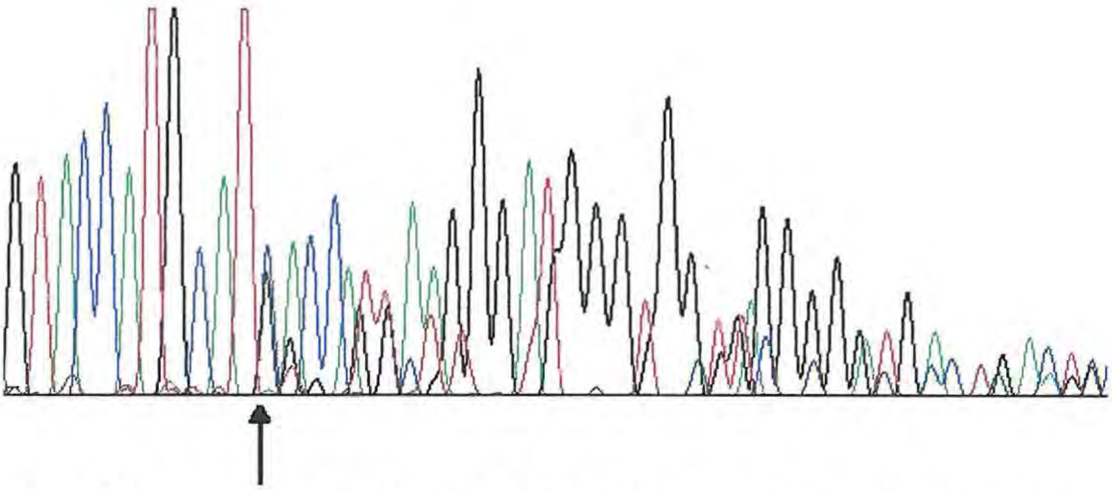

Figure 2.4.12: Sequence trace file showing nucleotide additivity at multiple bases in a *C. incana* sample beyond a certain point in ITS-1 (to the right of the arrow).

In an effort to obtain sequences of the different paralogues, attempts at cloning were made. Unfortunately, these met with no success. Manual comparison of the sequences with a consensus sequence from other *Chrysanthemoides* ESUs proved to be successful in retrieving up to 80 bases from the point of misalignment commencement of each sequence. From this approach it was apparent that the second paralogue has a three-base deletion that caused the misalignment in the two paralogues when they were sequenced (see Figure 2.2.6). This three-base insert is present in none of the other *Chrysanthemoides* ESUs that were sampled and sequenced, suggesting that the origin of this particular paralogue is most likely from another unsampled entity and may be present in the *C. incana* genome due to a hybridization event. This foreign paralogue was omitted from the "Complete ITS" data set so that only the one sequence obtained (as outlined below) from *C. incana* was included.

No other sequence in the data set has this three base insert, so an obvious source of a hybrid origin (within the genus) of these variant multicopy genes is not immediately clear (i.e. if the sequence insert had occurred in another of the subspecies within the *Chrysanthemoides* genus, it could logically be pointed to as one of the parents in a hybridisation event.)

Longer and clearer sequence data were obtained by means of two allele-specific internal primers (Chrys-I(1) and Chrys-I(2); Table 2.2). However, the internal primers were located too close to the region of DNA aimed for and the resulting sequences obtained stopped ten bases short of the region that contained the insert (See figure 2.2.6) that was responsible for the misalignment in the two sequences. The possibility of further internal primer designs was curtailed due to time constraints.

Bailey *et al.* (2003) raised concerns that when multiple ITS paralogues are present in an individual, one may not be functional (due to lengthy deletions that can render the secondary structure non-functional), and that this non-functional copy (pseudogene), which will no longer be subject to functional constraints on the rate of sequence divergence, can complicate phylogenetic reconstructions. The presence of non-functional ITS copies can often be detected by the presence of comparable rates of mutation not only in the ITS-1 and ITS-2 regions, but also in the functionally constrained 5.8S region. If however, the rate of mutation in the 5.8S region is much lower than the spacers, then the paralogues most likely has retained its functionality (Bailey *et al.*, 2003). The 5.8S regions of both *C. incana* paralogues are identical, as are the 5.8S regions of all the *C. monilifera* samples.

Although the two different ITS copies were not sequenced to completion in all the C. *incana* samples, the presence of multiple paralogues in the ITS data, in combination with significant incongruence between the data sets from the two genomes is a strong indication that hybridization has taken place in the genus, and that *C. incana* is a result of a hybridization event. Yonemori *et al.,* (2002) considered the presence of several instances of nucleotide additivity (two different nucleotides at the same locus) as evidence of hybridization in the *Mangifera*. In addition to this, the presence of two ITS paralogues in *Mangifera* species strongly pointed to a hybrid origin of some species in the genus (Yonemori *et al.,* 2002). The presence of multiple ITS sequences has been used to clarify hybrid origin (due to maintenance of both parental ITS sequences) in other studies as well (Wissemann, 1999; Koch & Al-Shehbaz, 2000).

Incongruence between cpDNA and ITS

A single locus alone cannot generate a strong hypothesis of hybridization and a good hypothesis of hybridization should therefore derive from establishment of a correlation among several sources of data (Sang & Zhang, 1999). Combined use of cytoplasmic and nuclear markers allows introgression to be detected and distinguished from other phenomena, such as joint retention of the ancestral condition, clinal variation and convergence (Cros *et al.*, 1998). In particular, analysis of cpDNA variation has proven to be valuable for revealing possible cases of introgression (Cros *et al.*, 1998).

It should be noted that low-taxonomic-level incongruence can be caused by factors other than hybridization, such as lineage sorting (ancestral polymorphism) and paralogy (gene duplication) and a nuclear gene tree alone cannot distinguish among the three possible alternatives (Schilling & Panero, 1996; Sang & Zhang, 1999).

Harrison (1991) explains how both differential introgression and random sorting of ancestral polymorphisms can lead to discordance between gene trees and species trees. Introgression is defined by Harrison (1991) as hybridization between species (or gene exchange between populations) that may result in the incorporation (possibly fixation) of alleles from one species in the gene pool of the second species. Lineage sorting as defined by Harrison (1991) is "an occurrence of allelic variation within a species" and as with morphological characters, a taxon can share molecular markers of related taxa due to the joint retention of alleles following speciation in a polymorphic ancestor (Rieseberg, 1997). Lineage sorting may be problematic in that an allele phylogeny will not necessarily have the same topology as the species phylogeny. This is possible if polymorphisms persist through speciation events, leading to the probability that the gene tree and the species tree have the same topology may be quite small Harrison (1991).

Another possible reason for cpDNA and nrITS incongruence is that of chloroplast capture. Shaw (2002) suggests two processes that could explain the geographically local cpDNA capture: (i) rare hybridization events or (ii) frequent interspecific hybridization events. The first theory, rare hybridization events, followed by the selection of cpDNA haplotypes of the alternate species' genetic background, could cause the displacement of a species' cpDNA haplotype by that of its sympatric counterpart. Under a rare hybridization event scenario, the cpDNA molecule of each species is hypothesised to evolve by genetic drift for some time independently. Upon contact and hybridization, each species' cpDNA is exposed to the genetic backgrounds of the alternate species. One species' cpDNA variant may have higher fitness than the other, and a selective sweep across the interspecific boundary could result from this interaction. The second theory (frequent interspecific hybridization and selection against hybrids) could provide a continual conduit for cpDNA across the species boundary. Genetic drift could then cause selectively neutral cpDNA variants to fix in the genetic background of the other species. This process does require strong reinforcing selection to resurrect the parental genotypes in the zone of hybridization and could be assisted by biases in the direction of interspecific mating. Shaw (2002) states that definitive data to test between these two alternative theories do not exist but notes that a presence of intermediate hybrids in nature makes the hypothesis of frequent hybridization more tenable.

Rieseberg (1997) notes that while discordant organellar and nuclear phylogenies apparently due to hybridization are being reported with increasing frequency (Rieseberg *et al.*, 1996), the evolutionary outcome of most ancient cases of hybridization appear to be introgression, rather than hybridization. The distinction between introgression and hybrid speciation can be difficult to distinguish. Incongruence between cytoplasmic- and nuclear-based phylogenetic trees, for example, suggests that hybridization played a role in the evolution of the wild cotton, *Gossypium bickii* (Wendel *et al.*, 1991; Rieseberg, 1997). However, it is not clear whether this ancient hybridization event was important in its origin, as *G. bickii* does not appear to have a biparental nuclear genome typical of hybrid species (Rieseberg, 1997), something that is evident in the *C. incana* samples. Convergence of synthetic and ancient hybrid lineages toward a similar set of gene combinations also suggests that hybrid speciation may be more repeatable than previously believed (Coyne, 1996).

Sang & Zhang's (1999) investigation into *Paeonia* species revealed several instances of both incongruence between different data sets, as well as sequence polymorphism in nrITS regions. These lines of evidence were used to infer the hybrid origin of a number of *Paeonia* species. The polymorphic sequences presumably reflect the co-existence derived from those of the derived parents and the slowing of concerted evolution that should have lead to homogenization of the ITS sequences was attributed to longer generation times in *Paeonia* (Sang, 1996). The hybrid origins of several additional *Paeonia* species, which did not show ITS polymorphism, were inferred on the basis that they had discordant positions between the ITS and the cpDNA phylogenies, and this discordance was considered to have resulted from fixation of ITS sequences from one parent and inheritance of cpDNA from the other parent (Sang & Zhang, 1999).

Harrison (1991) notes that in the case of speciation by hybridization, recent origins appear to be the rule for detecting hybridization events as the homogenization of ITS sequences by forces of concerted evolution can occur quickly. There is a relatively narrow window of time in which hybrid species can be identified (Morell & Rieseberg, 1998) as artificial hybrids can show concerted evolution in a single generation (Coleman, 2003).

The *C. incana* samples have incongruent cpDNA and nrITS phylogenies, as evidence by the presence of significant incongruence between the ITS and the cpDNA data sets ($p>0.33$) detected by the ILD test. Furthermore, the disagreement that is obvious between the two sets of genetic information (nuclear vs. chloroplast) is often taken as an indication of a hybridization event at specific levels (Cros *et al.*, 1998; Fuertes Aguilar *et al.*, 1999; Soliva *et al.*, 2001; Lee *et al.*, 2002). This, as well as heterozygous ITS sequences, strongly point to a hybrid origin of *C. incana*. For *Chrysanthemoides incana*, the presence of more than one paralogue of ITS suggests that this "species" arose from a recent hybridization event

The fact that the species *Chrysanthemoides monilifera* is rendered paraphyletic (in analyses of both the ITS and the cpDNA) by the inclusion of the *C. incana* samples may be explained by the following scenario: *C. incana* originated by hybridization whereby one parent (probably paternal) would have come from the *C. monilifera rotundata* ESU. This is evidenced by the close relationship between the *C. incana* samples and *C. monilifera rotundata*. The maternal lineage is difficult to determine, but probably has a 14-base insertion in the *psbA-trnH* spacer, which is unique among the samples sequenced here (Figure 2.3.27), suggesting the maternal origin is not included in the samples studied. As all the entities of *C. incana* identified by Griffioen have not been incorporated, it is possible that the final identity of one (or both) parents might still await discovery, or from among the closely related species of *Osteospermum*.

Hybridization and hybrid speciation

The term hybrid can be restricted to organisms formed by cross-fertilization between individuals of different species, or it can be defined more broadly as the offspring between individuals from populations which are distinguishable on the basis of one or more heritable characters (Rieseberg, 1997). Study of phylogenetic reticulation is of particular interest because the frequency of natural hybridization in one of the most distinctive yet poorly understood elements of flowering plant diversity (Schilling & Panero, 1996). Botanists have speculated that the morphological and ecological novelty created by hybridization might allow hybrid populations to spread onto previously unoccupied adaptive peaks (Arnold, 1997; Rieseberg, 1997). Hybridization may have several evolutionary consequences, including (i) increasing intraspecific genetic diversity, (ii) the origin and transfer of genetic adaptation, (iii) the origin of new ecotypes or species and (iv)the reinforcement or breakdown of reproductive barriers (Rieseberg, 1997). Interspecific hybridization is often regarded as an important speciation mechanism and may provide the raw materials for adaptive evolution in rapidly changing environments (Rieseberg, 1997; Bleeker, 2003).

Rieseberg (1997) notes that homoploid hybrid speciation is unusual because not only does it involve hybridization between taxa at the same ploidal level, but it also represents a type of sympatric speciation, as the parental species must co-occur geographically to produce hybrids. An accurate reconstruction of hybrid speciation has long been difficult due to the lack of phylogenetic methods for reconstructing reticulate evolution (Sang & Zhang, 1999). If the new hybrid lineage is more fit than either parent in all habitats, it quickly replaces the parental species. However, if the fitness advantage of the hybrid lineages is restricted to a divergent habitat, then it must co-exist with its parent species and it is selection rather than chance that largely governs hybrid species formation (Rieseberg, 1997). Although there is not complete sampling for all the *C. incana* ESUs, all the samples that have been collected and sequenced are hybrids (according to the presence of paralogues). The *C. incana* subspecies (as defined by Griffioen, 1995) all grow in very low rainfall areas (50-500 mm pa) along the coast lines, with erratic rainfall patterns (with C. *incana incana rangei* deriving most of its moisture from regular sea fogs; Griffioen, 1995). This suggests that this hybrid represents an expansion into a niche that is not available to other C. *monilifera* ESUs. This ability to exploit the low rainfall, coastal environment may be due to an inherited genetic constitution from one or both of its parents.

C. *monilifera rotundata* occupies the Eastern Coast, and C. *incana* the Western Coast, in very similar coastal conditions. The placement of the *C. incana* clade (in both the ITS data and the cpDNA data) as derived from *C. monilifera rotundata* (which occupies sandy soil coastal areas with low rainfall (less than 500 mm pa)) could account for *C. incana*'s ability to occupy its current ecological niche. The current distribution of *C. monilifera rotundata* does not reach more westerly than Tsitsikama (at the farthest point of its range), but in the past, with lower sea levels and a different climate, the C. *monilifera rotundata* ESU may have been brought into contact with another ESU, and the subsequent hybridization that took place lead to the current *C. incana* ESU.

Both the ITS sequence data and the similar ecological niches that the *C. incana* and *C. monilifera rotundata* ESUs occupy point to the successful expansion of the novel hybrid between (possibly) C. *monilifera rotundata* and an unknown ESU into the Western Cape coastal niche that C. *monilifera rotundata* does not (or can not) currently occupy at present. The fact that C. *incana*, despite its purely Western distribution, groups in the "East" branch of the phylogenetic ITS trees (and is derived from a C. *monilifera pisifera* or C. *monilifera rotundata* ancestor in the cpDNA data) also supports this theory of novel niche occupation, rather than speciation from one of the "West" branch clades. This study has thus proposed more questions than answers in terms of *C. incana*, and a substantial amount of additional sampling is required before any further conclusions can be drawn.

The taxonomic status of *C. m. canescens* and *C. m. septentrionalis* and speciation by vicariance

The basal placement of the *C. monilifera canescens* and *C. monilifera septentrionalis* ESUs in the ITS trees and their strongly distal placement in the halpotype trees from the cpDNA data suggests that in terms of genetic distinction, these two ESUs should perhaps be reinstated to full species status rather than subspecies. The presence of incongruence between the two genomes in C. *incana*, present in these two ESUs as well, could indicate hybridization events in their phylogenetic histories. Their disjunct mountainous geographical ranges, distant from the rest of the *Chrysanthemoides* subspecies, also adds weight to this argument, as the isolation they are subject to would be a strong supporting factor in the advancement of their speciation. Traditionally, the most important factor causing the speciation is believed to be geographic isolation (Barraclough *et al.*, 1998).

There is still evidence of continued interbreeding at the edges of the range of *C. monilifera canescens* (Griffioen, 1995; rendering it a subspecies under the BSC), but the strong genetic distinction of the clade suggests species status is still viable. There is not yet enough strongly disjunct phylogenetic definition for other ESUs ("subspecies") to accord them species status.

The discordant placement of these two taxa in the trees from the different genomes need not necessarily be problematic. Fuertes Aguilar *et al.*, (1999b) noted the presence of a geographical, rather than a purely taxonomic, pattern in the ITS tree of *Armeria villosa* subsp. *longiaristata*. Samples of the subspecies were collected from three different study sites around Spain, and in the resulting ITS phylogenetic tree, these samples appeared in three of the five major clades, all with bootstrap support (Fuertes Aguilar *et al.*, 1999b).These grouping were more closely associated with geographical distribution, rather than taxonomic designation. Fuertes Aguilar *et al.*, (1999b) attribute this pattern to introgression and reticulate hybridization in combination with concerted evolution, resulting in the samples of *A. v. langiaristata* that have hybridized possessing ITS sequences that have experienced rapid concerted evolution to match the other parental genotype.

This correlation between ITS data and geographic patterns is certainly equally applicable to the *Chrysanthemoides* "Monilifera ITS" data set, with geographical groupings taking precedence over taxonomic designations (e.g. the splits in *C. monilifera floribunda* and *C. monilifera pisifera pisifera*). This suggests that the basal placement of the C. *monilifera canescens* and C. *monilifera septentrionalis* clades is a reflection of their disjunct and distant distributions, rather than purely a taxonomic theory of the temporal order of the derivation of each subspecies. The cpDNA, by contrast, traces the matriarchal lineages (Palmer, 1987) of the subspecies, illustrating the derivation of *C. monilifera canescens* and *C. monilifera septentrionalis* from a *C. monilifera floribunda* (or *C. monilifera pisifera angustifolia*) ancestor, which was in turn derived from a *C. m. pisifera* ancestor.

Whatever their origin, *C. monilifera canescens* and *C. monilifera septentrionalis* can be clearly identified as ESUs (according to the criteria of Moritz, 1994) using both cpDNA and nrDNA, as well as showing distinction based on morphological and geographical distribution data. They also satisfy Avise's (2000) quest for deep branch separation in a gene tree as a qualifier for ESU status, and fulfil the PSC-WT criteria for species status, by possessing diagnosably distinct genetic characters.

Speciation by vicariance

Levin (1993) notes that speciation occurs within a geographical context and that a geographically defined group of individuals diverges from others and accumulates, through selective and stochastic processes, the numerous attributes (genetic and phenotypic) that results in it being distinguished as a new species. He also notes that evolutionary biologists have traditionally classified speciation events on the basis of the geographic context in which initial divergence occurs. Allopatric, parapatric and sympatric models of speciation describe situations in which divergence occurs in geographically isolated populations, in contiguous but non-overlapping populations, or between subpopulations at a single locality (Levin 1991). The geographical race and the local population are thought to be the prime units of speciation in plants (Levin, 1993). Beyond the speciation event, two general processes will cause a group of organisms to evolve together (to have species cohesion): gene exchange and ecological equivalence (Hey *et al.,* 2003). Both processes, alone or together, can cause genetic drift and adaptations to be shared by a group of organisms, and cause that group to evolve cohesively and separately from other such groups (Hey *et al.,* 2003).

Levin (1993) explains the process of speciation as follows: isolates arise from the fragmentation of contraction of a species' range due to deterioration of the environment or through dispersal from the main geographic range of the species. Most isolates are likely to go extinct in a short period of time, especially if they occupy marginal conditions. Others will survive for hundreds or thousands of generations, and retain their character or change in minor ways. A minute fraction will evolve into a new species.

An example of speciation by subdivision (vicariance) in *Chrysanthemoides* could be that of *C. monilifera septentrionalis*. This ESU occurs in tropical montane areas at altitudes of 1500-2400m above sea level in northern South Africa, Zimbabwe, Malawi, Southern Kenya and Tanzania (See Figure 2.1.5 for distribution map; Griffioen, 1995). It has a disjunct distribution from the other ESUs, and it is also disjunct within its own population distributions. This ESU comes out as basal in the ITS sequence data, which agrees with the logical theory that the taxon that has been separated the longest would be the most divergent from the others, whilst the more closely distributed ESUs (those in the Western and Eastern Cape) would have far less divergence between them. The disjunction of populations within the *C. monilifera septentrionalis* taxon would certainly be very interesting to investigate, certainly in terms of degree of divergence between the various populations, as well as determining if there is a total disjunction, or if birds that fly long distances can disperse the seeds between populations.

Taxonomic status of the remaining C. *monilifera* ESUs and sympatric speciation

As for the remaining ESUs identified by the ITS phylogeny, there is not yet enough strongly disjunct phylogenetic definition for other ESUs ("subspecies") to accord them species status. The assignment of subspecies status to each ESU based solely on ITS data is problematic when there is discord with the cpDNA phylogeny, especially when the split of *C. monilifera floribunda* and *C. monilifera pisifera pisifera* that is observed in the ITS data is not present in the cpDNA data. Whilst there is a clear geographical split between the two forms within the two subspecies, the existing morphological similarity and the conflicting cpDNA phylogeny suggest that elevating the "forms" of *C. monilifera floribunda* and *C. monilifera pisifera pisifera* to "subspecies" level based solely on ITS data is not justification enough.

The lack of cpDNA differentiation between *C. monilifera pisifera pisifera* and *C. monilifera rotundata* and the lack of significant support for the *C. monilifera rotundata* clade from ITS data (only 63% in the "*Monilifera* ITS" data set, and none in the "Complete ITS" data set) suggests that the "subspecies" status of *C. monilifera rotundata* should remain as it currently stands.

Sympatric speciation

Levin (1993) summarizes the process of sympatric speciation: Virtually all scenarios for sympatric divergence start with a stable polymorphism affecting performance on different resources or in different habitats. Speciation involves the evolution of an association between this polymorphism and alleles at loci affecting sortative mating (often mediated by habitat or resource selection). If sympatric speciation occurs at a single locality within the range of a widespread ancestral type, the daughter species, having shifted to a new habitat or resource can then spread.

An example of this shift into a new habitat could be illustrated by *C. monilifera rotundata.* This ESU has a distribution limited to the Eastern Coast beachfronts, never more than 150m above sea level (See Figure 1.8 for distribution map; Griffioen, 1995). This could have occurred by the spreading of a new form of *Chrysanthemoides* into the beachfront areas near Port Elizabeth (which is nearly the Eastern most coastal occurrence of other each dwelling subspecies of *C. monilifera*). This new form of *Chrysanthemoides* could then spread up along the Eastern coast without any competition from other *Chrysanthemoides* ESUs, although there would be some competition from other beach front plants (e.g. *Scaevola*, with which *C. monilifera rotundata* shares certain ecological and growth form traits).

Alternatively, in cases of sister species sympatric over a broad area, current distributions may reflect multiple speciation events (Levin, 1991). This could apply to the current complex distributions of the many ESUs of *C. incana* and *C. monilifera* in the Western Cape. The incidence and occurrence of hybridization is most likely very high, and Griffioen (1995) observed many instances of hybrid populations between several of the ESUs (C. m. *pisifera pisifera* form 1 X C. m. *rotundata*; C. m. *rotundata* X C. m. *canescens*; C. m. *floribunda* form 2 X C. m. *pisifera pisifera* form 1; C. m. *pisifera pisifera* form 2 X C. m. *canescens*). This ease of hybridization and the evidence of occurrence of hybrids in the wild (Griffioen, 1995) provides both an possible origin of new ecotypes or species, but may also indicate the possible slowdown of incipient speciation events due to high gene flow between populations. The confusing and intermingled *C. monilifera floribunda* and *C. monilifera pisifera* clades in the "West" branch of Figure 2.3.3 suggests evidence of the occurrence of either ongoing gene flow, despite evidence of population structure maintenance (Griffioen, 1995), or incomplete lineage sorting from recent speciation events. These two ESUs occur in the Western Cape, where repeated episodes of introgression and/or hybridization have most likely introduced additional variation, and where reproductive isolation is not biologically or geographically enforced (birds may carry seeds long distances between populations). Several authors have expressed skepticism concerning the use of quantitative phenotypic data to identify hybrids in the absence of information regarding the genetic basis of the characters being scored (Gottlieb, 1972; Heiser, 1973; Rieseberg, 1997).

It should be noted that morphological intermediacy can arise from forces other than hybridization, such as convergent morphological evolution (Rieseberg, 1997). The high morphological plasticity of *Chrysanthemoides* as noted by Griffioen (1995) should be remembered in light of this comment.

Comparisons to Griffioen's morphologically derived phylogeny

Figures 1.1 and 1.2 illustrate the morphologically derived taxonomy of the taxonomic entities recognized by Griffioen (1995) in this thesis. When Figure 1.2 is compared to the genetics based phylogenies (based on ITS data, Figure 2.3.2), some things are common to both trees: (i) *C. monilifera septentrionalis* and *C. monilifera canescens* are placed basal to the rest of the *Chrysanthemoides* entities (although *C. monilifera monilifera* is also placed as basal in the morphology-derived tree, but as more derived in the genetics-derived ITS tree); and (ii) *C. incana* is placed as derived from the *C. monilifera* entities. Figure 1.1 has a very different topology to that derived from genetic data: (i) *C. incana* is a sister clade to the *C. monilifera* clade, not derived from within it; (ii) *C. monilifera rotundata* is placed basal to the rest of the *C. monilifera* taxa, not derived; (iii) the two forms of *C. monilifera floribunda* and *C. monilifera pisifera pisifera* group by taxonomic division, not by geographical location.

These results suggest that whilst morphological data may provide taxonomic classification, it can not always provide a means to determine phylogenetic history.

The utility of ITS and plastid DNA in plant phylogeography, and comments on analytical methods

The ITS sequence data certainly provide enough phylogenetic signal at the intraspecific level in Chrysanthemoides to determine phylogeographical patterns, especially when distance-based analytical methods are applied. This supports Grube et al.'s (2000) statement that in species level studies (of fungi, in Grube's paper) discrete clades of ITS alleles are corroborated by morphological differences.

At the intraspecific level, Parsimony analysis struggles to provide resolution with relatively few phylogenetically informative characters. When the number of taxa is reduced (from 76 to 19) the phylogenetic resolution becomes clearer: compare Figure 2.3.1 (with 76 taxa and little phylogenetic resolution) and Figure 2.3.7 (with 19 taxa and a much clearer phylogenetic structure that is very similar to the NJ tree in Figure 2.3.2). This may be due to high levels of homoplasy in the data (an accusation levelled at the ITS region by Alvarez and Wendel (2003) across multiple studies), possibly due to fast evolution of the ITS region. This homoplasy can mask the phylogenetic signal, and the presence of reticulation may also obscure the phylogenetic signal.

Maximum Likelihood provides slightly more resolution, but still suffers from some lack of resolution at the terminal areas of the trees. Distance methods, specifically Neighbor Joining, provided the clearest results for the ITS data. The resultant phylogeny not only elucidated correlations in the data with morphological classifications that were not clearly shown by the other two analytical methods; but also with geographical distributions that were also not so clearly evident in results from the other two methods. Huelsenbeck & Hillis (1993) expressed some concern that the Jukes-Cantor corrected distances methods are inconsistent when the Jukes-Cantor assumptions of DNA sequence evolution are violated. Huelsenbeck (1995) did later note that it was almost impossible to exhaustively examine the combined effects of model violations, and that in all likelihood, all of the assumptions of every phylogeographic method are violated to some extent with real data.

ITS data

Because rRNA is divided into domains where bases are either paired or unpaired, it is likely that different evolutionary constraints operate in each case, leading to some evolutionary constraints in the mutational rates of ITS sequences related to the maintenance of specific secondary structures that provide functionality (Alvarez & Wendel, 2003, Goertzen *et al.*, 2003). The large eight-base deletion that defines the "West" branch of the ITS NJ phylogenetic trees occurs in the loop of Helix 2A (naming of helixes and secondary structure of ITS from Goertzen *et al.*, 2003), an unpaired region that shows a very high level of mutation in the ITS region (there is almost no sequence consensus for any of the bases in this region within the tribes of Asteraceae), illustrating that multi-base indels do not necessarily render the ITS region non-functional, as was a concern of Bailey *et al.* (2003).

Burban and Petit (2003) note that the phylogeographic resolution from sequence data will be greater if the genomes targeted are subject to limited gene flow. The strong geographical structuring present in some of the ITS data suggests that some limitation of inter-ESU gene flow is occurring, allowing genetic differentiation to build up within the ESUs. The fact that hybridization does still occur (indicated by Griffieon (1995) as morphologically intermediate specimens located at the boundaries between two ESU distributions) suggests that isolation mechanisms have not yet completely developed. The presence of hybrids between *C. monilifera canescens* (a strongly distinct ESU at the genetic level, with a disjunct distribution) and other C. *monilifera* subspecies at the edges of their respective distributions (Griffieon, 1995) suggests that in the absence of sympatry, the evolution of isolating mechanisms has little functional push to evolve, and may only occur by chance mutation, or a pollinator or environmentally driven change in breeding seasonality caused by differing rainfall patterns. Hare (2001) notes that a polyphyletic relationship among recently isolated populations (present in the West branch of the phylogenetic trees) is often caused by the persistence of ancestral polymorphisms, but similar genealogical patterns can also be caused by moderate gene flow over longer divergence times. Avise (2000) also notes that nuclear DNA sequences can have experienced recent recombination and this can lead to the phylogenetic history of mutational events being garbled by recombinational swaps between different branches in the intraspecific tree.

cpDNA data

Gielly & Taberlet (1994) note that insertions/deletions make up an important part of the sequence divergence observed at the intra- and intergeneric levels in chloroplast DNA. McDade & Moody (1999) observed a relatively high frequency of length mutations in cpDNA in their study of the Acanthaceae, especially in the *trnL-trnF* spacer. The presence of length mutations in the *Chrysanthemoides* cpDNA regions is much higher in the *psbA-trnH* region than the *trnL-trnF* region. The length mutations that are present in the *trnL-trnF* sequences (when compared to the *Osteospermum* outgroup) are insertions and deletions of single bases, whereas the *psbA-trnH* region has more indels (of longer lengths).

McDade & Moody, (1999) also observed in their study of the Acanthaceae that almost all length mutations in the spacer regions are deletions, whereas those in the intron are more balanced between insertions and deletions, suggesting that different evolutionary processes may be operating in these regions. The *psbA-trnH* region of *Chrysanthemoides* shows a similarity to the Acanthaceae, where the major changes from the outgroup sequence are multi-base deletions. The *psbA-trnH* region has 15 deletions, some of which are fairly large (four, six or even 10 bases in length), and two single-base inserts and one major 14-base insert (in the *C. incana* samples). The *trnL-trnF* region, however, has no length changes, and the two single-base deletions are balanced by two single-base insertions. (The ITS region of *Chrysanthemoides* has six single base insertions, and six deletions, including one eight-base deletion. There is also a three-base insertion and a three-base deletion in the *C. incana* multicopy samples).

Sang *et al.* (1997) also observed in the Paeoniaceae that the *psbA-trnH* intergenic spacer has a higher rate of nucleotide substitutions than the other chloroplast regions used in the study. This certainly holds true for *Chrysanthemoides*, with *psbA-trnH* having 21 substitutions (9 transitions and 12 transversions), and *trnL-trnF* having 17 (10 transitions, 7 transversions).

Due to the *psbA-trnH* region having the highest percentage of phylogenetically synapomorphic sites among informative sites, and *trnL-trnF*, the lowest percentage (in the Paeoniaceae at least), Sang *et al.*, (1997) consider the *psbA-trnH* spacer more useful for studies at the lower taxonomic levels, as it evolves more rapidly and provides the best synapomorphic information. In the *Chrysanthemoides* data: (i) the *psbA-trnH* data are more variable than the *trnL-trnF* data (3.7% vs 2.2% with gaps were ignored, but 10.4% vs. 2.8% when gaps were coded as useful data), and of those variable bases, (ii) *psbA-trnH* had a higher percentage of potentially parsimony-informative characters (1.4% vs 0.7% when gaps were ignored, but 7.7% vs. 1% when gaps were considered as useful data) and (iii) of those variable characters, more *psbA-trnH* characters than *trnL-trnF* characters were potentially parsimony-informative (38% vs 33% when gaps were ignored, but 75% vs. 34% when gaps were considered as useful data). In the chloroplast data the gap characters were most certainly phylogenetically informative.

The strange groupings that occur from *trnL-trnF* data (Figure 2.3.14 and 2.3.15), combined with the very low level of parsimoniously informative character (0.7% of all the characters were informative) and the low bootstrap support for the trees, suggests that at the intraspecific level in *Chrysanthemoides* at least, the spacer region, by itself, may not be all that useful for phylogenetic purposes. Any signal that is present in the data may be swamped by homoplasy at this low level. In combination with other cpDNA regions, though, it may lend support to phylogenetic signal and aid in defining other regions of phylogenetic trees.

Although the combination of the two data sets does provide more definition at the lower levels of the phylogenetic trees, the CI and RI values, whilst still very high, are lower than *psbA-trnH* data alone, and this, in combination with the lower bootstrap support for some branches suggests that the *trnL-trnF* data is rather homoplasious. This is a view shared by McDade & Moody (1999) who state that the *trn*L-*trnF* locus is primarily useful at genus level and above. In the chloroplast TCS gene trees, however, the use of both sets of data results in the best tree. Ritz *et al*., (2003) found the *trnL-trnF* region of *Euphorbia* to be lacking in phylogenetically informative characters at the genus level and thus they avoided the use of cpDNA data altogether. As a comparison with another genus level study (investigating the species of *Pelargonium*) by Bakker *et al.* (1999), the *trnL-trnF* region in that study had 13% variable characters (and of those, 16% were informative) and also had 22 phylogenetically informative indels ranging in length from two to 56 bases. By comparison, the *Chrysanthemoides* genus *trnL-trnF* region is far less informative (in terms of informative characters and gap density). While *trnL-trnF* may be relatively fast evolving compared to coding DNA sequences (Soliva *et al*., 2001), it is certainly not the fastest evolving spacer in the *Chrysanthemoides* chloroplast genome.

Genomic data comparisons (ITS vs. cpDNA)

The higher level of variability of ITS over that of the chloroplast genes agrees with McDade's (2000) observation that

> "comparisons (in the Acanthaceae) suggest that the ITS region as a whole is nearly twice as variable as the cpDNA loci studied",

although this is certainly even more so in this genus (nearly 10 times more variable).

McDade (2000) also notes that despite the high substitution rate, the ITS data in her study of the Acanthaceae was not markedly homoplastic. But she also notes that the nrITS data were, however, considerably more homoplasious than the *trnL-trnF* data (McDade, 2000). This comment on homoplasy in *trnL-trnF* is contradicted by Sang *et al.* (1997), who note that, in the Paeoniaceae, this

> "most frequently used intergenic spacer [evolves very] slowly and homoplasiously, and thus its phylogenetic utility at the intrageneric level is questionable".

Gielly *et al.* (1996) observe that in the Asteraceae, nuclear DNA evolves about twice as fast as cpDNA. They also judge the relative informativeness of cpDNA to be lower at specific levels than ITS. This is a view shared by McDade & Moody (1999), and it suggests that future phylogeographic studies should utilise *psbA-trnH* as a region of preference.

Conclusions

The ITS region of nrDNA has been proven to provide enough sequence variability to render it a useful tool at subspecies level in *Chrysanthemoides*. The genetic entities derived from the ITS phylogeny (using Neighbor Joining analysis) not only correlate with morphologically defined taxonomic entities, but also correlate with geographical distribution patterns, thus warranting the designation of ESU status. The patterns of geographical variation that are present in the ITS data can also be used to investigate the effects that environmental heterogeneity can have on speciation.

The cpDNA regions utilised provide a different phylogeny to that provided by the ITS sequences, revealing maternal inheritance patterns (rather than phylogeographical interactions), and thus allow for theorising of the probable successive derivation of each ESU from a putative ancestor ESU.

This agrees with Hey *et al.*'s (2003) comment that, in principle, species taxa that are used as hypotheses might be simply confirmed or rejected, but more typical outcomes are likely to be fuller descriptions of the evolutionary processes that occur among the organisms that would be identified as members of a taxon.

The combined use of the two differentially inherited genomes allowed for an investigation into the processes of hybridization and speciation that may have given rise to the *Chrysanthemoides incana* taxonomic entity. The disagreement that is obvious between the two sets of genetic information (nuclear vs chloroplast) is often taken as an indication of a hybridisation event (Fuertes Aguilar *et al.*, 1999) and the placement of the *C. incana* sequences in the ITS tree suggests a hybridisation event between *C. monilifera rotundata* and another lineage. As to the identity of this other parent lineage, a member of the *Osteospermum* genus is a logical suggestion, owing to the close taxonomic affiliation of the two genera. However, with a large number of species in *Osteospermum*, the source of this parentage will take some effort to locate (assuming that it is even in *Osteospermum* at all). The choice of *Osteospermum* is a logical assumption, given the past taxonomic placement of *Chrysanthemoides* in that genus. A paucity of *Osteospermum* species sequence data publicly available for comparative searches is, however, a hindering factor.

Harrison (1991) asks a number of practical questions in regard to speciation: (i) Does speciation leave a distinctive signature on patterns of molecular genetic variation? (ii) If so, can we use variation in DNA sequences to gain insights into either the geography of speciation or the evolutionary forces that have been operating? (iii)Can we use estimates of genetic distance to make judgments about species status?

Answers to these questions can be phrased in the context of *Chrysanthemoides* genetic data: (i) Yes, evident in both the incongruence between cpDNA and nrITS as well as the heterozygous nature of the nrITS sequences of *C. incana;* (ii) Yes, the variation evident in the nrITS sequences is clearly correlated with morphologically and geographically defined entities in most cases (exhibiting geographically and climatically defined disjunctions), and where it is not as clearly defined, the samples are from areas that are geographically overlapping and experience intermingling gene flow; (iii) Yes, but only for those entities that have diverged sufficiently; e.g. *C. monilifera septentrionalis* and *C. monilifera canescens* should (on the basis of geographical distribution and nrITS data) be re-elevated to species level, despite relictual interbreeding ability with other species of *Chrysanthemoides*.

Chapter 3.

Ecophysiology of selected *Chrysanthemoides* ESUs

Introduction

Genetic framework for the further investigation of ESU status

Previous taxonomies have relied on morphological data to define the species and subspecies of *Chrysanthemoides monilifera*. The applicability of genetic sources of data to define ESUs has been clearly proven in Chapter 2, but researchers have pointed out that basing ESUs on genetic data only (particularly neutral data) may not fully address the problems of conservation and further factors, especially ecological ones, must also be taken into account (Cavers *et al*., 2003). Whitten *et al*., (2000) note that the real importance of DNA patterns is not just in clarifying generic limits, but rather as the foundation for examination of evolutionary effects of life history traits (e.g. Ecophysiology).

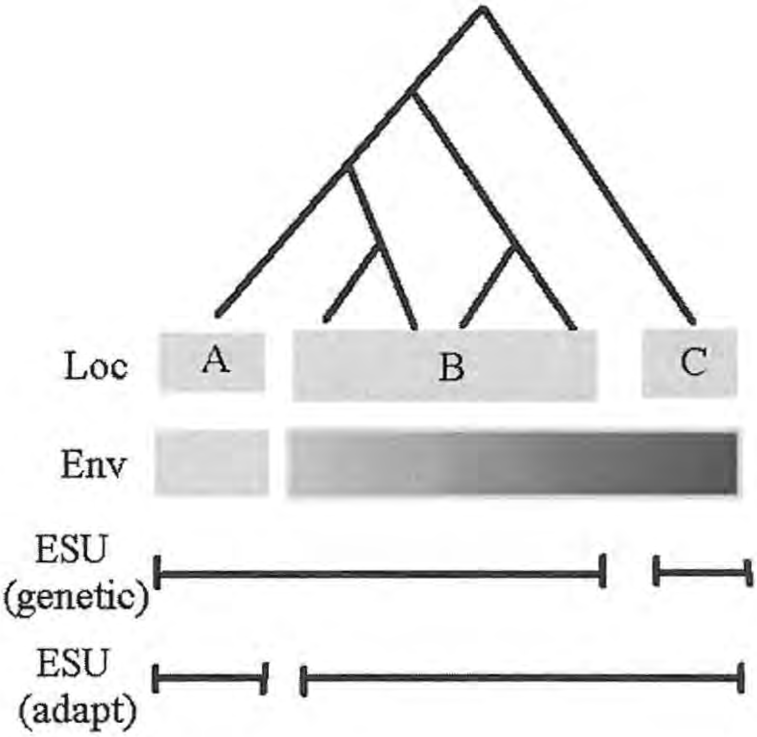

Figure 3.1.1: Schematic illustration of continua for genealogy and environmental conditions that accompany divergence of populations. If one were to base ESU assessment on genealogy, then A+B would form one ESU and C another. If one were to base ESU assessment on adaptive criteria, then A would form one ESU and B+C another (image adapted from Moritz, 2002).

Chapter 2 provides a "subspecies"/ESU phylogeny of *Chrysanthemoides* to work from. The next step was to select sample individuals of ESUs ("subspecies") from different geographical regions (preferably ones that showed divergent climatic characteristics). It would be a logical step that any correlations between genetic and morphological characteristics that help define ESUs should be compared to biogeographical and climatic data in an effort to further elucidate these patterns of correlation.

Researchers have made attempts to identify environmental variables that dictate range margins as well as attempting to identify characteristics that may limits species distributions at its range margins (Jonas *et al.,* 1999). Patterson *et al.* (1997) make mention of the fact that environmental stresses, and the trade-offs plants face that are unavoidable in adapting to them, are the principle factors responsible for both generating and maintaining the distribution of plant species. Barraclough *et al.* (1999) note that ecomorphology appears to evolve in concert with changes in the habitat occupied by species, suggesting that the environment determines ecomorphology, rather than species interactions. They further suggest that lineages may simply be separated by geographic division in parts of their ancestor's range, but an alternative is that ecological differentiation is a necessary component of the build up of species within clades.

The concept of the ecological niche has been described by Colwell and Fuentes (1975) as a useful device for the concise description of the patterns and limits of the response of organisms to their environment and they also note that the response of organisms to different environments is an essential component of the niche. A niche can be defined as

> "... a phenotypic attribute of a population of conspecific individuals, a statistical entity that changes whenever the members of the population change in their response to the biotic and abiotic environment." (Colwell and Fuentes, 1975, p. 282)

Although the niche concept is most often emphasised in the context of community ecology, the concept is also used by ecologists working at most levels of ecological organisation (Leibold, 1995).

Populations in nature often do not occupy as wide a variety of habitats as they are physiologically capable of occupying (Colwell and Fuentes, 1975). This difference is between the fundamental niche of an organism (the habitats that it is physiologically capable of inhabiting) and the realised niche of an organism (the actually habitats in which the organism occurs).

The choice to sample regions with different climatic factors was made to investigate if the genetically distinguishable ESUs showed equally distinct ecophysiological adaptations to their different environments (i.e. they each occupy a different ecological niche). A further subject of interest was whether any unique ecophysiological characteristics that may be found to be present in the field study sites are maintained when individuals from each ESU are subjected to identical greenhouse conditions.

Ecophysiological adaptations

Plant ecophysiology has progressed considerably as an interdisciplinary research area that attempts to explain the distribution and performance of plants in nature though experimental studies on functional properties related to basic plant traits and environmental conditions (Duarte *et al.*, 1995). It has traditionally focused on how a single species copes with natural conditions, rather than using comparative techniques (Duarte *et al.*, 1995). Functional plant ecology has recently emerged as an essentially comparative science concerned with the elucidation of the range of variation in functional properties among plants and the search for patterns and functional laws accounting for this variation (Duarte, 1999). The comparative approach is also a powerful tool to examine the effect of environmental conditions *in situ* (Duarte, 1999).

For comparative studies that explicitly address adaptational or evolutionary issues, the selection of species should be guided by phylogenetic considerations (Huey, 1987). Colwell and Fuentes (1975) note that, ideally, an experimental study should examine the response of genetically similar organisms to different environments, and Koehn (1987) agrees, stating that the relationships among genetic variation, physiological performance and fitness (i.e. adaptation) can be pursued only *within* species (however, in many cases, much like *Chrysanthemoides*, the definition of species is difficult, and in these cases ESUs can be a more useful concept). This investigation below species level is due to the possibility that the genotype of an individual can have a measurable and biologically significant effect upon physiological performance (Koehn, 1987). A comparative study is in effect an experiment over historical time and comparisons of close relatives are less likely to be confounded by phylogenetic artefacts and observed differences are more likely to reflect adaptations to the environment (Huey, 1987). The use of ESUs that experience the same rainfall season (all three ESUs are from the "East" branch of the phylogeny in Chapter 2, theorised to be split on the basis of rainfall seasonality) also allows for the minimization of extraneous environmental factors on differences in photosynthetic functioning.

The phylogeny from chapter 2 allows one to raise the question of whether these genetically identifiable ESUs can also be demarcated by unique ecophysiological traits. This requires the presence of one or more physiological traits to be present across the ESU's entire range of distribution. However, this raises the question of whether this uniqueness is related to climatic or environmental factors within the ESU's range of distribution, rather than genetic factors. Jonas *et al.* (1999) point out much the same: that adaptive explanations of trait variation across environmental gradients become more convincing (a) when the same pattern of trait variation in relation to an environmental gradient is found in different parts of a species' range, and (b) when functionally related traits vary in co-ordinated fashion across an environmental gradient.

Linhart *et al.* (1996) point out that in early studies of geographic variation, the focus of the analyses was to determine if variation showed a continuous (clinal) pattern or a discontinuous (ecotypic) one. They indicate that detailed analyses of variation show that, within the same species, some characters can vary gradually, others discontinuously, depending on many things, such as: gene flow, intensity of selection, number of genes involved, and terrain configuration.

If the field study sites yield significantly different ecophysiological traits for each ESU, are they merely a plastic response to environmental conditions that the plants have grown in (or are still growing in)? Phenotypic plasticity is the ability to respond to change in environmental conditions by alteration of morphology or physiology, and is widely recognised as an important potential means of adapting to these changing environments (Schlichting, 1986; West-Eberhard, 1986; Stearns, 1989; Schlichting & Pigliucci, 1995). It has been suggested that this phenotypic plasticity is due to a generalised stress response system (under simple genetic control) that plants possess, which allows for some phenotypic adaptation to a variety of stresses (Chapin, 1991).

On the island of Hawaii, *Metrosideros polymorpha* is abundant across gradients of substrate age, soil moisture, elevation, precipitation and temperature, from sea level to 2500 m (Cordell *et al.*, 1998). Some of the morphological variation that is evident in the species has been associated with elevation and substrate age. Investigations by Cordell *et al.* (1998) into physiological variation in *M. polymorpha* along an altitudinal gradient found that several characteristics, (such as ecophysiological behavior and anatomical features) were largely induced by the environment. But other characteristics, particularly leaf morphology, appeared to be genetically determined.

If the unique ecophysiological trait is genetically determined, then could the distribution of the members of the ESU be related to gradients (of water use efficiency or nitrogen use efficiency, for example)? Jonas *et al.* (1999) report on a study in which leaf carbon isotope analysis of greenhouse grown *Clarkia unguiculata* plants from populations at two differing latitudes suggested that rates of physiology were negatively correlated with elevation. The fact that the plants were greenhouse grown suggests that the difference in physiology may have been genetically predetermined, rather than a response to immediate environmental conditions.

For species with a wide ecological distribution (such as *M. polymorpha*, cited above, or *C. monilifera,* studied here), variation in physiological, morphological and anatomical characteristics may be achieved by a combination of genotypic differentiation and phenotypic plasticity (Cordell *et al.*, 1998). It has been further suggested by Cordell *et al.* (1998) that phenotypic plasticity should be high for characters that are adaptations to environmental conditions that are likely to change multiple times during the lifespan of an organism. In contrast, genetic differentiation is likely for characters that are adaptations to environmental conditions that are likely to remain fairly constant during the organism's lifespan (Cordell *et al.*, 1998).

The question of whether the plant is only plastic during early development or whether the plant can continue to alter its physiology at adult stages of growth can be investigated by several methods. If reciprocal transplant experiments result in plants showing ecophysiological traits correlated only with locality, and not genotype, the ecophysiological traits are evidently a result of environment, and not inherent in the ESU. If the plants show functional differences in the field, and also maintain those differences in either reciprocal transplant experiments or under greenhouse conditions, it would indicate that the functional differences are inherent and nonplastic. If there are differences in the field, but the plants become uniform in their ecophysiological traits in the greenhouse, this can indicate that the period of plasticity is not limited to seedling stage, but remains throughout the rest of the plant's life. A reciprocal transplant experiment was beyond the time frame of this project (though it would indeed be vastly informative) and thus one is limited to greenhouse data as a means of comparison.

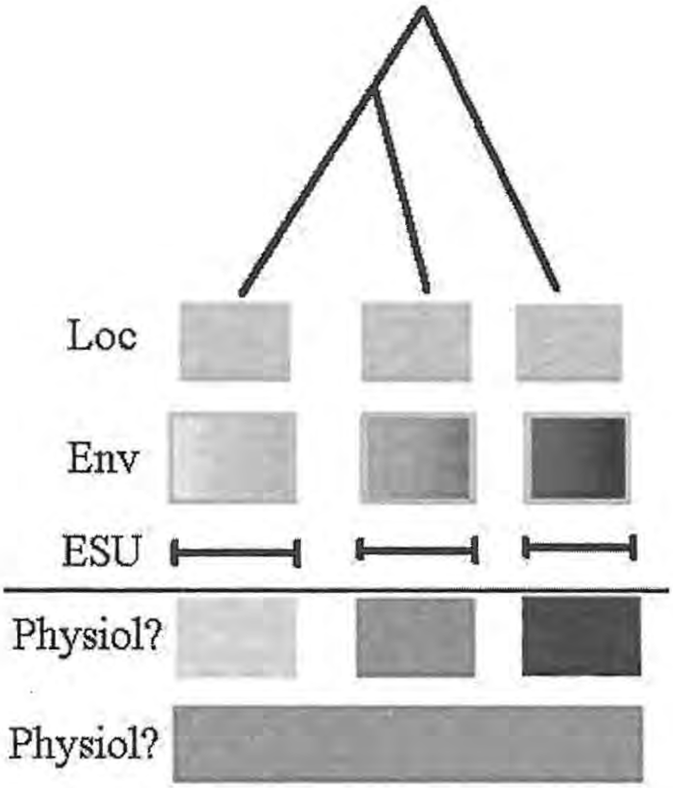

Figure 3.1.2: Schematic illustration genealogy and environmental conditions that exist for *Chrysanthemoides* ESU sample sites, in conjunction with potential ecophysiological data. Are the ecophysiological characters of each ESU functionally unique or do they show no differences? (Adapted from Moritz, 2002).

Ecological differentiation may be important in promoting the early stages of speciation or it may play a role in the long-term persistence and subsequent radiation of lineages (Barraclough *et al.*, 1999). Thus, the presence of a unique ecophysiological functional trait does not just provide another characteristic to aid in ESU identification, but may, in combination with a biogeographic or climatic pattern, point to the driving selective pressures that have resulted or may yet result in speciation within the *Chrysanthemoides* genus.

Methods available for the investigation of ecophysiological traits

There are many physiological factors that can be measured to illuminate ecophysiological adaptations. Two of the most commonly measured physiological traits are photosynthesis and plant water status, as variation in these parameters, in association with morphological characters, can play an important role in the adaptability of species to environmental constraints (Sandquist & Ehleringer, 1997). Another commonly used technique is that of measuring chlorophyll fluorescence, which has been widely used to measure the properties and status of the photosynthetic apparatus of plants (reviews of studies can be found in

Horton & Bowyer, 1990; Krause & Weis, 1991, Bolhar-Nordenkampf *et al.*, 1994; Schreiber *et al.*, 1994; Lichtenthaler, 1996).

The first two physiological parameters, in conjunction with chlorophyll fluorescence, have been put to use in a wide variety of studies in both the field and the laboratory. Leaf gas exchange and chlorophyll fluorescence have often been used in combination to investigate plant responses to changing environmental conditions (Nagy *et al.*, 1998; Thomas & Turner, 2001; Mielke *et al.*, 2003). Munné-Bosch *et al.* (1999) used all three techniques to investigate diurnal variations of *Lavandula stoechas* and *Rosmarinus officinalis* in Mediterranean field conditions, whilst Flexas *et al.* (2000) used these to investigate responses of grapevines (*Vitis vinifera*) to water stress.

Leaf water potential

The leaf is an organ that permits CO_2 uptake at a rate required for maintenance of photosynthesis while at the same time keeping water evaporation from leaves at a reasonably low rate. All gas exchange into and out of the leaf occurs via stomata, the guard cell of which control diffusion of CO_2 into the leaf and water vapor out of the leaf (Tyree, 1999). The physiology of the guard cells has evolved to optimize photosynthesis when conditions are right for photochemical reactions and to minimize water loss when conditions are unfavourable for photosynthesis (stomata remain closed or partly open when soils are dry; Tyree, 1999).

Whenever leaves lose water faster than the rate of water uptake by roots, the water potential of the leaf cells begins to drop. This drop can be measured by means of a pressure bomb, which is a metal chamber into which an excised shoot is placed. When gas is pumped into the chamber, the pressure of the gas increases the pressure of the fluid in the leaf cells. When the pressure of the gas is at the balance pressure, xylem sap is squeezed out the end of the branch protruding outside the chamber (the balance pressure is equal to the negative leaf water potential ($-\psi_{leaf}$); Tyree, 1999).

Leaf Fluorescence

There are several mechanisms that allow excess excitation energy to be quenched, and hence to prevent damage to the photosynthetic apparatus. Among these mechanisms are the antioxidant system, fluorescence, photochemistry and electron cycling around Photosystem I (PSI) and Photosystem II (PSII) (Horton *et al.*, 1994).

Chlorophyll fluorescence is a sensitive indicator of any disruption of chloroplast membrane integrity and its functions. PS II in the thylakoid membrane is more sensitive than other parts of the light harvesting apparatus and is the primary site of damage. The kinetics of the chlorophyll a fluorescence can be used as an indicator of the site of a disruption or perturbation in the mechanism of light harvesting. Chlorophyll *a* fluorescence arises almost exclusively from PS II (at room temperature; Critchley, 1988). Chlorophyll fluorescence can be measured with a chlorophyll fluorimeter, and PSII efficiency can be calculated. A hypothetical PSII efficiency of 100% would correspond to a ratio of variable to maximum chlorophyll fluorescence (F_v:F_m) ratio of 1.0, and most healthy plants have values of about 0.8 (Demmig-Adams *et al.*, 1996). Chlorophyll fluorescence can be used to distinguish photochemical and non-photochemical quenching (Horton *et al.*, 1994).

Aim

The aim of this study was to investigate the whether genetically distinguishable ESUs also show any measurable functional ecophysiological distinctions, both in the field and later after a sustained period of growth in controlled greenhouse conditions.

Methods

Field Study sites

Huey (1987) states that in attempting to study physiological adaptations to an environmental factor, comparisons should involve three, not just two species (if only for statistical comparative reasons). In order to investigate the potential functional differences in ecophysiological processes across several different climatic types, three study sites were chosen where different, genetically identifiable ESU grew (see Figures 3.2.1 and 3.2.2 below).

The first study site was 20 km east of the town of Joubertina, by the side of the R62 road (where *C. monilifera pisifera* occurs). The area is in the rain-shadow of the Tsitsikama Mountains, and receives an average of no more than 400mm of rain per year. The second study site was in a timber lot in Tsitsikama near Knysna (where C. *monilifera floribunda* form 2 occurs), on the seaward side of the Tsitsikama Mountains, and receives an average of 800-900 mm of precipitation per year. The third study site was on the beach front in St Francis Bay, on dunes about 30m from the high-tide mark (where C. *monilifera rotundata* occurs). Field data were collected from the 22nd to the 24th of January 2003.

Climate Diagrams

The South African Weather Service provided data from a variety of weather stations, within the ranges of each subspecies. These data consisted of mean monthly rainfall, mean monthly minimum and maximum temperatures for the last 10 years. The minimum and maximum temperatures were averaged and used as one axis of the diagram, while the precipitation was used as the other axis. These climograms are a means of easily representing and exploring the seasonal and climatic differences in seasonal variation in temperature and rainfall for different study areas.

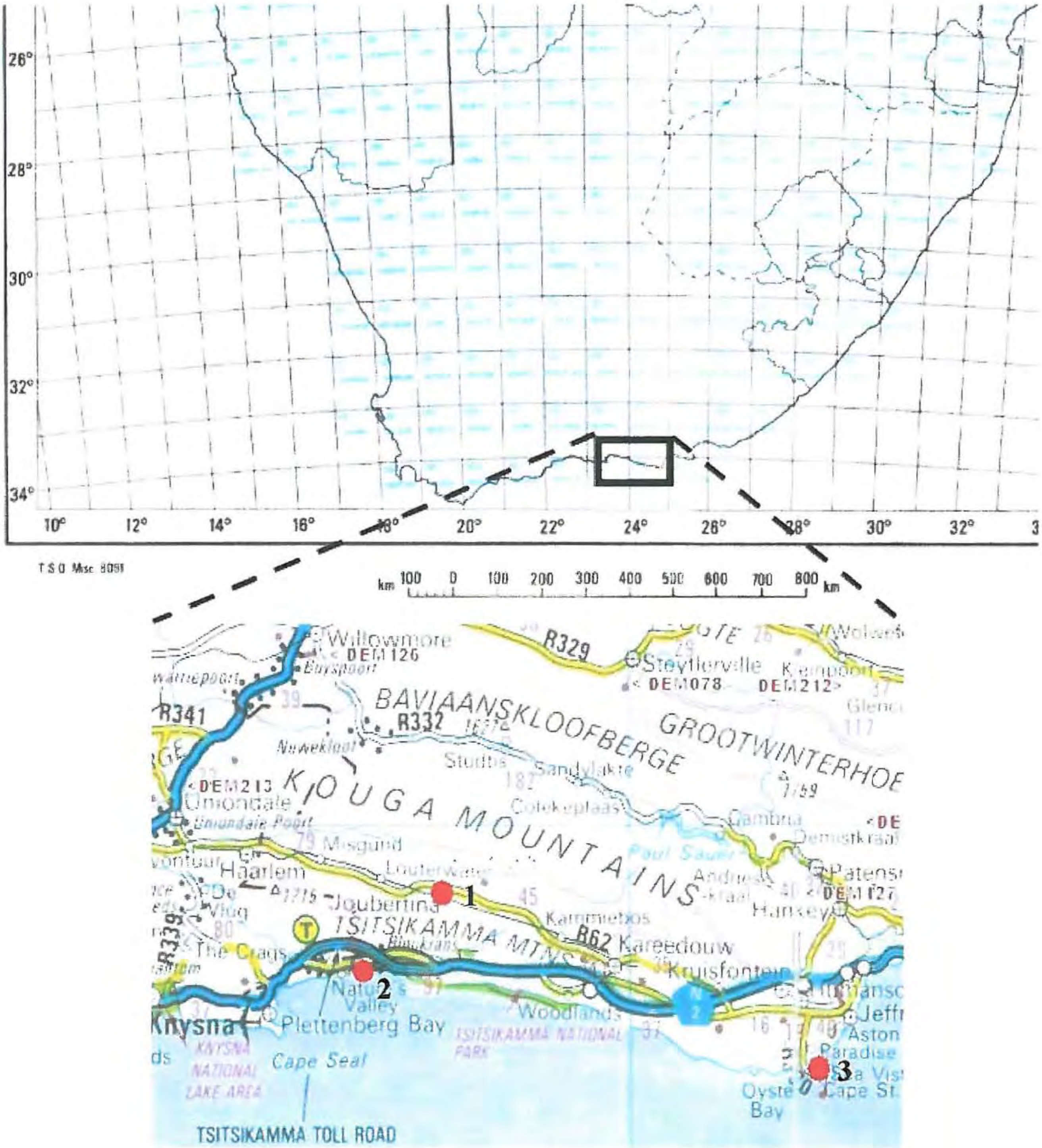

Figure 3.2.1: Map of South Africa, showing study sites in detailed map [1 = Joubertina (33° 49' S; 23° 52' E) ; 2 = Tsitsikama (33° 58' S; 23° 38' E); 3 = St Francis Bay (34° 12' S; 24° 49' E)]

Plant collections

Approximately 30 seedlings and cuttings were taken from each study site, and planted in 30 cm wide pots containing potting soil, with a layer of sand on top. They were watered daily or every two days, as soil moisture levels required, with slow release fertiliser added as necessary. When final measurements were taken, five healthy individuals of each population were used. The plants had been subjected to greenhouse conditions for more than 6 months, ensuring that any leaf that was measured had grown under greenhouse conditions.

Leaf Age measurements

Initial concerns about the effects of leaf age on the ecophysiological measurements were investigated in the field. As leaves grow and age, their photosynthetic responses and capacity can change. In order to minimise the effects of leaf age on photosynthetic measurements that were going to be taken, spot measurements were taken from the top leaf on the stem (nearest the apical bud) that had a surface area sufficient to fill the IRGA chamber 6cm^2 (this leaf was numbered 1) down the stem until the rate of photosynthesis became constant. For the purposes of the gas exchange studies that follow, the leaf numbered 17±1 was used.

Gas exchange studies

A:Ci curves can yield several significant points, these points include: The carboxylation efficiency (given by the initial slope of the curve), the rate of RuBP regeneration (given by the saturated assimilation rate), the CO_2 compensation point (the x-intercept, where y=0) and the rate of respiration in the light (the y-intercept, where x=0). Light response curves can also yield several significant points, these include: the light compensation point (the x-intercept), the efficiency of utilisation of incident light (given by the initial slope of the curve), the maximum assimilation rate (given by the saturated assimilation rate) and the respiration rate (the y-intercept). These points can be calculated using a curve fitted to the data, based on an equation from Causton and Dale (1990).

Measurements of CO_2 uptake rates were made with a LICOR 6400 portable photosynthesis system (LICOR Inc., USA). In all cases, relevant values were recorded when assimilation, transpiration, light and CO_2 values were stable. Reference water vapour concentrations were not controlled, but rather allowed to remain at ambient levels: 14.4 ± 0.9 mmol $H_2O.mol^{-1}$ at

Joubertina, 13.9 ± 1.93 mmol H_2O.mol^{-1} at Tsitsikama, 18.75 ± 1.15 mmol H_2O.mol^{-1} at St Francis Bay for A:Ci curves; and 14.5 ± 1.88 mmol H_2O.mol^{-1} at Joubertina, 13.6 ± 2.03 mmol H_2O.mol^{-1} at Tsitsikama, 18.4 ± 1.16 mmol H_2O.mol^{-1} at St Francis Bay for light response curves. For all greenhouse curves, reference water vapour concentrations remained at an ambient level of 7.4 ± 2.2 mmol H_2O.mol^{-1}.

For the field study A:Ci curves, light intensity was maintained at 1800 μmol.m^{-2}.s^{-1}, and CO_2 concentrations were altered in the following sequence: 400, 600, 800, 1000, 1200, 400, 200, 100, 50 μmol.mol^{-1}. During the construction of field study light response curves, reference CO_2 concentrations were maintained at 370 ± 1.5 μmol.mol^{-1}, and light intensities (PPFD) were altered in the following sequence: 1500, 1000, 800, 600, 400, 200, 100, 50, 0 μmol. m^{-2}.s^{-1}.

Diurnal spot measurements of photosynthetic rate were made at hourly intervals, concurrent with the collection of leaf water potential measurements (by pressure bomb). The light intensity was correlated with an external light sensor and reference CO_2 was maintained at 360 ± 1 μmol.mol^{-1}.

For the greenhouse study A:Ci curves, light intensity was maintained at 1000 μmol m^{-2} s^{-1} (the lower level was due to the slightly lower light levels that were experience by the plants for 6 months in the greenhouse) and CO_2 concentrations were altered in the following sequence: 400, 600, 800, 1000, 1200, 1500, 400, 200, 100, 50, 20 μmol mol^{-1}. During the construction of greenhouse study light response curves, reference CO_2 concentrations were maintained at 360 ± 1.3 μmol mol^{-1}, and light intensities were altered in the following sequence: 100, 200, 300, 400, 650, 800, 1000, 1250, 1500, 100, 50, 0 μmol.m^{-2}.s^{-1}.

Leaf water potential

During the field study, pressure bomb measurements of leaf water potential on a shoot taken from each study plant were made concurrent with the diurnal photosynthetic measurements (approximately every hour).

Leaf Fluorescence

Chlorophyll fluorescence was measured with a Plant Efficiency Analyser (Hansatech, UK). Three leaves on each of the five plants were dark adapted for half an hour, and then the fluorescence transients upon illumination measured. Further readings were taken after a 30-second dark adapt period. Data were analysed using Biolyzer version 3.0.6 (Laboratory of Bioenergetics, University of Geneva, Switzerland) incorporating the JIP test of Strasser *et al.* (1996).

The JIP test

All oxygenic photosynthetic material show a polyphasic fluorescence rise during the first second of illumination, the phases of which have been labelled O, J, I and P (Strasser *et al.* 1996). The JIP-Test is based on a simple model of how photon flux absorbed by the antenna pigments (ABS) is dissipated as heat (DI) and fluorescence, or channelled as trapping flux (TR) to the reaction centres to be converted to redox energy by reducing Q_A to Q_A^-. Q_A^- is then re-oxidised to Q_A and creates an electron transport that leads to CO_2 fixation (Strasser and Tsimilli-Michael, 2001). This OJIP-Transient changes its shape according to environmental variables, such as temperature, light intensity and drought. From the first second of data, nine values can be retained and normalised to obtain further biophysical data (Strasser et al. 1996), including maximum fluorescence intensity (FM), the fluorescence intensity at 50 µs (F_O), 300 µs (F_{300}), 2 ms (F_J) and 30 µs (F_I), as well as the area between the fluorescence curve and F_M (Area).

From these initial data, formulae (given in Appendix 4) can be used to calculate certain aspects of PSII behaviour at the onset of fluorescence induction (indicated by the subscript "0") that can be quantified. These include: (i) the specific energy fluxes (per reaction centre) for absorption (ABS/RC), trapping (TR_0/RC), dissipation (DI_0/RC) and electron transport (ET_0/RC); (ii) the flux ratios (yields): the maximum quantum yield of primary photochemistry ($\varphi_{P0} = TR_0/ABS$), the efficiency with which a trapped exciton can move an electron into the electron transport chain further than Q_A^- ($\Psi_0 = ET_0/TR_0$) and the quantum yield of electron transport ($\varphi_{E0} = ET_0/ABS = \varphi_{P0}\ \Psi_0$). The performance index (PI_{ABS}) was also calculated. The fluorescence parameters are represented as proportions relative to the control values for St Francis full dark adapted data. The proportions were calculated as: (sample value/control value) – 1.

Specific leaf Area

Specific leaf area can be defined as the area of the leaves, divided by the weight of the dry mass, measured in $cm^2.g^{-1}$. This is a measure of the leafiness of the plant, on a dry weight basis (Beadle, 1985). This can help highlight differences in carbon investment in the plants.

Two 30 – 45 cm shoots from each study plant were collected in the field, five leaves from each plant were outlined on paper, numbered and collected along with the other leaves and the stems in paper bags. They were dried for 10 days in an oven at 60°C. The leaf areas on paper were measured using a leaf area analyzer (WinDIAS Leaf Area Meter, Delta-T devices) and the dried leaves weighed. The weight and area of each leaf were compared to determine a function that would predict the total area of the leaf material on the shoots, without having to measure each and every leaf.

Statistical analyses

Significant differences in data were analysed using the ANOVA test following a test for homogeneity (both implemented in STATISTICA; Version 6.1, StatSoft, Inc).

Results

Climate Diagrams

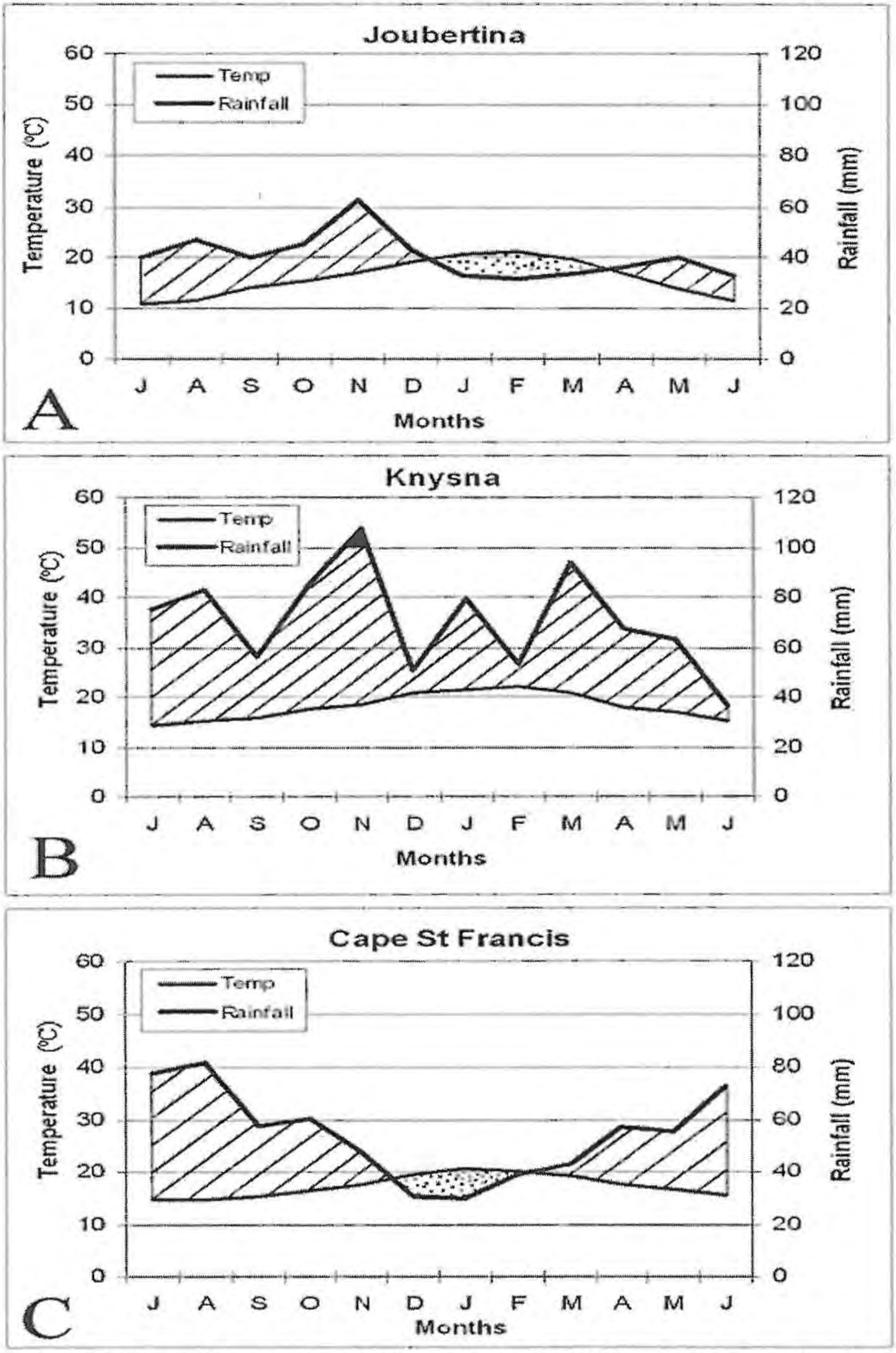

Figure 3.3.1: Climate Diagrams for study sites: (a) Joubertina, (b) Knysna, (c) Cape St Francis. (Stripes indicate humid periods, stipples indicate arid periods and black indicates precipitation above 100 mm). Knysna is the closest weather station to Tsitsikama that has both rainfall and temperature data for a sufficient time period to be informative.

As can be seen from the climate diagrams, Joubertina has relatively little precipitation, Knysna has a relatively wet climate all year round, and St Francis has clearly defined dry and wet seasons.

Climographs

While the climate diagrams (Fig 3.3.1) clearly show the seasonal variation in precipitation for each locality, the climographs (Fig 3.4.2) illustrate the different climatic "space" that organisms from each locality are adapted to. The climographs also factor in time as another element to the analysis, and though there is some overlap at certain times of the year, the three sites differ during other times of the year, e.g. Joubertina and St Francis are very similar during January but completely different during August, and St Francis and Tsitsikama are nearly identical during August, but completely different during January.

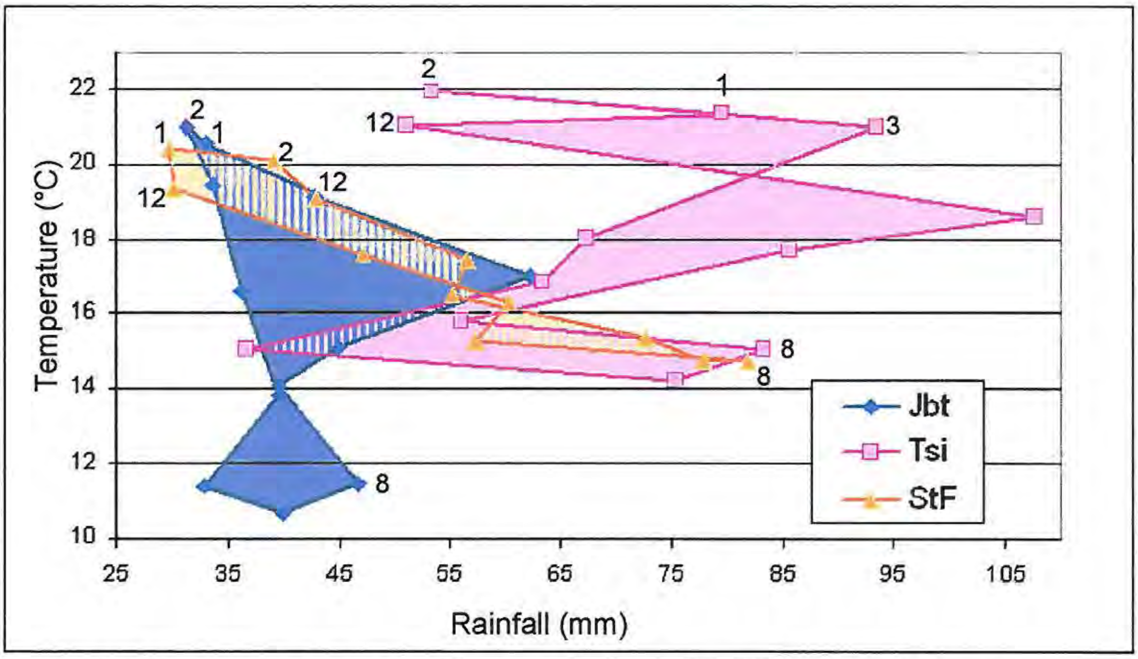

Figure 3.3.2: Climographs of study sites. Numbers indicate months of the year (i.e. 1= January, 2=February, 8=August).

Leaf age measurements

The graph in Figure 3.3.3 illustrated the response of photosynthetic rate to leaf age. Leaves selected for photosynthetic measurements in the field were typically the leaf numbered 17±1.

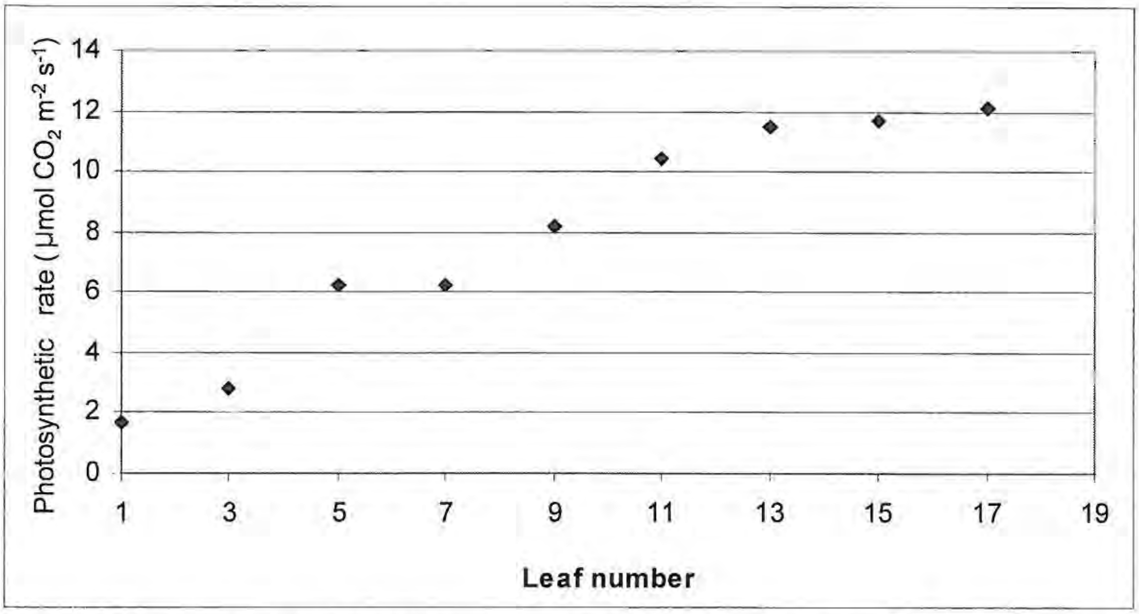

Figure 3.3.3: Plot of Photosynthetic rate against leaf age (indicated as position number of leaf down stem from apical bud, N=1).

Gas exchange studies

Light response curves

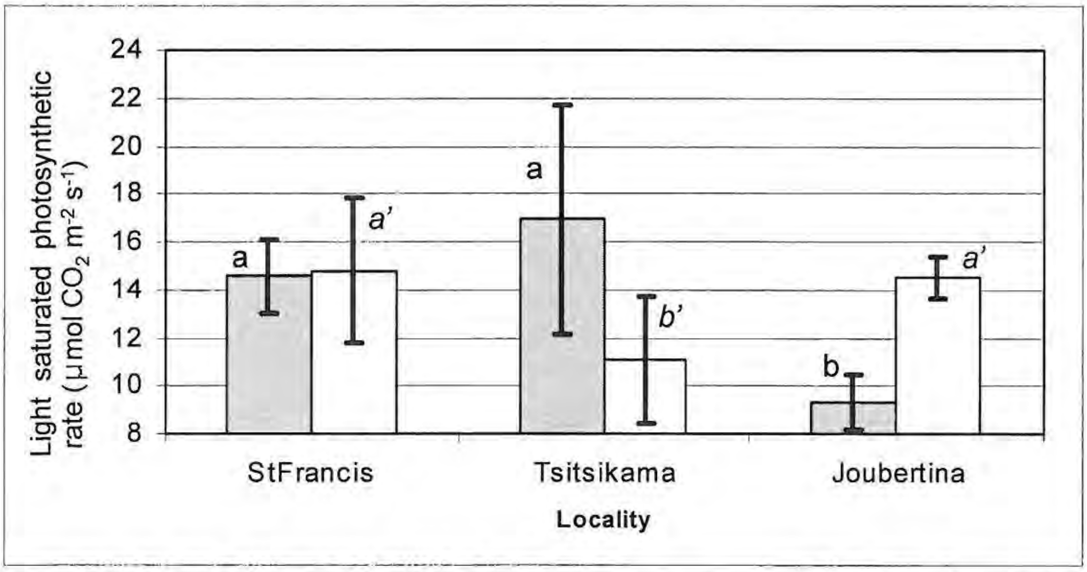

Figure 3.3.4: Means (with standard deviations) of light saturated photosynthesis per locality. Shaded columns are field study data; white columns are greenhouse study data. N=5; For either field study or greenhouse data bars labelled with the same letters indicate statistically homogenous groups at the 90% confidence level (i.e. "a" is not comparable to "a′ "). Field study data shows a significant difference of $p<0.02$; Greenhouse data shows a significant difference of $p<0.05$.

As Figure 3.3.4 illustrates, at the Joubertina study site, plants have significantly lower light-saturated photosynthetic levels than at the other two sites, but the difference between plants from the Joubertina and the St Francis sites is no longer apparent in the greenhouse study. In the greenhouse study, it is the Tsitsikama plants that have significantly lower photosynthetic levels than the plants from the other two study sites. For both sets of data, differences were significant at the 90% confidence level, but not at the 95% confidence level, indicating that any difference that does exist is minor.

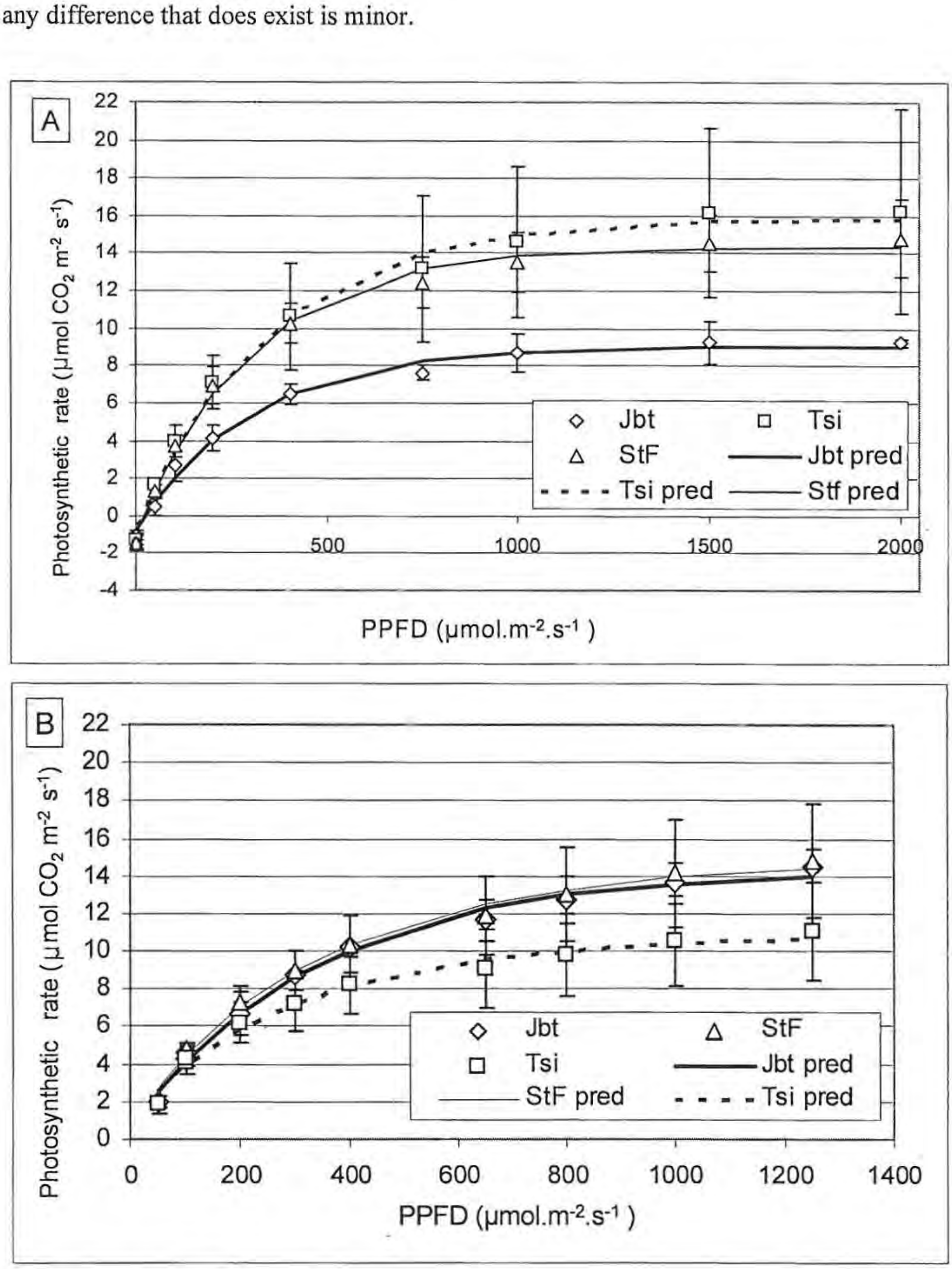

Figure 3.3.5: Means (with standard deviations) of light response curves (lines fitted according predicted values according to Causton & Dale, 1990) for field (A) and greenhouse (B) plants (N=5).

Figure 3.3.5 shows the light response curves from the field and greenhouse plants, and Figure 3.3.6 illustrates the significant points derived from the light curves. The means and standard deviations shown in Figure 3.3.5-B illustrate the lack of difference between the plants taken from the Joubertina and the St Francis field study sites. Figure 3.3.6 below illustrates the significant points taken from the light response curves for both field and greenhouse experiments. The light compensation points of the field plants at the Tsitsikama and Joubertina study sites were significantly different from each other ($p < 0.03$, Figure 3.3.6 A). This difference was not retained when plants were grown under greenhouse conditions. The slightly lower rate of light saturated photosynthesis shown by the Tsitsikama greenhouse grown plants is significantly different ($p < 0.02$). Neither dark respiration rates nor efficiency of utilization of incident PPFD were significantly different between plants from different localities and remained unaffected when plants were grown in the greenhouse.

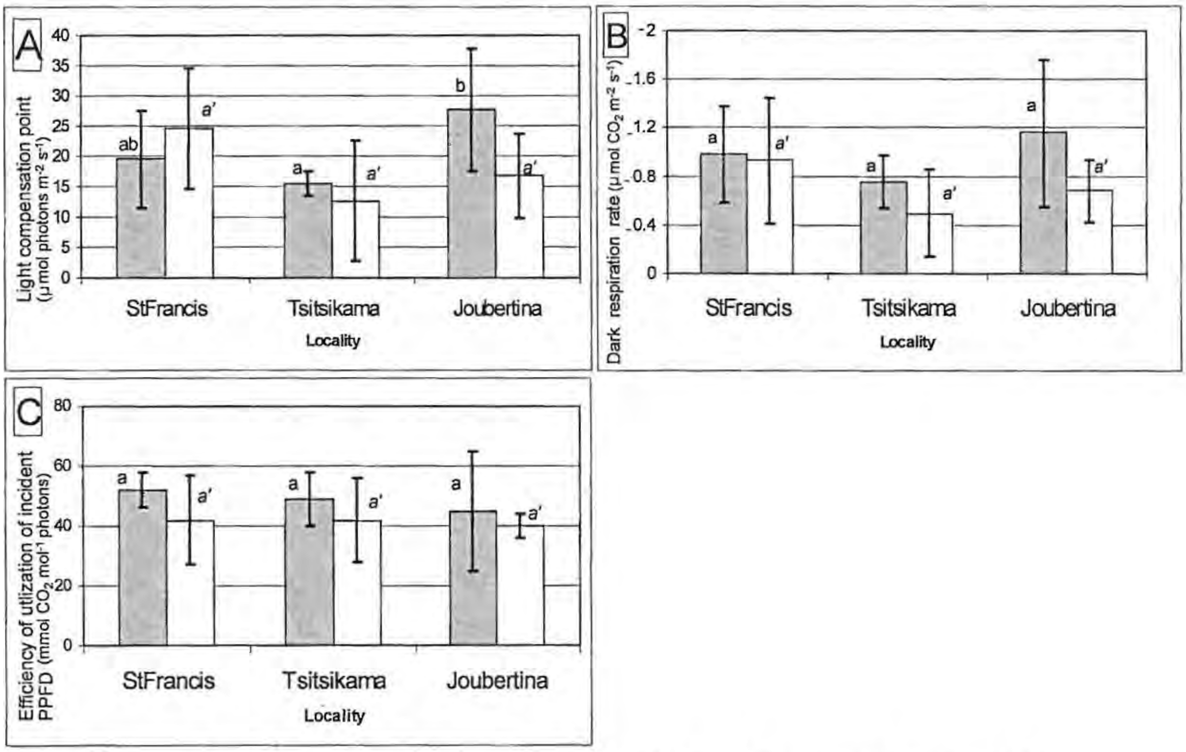

Figure 3.3.6: Means (with standard deviations) per locality of (A) Light Compensation point, (B) Dark respiration Rate, and (C) Efficiency of utilization of incident PPFD. Shaded columns are field study data; white columns are greenhouse study data. N=5; For either field study or greenhouse data bars labelled with the same letters indicate statistically homogenous groups at the 95% confidence level (i.e. "a" is not comparable to "a′ ").

A:Ci curves

Figure 3.3.7 below illustrates the A:Ci curves for field and greenhouse plants. The significant points derived from the A:Ci response curves for field and greenhouse plants are illustrated in Figure 3.3.8. RuBP regeneration rates for field grown plants were significantly different between Joubertina plants and Tsitsikama plants, but these differences were not retained when plants were transferred to the greenhouse. The carboxylation efficiencies of the field grown Joubertina plants were significantly lower (but only at the 80% confidence level, not the 95% confidence level) than those of the plants from the other two study sites, but this difference was also not retained after transferral to the greenhouse. Neither the CO_2 compensation point nor the photorespiration rate was significantly different in the field or the greenhouse grown plants.

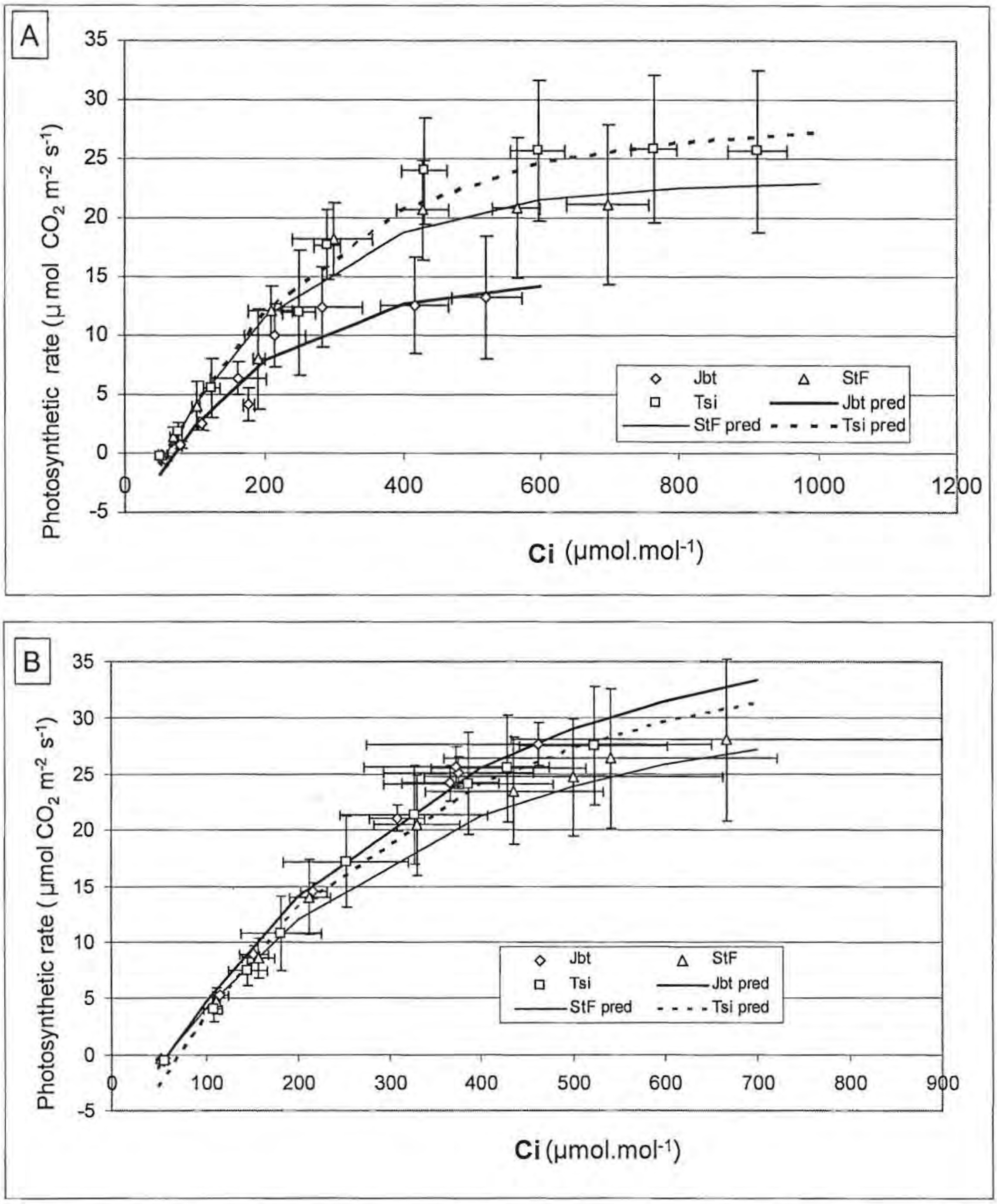

Figure 3.3.7: Means (with standard deviations) of A:Ci curves (lines fitted to predicted values according to Causton & Dale, 1990) for field (A) and greenhouse (B) plants (N=5).

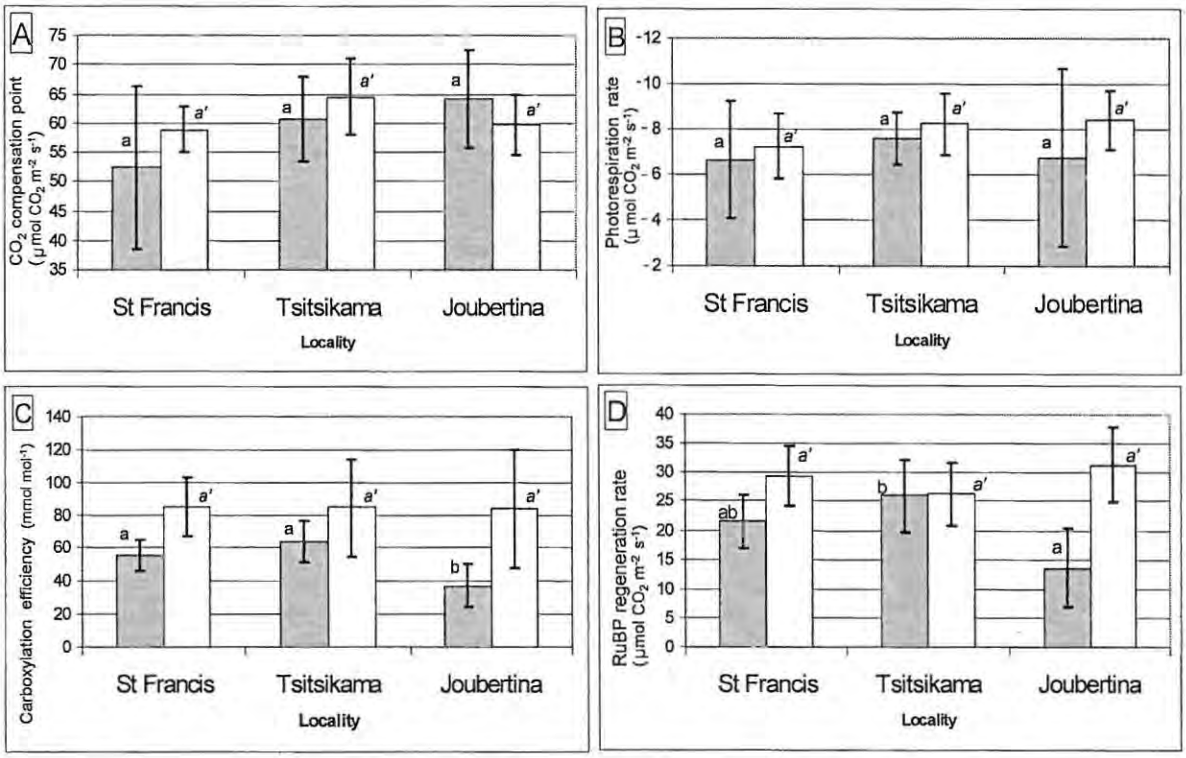

Figure 3.3.8: Means (with standard deviations) per locality: (A) CO_2 compensation point; (B) Photorespiration rate; (C) Carboxylation efficiency (Joubertina field data only significantly different at the 80% confidence level; $p<0.05$); (D)RuBP regeneration rates ($p < 0.02$ for field study); Shaded columns are field study data; white columns are greenhouse study data. N=5; For either field study or greenhouse data (except for field data for graph C) bars labelled with the same letters indicate statistically homogenous groups at the 95% confidence level (i.e. "a" is not comparable to "a' ").

Diurnal Measurements

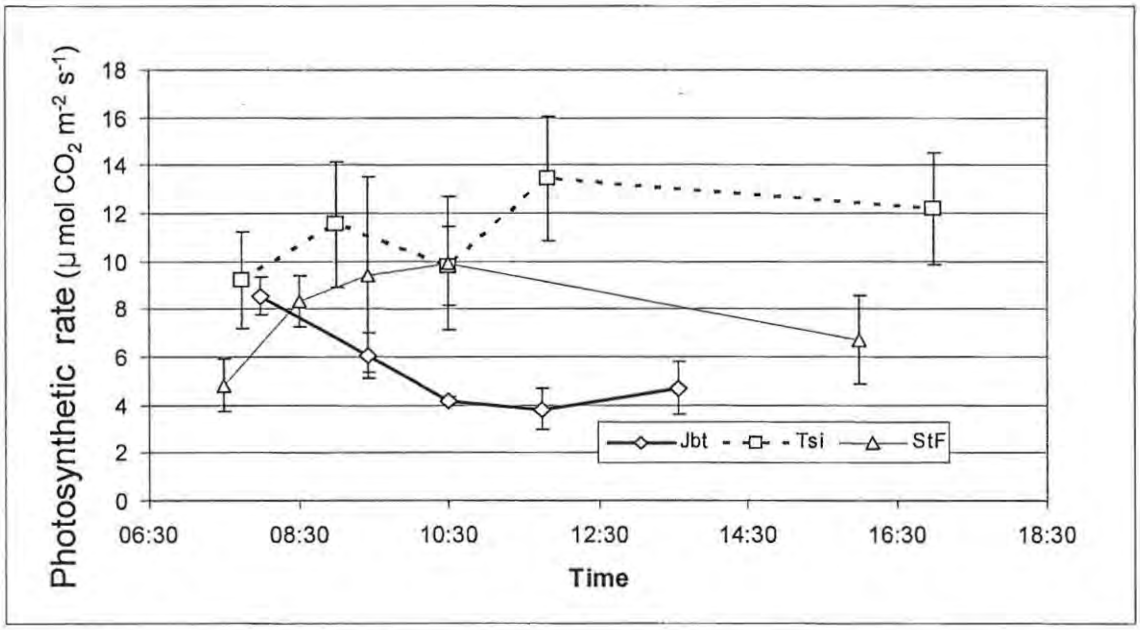

Figure 3.3.9: Means (with standard deviations) of diurnal response in photosynthetic rates for plants growing at each locality.

Figure 3.3.9 shows a clear difference between the Joubertina plant diurnal response and those from the plants at the other two study sites. The Joubertina plants show a steady decline in photosynthetic rates starting early in the morning and continuing through the day. This was not due to cloud cover as the skies were clear throughout the measuring period, and in fact the level of light continued to rise throughout the day (Figure 3.3.10). The decline in the Tsitsikama plant measurements at 10.30 am is most likely due to a slight dip in light levels due to cloud cover.

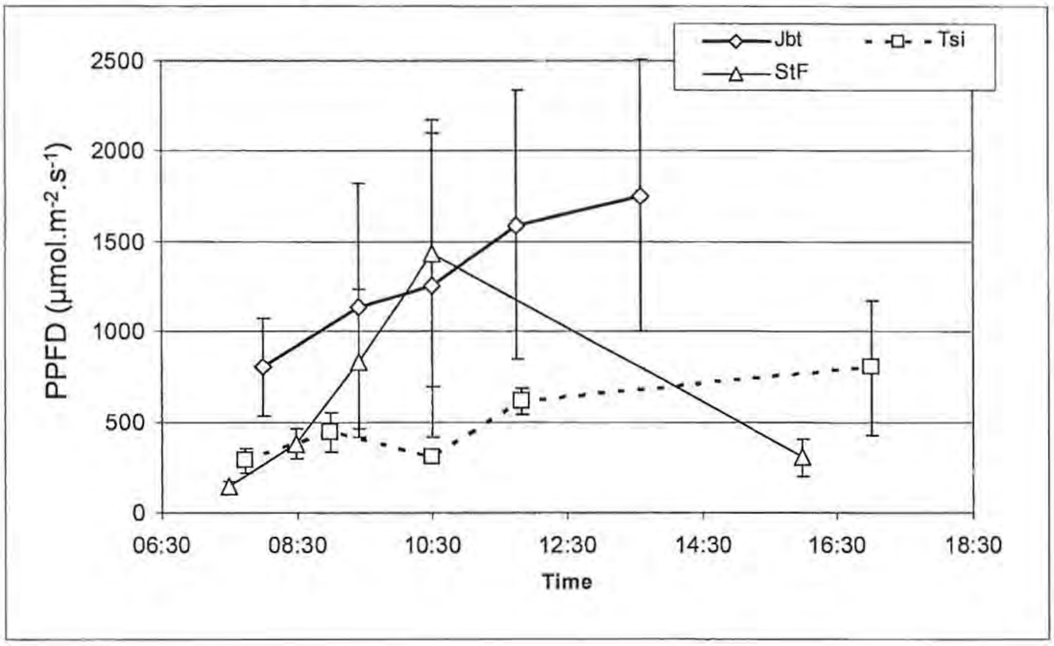

Figure 3.3.10: Means (with standard deviations) of diurnal ambient light levels at each study site.

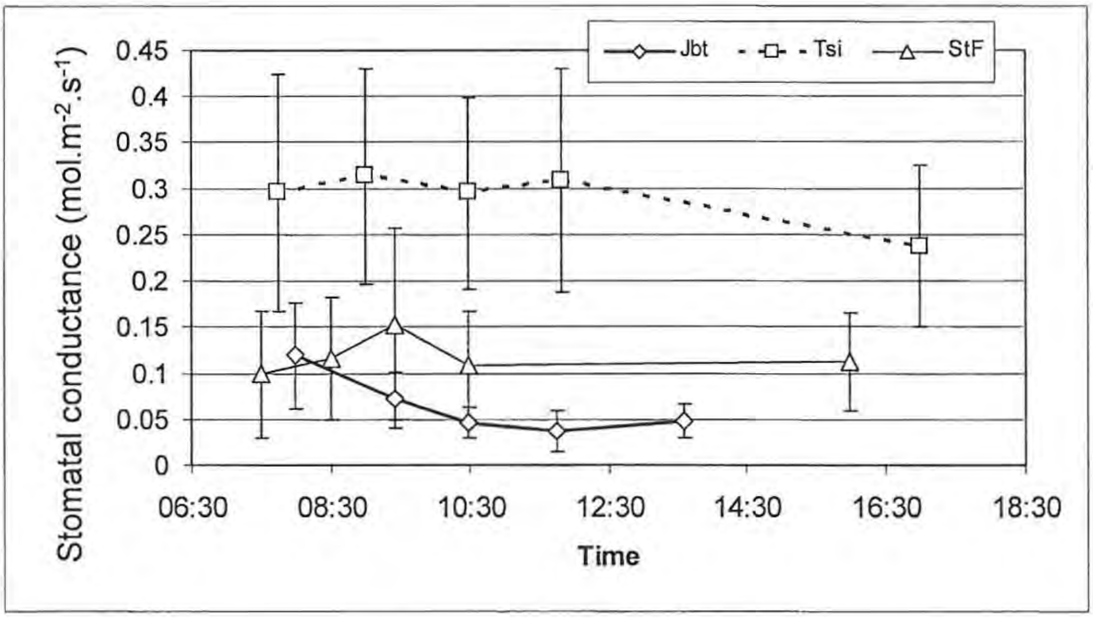

Figure 3.3.11: Means (with standard deviations) of diurnal response in stomatal conductance for plants grown at each locality.

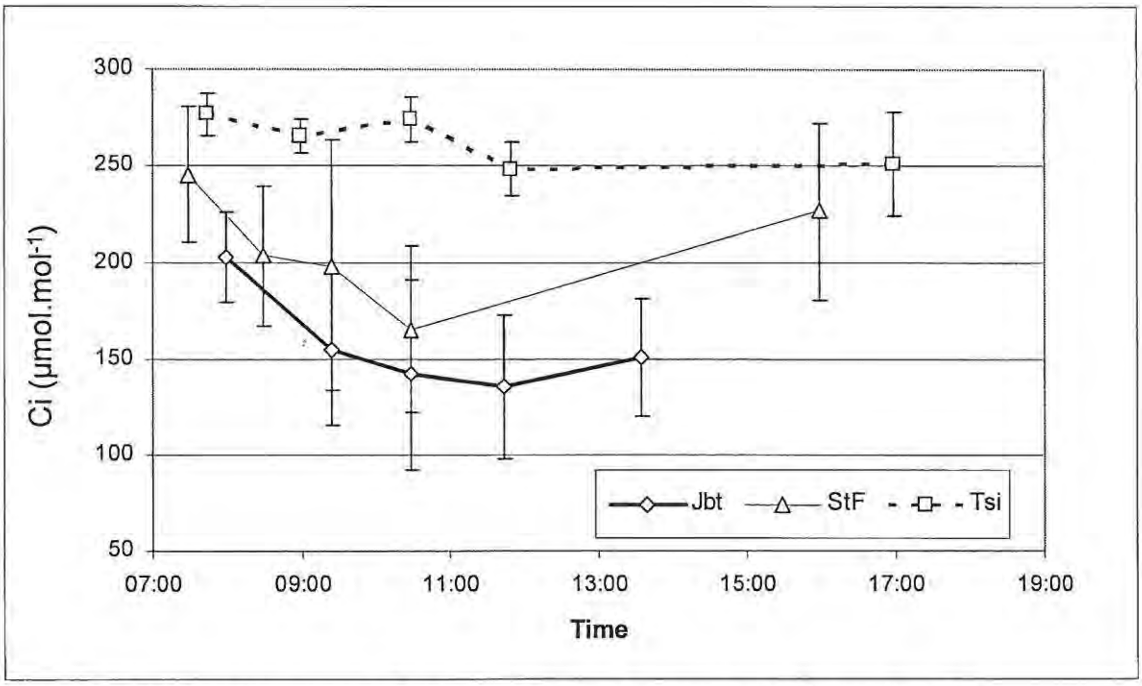

Figure 3.3.12: Means (with standard deviations) of diurnal response in leaf intercellular CO_2 concentration (Ci) values for plants growing at each locality.

Figure 3.3.10 and 3.3.11 illustrate that for the plants from Joubertina there is a closer correlation between the photosynthetic rate and the stomatal conductances of the leaves measured than with the ambient light levels experienced by the plant. Figure 3.3.12 also illustrates the changes in Ci throughout the day. The higher Ci in the Tsitsikama plants is likely due to the fact that the ambient PPFD is less than the level required for light saturated photosynthesis (see Figure 3.3.10), and while the stomata remain conductive, the CO_2 in the leaf is not being used up in photosynthetic reactions. This also explains why the Tsitsikama plants' photosynthetic rates are more strongly correlated to ambient light levels than to stomatal conductance.

A comparison of photosynthetic rate versus PPFD shows a stronger correlation for the plants from Tsitsikama and St Francis ($R^2 = 0.52$ and $R^2 = 0.50$ respectively) than the plants from Joubertina (only $R^2 = 0.03$). Whereas a comparison of photosynthetic rate versus stomatal conductance shows very clearly that the Joubertina plants' photosynthetic rates are more strongly correlated ($R^2 = 0.88$) with stomatal conductance (and potentially plant water status) than with ambient light levels. This correlation was not as strong for the other two study sites ($R^2 = 0.50$ for Tsitsikama and $R^2 = 0.41$ for St Francis). These graphs indicate that whilst the photosynthetic rates of the plants from Joubertina are governed by stomatal conductance, the photosynthetic rates of the plants from the other two sites are equally responsive to both stomatal conductance and light levels.

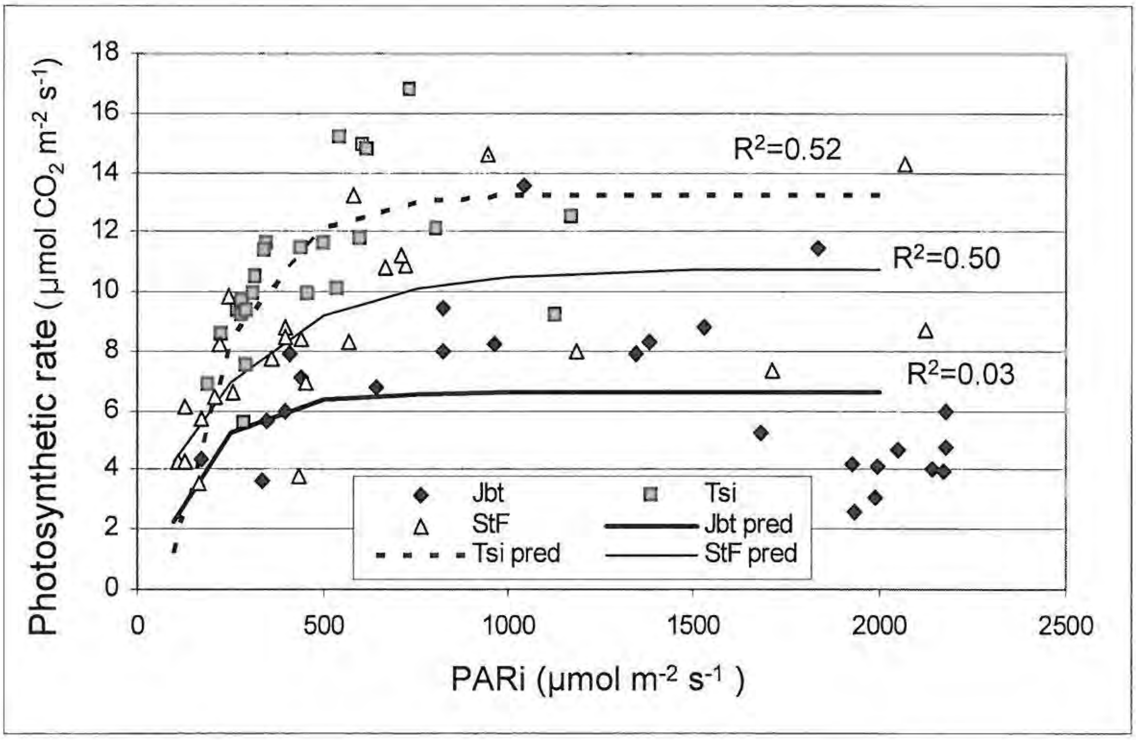

Figure 3.3.13: Photosynthetic rate versus PPFD for plants at three localities. Trend lines fitted to predicted values, R^2 values to the right of each line (N=5).

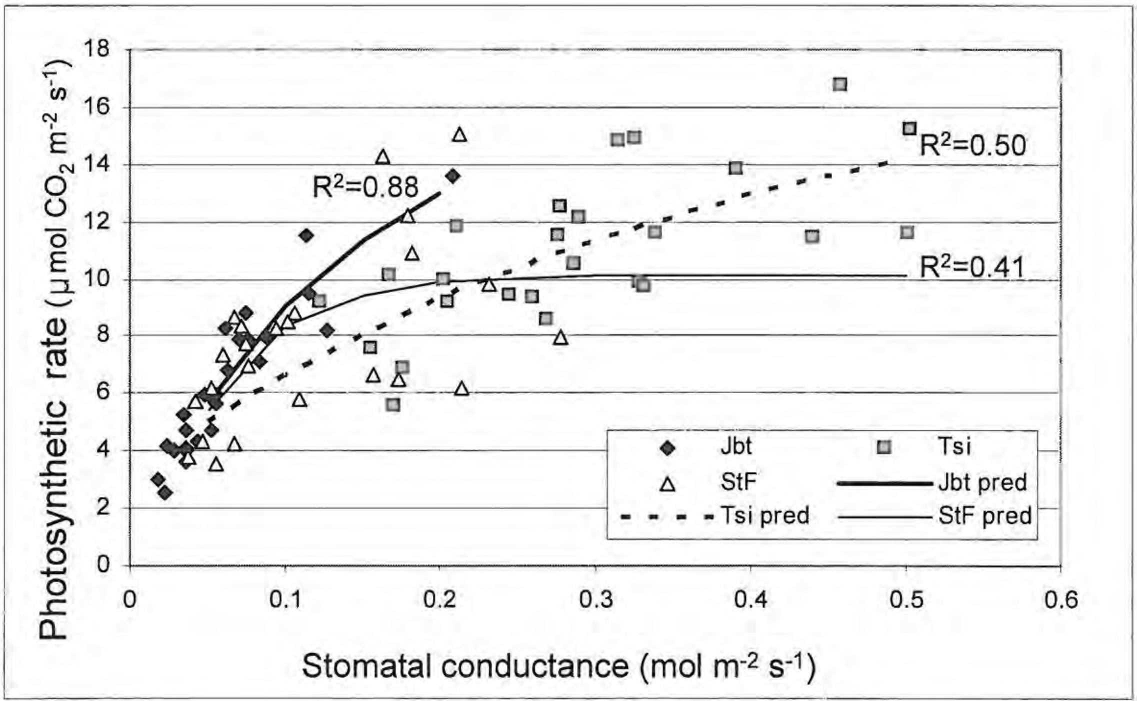

Figure 3.3.14: Photosynthetic rate versus stomatal conductance for plants at three localities. Trend lines fitted to predicted values, R^2 values next to each line (N=5).

Leaf Water Potential

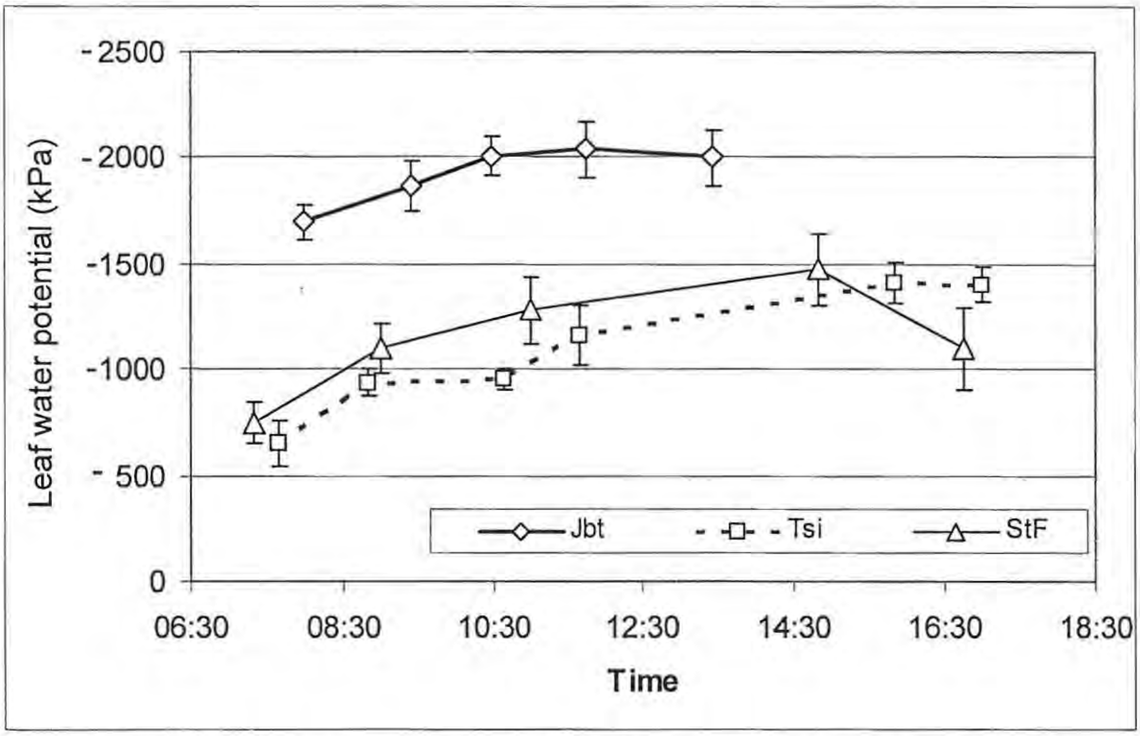

Figure 3.3.15: Means (with standard deviations) of diurnal response in leaf water potential plants from three localities.

Figure 3.3.15 shows clearly the difference between the leaf water potentials for the plants from Joubertina versus the plants from the other two study sites. The midday leaf water potentials are significantly different ($p<0.01$) between each site. Leaf water potential is a result of the interaction between the inflow supply of water to the leaf and the transpiration rates of water from the leaf. Joubertina maintains much lower leaf water potentials, indicating either a much drier climate (less water entering the leaf) or higher transpiration rates (more water exiting the leaf). As the transpiration rate of the plants from Joubertina is known (Table 3.3.1) and is not statistically different from the rate shown by the plants from St Francis, we can conclude that the difference is due to the limited supply of water to the leaf, the result of limited soil water and/or hydraulic conductance.

Table 3.3.1 is a summary table of the means of various parameters extracted from diurnal gas exchange and plant water potential measurements, calculated for the approximate times 8 am, 10.30 am and the final time available from each day's measurements. At the 10.30 am measurement, the plants from Joubertina are significantly different from those from Tsitsikama for all parameters presented in the table, with the Joubertina plants displaying much less conductive stomata, much lower photosynthetic rates (despite higher PPFDs), lower transpiration rates and much lower leaf water potentials. The Joubertina and St Francis plants, despite displaying no significant difference in transpiration rate or stomatal conductance at 10.30 am, show very different photosynthetic rates, with Joubertina plants displaying a photosynthetic rate approximately half that of the St Francis plants, despite being exposed to similar PPFDs measured at each study site.

At 8 am Joubertina plants display much lower leaf water potentials than plants at the other two sites, and they also do not show a recovery in leaf water potential at the last measurement point of the day.

Table 3.3.1: Table of means (with standard deviations) of ambient light intensity (PPFD, in μmol.m^{-2}.s^{-1}.), photosynthetic rate (A, in μmol CO_2.m^{-2}.s^{-1}), stomatal conductance (g_{ST}, in mol.m^{-2}.s^{-1}), transpiration rate (E, in mmol H_2O.m^{-2}.s^{-1}) and leaf water potential (LWP, in kPa) for three points in diurnal data. (i) First point: taken near 8 am; (ii) Middle point: taken near 10.30 am; (iii) Last: taken at 1.30 pm for Joubertina plants, 4 pm for St Francis plants, 4.45 pm for Tsitsikama plants. (Letters indicate significant differences at the 95% confidence level).

Local	PPFD	A	g_{ST}	E	LWP
	ca. 8 am				
Joubertina	803 ± 269 ^{a}	8.5 ± 0.8 ^{a}	0.119 ± 0.058 ^{a}	1.82 ± 0.67 ^{a}	-1698 ± 86 ^{a}
Tsitsikama	285 ± 65 ^{b}	9.2 ± 2.0 ^{a}	0.296 ± 0.129 ^{b}	3.49 ± 0.81 ^{b}	-648 ± 108 ^{b}
St Francis	142 ± 27 ^{b}	4.8 ± 1.1 ^{b}	0.098 ± 0.069 ^{a}	1.15 ± 0.59 ^{a}	-748 ± 99 ^{b}
	ca. 10.30 am				
Joubertina	1258 ± 846 ^{a}	4.1 ± 0.2 ^{a}	0.045 ± 0.017 ^{a}	1.25 ± 0.38 ^{a}	-2007 ± 91 ^{a}
Tsitsikama	306 ± 24 ^{b}	9.7 ± 1.7 ^{b}	0.295 ± 0.104 ^{b}	3.34 ± 0.71 ^{b}	-950 ± 50 ^{b}
St Francis	1437 ± 743 ^{a}	9.9 ± 2.8 ^{b}	0.108 ± 0.059 ^{a}	1.50 ± 0.63 ^{a}	-1280 ± 160 ^{c}
	Last time of day from diurnal				
Joubertina	1753 ± 755 ^{a}	4.7 ± 1.1 ^{a}	0.048 ± 0.018 ^{a}	1.39 ± 0.42 ^{a}	-2000 ± 127 ^{a}
Tsitsikama	800 ± 371 ^{b}	12.2 ± 2.3 ^{b}	0.237 ± 0.087 ^{b}	2.99 ± 0.75 ^{b}	-1404 ± 80 ^{b}
St Francis	304 ± 106 ^{b}	6.7 ± 1.9 ^{a}	0.112 ± 0.054 ^{a}	1.16 ± 0.44 ^{a}	-1100 ± 197 ^{c}

Leaf fluorescence

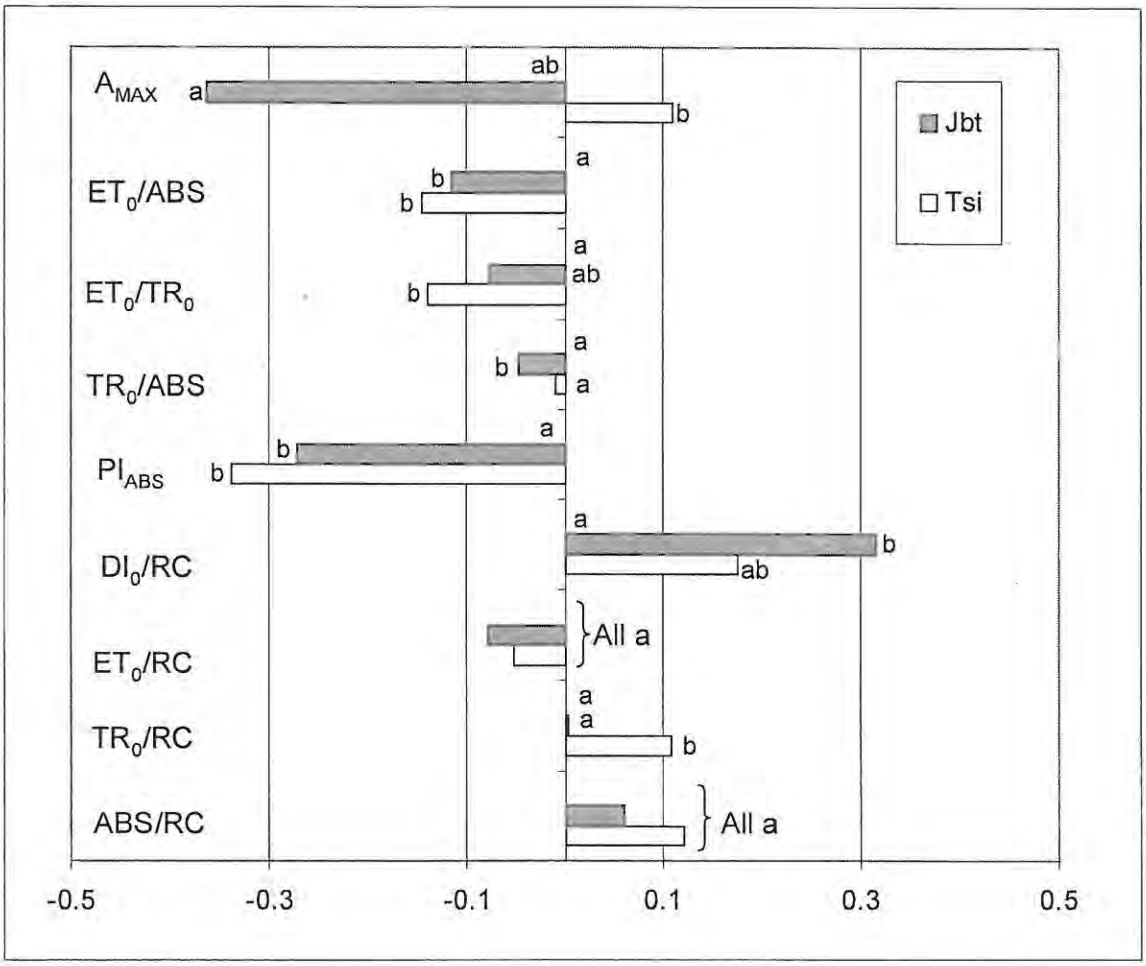

Figure 3.3.16: Graph illustrating proportional data for light saturated photosynthesis (A_{MAX}) and parameters derived from 30-minute dark adapted fluorescent transients, means plotted as a proportion of the data for St Francis bay. Bars labelled with the same letters indicate statistically homogenous groups at the 95% confidence level. The blank column at the top of each set of bars indicates the position of St Francis bay control data.

Figure 3.3.16 illustrates the results of the analysis of the data from the leaf fluorescence study. The specific energy fluxes per reaction centre for electron transport (ET_0/RC) shows no significant difference between sites, and neither did the specific energy fluxes for absorption per reaction centre (ABS/RC). The maximum quantum yield of photochemistry (TR_0/ABS) showed a significant difference ($p<0.01$) between the plants from Joubertina (which were lower) and those from the other two study sites. The ratios of trapping per reaction centre (TR_0/RC) were significantly higher ($p=0.03$) in the plants from Tsitsikama than those from the other two sites.

The performance indices (PI_{ABS}) of the plants from the three study sites were significantly lower ($p=0.02$) in the plant from Tsitsikama and Joubertina than those from St Francis. This is also seen in the quantum yield of electron transport (ET_0/ABS), which also shows significantly lower ($p=0.01$) proportions in the plants from Joubertina and Tsitsikama.

The specific energy fluxes per reaction centre for energy dissipation (DI_0/RC) in the Joubertina site plants are significantly higher ($p=0.04$) than those of the plants from St Francis, indicating that the Joubertina plants are dissipating more energy through heat and fluorescence, rather than shunting the energy to electron transport. The specific energy fluxes per reaction centre for energy dissipation for the Joubertina plants and the St Francis plants were not, however, significantly different from those of the plants from Tsitsikama.

The ratios of the efficiency with which a trapped exciton can move an electron into the transport chain further than Q_A^- (ET_0/TR_0) is significantly lower ($p=0.01$) for the Tsitsikama site plants, but the Joubertina plants were not significantly different from either of the other two sites.

The Joubertina plants showed slightly less efficient energy trapping (TR_0/ABS), slightly increased energy trapping per reaction centre, and dissipated significantly more energy as heat or fluorescence (DI_0/RC). This is possibly connected to the stomatal limitation of photosynthesis and hence a greater requirement for the dissipation of absorbed energy non-photochemically. The plants from Joubertina and Tsitsikama both showed significantly decreased electron transport efficiency (ET_0/ABS) and significantly lower performance indices (PI_{ABS}).

The chlorophyll transients are tabulated in Appendix 5.

Specific Leaf Area

Figure 3.3.17 below shows that the mean SLAs are significantly different between Joubertina and the other two sites ($p < 0.01$), indicating a difference in the general carbon allocation of the plants from the three study sites.

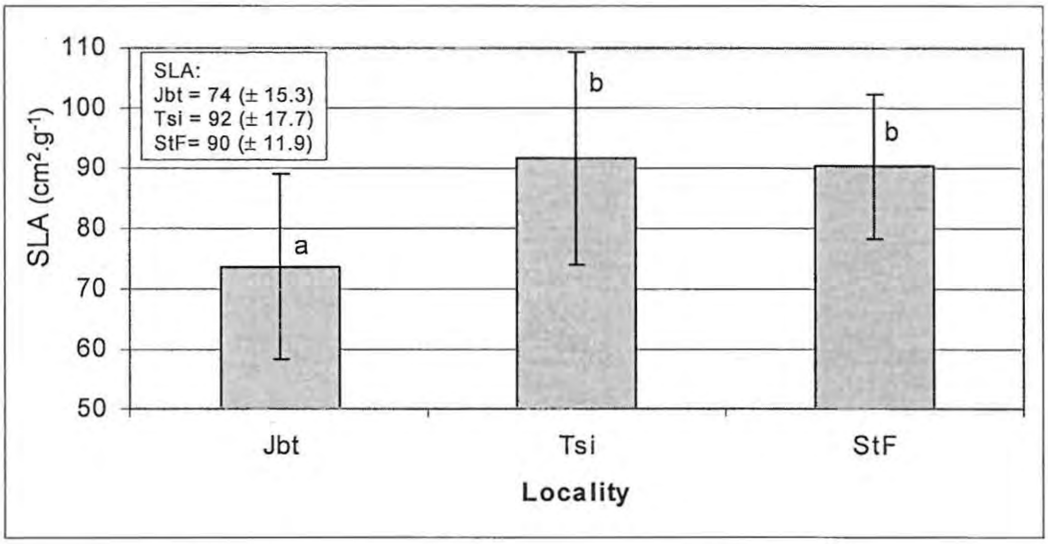

Figure 3.3.17: Means (with standard deviations) of Specific Leaf Areas of plants at each locality (values given in insert Box).

Leaf plasticity in the greenhouse study

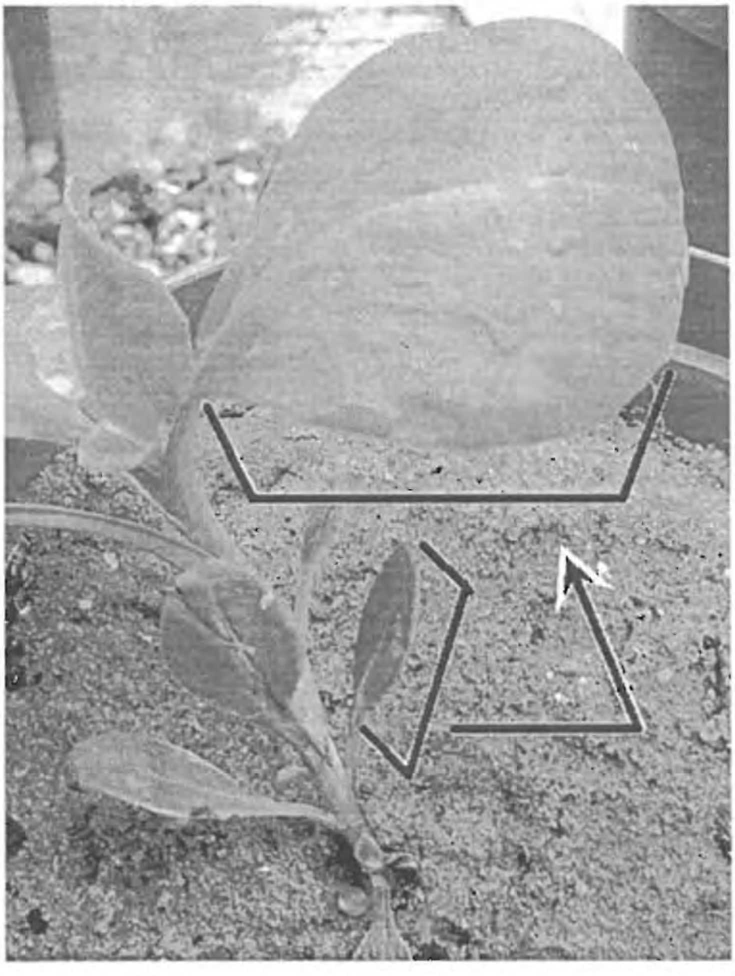

Figure 3.3.18: Example of leaf plasticity shown in greenhouse plants. The large increase in leaf size was experienced by all plants, regardless of age.

The plants grown in the greenhouse exhibited a high degree of leaf plasticity, the illustration in Figure 3.3.18 shows the massive increase in leaf size from a field grown leaf (the smaller leaf) to the greenhouse grown leaf (the larger one). This was observed in leaves of different ages and in all three plant study groups.

Discussion

Climate diagrams

A comparison of climate diagrams yields a similar pattern of climate differentiation to climographs. Hoffman *et al.* (2002) state that on a global analytical scale, distribution ranges of species may be considered as a function of the physiological constitution of the species and environmental factors. In an effort to illustrate this, Hoffman *et al.*(2002) suggested Climatic Mean Diagrams (Figure 3.4.1). These are based on climographs (Pianka, 1988), that use mean monthly rainfall as the x-axis and mean monthly temperature as the y-axis and can illustrate in a somewhat simplified but easily comparable way the climatic preferences of different plant species. As the climatic spaces of the species become directly comparable, this may offer a way for understanding evolutionary shifts in the ecogeographical constitution of closely related species. There are several possibilities of niche-occupation by species (as illustrated by Figure 3.4.1 below).

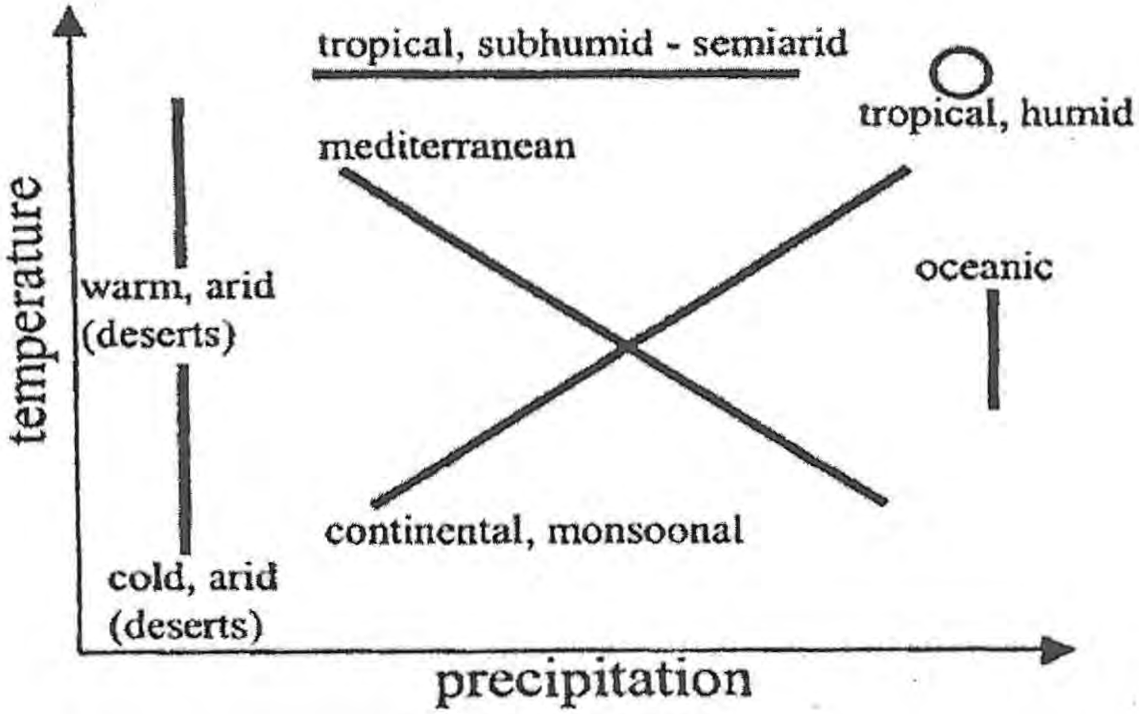

Fig 3.4.1: Theoretical Climatic Mean Diagram, showing several possibilities of niche-occupation in a variety of climates (from Hoffman *et al.*, 2002).

Hoffman *et al.* (2002) used this technique to study species level comparisons across the European continent. In *Chrysanthemoides* the span of the climatic variation is across South Africa (in one case, opposite sides of a mountain chain), and at the subspecies level. If one compares the placement of the climatic zones occupied by each ESU in Figure 3.3.2 with the models of niche-occupation in Figure 3.4.1, it can be seen that the Joubertina plants inhabit the zone near the "desert" niche, and the St Francis plants run parallel with the line titled "mediterranean". The Tsitsikama plants inhabit a much less clearly defined niche, with the November to March part of the year in the "tropical, humid" niche area, and the April-May and October period running parallel to the "continental, monsoon" niche and the climate in the July-August-September period running parallel to the "mediterranean" line.

Ecological differences between field study sites

The Joubertina study site sits in the inland rain-shadow of the Tsitsikama Mountains and receives less than 400 mm of precipitation per year, with a three-month arid period from January to March. The Tsitsikama study site is on the seaward side of the Tsitsikama Mountains; it receives up to 900 mm of rain a year, and has no arid periods in its climate profile. Despite the big difference in the climates of these two study sites, they are about 30 km (in a straight line) apart, and Joubertina is only about 20 km (in a straight line) from the sea. The St Francis Bay study site has both moderately high annual rainfall (650 mm a year) and a three-month arid period (from December to February), providing an intermediate of sorts between the two other study sites, in terms of precipitation patterns at least. The Joubertina study site, without the temperature buffering provided by the presence of the sea at the other two study sites, experiences much lower winter temperatures (about 5°C lower) than the other two sites.

The three study sites occupy different, but marginally overlapping area in climatic space (See Figure 3.3.2 for illustration). Logic would thus dictate that the plants that occur in the drier Joubertina study site would in some way be adapted to cope with much lower precipitation levels than the other two sites, and this is shown in the diurnal data (in the lower leaf water potentials, the reduced stomatal conductance and the higher level of energy dissipation by non-photochemical means).

Photosynthetic characteristics of response curves

The A:Ci curves in the field show significantly lower carboxylation efficiency in the Joubertina plants relative to the plants at the other two study sites. The other significant difference was between the RuBP regeneration rates of plants at Joubertina and the plants at Tsitsikama. The decrease in both RuBP regeneration rate and carboxylation efficiency that was noted for the Joubertina plants has been observed by Panković *et al.* (1999) in field grown sunflowers responding to drought. The greenhouse data yield no significant differences between the plants taken from each of the three study sites, indicating that any difference that existed in the field disappeared when the samples of the three ESUs were subjected to identical conditions.

The light response curves in the field study sites showed a significantly lower rate of light saturated photosynthesis in the plants at Joubertina, as compared to the plants at the other two study sites (the light compensation points from the Joubertina plants was also significantly higher than the LCPs from the Tsitsikama plants). Panković *et al.* (1999) also observed a similar drop in light saturated photosynthetic rate with increasing drought in field grown sunflowers.

In the greenhouse study, the light saturated photosynthetic rates from the light response curves for the plants from Tsitsikama were significantly lower than the rates for the plants taken from the other two study sites. This change in photosynthetic rates provides evidence that the photosynthetic functioning of the *Chrysanthemoides* plants is plastic and can change in accordance with altering environmental conditions.

Had the plants taken from Joubertina maintained their significantly lower photosynthetic rates even in the greenhouse, and the photosynthetic rates of the plants from Tsitsikama not dropped below those of the plants from Joubertina after transferral to the greenhouse, one could have drawn the conclusion that the photosynthetic functioning of the plants was genetically predetermined, which does not appear to be the case.

Diurnal data

The first indication that the diurnal data may provide a clue as to the ecophysiological functioning of the plants studies was seen in the photosynthetic rates, where, despite increasing light levels as the day progressed, the Joubertina plants showed a decrease in photosynthetic rate. Neither the Tsitsikama plants nor the St Francis Bay plants show this decrease. The St Francis Bay plants show a 200% increase in photosynthetic rate from 7.30 am level of 4.8 ± 1.1 $\mu mol.mol^{-1}$ to the 10.30 am level of 9.9 ± 2.8 $\mu mol.mol^{-1}$ followed by a slight drop to an end of the day level of 6.7 ± 1.9 $\mu mol.mol^{-1}$. The Tsitsikama plants show a similar diurnal pattern in photosynthetic rates (except for a slight dip in the 10.30 am measurement, due to cloud cover). In contrast to the plants from Joubertina, the photosynthetic rates of the plants from Tsitsikama and St Francis Bay show a trend of following the ambient light levels. The Tsitsikama study site, despite having the lowest light levels, showed the highest diurnal photosynthetic rates.

The comparison of photosynthetic rate with stomatal conductance and light level for the three study site plants is revealing: The Joubertina plants show a strong correlation ($R^2 = 0.88$) between photosynthetic rate and stomatal conductance, and a much weaker correlation with incident PPFD ($R^2 = 0.03$). The plants from the other two study sites show a slightly higher correlation between photosynthetic rate and PPFD than between photosynthetic rate and stomatal conductance ($R^2 = 0.52$ vs. $R^2 = 0.50$ for the Tsitsikama plants and $R^2 = 0.50$ vs. $R^2 = 0.41$ for the St Francis plants).

The reason for this close correlation between photosynthetic rate and stomatal conductance of the Joubertina plants may be apparent when the leaf water potentials for plants from the three sites are compared. The Joubertina plants operate at a much lower tissue water potential (nearly 200% lower than that of the other two sites at most measurement points; Table 3.3.1 and Figure 3.3.10) suggesting that a water limitation is the cause for reduced stomatal conductance. The decrease in stomatal conductance, and correlating drop in Ci and photosynthetic rate, (despite increasing light levels) is most likely a water conservation strategy for the drought-tolerant Joubertina plants (the diurnal measurements were taken near the end of the first month of the most arid part of the Joubertina yearly rain cycle). Chaves (1991) and Chaves and Pereira (1992) note in their reviews that reduced stomatal conductance is primarily responsible for limiting photosynthesis under mild to moderate water stress. This is supported by the fact that the Joubertina plants in general operate at lower stomatal conductances than the other two sites during the middle of the day, and also show a greater increase in photosynthetic rate for an equivalent increase in stomatal conductance than the other two sites.

The Joubertina and St Francis plants were both near the middle of the arid parts of their yearly rain cycles when the data were collected and both sets of plants are operating at similar transpiration rates and stomatal conductances (Table 3.3.1). At 10.30 am, the St Francis plants achieved a much higher photosynthetic rate at equivalent light levels, despite maintaining nearly identical stomatal conductances and transpiration rates. The two also have similar Ci levels (Figure 3.3.12) and when one refers to the CO_2 response curves, the relatively low Ci values indicate that the photosynthetic rates of the plants are not being subjected to limitation by their RuBP regeneration rates, but rather show a difference in photosynthetic rates due to their carboxylation efficiencies (Joubertina plants show a significantly lower carboxylation efficiency than either the Tsitsikama or St Francis plants). This suggests that the lower photosynthetic rates of the Joubertina plants are due to both stomatal and non-stomatal factors, as has been noted in studies of other Asteraceae (Panković *et al.*, 1999; and papers listed therein). The reduction in photosynthetic rate in the Joubertina plants that can be attributed to reduced carboxylation efficiency has also been observed in another member of the Asteraceae by Wise *et al.* (1991), who also noted a mid-day decline in photosynthetic rate in field grown sunflowers (*Helianthus annuus*).

Another major difference between the plants from St Francis and Joubertina lies in the leaf water potential levels: the St Francis plants have leaf water potentials nearly half as large as those of the Joubertina plants; suggesting that the Joubertina plants, whilst losing water at an equivalent rate, are taking up much less water from the soil. The Tsitsikama plants operate with a higher stomatal conductance and with transpiration rates more than double those of the plants at the other two study sites, and this, in combination with the high rainfall levels for the study site, suggest that the Tsitsikama plants do not limit their water loss as strongly as the plants at the other two study sites do.

Chaves (1991) notes that the photosynthetic apparatus shows a "remarkable resistance" to dehydration, and suggests that plant survival under drought results partly from the maintenance of full photosynthetic capacity, allowing for a rapid recovery after rehydration. The potential for damage to the photosynthetic apparatus of the Joubertina plants, from the high light levels, low transpiration rates and low photosynthetic rates, can be investigated in the chlorophyll fluorescence data.

The Joubertina plants show significantly lower performance indices (nearly 30% lower than the control, although the Tsitsikama plants show this as well), accompanied by lower electron transport efficiency (just over 10% lower, also shown by the Tsitsikama plants), less efficient energy trapping (only around 5% lower), increased energy trapping per reaction centre (10% higher), and they also dissipated significantly more energy as heat or fluorescence (just over 30% more than the control). This greater capacity for energy dissipation (either as heat or fluorescence) may indicate (in conjunction with a slightly higher light compensation point) a possible adaptation to the high-light environment that the Joubertina plants occur in.

Ecophysiological uniqueness for each ESU

The aim of this study was to investigate if each of the three genetically identified ESUs possessed unique ecophysiological functional traits, and to determine if these traits are malleable or genetically predetermined.

The three ESUs all occur in different ecological habitats, and as such are subject to different selective pressures. The Joubertina ESU, subject to the lowest precipitation and the largest seasonal temperature differences, certainly possesses some unique ecophysiological traits in the field, such as operational leaf water potentials, diurnal stomatal closure and associated photosynthetic responses, lower RuBP regeneration rates and lower carboxylation efficiencies, as well as higher non-photochemical energy dissipation in the photosynthetic apparatus. All of these clearly indicate ecophysiological adjustment to a drier environment, which may allow this ESU to occupy drier climatic areas that other *Chrysanthemoides* ESUs in the Eastern areas are not adapted to inhabit.

The Tsitsikama and St Francis ESUs, despite occupying different climatic zones, showed fewer significant differences in their ecophysiological traits (although it should be noted that the St Francis plants were measured on a sunny day, and the Tsitsikama plants were measured on a cloudy day). Those traits that did differ between the two coastal ESUs included diurnal stomatal conductances and transpiration rates, as well as a significantly different photosynthetic performance index from the leaf fluorescence study. These two ESUs also show very different branching and growth patterns, with the plants from Tsitsikama showing an upright, bushy form, and those from St Francis showing a far more prostrate, scrambling form (Griffioen, 1995; and personal observation).

When grown in the greenhouse, all the ecophysiological uniqueness that each ESU possessed disappeared, (except for a new lower light saturate photosynthetic rate shown by the plants from Tsitsikama, which was not seen in the field), clearly illustrating that under high precipitation conditions, the three ESUs appear to function very similarly. Were the three ESUs to be subjected to greenhouse conditions similar to those that the Joubertina ESU experiences in the field, the three ESUs may display very different responses. The ecophysiological traits shown by the Joubertina ESU may well be unique to that particular ESU, as evidence of adaptation to the drier climate that it grows in.

Thus the two potential ESU designations from Figure 3.1.2 becomes Figure 3.4.2 below, where the ecophysiological traits unique to each ESU, observed in the field, become a uniform response under greenhouse conditions.

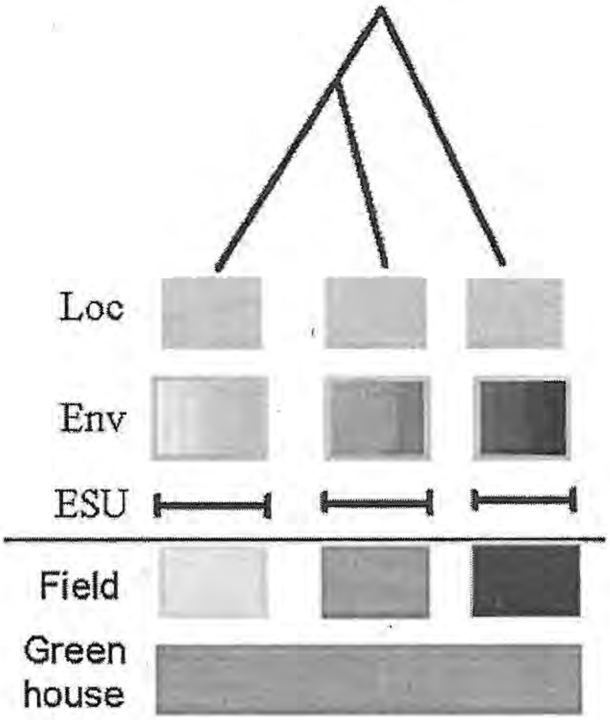

Figure 3.1.2: Schematic illustration of genealogy and environmental conditions that exist for *Chrysanthemoides* ESU sample sites, in conjunction with ecophysiological data from the field (where each ESU possesses unique functional ecophysiological characteristics) and the greenhouse (where those unique traits disappear under high precipitation conditions). (Adapted from Moritz, 2002).

Further avenues of study

Apart from greater sampling across seasonal variations (i.e. sampling these same three sites during high-rainfall periods), more insights into ecophysiological adaptation in the *Chrysanthemoides* genus could be gained by sampling from more ESUs from *C. monilifera*, and sampling of subspecies from *C. incana*, as well as measuring other parameters (e.g. carbon and nitrogen isotope studies, which have revealed patterns of water and nitrogen usage across rainfall and other environmental gradients; Jonas *et al.*, 1999).

Transplant experiments

As the Tsitsikama plants do not generally encounter arid periods in their rainfall cycles, the capacity for the Tsitsikama plants to deal with water deficits and high light levels could only be investigated by means of a reciprocal transplant experiment (or greenhouse experiments). If the Tsitsikama plants exhibit the same pattern of diurnal decrease in photosynthetic rate and stomatal conductance, then the Joubertina plants may be displaying an inherent stress-coping mechanism that is present in the C. *monilifera* species. If, however, the Tsitsikama plants fail to show this pattern (and this would most likely lead to their demise in the arid Joubertina area), the Joubertina plants may be displaying a novel ecophysiological adaptation to drier rainfall areas, that could have allowed them to occupy a niche that other *C. monilifera* ESUs could not access. If the Joubertina plants, relocated to the Tsitsikama area, lose their diurnal pattern of down-regulation (which seems a likely possibility, considering the increased rate of photosynthesis seen in the greenhouse light response curves) then they remain plastic to their environment; if however, they continue to show the pattern of stomatal control, then they have become genetically inherently adapted to their more arid environment.

This experiment could be further extended to include growing seed from the Tsitsikama area in the Joubertina area, and investigating if growth in the arid environment can prompt the development of arid adaptations seen the Joubertina plants.

Phenology

An investigation into the differences in peak flowering time and seed production (both timing and volume of peak production) in relation to rainfall peaks between different ESUs (especially as a comparison between ESUs from the "East" and "West" branches of the phylogeny derived in Chapter 2) would be highly informative as to the effects the differing rainfall patterns have on geographical reproductive separation of existing ESUs.

References

Aldrich, J., Cherney, B.W., Merlin, E., and Christopherson, L. 1988. The role of insertions/deletions in the evolution of the intergenic region between psbA and trnH in the chloroplast genome. *Current Genetics* 14: 137-146.

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

Anderson, E. 1948. Hybridization of the habitat. *Evolution*. 2:1-9.

Arnold, M.L. 1997. *Natural hybridization and evolution*. Oxford University press, Oxford.

Asmussen, C.B., and Chase, M.W. 2001. Coding and noncoding plastid DNA in palm systematics. *American Journal of Botany*. 88(6): 1103-1107.

Asmussen, C.B., and Liston, A. 1998. Chloroplast DNA characters, phylogeny and classification of *Lathyrus* (Fabaceae). *American Journal of Botany*. 85: 387-401.

Avise, J.C. 1998. The history and purview of phylogeography: a personal reflection. *Molecular Ecology* 7: 371-379

Avise, J.C. 2000. *Phylogeography: the history and formation of species*. Harvard University Press, Cambridge.

Avise, J.C., Arnold, J., Ball, R.M., Bermingham, E., Lamb, T., Neigel, J.E., Reeb, C.A., and Saunders, N.C. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics*. 18: 489-522.

Bailey, C.D., Carr, T.G., Harris, S.A. and Hughes, C.E. 2003. Characterization of angiosperm nrDNA polymorphism, paralogy and pseudogenes. *Molecular Phylogenetics and Evolution*. 29: 435-455.

Baker, W.J., Hedderson, T.A. and Dransfield, J. 2000. Molecular phylogenetics of subfamily Calamoideae (Palmae) based on nrDNA ITS and cpDNA rps16 intron sequence data. *Molecular phylogenetics and Evolution*. 14(2): 195-217.

Bakker, F.T., Culham, A., Daugherty, L.C., and Gibby, M. 1999. A *trnL-F* based phylogeny for species of *Pelargonium* (Geraniaceae) with small chromosomes. *Plant Systematics and Evolution*. 216: 309-324.

Baldwin, B.G. 1992. Phylogenetic utility of the transcribed spacer of nuclear ribosomal DNA in plants: an example from the Compositae. *Molecular Phylogenetics & Evolution*. 1: 3-16.

Baldwin, B.G. 1993. Molecular phylogenetics of *Calycadenia* (Compositae) based on ITS sequences of nuclear ribosomal DNA. *American Journal of Botany* 80:222-238.

Baldwin, B.G., Sanderson, M.J., Porter, J.M., Wojciechowski, M.F., Campbell, C.S., and Donoghue, M.J. 1995. The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Annals of the Missouri Botanical Gardens.*, 82: 247-277.

Barker, N.P., Von Senger, I., Howis, S., Zachariades, C. and Ripley, B.S. 2003 (in press). The use of Internal Transcribed Spacer (ITS) Sequence data for plant phylogeography: two examples from the Asteraceae. Submitted to "*Plant species-level systematics: patterns, processes and new applications*". (Eds) by FT Bakker, LW Chatrou, B Gravendeel & PB Pelser. Koeltz, Königstein [Regnum Vegetabile 142].

Barraclough, T.G., Hogan, J.E., and Vogler, A.P. 1999. Testing whether ecological factors promotes cladogenesis in a group of tiger beetles (Coleoptera: Cicindelidae). *Proceedings of the Royal Society of London. B.* 266: 1061-1067.

Barraclough, T.G., Vogler, A.P., and Harvey, P.H. 1998. Revealing the factors that promote speciation. *Philosophical Transactions of the Royal Society of London. B.* 353: 241-249.

Bayer, C.F. and Starr, J.R. 1998. Tribal phylogeny of the Asteraceae based on two non-coding chloroplast sequences, the *trnL* intron and *trnL/trnF* intergenic spacer. *Annals of the Missouri Botanical Gardens*. 85: 242-256.

Beadle, C.L. 1985. Plant growth analysis. In: *Techniques in bioproductivity and photosynthesis.* 2nd Edition. (Eds) Coombs, J., Hall, D.O., Long, S.P., Scurlock, J.M.O. Pergamon press, Oxford.

Bergius, P.J. 1767. *Plantae Capensis.* Stockholme Typis et Impensis Direct. Laur. Salvii.

Bermingham, E., and Moritz, C. 1998. Comparative phylogeography: concepts and applications. *Molecular Ecology* 7: 367-369.

Bleeker, W. 2003. Hybridization and *Rorippa austriaca* (Brassicaceae) invasion in Germany. *Molecular Ecology* 12: 1831-1841.

Bolhar-Nordenkampf, H. R., Long, S. P., Baker, N. R., Öquist, G., Schreiber, U., and Lechner, E. G. 1989. Chlorophyll fluorescence as a probe of the photosynthetic competence of leaves in the field: A review of current instrumentation. *Functional Ecology*. 3:497-514.

Burban, C., and Petit, R.J. 2003. Phylogeography of maritime pine inferred with organelle markers having contrasted inheritance. *Molecular Ecology* 12: 1487-1495.

Burmanni, J. 1768. *Thesaurus Zeylonicus*. Amstelaedami, Jonssonio-Waesbergios & Salomonem Schouten.

Cassini. 1818. *Bul. Sci. Sco. Philom.* Paris. 142.

Causton, D.R., and Dale, M.P. 1990. The monomolecular and rectangular hyperbola as empirical models of the response of photosynthetic rate to photon flux density, with applications to three Veronica Species. *Annals of Botany* 65:389-394.

Cavers, S., Navarro, C and Lowe, A.J. 2003. A combination of molecular markers identifies evolutionarily significant units in *Cedrela odorata* L. (Meliaceae) in Costa Rica. *Conservation Genetics* 4: 571–580.

Chandler, G.T., Bayer, R.J., and Crisp, M.D. 2001. A molecular phylogeny of the endemic Australian genus *Gastrolobium* (Fabaceae: Mirbelieae) and allied genera using chloroplast and nuclear markers. *American Journal of Botany* 88(9):1675-1687.

Chapin, F.S. 1991. Integrated responses of plants to stress. *BioScience*. 41: 29-36.

Chase, M.W., and Cox, A.V. 1998. Genes sequences, collaboration and analysis of large data sets. *Australian Systematic Botany* 11: 215-229.

Chase, M.W. and Hillis, H.H. 1991. Silica gel: an ideal material for field preservation of leaf samples for DNA studies. *Taxon*. 40: 215-220.

Chatterton, N.J., Hsiao, C., Asay, K.H., Wang, R.-C. and Jensen, K.B. 1992. Nucleotide sequence of the internal transcribed spacer region of rDNA in wheat, *Triticum aestivum* L. (Gramineae). *Plant molecular biology*. 20: 159-160.

Chaves, M.M. 1991. Effects of water deficits on carbon assimilation. *Journal of Experimental Botany*. 42: 1-16.

Chaves, M.M. and Pereira, J.S. 1992. Water stress, CO_2, and climate change. *Journal of Experimental Botany*. 43: 1131-1139.

Clegg, M.T., Brown, A.D.H., and Whitfeld, P.R. 1984a. Chloroplast DNA diversity in wild and cultivated barley: implications for genetic conservation. *Genetic Research* 43: 339-343.

Clegg, M.T., Rawson, J.R.Y., and Thomas, K. 1984b. Chloroplast DNA variation in pearly millet and related species. *Genetics*. 106: 449-461.

Clegg, M.T., Gaut, B.S., Learn, G.H., and Morton, B.R. 1995. Rates and patterns of chloroplast DNA evolution. In: *Tempo and Mode in evolution: Genetics and Paleontology 50 years after Simpson*. (Eds) W.A. Fitch & F.J. Ayala. National Academy of Sciences.

Clement, M., Posada, D., and Crandall, K.A. 2000. TCS: a computer program to estimate gene genealogies. *Molecular Ecology*. 9(10): 1657-1660.

Coleman, A.W. 2003. ITS2 is a double-edged tool for eukaryote evolutionary comparisons. *Trends in Genetics*. 19(7): 370-375.

Colwell, R.K. and Fuentes, E.R. 1975. The experimental studies of the niche. *Annual Review of Ecology and Systematics*. 6: 218-310.

Compton & Pillans. 1931. Trans. Roy. Soc. S. Afri.

Conti, E., Soltis, D.E., Hardig, T.M., and Schneider, J. 1999. Phylogenetic relationships of the silver saxifrages (*Saxifraga*, sect. Ligulatae Haworth): implications for the evolution of substrate specificity, life history and biogeography. *Molecular Phylogenetics & Evolution*. 13: 536-555.

Conti, E., Suring, E., Boyd, D., Jorgensen, J., Grant, J., and Kelso, S. 2000. Phylogenetic relationships and character evolution in *Primula* L.: the usefulness of ITS data. *Plant Biosystematics* 134: 385-392.

Cordell, S., Goldstein, G., Mueller-Dombois, D., Webb, D., and Vitousek, P.M. 1998. Physiological and morphological variation in *Metrosideros polymorpha*, a dominant Hawaiian tree species, along an altitudinal gradient: the role of phenotypic plasticity. *Oecologia* 113: 188-196.

Cowling, R.M. and Holmes, P.M. 1992. Endemism and speciation in lowland flora from the Cape Floristic Region. *Biological Journal of the Linnean Society*. 47: 367-383.

Coyne, J.A. 1996. Speciation in action. *Science*. 272: 700-701.

Cracraft, J. 1989. Speciation and its ontology: the empirical consequences of alternative species concepts for understanding patterns and processes of differentiation. In: *Speciation and its consequences*. (Eds) Otte, D., Endler, J.A. Sinauer. Sunderland.

Cracraft, J. 2000. Species concepts in theoretical and applied biology: a systematic debate with consequences. In: *Species concepts and phylogenetic theory - a debate*. (Eds) Q.D. Wheeler & R. Meier. Columbia University Press, New York.

Crandall, K.A. 1994. Intraspecific cladogram estimation: accuracy at higher levels of divergence. *Systematic Biology*. 43: 222-235.

Crandall, K.A., and Templeton, A.R. 1996. Applications of intraspecific phylogenetics. In: *New uses for new phylogenies*. (eds) Harvey, P.H., Brown, A.H.L., Smith, J.M., Nee, S. Oxford University Press, Oxford.

Critchley, C. 1988. The molecular mechanism of photoinhibition -- facts and fiction *Aust J Plant Physiology* 15: 27-41.

Crozier, R. H. 1997. Preserving the information content of species: genetic diversity, phylogeny, and conservation worth. *Annual Review of Ecology and Systematics* 28:243-268

Davis, J.I., and Manos, P.S. 1991. Isozyme variation and species delimitation in the *Puccinellia nuttalliana* complex (Poaceae): an application of the phylogenetic species concept. *Systematic Botany* 16(3): 431-445.

De Candolle, A.A.P. 1836. *Prodromus Systematis Naturalis*. Parisiis.

de Meeus, T., Durand, P., and Renaud, F. 2003. Species concepts: what for? *Trends in parasitology*. 19(10): 425-427.

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

Demmig-Adams, B., and Adams, W.W. 1996. Xanthophyll cycle and light stress in nature: uniform response to excess direct sunlight among higher plant species. *Planta*. 198: 460-470.

de Queiroz, A. 1993. For consensus (sometimes). *Systematic Biology*. 42(3): 368-372.

de Queiroz, A., Donoghue, M.J. and Kim, J. 1995. Separate versus combined analysis of phylogenetic evidence. *Annual review of Ecology and Systematics*. 26: 657-681.

Dong, J., and Wagner, D.B. 1994. Paternally inherited chloroplast polymorphism in Pinus: estimation of diversity and population subdivision, and tests of disequilibrium with a maternally inherited mitochondrial polymorphism. *Genetics*. 136: 1187-1194.

Downie, S.R., Hartman, R.L., Sun F.-J. and Katz-Downie, D.S. 2002. Polyphyly of the spring-parsleys (*Cymopterus*): molecular and morphological evidence suggests complex relationships among the perennial endemic genera of western North American Apiaceae. *Canadian Journal of Botany*. 80(12): 1295-1324.

Doyle, J.J. and Doyle, J.L. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf material. *Phytochemical bulletin*. 19(1): 11-15.

Drège, J.F. 1843. *Zwei Pflanzengeographische Documente*. Besondere Beigade Zur Flora.

Duarte, C.M. 1999. Methods in comparative functional ecology. In: *Handbook of functional plant ecology*. (Eds) Pugnaire, F.I., Valladares, F. Marcell Dekker, Inc. NewYork.

Duarte, C.M., Sand-Jensen, K., Nielsen, S.L., Enriquez, S., and Agusti, S. 1995. Comparative functional plant ecology: rationale and potentials. *Trends in Ecology and Evolution* 10(10): 418-421.

Dubouzet, J.G. and Shinoda, K. 1998. Phylogeny of *Allium* L. subg. *Malanocrommyum* (Webb et Berth.) Rouy based on DNA sequence analysis of the internal transcribed spacer region of nrDNA. *Theoretical and Applied Genetics*. 97: 541-549.

Edwards, K.J. and Gadek, P.A. 2001. Evolution and biogeography of *Alectryon* (Sapindaceae). *Molecular Phylogenetics and Evolution*. 20(1): 14-26.

Eernisse, D.J. and Kluge, A.G. 1993. Taxonomic congruence versus total evidence, and Amniote phylogeny inferred from fossils, molecules and morphology. *Molecular Biology and Evolution*. 10(6): 1170-1195.

Ellstrand, N.C., Whitkus, R., and Rieseberg, L.H. 1996. Distribution of spontaneous plant hybrids. *Proceedings of the National Academy of Science USA* 93: 5090-5093.

Engler, A. 1892. *Hochgebirgsflora des Tropischen Africa*. Berlin, Verlag Der Konigl Akademie der Wissenschaften.

Farris, J.S., Källersjö, M., Kluge, A.G., and Bult, C. 1994. Testing significance of incongruence. *Cladistics*. 10: 315-319.

Ferguson, C.J., and Jansen, R.K. 2002. A chloroplast DNA phylogeny of eastern *Phlox* (Polemoniaceae): implications of congruence and incongruence with the ITS phylogeny. *American Journal of Botany* 89(8): 1324-1335.

Flexas, J., Briantais, J.-M., Cerovic, Z., Medrano, H., and Moya, I. 2000. Steady-state and maximum chlorophyll fluorescence response to water stress in grapevine leaves: a new remote sensing system. *Remote Sensing of Environment*. 73: 283-297.

Francisco-Ortega, J., Santos-Guerra, A., Hines, A., and Jansen, R.K. 1997. Molecular evidence for a Mediterranean origin of the Macronesian endemic genus *Argyranthemum* (Asteraceae). *American Journal of Botany* 84(11): 1595-1613.

Freudenstein, J.V., and Doyle, J.J. 1994. Plastid DNA, morphological variation, and the phylogenetic species concept: the *Corallorhiza maculata* (Orchidaceae) complex. *Systematic Botany* 1994. 19(2): 273-290.

Fuertes Aguilar, J., Fryxell, P.A., and Jansen, R.K. 2003. Phylogenetic relationships and classification of the Sida generic alliance (Malvaceae) based on nrDNA ITS evidence. *Systematic Botany* 28(2): 352-364.

Fuertes Aguilar, J., Rossello, J.A., and Feliner, G.N. 1999a. Nuclear ribosomal DNA (nrDNA) concerted evolution in natural and artificial hybrids of *Armeria* (Plumbaginaceae). *Molecular Ecology* 8: 1341-1346

Fuertes Aguilar, J., Rossello, J.A., and Feliner, G.N. 1999b. Molecular evidence for the Compilospecies model of a reticulate evolution in *Armeria* (Plumbaginaceae). *Systematic Biology.* 48(4): 735-754.

Gaut, B.S., Muse, S.V., and Clegg, M.T. 1993. Relative rates of nucleotide substitution in the chloroplast genome. *Molecular Phylogenetics & Evolution* 2(2): 89-96.

Gielly, L., and Taberlet, P. 1994. The use of chloroplast DNA to resolve plant phylogenies: noncoding versus rbcL sequences. *Molecular Biology & Evolution.* 11(5): 769-777.

Gielly, L., Yuan, Y.-M., Kupper, P., and Taberlet, P. 1996. Phylogenetic use of noncoding regions in the genus *Gentiana* L.: chloroplast *trnL*(UAA) intron versus nuclear ribosomal internal transcribed spacer sequences. *Molecular Phylogenetics & Evolution* 5(3): 460-466.

Goertzen, L.R., Cannone, J.J., Gutell, R.R., and Jansen, R.K. 2003. ITS secondary structure derived from comparative analysis: implications for sequence alignment and phylogeny of the Asteraceae. *Molecular Phylogenetics and Evolution.* 29: 216-234.

Goldblatt, P. and Manning, J.C. 2002. Plant diversity of the Cape region of southern Africa. *Annals of the Missouri Botanical Gardens.* 89: 281-302.

Gottlieb, L.D. 1972. Levels of confidence in the analysis of hybridization in plants. *Annals of the Missouri Botanical Gardens* 59: 435-446.

Griffioen, R.C. 1995. *A taxonomic study of Chrysanthemoides Tourn. ex Medik. (Compositae).* Thesis, University of Cape Town.

Grube, M., and Kroken, S. 2000. Molecular approaches and the concept of species and species complexes in lichenized fungi. *Mycological Research.* 104(11): 1284-1294.

Hall, B.G. 2001. Additional methods for creating trees. In: *Phylogenetic trees made easy.* Sinaur Associates Inc. Sunderland, Massachusetts.

Hare, M.P. 2001. Prospects for nuclear gene phylogeography. *Trends in ecology and evolution* 16(12): 700-706.

Harrison, R.G. 1991. Molecular changes at speciation. *Annual Review of Ecology and Systematics* 22: 281-308.

Harvey, W.H. and Sonder, O.W. 1865. *Flora Capensis Vol 3.* Hodges, Smith and Co., Dublin.

Hasegawa, M., Kishino, H. and Yano, T. 1985. Dating the human-ape split by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution.* 22:160-174.

Heiser, C.B. 1973. Introgression re-examined. *Botanical Reviews* 39: 347-366.

Hendry, A.P., Vamosi, S.M., Latham, S.J., and Heilbuth, J.C.,& Day, T. 2000. Questioning species realities. *Conservation Genetics* 1: 67–76.

Hey, J. 2001. The mind of the species problem. *Trends in ecology and evolution* 16(7): 326-329.

Hey, J., Waples, R.S., Arnold, M.L., Butlin, R.K., and Harrison, R.G. 2003. Understanding and confronting species uncertainty in biology and conservation. *Trends in ecology and evolution* 18(11): 597-603.

Horton, P., and Bowyer, J.R. 1990. Chlorophyll fluorescence transients. In: *Methods in Plant biochemistry Volume 4.* (Eds) Bowyer, J.R. Harwood. J. Academic Press, London.

Horton, P. Ruban, A.V., and Walters, R.G. 1994. Regulation of light harvesting in green plants. *Plant Physiology* 106: 415-420.

Hudson, R.R., and Turelli, M. 2003. Stochasticity overrules the 'three-times rule': genetic drift, genetic draft, and coalescence times for nuclear loci versus mitochondrial DNA. *Evolution*. 57: 182-190.

Huelsenbeck, J.P. 1995. Performance of phylogenetic methods in simulation. *Systematic Biology*. 44(1): 17-48.

Huelsenbeck, J.P., Bull, J.J. and Cunningham, C.W. 1996. Combining data in phylogenetic analysis. *Trends in Ecology and Evolution*. 11(4): 152-158.

Huelsenbeck, J.P. and Hillis, D.M. 1993. Success of phylogenetic methods in the four-taxa case. *Systematic Biology*. 42(3): 247-264.

Huelsenbeck, J.P., and Kirkpatrick, M. 1996. Do phylogenetic methods produce trees with biased shapes? *Evolution* 50(4): 1418-1424.

Huey, R.B. 1987. Phylogeny, history and the comparative method. In: *New directions in ecological physiology*. (Eds) Feder, M.E., Bennett, A.F., Burggren, W.W., Huey, R.B. Cambridge University Press. Cambridge.

Hufford, K.M., and Mazer, S.J. 2003. Plant ecotypes: genetic differentiation in the age of ecological restoration. *Trends in ecology and evolution* 18(3): 147-155.

Jacquin, N.J. 1798. *Plantarum Rariorum Hortus Caesarei Schoenbrumensis*.

Jonas, C.S. and Geber, M.A. 1999. Variation among populations of *Clarkia unguiculata* (Onagraceae) along altitudinal and latitudinal gradients. *American Journal of Botany* 86:333-343

Johnson, L.A., and Soltis, D.E. 1995. Phylogenetic inference in Saxifragaceae *sensu stricto* and *Gilia* (Polemoniaceae) using matK sequences. *Annals of the Missouri Botanical Gardens*. 82: 149-175.

Jordan, W.C., Courtney, M.W., and Neigel, J.E. 1996. Low levels of intraspecific genetic variation at a rapidly evolving chloroplast DNA locus in North American duckweek (Lemnaceae). *American Journal of Botany* 83(4): 430-439.

Jukes, T.H., and Cantor, C.R. 1969. Evolution of protein molecules. In: *Mammalian protein metabolism*. (eds) Munro, H.N. Academic Press, New York.

Karl, S.A., and Bowen, B.W. 1999. Evolutionary significant units versus geopolitical taxonomy: molecular systematics of an endangered sea turtle (genus *Chelonia*). *Conservation biology*. 13(5): 990-999.

Kelchner, S.A. 2000. The evolution of non-coding chloroplast DNA and its application in plant systematics. *Annals of the Missouri Botanical Gardens*. 87: 482-498.

Kellogg, E.A., Appels, R., and Mason-Gamer, R.J. 1994. When genes tell different stories: the diploid genera of Triticeae (Gramineae). *Systematic Botany*. 21(3): 321-348.

Kim, S.-C., Crawford, D.J., Jansen, R.K., and Arnoldo, S.-G. 1999. The use of a non-coding region of chloroplast DNA in phylogenetic studies of the subtribe Sonchinae (Asteraceae: Lactuceae). *Plant Systematics and Evolution* 215: 85-99.

Koch, M. and Al-Shehbaz, I.A. 2000. Molecular systematics of the Chinese *Yinshania* (Brassicaceae): Evidence from plastid and nuclear ITS DNA sequence data. *Annals of the Missouri Botanical Gardens*. 87: 246-272.

Koehn, R.K. 1987. The importance of genetics to physiological ecology. In: *New directions in ecological physiology*. (Eds) Feder, M.E., Bennett, A.F., Burggren, W.W., Huey, R.B. Cambridge University Press. Cambridge.

Kollipara, K.P., Singh, R.J., and Hymnowitz, T. 1997. Phylogenetic and genomic relationships in the genus *Glycine* Willd. based on sequences from the ITS region of nuclear rDNA. *Genome*. 40: 57-68.

Krause, G.H., and Weis, E. 1991. Chlorophyll fluorescence and photosynthesis: the basics. *Annual Review of Plant Physiology and Plant Molecular Biology*. 42: 313-349.

Lashermes, P., Combes, M.C., Trouslot, P., and Charrier, A. 1997. Phylogenetic relationships of coffee-tree species (*Coffea* L.) as inferred from ITS sequences of nuclear ribosomal DNA. *Theoretical and Applied Genetics*. 94: 947-955.

Lavin, M., Mathews, S., and Hughes, C. 1991. Chloroplast DNA variation in *Gliricida sepium* (Leguminosae): intraspecific phylogeny and tokogeny. *American Journal of Botany* 78: 1576-1585.

Lee, J.-Y., Mummenhoff, K., and Bowman, J.L. 2002. Allopolyploidization and evolution of species with reduced floral structures in *Lepidium* L. (Brassicaceae). *Proceedings of the National Academy of Sciences. USA* 99(26): 16835-16840.

Leibold, M.A. 1995. The niche concept revisited: mechanistic models and community context. *Ecology.* 76(5): 1371-1382.

Levin, D.A. 1993. Local speciation in plants: the rule not the exception. *Systematic Botany.* 18(2): 197-208.

Lichtenthaler, H. K. 1996. Vegetation stress: An introduction to the stress concept in plants. *Journal of Plant Physiology.* 148:4–14.

Linder, H.P., and Midgley, J.J. 1994. Taxonomy, compositional biodiversity and functional biodiversity of fynbos. *South African Journal of Botany.* 90: 329-332.

Linder, H.P. and Vlok, J.H. 1991. The morphology, taxonomy and evolution of *Rhodocoma* (Restionaceae). *Plant Systematics and Evolution.* 175: 139-160.

Linhart, Y.B., and Grant, M.C. 1996. Evolutionary significance of local genetic differentiation in plants. *Annual Review of Ecology and Systematics* 27:237-277

Luckow, M. 1995. Species concepts: assumptions, methods and applications. *Systematic Botany.* 20(4): 589-605.

Linnaeus, C. 1753. *Species Plantarum.* Vol. 1.

Lloyd, D.G. & Calder, V.L. 1991. Multi-residue gaps, a class of molecular characters with exceptional reliability for phylogenetic analyses. *Journal of Evolutionary Biology.* 4: 9-21.

Maddison, D.R. 1991. The discovery and importance of multiple islands of most-parsimonious trees. *Systematic Biology.* 40: 315-328.

Maddison, W.P. 1993. Missing data versus missing characters in phylogenetic analysis. *Systematics Biology*. 42: 576-581.

Mason-Gamer, R.J., Kellogg, E.A. 1996. Testing for phylogenetic conflict among molecular data sets in the tribe Triticeae (Gramineae). *Systematic Biology*. 45(4): 524-545.

Matos, J.A. and Schaal, B.A. 2000. Chloroplast evolution in the *Pinus montezumae* complex: II. a coalescent approach to hybridization. *Evolution*. 54(4): 1218-1233.

Mayden, RL. 1997. A hierarchy of species concepts: the denouement in the saga of the species problem. In: *Species: The units of diversity*, (Eds) Claridge, M., Dawah, H. and Wilson, M. Chapman and Hall, London.

Mayer, M.S., and Soltis, P.S. 1994. The evolution of serpentine endemics: a chloroplast DNA phylogeny of the *Streptanthus glandulosus* complex (Cruciferae). *Systematic Botany*. 19: 557-574.

Mayr, E. 1940. Speciation phenomena in birds. *American Naturalist*., 74, 249–278.

Mayr, E. 1963. *Animal species and evolution*. Belknap Press, Cambridge.

McDade, L.A. 1992. Hybrids and phylogenetic systematics II: the impact of hybrids on cladistic analysis. *Evolution*. 46(5): 1329-1346.

McDade, L.A., Masta, S.E., Moody, M.L., and Waters, E. 2000. Phylogenetic relationships among *Acanthaceae*: evidence from two genomes. *Systematic Botany* 25(1): 106-121.

McDade, L.A., and Moody, M.L. 1999. Phylogenetic relationships among Acanthaceae: evidence from non-coding *trnL-trnF* chloroplast DNA sequences. *American Journal of Botany* 86: 70-80.

Mes, T.H.M. and t'Hart, H. 1994. *Sedum surculosum* and *S. jaccardianum* (Crassulaceae) share a unique 70bp deletion in the chloroplast DNA trnL(UAA)-trnF(GAA) intergenic spacer. *Plant Systematics and Evolution*. 206: 411-420.

Mickevich, M.F. and Farris, J.S. 1981. Implications of congruence in *Menidia. Systematic Zoology*. 30: 351-320.

Mielke, M.S., de Almeida, A.-A., Gomes, F.P., Aguilar, M.A.G., and Mangabeira, P.A.O. 2003. Leaf gas exchange, chlorophyll fluorescence and growth responses of *Genipa americana* seedlings to soil flooding. *Environmental and Experimental Botany*. 50: 221-231.

Miller, J.T., Grimes, J.W., Murphy, D.J., Bayer, R.J., and Ladiges, P.Y. 2003. A phylogenetic analysis of he Acacieae and Ingeae (Mimosoideae: Fabaceae) based on rnK, matK, psbA-trnH and trnL/trnF sequence data. *Systematic Botany*. 28(3): 558-566.

Milligan, B.G. 1991. Chloroplast DNA diversity within and among populations of *Trifolium pratense*. *Current Genetics* 19: 411-416.

Mishler, B.D., and Theriot, E.C. 2000. A critique from the Mishler and Theriot phylogenetic species concept perspective. In: *Species concepts and phylogenetic theory - a debate*. (Eds) Q.D. Wheeler & R. Meier. Columbia University Press, New York.

Moritz, C. 1994. Defining 'Evolutionarily Significant Units' for conservation. *Trends in Ecology and Evolution* 9: 373–375.

Morrell, P.L. and Rieseberg, L.H. 1998. Molecular tests of the proposed diploid hybrid origin of *Gilia achillefolia* (Polemoniaceae). *American Journal of Botany* 85(10): 1439-1453.

Mort, M.E., Soltis, D.E., Soltis, P.S., Francisco-Ortega, J., and Santos-Guerra, A. 2002. Phylogenetic and evolution of the Macronesian clade of Crassulaceae inferred from nuclear and chloroplast sequence data. *Systematic Botany*. 27(2): 271-288.

Mummenhoff, K., Bruggermann, H., and Bowman, J.L. 2001. Chloroplast DNA phylogeny and biogeography of *Lepidium* (Brassicaceae). *American Journal of Botany* 88(11): 2051-2063.

Munné-Bosch, S., Nogués, S., and Alegre, L. 1999. Diurnal variations of photosynthesis and dew absorption by leaves in two evergreen shrubs growing in Mediterranean field conditions. *New Phytologist*. 144: 109-119.

Muschler, R. 1910. In: *Bot. Jahrb.* (Ed) Engler, A.

Nagy, Z., Takács, Z., Szente, K., Csintalan, Z., Lichtenthaler, H.K., and Tuba, Z. 1998. Limitations of net CO_2 uptake in plant species of a temperate dry loess grassland. *Plant Physiology and Biochemistry*. 36(10): 753-758.

Nixon, K.C., and Wheeler, Q.D. 1990. An amplification of the phylogenetic species concept. *Cladistics* 6: 211-223.

Noor, M.A.F. 2002. Is the biological species concept showing its age? *Trends in Ecology and Evolution* 17(4): 153-154.

Norlindh, T. 1943. *Studies in the Calenduleae I. Monographs of the Genera Dimorphotheca, Castalis, Osteospermum, Gibbaria and Chrysanthemoides*. Lund. Carl Blom Botryckeri, CWK Gleerup. pp. 367-403

Norlindh, T. 1977. Calenduleae – systematic review. In: *The biology and chemistry of the Compositae* Vol 2. (Eds) Heywood, V.H., Harborne, J.B., Turner, B.L. Academic Press, London.

Olmstead, R.G., Jansen, R.K., Michaels, H.J., Downie, S.R., and Palmer, J.D. 1990. Chloroplast DNA and phylogenetic studies in the Asteridae. In: *Biological approaches and evolutionary trends in plants*. (Ed) Kawano, S. Academic Press, London.

Olmstead, R.G., and Palmer, J.D. 1992. A chloroplast DNA phylogeny of the Solanaceae: subfamilial relationships and character evolution. *Annals of the Missouri Botanical Gardens* 79: 346-360.

Olmstead, R.G., and Palmer, J.D. 1994. Chloroplast DNA systematics: a review of methods and data analysis. *American Journal of Botany* 81(9): 1205-1224.

Oxelman, B., Liden, M., Berglund, D. 1997. Chloroplast rps16 intron phylogeny of the tribe Sileneae (Caryophyllaceae). *Systematic Evolution.* 206: 393-410.

Palmer, J.D. 1987. Chloroplast DNA evolution and biosystematic uses of chloroplast DNA variation. *American Naturalist.* 130: S6-S29.

Palmer, J.D., and Zamir, D. 1982. Chloroplast DNA evolution and phylogenetic relationships in *Lycopsericon. Proceedings of the National Academy of Sciences USA* 79: 5006-5010.

Palmer, J.D., Jansen, R.K., Michaels, H.J., Chase, M.W., and Manhart, J.W. 1988. Chloroplast DNA variation and plant phylogeny. *Annals of the Missouri Botanical Gardens.* 75: 1180-1206.

Palmer, J.D., Nugent, J.M., and Hebron, L.A. 1987. Unusual structure of geranium chloroplast DNA: a triple-sized inverted repeat, extensive gene duplications, multiple inversions, and two repeat families. *Proceedings of the National Academy of Sciences USA.* 84: 769-773.

Panković, D., Sakač, Z., Keveršan, S. and Plesničar, M. 1999. Acclimation to long-term water deficit in the leaves of two sunflower hybrids: photosynthesis, electron transport and carbon metabolism. *Journal of Experimental Botany.* 50 (330): 127-138.

Patterson, T.B., Guy, R.D., and Dang, Q.L. 1997. Whole-plant nitrogen- and water-relations traits, and their associated trade-offs, in adjacent muskeg and upland boreal spruce species. *Oecologia.* 110: 160-168.

Paul, R. 2002. Species concepts versus species criteria. *Trends in parasitology.* 18(10): 439.

Paulo, O., Jordan, W.C., Bruford, M.W., and Nichols, R.A. 2002. Using nested clade analysis to assess the history of colonization and the persistence of populations of an Iberian lizard. *Molecular Ecology.* 11:809-819.

Perl-Treves, R., and Galun, E. 1985. The *Cucumis* plastome: physical map, intrageneric variation and phylogenetic relationships. *Theoretical & Applied Genetics* 71: 417-429.

Pianka, E.R. 1988. *Evolutionary Ecology.* Harper & Raw, Cambridge.

Platnick, N.I., and Wheeler, Q.D. 2000. A defense of the Phylogenetic species concept (*sensu* Wheeler and Platnick). In: *Species concepts and phylogenetic theory - a debate.* (Eds) Q.D. Wheeler & R.Meier. Columbia University Press, New York.

Popp, M. and Oxelman, B. 2001. Inferring the history of the polyploid *Silene aegaea* (Caryophyllaceae) using plastid and homoeologous nuclear DNA sequences. *Molecular Phylogenetics and Evolution.* 20(3): 474-481.

Posada, D., and Crandall, K.A. 1998. ModelTest: testing the model of DNA substitution. *Bioinformatics* 14: 817-818.

Prentice, H.C., Malm, J.U., Mateu-Andres, I., and Segarra-Moragues, J.G. 2003. Allozyme and chloroplast DNA variation in island mainland populations of the rare Spanish endemic, *Silene hifacensis* (Caryophyllaceae). *Conservation Genetics* 4: 543-555.

Rajakaruna, N., and Bohm, B.A. 1999. The edaphic factor and patterns of variation in *Lasthenia californica* (Asteraceae). *American Journal of Botany* 86:1576-1596

Reboud, X., and Zeyl, C. 1994. Organelle inheritance in plants. *Heredity.* 72: 132-140.

Riddle, B.R., and Hafner, D.J. 1999. Species as units of analysis in ecology and biogeography: time to take the blinders off. *Global Ecology & Biogeography.*, 8, 433–441.

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

Rieseberg, L.H., and Whitton, J., Linder, R. 1996. Molecular marker discordance in plant hybrid zones and phylogenetic trees. *Acta Bot. Neerl.* 45: 243-262.

Ritz, C.M., Zimmerman, N.F.A., Hellwig, F.H. 2003. Phylogeny of subsect. Meleuphorbia (A. Berger) Pax & Hoffm. (Euphorbia L.) reflects the climatic regime in South Africa. *Plant Systematics and Evolution.* 241: 245-259.

Robinson, D.E., Wagner, R.G., Bell, F.W., and Swanton, C.J. 2001. Photosynthesis, nitrogen-use efficiency, and water-use efficiency of jack pine seedlings in competition with four boreal forest plant species. *Canadian Journal of Forestry Research* 31: 2014–2025

Rodriguez, F., Oliver, J.F., Marin, A., and Medina, J.R. 1990. The general stochastic model of nucleotide substitution. *Journal of Theoretical Biology.* 142: 485-501.

Rumpunen, K., and Bartish, I.V. 2002. Comparison of differentiation estimates based on morphometric and molecular data, exemplified by various leaf shape descriptors and RAPDs in the genus *Chaenomeles* (Rosaceae). *Taxon.* 51: 69-82.

Ryder, O.A. 1986. Species conservation and systematics: the dilemma of subspecies. *Trends in Ecology and Evolution* 1(1): 9-10.

Sandquist, D.R. and Ehleringer, J.R. 1997. Intraspecific variation of leaf pubescence and drought response in *Encelia farinosa* associated with contrasting desert environments. *New Phytologist.* 135: 635-644.

Sang, T. 1996. *Phylogeny and biogeography of Paeonia (PAEONIACEAE).* Thesis, Ohio State University.

Sang, T., Crawford, D.J., Kim, S.-C., and Steussy, T.F. 1994. Radiation of the endemic genus *Dendroseris* (Asteraceae) on the Juan Fernandez Islands: evidence from sequences of the ITS regions of nuclear ribosomal DNA. *American Journal of Botany.* 81(11): 1494-1501.

Sang, T., Crawford, D.J., and Stuessy, T.F. 1997. Chloroplast DNA phylogeny, reticulate evolution and biogeography of *Paeonia* (Paeoniaceae). *American Journal of Botany* 84(8): 1120-1136.

Sang, T., and Zhang, D. 1999. Reconstructing hybrid speciation using sequences of low copy nuclear genes: hybrid origins of five *Paeonia* species based on *Adh* gene phylogenies. *Systematic Biology* 24(2): 148-163.

Santiago-Valentin, E., and Olmstead, R.G. 2003. Phylogenetics of the Antillean Goetzeoideae (Solanaceae) and their relationships within the Solanaceae based on chloroplast and ITS DNA sequence data. *Systematic Botany.* 28(2): 452-460.

Schaal, B.A., Hayworth, D.A., Olsen, K.M., Rauscherm J.T., and Smith, W.A. 1998. Phylogeographic studies in plants: problems and prospects. *Molecular Ecology* 7:465-474.

Schilling, E.E., and Panero, J.L. 1996. Phylogenetic reticulation in subtribe Helianthinae. *American Journal of Botany* 83(7): 939-948.

Schlichting, C.D. 1986. The evolution of phenotypic plasticity in plants. *Annual Review of Ecology and Systematics.* 17: 667-693.

Schlichting, C.D., and Pigliucci, M. 1995. Lost in phenotypic space: environment-dependent morphology in *Phlox drummondii* (Polemoniaceae). *International Journal of Plant Science.* 156(4): 542-546.

Schreiber, U., Bilger, W., and Neubauer, C. 1994. Chlorophyll fluorescence as a nonintrusive indicator for rapid assessment of in vivo photosynthesis. In: *Ecophysiology of Photosynthesis.* (Eds) Schulze, E-D., Caldwell, M.M. Springer- Verlag, Berlin.

Schwartz, M.W. 1999. Choosing the appropriate scale of reserves for conservation. *Annual Review of Ecology and Systematics* 30:83-108.

Scott, J.K. 1996. Population ecology of *Chrysanthemoides monilifera* in South Africa: implications for its control in Australia. *Journal of Applied Ecology.* 33: 1496-1508.

Shaw, K.L. 1998. Species and the diversity of natural groups. In: *Endless forms: Species and speciation.* (Eds) Howard, D.J. & Berlocher, S.H. Oxford University Press, New York.

Shaw, K.L. 2002. Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: what mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. *Proceedings of the National Academy of Sciences USA* 99(25): 16122-16127.

Sigiura,M. 1989. The chloroplast chromosomes in land plants. *Annual Review of Cell Biology.* 5: 51-70.

Simmons, M.P. & Ochoterena, H. 2000. Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology.* 49(2): 369-381.

Soliva, M., Kocyan, A., and Widmer, A. 2001. Molecular phylogenetics of the sexually deceptive orchid genus *Ophrys* (Orchidaceae) based on nuclear and chloroplast DNA sequences. *Molecular Phylogenetics & Evolution* 20(1): 78-88.

Soltis, D.E., and Soltis, P.S. 1998. Choosing an approach and an appropriate gene for phylogenetic analysis. In: *Plant Molecular Systematics II,* (Eds) Soltis, D. E., Soltis, P. S., & Doyle, J.J., Ch 2, Kluwer, Boston.

Soltis, D.E., Soltis, P.S., and Milligan, B.G. 1992. Intraspecific chloroplast DNA variation: systematic and phylogenetic implications. In: *Molecular systematics of plants*. (Eds) Soltis, P.S., Soltis, D.E., and Doyle, J.J. Chapman and Hall, New York.

Soltis, D.E., Soltis, P.S., Mort, M.E., Chase, M.W., Savolainen, V., Hoot, S.B., and Morton, C.M. 1998. Inferring complex phylogenies using parsimony: an empirical approach using three large DNA data sets for angiosperms. *Systematic Biology* 47:32-42.

Soltis, D.E., Soltis, P.E., Ranker, T.A., and Ness, B.D. 1989. Chloroplast DNA variation in a wild plant, *Tolmiea mensiesii*. *Genetics* 121: 819-826.

Stearns, S.C. 1989. The evolutionary significance of phenotypic plasticity. *BioScience*. 39: 436-445.

Stebbins, G.L. 1950. *Variation and evolution in plants*. Columbia University Press, New York.

Stehlik, I., Blattner, F.R., Holderegger, R. And Bachmann, K. 2002. Nunatak survival of the high Alpine plant Eritrichium nanum (L.) Gaudin in the central Alps during the ice ages. *Molecular Ecology*. 11: 2027-2036.

Strasser, B.J., and Strasser, R.J. 1995. Measuring fast fluorescence transients to address environment questions: the JIP-test. In: *Photosynthesis: from light to biosphere*. Ed: Mathis, P. Vol V: 977-980.

Strimmer, K., and von Haeseler, A. 1996. Accuracy of Neighbor Joining for n-taxon trees. *Systematic Biology*. 45(4): 516-523.

Suh, Y., Thien, L.B., Reeve, H.E., and Zimmer, E.A. 1993. Molecular evolution and phylogenetic implications of internal transcribed spacer sequences of ribosomal DNA in Winteraceae. *American Journal of Botany* 80: 1042-1055.

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

Tautz, D., Arctander, P., Minelli, A., Thomas, R.H., and Vogler, A.P. 2003. A plea for DNA taxonomy. *Trends in ecology and evolution* 18(2): 70-74.

Taylor, J.W., Jacobson, D.J., Kroken, S., Kasuga, T., Geiser, D.M., Hibbett, S., and Fisher, M.C. 2000. Phylogenetic species recognition and species concepts in fungi. *Fungal Genetics and Biology*. 31: 21-32.

Templeton, A.R. 1989. The meaning of species and speciation: a genetic perspective. In: *Speciation and its consequences*. (Eds) Otte, D., Endler, J.A. Sinauer Associates, Massachusetts.

Templeton, A.R. 1998. Nested clade analysis of phylogeographic data: testing hypotheses about gene flow and population history. *Molecular Ecology*. 7: 381-397.

Templeton, A.R., Crandall, K.A., and Sing, C.F. 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. III. Cladogram estimation. *Genetics*. 132: 619-633.

Templeton, A.R., Routman, E., and Phillips, C.A. 1995. Separating population structure from population history: a cladistics analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander *Ambystoma tigrinum*. *Genetics*. 140: 767-782.

Templeton, A.R., and Sing, C.F. 1993. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. IV. Nested analyses with cladogram uncertainty and recombination. *Genetics*. 134: 659-669.

Terauchi, R., Terachi, T., and Tsunewaki, K. 1991. Intraspecific variation of chloroplast DNA in *Dioscorea bulbifera* L. *Theoretical and Applied Genetics*. 81: 461-470.

Thomas, D.S., and Turner, D.W. 2001. Banana (*Musa* sp.) leaf gas exchange and chlorophyll fluorescence in response to soil drought, shading and lamina folding. *Scientia Horticultuae*. 90: 93-108.

Thunberg, C.P. 1823. *Flora Capensis*. Stuttgardtiae.

Tyree, M.T. 1999. Water relations and hydraulic architecture. In: *Handbook of functional plant ecology*. (Eds) Pugnaire, F.I., Valladares, F. Marcell Dekker, Inc. NewYork.

van Dijk., M.A.M., Paradis, E., Catzeflis, F. & de Jong, W.W. 1999. The virtues of gaps: Xenarthran (Edentate) monophyly supported by a unique deletion in α A-Crystallin. *Systematic Biology*. 48(1): 94-106.

van Valen, L. 1976. Ecological species, multispecies and oaks. *Taxon*. 25: 233-239.

Vogler, A.P., Knisley, C.B., Glueck, S.B., Hill, J.M., and DeSalle, R. 1993. Using molecular and ecological data to diagnose endangered populations of the puritan tiger beetle *Cicindela puritana. Molecular Ecology*. 2: 375-383.

Wagner, D.B., Furnier, G.R., Saghai-Maroof, M.A., Williams, S.M., Dancik, B.P., and Allard, R.W. 1987. Chloroplast DNA polymorphisms in lodgepole and jack pines and their hybrids. *Proceedings of the National Academy of Sciences USA* 84: 2097-2100.

Wanntorp, L., Wanntorp, H.-E. 2003. The biogeography of *Gunnera* L.: vicariance and dispersal. *Journal of Biogeography*. 30(7): 979-987.

Wanntorp, L., Wanntorp, H.-E., Oxelman, B., Kallersjo, M. 2001. Phylogeny of *Gunnera*. *Plant Systematics and Evolution*. 226: 85-107

Welker, J.M. 1993. Carbon isotope discrimination and water use efficiency. In: *Methods in comparative plant ecology*. (Eds) Hendry, G.A.F., and Grime, J.P. Chapman & Hall. London.

Wendel, J.F., Stewart, J.M., and Rettig, J.H. 1991. Molecular evidence for homoploid reticulate evolution among Australian species of *Gossypium*. *Evolution* 45: 694-711.

West, J.G., and Noble, I.R. 1984. Analyses of digitised leaf images of the *Dodonaea viscosa* complex in Australia. *Taxon*. 33(4): 595-613.

West-Eberhard, M.J. 1989. Phenotypic plasticity and the origins of diversity. *Annual Review of Ecology and Systematics*. 20: 249-278.

Wheeler, Q.D., and Platnick, N.I. 2000. The Phylogenetic Species Concept (*sensu* Wheeler & Platnick). In: *Species concepts and phylogenetic theory - a debate*. (Eds) Wheeler, Q.D. and Meier, R. Columbia University Press, New York.

White, T.J., Bruns, T., Lee, S. and Taylor, J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: *PCR protocols: a guide to methods and applications*, (Eds) Innis, M.A., Gelfand, D.H., Sninsky, J.J., White, T.J. Academic Press, Toronto.

Whitten, W.M., Williams, N.H., and Chase, M.W. 2000. Subtribal and generic relationships of Maxillarieae (Orchidaceae) with emphasis on Stanhopeinae: combined molecular evidence. *American Journal of Botany* 87(12): 1842-1856.

Wiens, J.J., and Penkrot, T.A. 2002. Delimiting species using DNA and morphological variation and discordant species limits in spiny lizards (*Sceloporus*). *Systematic Botany*. 51(1): 69-91.

Williams, D.M. 1994. Combining trees and combining data. *Taxon* 43:449-455.

Wise, R.R., Sparrow, D.H., Ortiz-Lopez, A. and Ort, D.R. 1991. Biochemical regulation during mid-day decline in field-grown sunflower. Plant Science.74: 45-52.

Wissemann, V. 1999. Genetic constitution of *Rosa* sect. Caninae (*R. canina, R. jundzillii*) and sect. Gallicanae (*R. gallica*). *Journal of Applied Botany*. 73: 191-196.

Wojciechowski, M., Sanderson, M., Baldwin, B., and Donoghue, M. 1993. Monophyly of aneuploid *Astragalus* (Fabaceae): evidence from nuclear ribosomal DNA internal transcribed spacer sequences. *American Journal of Botany* 80: 711-722.

Wolf, K.H., Li, W.-H., and Sharp, P.M. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast and nuclear DNAs. *Proceedings of the National Academy of Sciences USA* 84: 9054-9058.

Yang, F.-S., Wang, X.-Q., and Hong, D.-Y. 2003. Unexpected high divergence in nrDNA ITS and extensive parallelism in floral morphology of *Pedicularis* (Orobanchaceae). *Plant Systematics and Evolution*. 240: 91-105.

Yonemori, K., Honsho, C., Kanzaki, S., Eiadthong, W., and Sugiura, A. 2002. Phylogenetic relationships of *Mangifera* species revealed by ITS sequences of nuclear ribosomal DNA and a possibility of their hybrid origin. *Plant Systematics and Evolution*. 231: 59-75.

Zhang, D.-X., and Hewitt, G.M. 2003. Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Molecular Ecology* 12: 563-584.

Appendix 1: DNA Extraction protocol

1) Grind $2mm^2$ piece of tissue in1ml CTAB extraction buffer (see bottom for details) and 10μl β-Mercaptoethanol.
2) Decant to 1.5ml Eppindorf tube and incubate in 60°C water bath for 10-30 minutes
3) Add 500μl CIA (see bottom of page for details) and shake.
4) Centrifuge for 1 min at 13,000 rpm.
5) Remove 600μl of clear supernatant to new Eppindorf.
6) Add 400μl of Isopropanol; shake, leave on ice for 10 (or overnight if desired).
7) Centrifuge for 10minutes at 13,000 rpm.
8) Pour off supernatant.
9) Add 700μl of 70% Ethanol.
10) Pour off and dry pellet.
11) Resuspend in 300μl distilled water.

Solutions

CIA
Chloroform: isoamyl alcohol (24:1)

CTAB extraction buffer
For 100ml:

- 10ml 1M Tris [Tris(Hydroxymethyl)aminomethane]
- 28ml 5M NaCl
- 4ml 0.5M EDTA [Ethylene Diamine Tetra-acetic acid Di-sodium salt]
- 2g CTAB powder [Cetyltrimethylammonium bromide]
- 1g PVP [Polyvinyl pyrrolidone]
- 57ml water

[pH = 8]

Appendix 2: PCR reagents

Table 1: Table of PCR reagents and their volumes (in µl) in the PCR reactions.

Mg concent.	H_2O	10x	dNTPs	Primer 1	Primer 2	BioTaq	DNA	$MgCl_2$
1]	78	10	4	1	1	0.25	4	2
2]	76	"	"	"	"	"	"	4
3]	74	"	"	"	"	"	"	6
4]	72	"	"	"	"	"	"	8

Mg concent.= Magnesium concentration;
H_2O = distilled PCR-quality water;
10x = 10x Bioline NH_4 Dilution Buffer ($MgCl_2$ free);
BioTaq = Bioline Taq Polymerase;
DNA = DNA template;
$MgCl_2$ = 50mM solution of $MgCl_2$ provided with enzyme and 10x buffer.

Appendix 3: Final Sequence Alignments

<u>*trnL-trnF*</u>

```
                                                                               60
SH44_Oju        GGGCAATCCT GAGCCAAATC ACGTTTTCCA AAAATTAAGA AAGCGAAAAT AAAAAAGGAT
SH50            ?????????? GAGCCAAATC ACGTTTTCCA AAAATTAAGA AAGCGAAAAT AAAAAAGGAT
SR178           GGGCAATCCT GAGCCAAATC ACGTTTTCCA AAAATTAAGA AAGCGAAAAT AAAAAAGGAT
NPB1831         GGGCAATCCT GAGCCAAATC ACGTTTTCCA AAAATTAAGA AAGCGAAAAT AAAAAAGGAT
SH40            GGGCAATCCT GAGCCAAATC ACGTTTTCCA AAAATTAAGA AAGCGAAAAT AAAAAAGGAT
SH57            GGGCAATCCT GAGCCAAATC ACGTTTTCCA AAAATTAAGA AAGCGAAAAT AAAAAAGGAT
SH58            ?GGCAATCCT GAGCCAAATC ACGTTTTCCA AAAATTAAGA AAGCGAAAAT AAAAAAGGAT
MP452           ?????????? ?????????? ?????????? ?????????? ?????????? ??????GGAT
CP491           GGGCAATCCT GAGCCAAATC ACGTTTTCCA AAAATTAAGA AAGCGAAAAT AAAAAAGGAT
SR173           GGGCAATCCT GAGCCAAATC ACGTTTTCCA AAAATTAAGA AAGCGAAAAT AAAAAAGGAT
SH75            ?????????? ?????????? ?????????? ?????????? ?????????? ??????GGAT
SH72            ?????????? ?????????? ?????????? ?????????? ?????????? ??????GGAT
SH111           ?????????? ?????????? ?????????? ?????????? ??GCGAAAAT AAAAAAGGAT
SH88            GGGCAATCCT GAGCCAAATC ACGTTTTCCA AAAATTAAGA AAGCGAAAAT AAAAAAGGAT
SH77            ?????????? ?????????? ?????????? ?????????? ?????AAAAT AAAAAAGGAT
SH52            GGGCAATCCT GAGCCAAATC ACGTTTTCCA AAAATTAAGA AAGCGAAAAT AAAAAAGGAT
NPB1820         GGGCAATCCT GAGCCAAATC ACGTTTTCCA AAAATTAAGA AAGCGAAAAT AAAGAAGGAT
AW20            ?????????? ?????????? ?????????? ?????????? ?????????? AAAAAAGGAT

                                                                              120
SH44_Oju        AGGTGCAGAG ACTCGATGGA AGCTGTTCTA ACGAATAGAG TTGATTGTCT TACATTGGTA
SH50            AGGTGCAGAG ACTCGATGGA AGCTGTTCTA ACGAATGGAG TTGATTGTCT TACATTGATA
SR178           AGGTGCAGAG ACTCGATGGA AGCTGTTCTA ACGAATAGAG TTGATTGTCT TACATTGGTA
NPB1831         AGGTGCAGAG ACTCGATGGA AGCTGTTCTA ACGAATAGAG TTGATTGTCT TACATTGGTA
SH40            AGGTGCAGAG ACTCGATGGA AGCTGTTCTA ACGAATAGAG TTGATTGTCT TACATTGGTA
SH57            AGGTGCAGAG ACTCGATGGA AGCTGTTCTA ACGAATGGAG TTGATTGTCT TACATTGATA
SH58            AGGTGCAGAG ACTCGATGGA AGCTGTTCTA ACGAATGGAG TTGATTGTCT TACATTGATA
MP452           AGGTGCAGAG ACTCGATGGA AGCTGTTCTA ACGAATGGAG TTGATTGTCT TACATTGATA
CP491           AGGTGCAGAG ACTCGATGGA AGCTGTTCTA ACGAATAGAG TTGATTGTCT TACATTGGTA
SR173           AGGTGCAGAG ACTCGATGGA AGCTGTTCTA ACGAATAGAG TTGATTGTCT TACATTGGTA
SH75            AGGTGCAGAG ACTCGATGGA AGCTGTTCTA ACGAATGGAG TTGATTGTCT TACATTGGTA
SH72            AGGTGCAGAG ACTCGATGGA AGCTGTTCTA ACGAATGGAG TTGATTGTCT TACATTGATA
SH111           AGGTGCAGAG ACTCGATGGA AGCTGTTCTA ACGAATAGAG TTGATTGTCT TACATTGGTA
SH88            AGGTGCAGAG ACTCGATGGA AGCTGTTCTA ACGAATGGAG TTGATTGTCT TACATTGATA
SH77            AGGTGCAGAG ACTCGATGGA AGCTGTTCTA ACGAATGGAG TTGATTGTCT TACATTGGTA
SH52            AGGTGCAGAG ACTCGATGGA AGCTGTTCTA ACGAATGGAG TTGATTGTCT TACATTGATA
NPB1820         AGGTGCAGAG ACTCGATGGA AGCTGTTCTA ACGAATGGAG TTGATTGTCT TACATTGATA
AW20            AGGTGCAGAG ACTCGATGGA AGCTGTTCTA ACGAATGGAG TTGATTGTCT TACATTGGTA

                                                                              180
SH44_Oju        GAGTAATCCT TCTATCGAAA CTTCAGAAAA GATGAAGGAT AACCTGTATA CATAATACAG
SH50            GAGTAATCCT TCTATCGAAA CTTCAGAAAA GATGAAGGAT AACCTGTATA CATAATACAG
SR178           GAGTAATCCT TCTATCGAAA CTTCAGAAAA GATGAAGGAT AACCTGTATA CATAATACAG
NPB1831         GAGTAATCCT TCTATCGAAA CTTCAGAAAA GATGAAGGAT AACCTGTATA CATAATACAG
SH40            GAGTAATCCT TCTATCGAAA CTTCAGAAAA GATGAAGGAT AACCTGTATA CATAATACAG
SH57            GAGTAATCCT TCTATCGAAA CTTCAGAAAA GATGAAGGAT AACCTGTATA CATAATACAG
SH58            GAGTAATCCT TCTATCGAAA CTTCAGAAAA GATGAAGGAT AACCTGTATA CATAATACAG
MP452           GAGTAATCCT TCTATCGAAA CTTCAGAAAA GATGAAGGAT AACCTGTATA CATAATACAG
CP491           GAGTAATCCT TCTATCGAAA CTTCAGAAAA GATGAAGGAT AACCTGTATA CATAATACAG
SR173           GAGTAATCCT TCTATCGAAA CTTCAGAAAA GATGAAGGAT AACCTGTATA CATAATACAG
SH75            GAGTAATCCT TCTATCGAAA CTTCAGAAAA GATGAAGGAT AACCTGTATA CATAATACAG
SH72            GAGTAATCCT TCTATCGAAA CTTCAGAAAA GATGAAGGAT AACCTGTATA CATAATACAG
SH111           GAGTAATCCT TCTATCGAAA CTTCAGAAAA GATGAAGGAT AACCTGTATA CATAATACAG
SH88            GAGTAATCCT TCTATCGAAA CTTCAGAAAA GATGAAGGAT AACCTGTATA CATAATACAG
SH77            GAGTAATCCT TCTATCGAAA CTTCAGAAAA GATGAAGGAT AACCTGTATA CATAATACAG
SH52            GAGTAATCCT TCTATCGAAA CTTCAGAAAA GATGAAGGAT AACCTGTATA CATAATACAG
NPB1820         GAGTAATCCT TCTATCGAAA CCTCAGAAAA GATGAAGGAT AACCTGTATA CATAATACAG
AW20            GAGTAATCCT TCTATCGAAA CTTCAGAAAA GATGAAGGAT AACCTGTATA CATAATACAG
```

```
                                                                          240
SH44_Oju    AAGAATTGGT GTGAATCGAT TCCATATTCA AGAAAGAATC AAATATTCAT TGATCAAATC
SH50        AAGAATTGGT GTGAATCGAT TCCATATTCA AGAAAGAATC AAA-ATTCAT TGATCAAATC
SR178       AAGAATTGGT GTGAATCGAT TCCATATTCA AGAAAGAATC AAATATTCAT TGATCAAATC
NPB1831     AAGAATTGGT GTGAATCGAT TCCATATTCA AGAAAGAATC AAATATTCAT TGATCAAATC
SH40        AAGAATTGGT GTGAATCGAT TCCATATTCA AGAAAGAATC AAATATTCAT TGATCAAATC
SH57        AAGAATTGGT GTGAATCGAT TCCATATTCA AGAAAGAATC AAA-ATTCAT TGATCAAATC
SH58        AAGAATTGGT GTGAATCGAT TCCATATTCA AGAAAGAATC AAA-ATTCAT TGATCAAATC
MP452       AAGAATTGGT GTGAATCGAT TCCATATTCA AGAAAGAATC AAATATTCAT TGATCAAATC
CP491       AAGAATTGGT GTGAATCGAT TCCATATTCA AGAAAGAATC AAATATTCAT TGATCAAATC
SR173       AAGAATTGGT GTGAATCGAT TCCATATTCA AGAAAGAATC AAATATTCAT TGATCAAATC
SH75        AAGAATTGGT GTGAATCGAT TCCATATTCA AGAAAGAATC AAATATTCAT TGATCAAATC
SH72        AAGAATTGGT GTGAATCGAT TCCATATTCA AGAAAGAATC AAA-ATTCAT TGATCAAATC
SH111       AAGAATTGGT GTGAATCGAT TCCATATTCA AGAAAGAATC AAATATTCAT TGATCAAATC
SH88        AAGAATTGGT GTGAATCGAT TCCATATTCA AGAAAGAATC AAA-ATTCAT TGATCAAATC
SH77        AAGAATTGGT GTGAATCGAT TCCATATTCA AGAAAGAATC AAATATTCAT TGATCAAATC
SH52        AAGAATTGGT GTGAATCGAT TCCATATTCA AGAAAGAATC AAA-ATTCAT TGATCAAATC
NPB1820     AAGAATTGGT GTGAATCGAT TCCATATTCA AGAAAGAATC AAATATTCAT TGATCAAATC
AW20        AAGAATTGGT GTGAATCGAT TCCATATTCA AGAAAGAATC AAATATTCAT TGATCAAATC

                                                                          300
SH44_Oju    ATTCACTCTA TAATCTGATA GATCTTTTGA AGAACTGATT AATCGGACGA GAATAAAGAT
SH50        ATTCACTCTA TAATCTGATA GATCTTTTGA AGAACTGATT AATCGGACGA GAATAAAGAT
SR178       ATTCACTCTA TAATCTGATA GATCTTTTGA AGAACTGATT AATCGGACGA GAATAAAGAT
NPB1831     ATTCACTCTA TAATCTGATA GATCTTTTGA AGAACTGATT AATCGGACGA GAATAAAGAT
SH40        ATTCACTCTA TAATCTGATA GATCTTTTGA AGAACTGATT AATCGGACGA GAATAAAGAT
SH57        ATTCACTCTA TAATCTGATA GATCTTTTGA AGAACTGATT AATCGGACGA GAATAAAGAT
SH58        ATTCACTCTA TAATCTGATA GATCTTTTGA AGAACTGATT AATCGGACGA GAATAAAGAT
MP452       ATTCACTCTC TAATCTGATA GATCTTTTGA AGAACTGATT AATCGGACGA GAATAAAGAT
CP491       ATTCACTCTA TAATCTGATA GATCTTTTGA AGAACTGATT AATCGGACGA GAATAAAGAT
SR173       ATTCACTCTA TAATCTGATA GATCTTTTGA AGAACTGATT AATCGGACGA GAATAAAGAT
SH75        ATTCACTCTA TAATCTGATA GATCTTTTGA AGAACTGATT AATCGGACGA GAATAAAGAT
SH72        ATTCACTCTA TAATCTGATA GATCTTTTGA AGAACTGATT AATCGGACGA GAATAAAGAT
SH111       ATTCACTCTA TAATCTGATA GATCTTTTGA AGAACTGATT AATCGGACGA GAATAAAGAT
SH88        ATTCACTCTA TAATCTGATA GATCTTTTGA AGAACTGATT AATCGGACGA GAATAAAGAT
SH77        ATTCACTCTA TAATCTGATA GATCTTTTGA AGAACTGATT AATCGGACGA GAATAAAGAT
SH52        ATTCACTCTA TAATCTGATA GATCTTTTGA AGAACTGATT AATCGGACGA GAATAAAGAT
NPB1820     ATTCACTCTA TAATCTGATA GATCTTTTGA AGAACTGATT AATCGGACGA GAATAAAGAT
AW20        ATTCACTCTA TAATCTGATA GATCTTTTGA AGAACTGATT AATCGGACGA GAATAAAGAT

                                                                          360
SH44_Oju    AGAGTCCCGT TCTACATGTC AATACCGGCA ACAATGAAAT TTATAGTAAG AGGAAAATCC
SH50        AGAGTCCCGT TCTACATGTC AATACCGGCA ACAATGAAAT TTATAGTAAG AGGAAAATCC
SR178       AGAGTCCCGT TCTACATGTC AATACCGGCA ACAATGAAAT TTATAGTAAG AGGAAAATCC
NPB1831     AGAGTCCCGT TCTACATGTC AATACCGGCA ACAATGAAAT TTATAGTAAG AGGAAAATCC
SH40        AGAGTCCCGT TCTACATGTC AATACCGGCA ACAATGAAAT TTATAGTAAG AGGAAAATCC
SH57        AGAGTCCCGT TCTACATGTC AATACCGGCA ACAATGAAAT TTATAGTAAG AGGAAAATCC
SH58        AGAGTCCCGT TCTACATGTC AATACCGGCA ACAATGAAAT TTATAGTAAG AGGAAAATCC
MP452       AGAGTCCCGT TCTACATGTC AATACCGGCA ACAATGAAAT TTATAGTAAG AGGAAAATCC
CP491       AGAGTCCCGT TCTACATGTC AATACCGGCA ACAATGAAAT TTATAGTAAG AGGAAAATCC
SR173       AGAGTCCCGT TCTACATGTC AATACCGGCA ACAATGAAAT TTATAGTAAG AGGAAAATCC
SH75        AGAGTCCCGT TCTACATGTC AATACCGGCA ACAATGAAAT TTATAGTAAG AGGAAAATCC
SH72        AGAGTCCCGT TCTACATGTC AATACCGGCA ACAATGAAAT TTATAGTAAG AGGAAAATCC
SH111       AGAGTCCCGT TCTACATGTC AATACCGGCA ACAATGAAAT TTATAGTAAG AGGAAAATCC
SH88        AGAGTCCCGT TCTACATGTC AATACCGGCA ACAATGAAAT TTATAGTAAG AGGAAAATCC
SH77        AGAGTCCCGT TCTACATGTC AATACCGGCA ACAATGAAAT TTATAGTAAG AGGAAAATCC
SH52        AGAGTCCCGT TCTACATGTC AATACCGGCA ACAATGAAAT TTATAGTAAG AGGAAAATCC
NPB1820     AGAGTCCCGT TCTACATGTC AATACCGGCA ACAATGAAAT TTATAGTAAG AGGAAAATCC
AW20        AGAGTCCCGT TCTACATGTC AATACCGGCA ACAATGAAAT TTATAGTAAG AGGAAAATCC
```

```
                                                                           420
SH44_Oju     GTCGATTTCA AAAATGGTGA GGGTTCAAGT CCCTCTATCC CCAAAAAGAC CATTTGACTC
SH50         GTCGATTTCA AAAATCGTGA GGGTTCAAGT CCCTCTATCC CCAAAAAGAC CATTTGACTC
SR178        GTCGATTTCA AAAATCGTGA GGGTTCAAGT CCCTCTATCC CCAAAAAGAC CATTTGACTC
NPB1831      GTCGATTTCA AAAATCGTGA GGGTTCAAGT CCCTCTATCC CCAAAAAGAC CATTTGACTC
SH40         GTCGATTTCA AAAATCGTGA GGGTTCAAGT CCCTCTATCC CCAAAAAGAC CATTTGACTC
SH57         GTCGATTTCA AAAATCGTGA GGGTTCAAGT CCCTCTATCC CCAAAAAGAC CATTTGACTC
SH58         GTCGATTTCA AAAATCGTGA GGGTTCAAGT CCCTCTATCC CCAAAAAGAC CATTTGACTC
MP452        GTCGATTTCA AAAATCGTGA GGGTTCAAGT CCCTCTATCC CCAAAAAGAC CATTTGACTC
CP491        GTCGATTTCA AAAATCGTGA GGGTTCAAGT CCCTCTATCC CCAAAAAGAC CATTTGACTC
SR173        GTCGATTTCA AAAATCGTGA GGGTTCAAGT CCCTCTATCC CCAAAAAGAC CATTTGACTC
SH75         GTCGATTTCA AAAATCGTGA GGGTTCAAGT CCCTCTATCC CCAAAAAGAC CATTTGACTC
SH72         GTCGATTTCA AAAATCGTGA GGGTTCAAGT CCCTCTATCC CCAAAAAGAC CATTTGACTC
SH111        GTCGATTTCA AAAATCGTGA GGGTTCAAGT CCCTCTATCC CCAAAAAGAC CATTTGACTC
SH88         GTCGATTTCA AAAATCGTGA GGGTTCAAGT CCCTCTATCC CCAAAAAGAC CATTTGACTC
SH77         GTCGATTTCA AAAATCGTGA GGGTTCAAGT CCCTCTATCC CCAAAAAGAC CATTTGACTC
SH52         GTCGATTTCA AAAATCGTGA GGGTTCAAGT CCCTCTATCC CCAAAAAGAC CATTTGACTC
NPB1820      GTCGATTTCA AAAATCGTGA GGGTTCAAGT CCCTCTATCC CCAAAAAGAC CATTTGACTC
AW20         GTCGATTTCA AAAATCGTGA GGGTTCAAGT CCCTCTATCC CCAAAAAGAC CATTTGACTC

                                                                           480
SH44_Oju     CCTACCCCAA TTATTTATAT CCTTTTGATT TATCCGTTTT TCGTTAGCGG TTCAAAACTC
SH50         CCTACCCCAA TTATTTATAT CCTTTTGATT TATCCGTTTT TCGTTAGCGG TTCAAAACTC
SR178        CCTACCCCAA TTATTTATAT CCTTTTGATT TATCCGTTTT TCGTTAGCGG TTCAAAACTC
NPB1831      CCTACCCCAA TTATTTATAT CCTTTTGATT TATCCGTTTT TCGTTAGCGG TTCAAAACTC
SH40         CCTACCCCAA TTATTTATAT CCTTTTGATT TATCCGTTTT TCGTTAGCGG TTCAAAACTC
SH57         CCTACCCCAA TTATTTATAT CCTTTTGATT TATCCGTTTT TCGTTAGCGG TTCAAAACTC
SH58         CCTACCCCAA TTATTTATAT CCTTTTGATT TATCCGTTTT TCGTTAGCGG TTCAAAACTC
MP452        CCTACCCCAA TTATTTATAT CCTTTTGATT TATCCGTTTT TCGTTAGCGG TTCAAAACTC
CP491        CCTACCCCAA TTATTTATAT CCTTTTGATT TATCCGTTTT TCGTTAGCGG TTCAAAACTC
SR173        CCTACCCCAA TTATTTATAT CCTTTTGATT TATCCGTTTT TCGTTAGCGG TTCAAAACTC
SH75         CCTACCCCAA TTATTTATAT CCTTTTGATT TATCCGTTTT TCGTTAGCGG TTCAAAACTC
SH72         CCTACCCCAA TTATTTATAT CCTTTTGATT TATCCGTTTT TCGTTAGCGG TTCAAAACTC
SH111        CCTACCCCAA TTATTTATAT CCTTTTGATT TATCCGTTTT TCGTTAGCGG TTCAAAACTC
SH88         CCTACCCCAA TTATTTATAT CCTTTTGATT TATCCGTTTT TCGTTAGCGG TTCAAAACTC
SH77         CCTACCCCAA TTATTTATAT CCTTTTGATT TATCCGTTTT TCGTTAGCGG TTCAAAACTC
SH52         CCTACCCCAA TTATTTATAT CCTTTTGATT TATCCGTTTT TCGTTAGCGG TTCAAAACTC
NPB1820      CCTACCCCAA TTATTTATAT CCTTTTGATT TATCCGTTTT TCGTTAGCGG TTCAAAACTC
AW20         CCTACCCCAA TTATTTATAT CCTTTTGATT TATCCGTTTT TCGTTAGCGG TTCAAAACTC

                                                                           540
SH44_Oju     CTTCATTCAC TACTAAACGG GTCTGAGCGG AAATGCTGTT CTCTTATCAC ATGTGATATA
SH50         CTTCATTCAC TACTAAACGG GTCTGAGCGG AAATGCTGTT CTCTTATCAC ATGTGATATA
SR178        CTTCATTCAC TACTAAACGG GTCTGAGTGG AAATGCTGTT CTCTTATCAC ATGTGATATA
NPB1831      CTTCATTCAC TACTAAACGG GTCTGAGTGG AAATGCTGTT CTCTTATCAC ATGTGATATA
SH40         CTTCATTCAC TACTAAACGG GTCTGAGTGG AAATGCTGTT CTCTTATCAC ATGTGATATA
SH57         CTTCATTCAC TACTAAACGG GTCTGAGCGG AAATGCTGTT CTCTTATCAC ATGTGATATA
SH58         CTTCATTCAC TACTAAACGG GTCTGAGCGG AAATGCTGTT CTCTTATCAC ATGTGATATA
MP452        CTTCATTCAC TACTAAACGG GTCTGAGCGG AAATGCTGTT CTCTTATCAC ATGTGATATA
CP491        CTTCATTCAC TACTAAACGG GTCTGAGTGG AAATGCTGTT CTCTTATCAC ATGTGATATA
SR173        CTTCATTCAC TACTAAACGG GTCTGAGTGG AAATGCTGTT CTCTTATCAC ATGTGATATA
SH75         CTTCATTCAC TACTAAACGG GTCTGAGCGG AAATGCTGTT CTCTTATCAC ATGTGATATA
SH72         CTTCATTCAC TACTAAACGG GTCTGAGCGG AAATGCTGTT CTCTTATCAC ATGTGATATA
SH111        CTTCATTCAC TACTAAACGG GTCTGAGTGG AAATGCTGTT CTCTTATCAC ATGTGATATA
SH88         CTTCATTCAC TACTAAACGG GTCTGAGCGG AAATGCTGTT CTCTTATCAC ATGTGATATA
SH77         CTTCATTCAC TACTAAACGG GTCTGAGCGG AAATGCTGTT CTCTTATCAC ATGTGATATA
SH52         CTTCATTCAC TACTAAACGG GTCTGAGCGG AAATGCTGTT CTCTTATCAC ATGTGATATA
NPB1820      CTTCATTCAC TACTAAACGG GTCTGAGCGG AAATGCTGTT CTCTTATCAC ATGTGATATA
AW20         CTTCATTCAC TACTAAACGG GTCTGAGCGG AAATGCTGTT CTCTTATCAC ATGTGATATA
```

```
                                                                          600
SH44_Oju    TATGATACAT GTACAAATGA ACATCTTTGA GCAAGGAATC CCCGTTTG-A ATGATTCACG
SH50        TATGATACAT GTACAAATGA ACATCTTTGA GCAAGGAATC CCCGTTTG-A ATGATTCACG
SR178       TATGATACAT GTACAAATGA ACATCTTTGA GCAAGGAATC CCCGTTTG-A ATGATTCACG
NPB1831     TATGATACAT GTACAAATGA ACATCTTTGA GCAAGGAATC CCCGTTTG-A ATGATTCKCG
SH40        TATGATACAT GTACAAATGA ACATCTTTGA GCAAGGAATC CCCGTTTG-A ATGATTCACG
SH57        TATGATACAT GTACAAATGA ACATCTTTGA GCAAGGAATC CCCGTTTG-A ATGATTCCCG
SH58        TATGATACAT GTACAAATGA ACATCTTTGA GCAAGGAATC CCCGTTTG-A ATGATTCACG
MP452       TATGATACAT GTACAAATGA ACATCTTTGA GCAAGGAATC CCCGTTTG-A ATGATTCACG
CP491       TATGATACAT GTACAAATGA ACATCTTTGA GCAAGGAATC CCCGTTTG-A ATGATTCACG
SR173       TATGATACAT GTACAAATGA ACATCTTTGA GCAAGGAATC CCCGTTTG-A ATGATTCACG
SH75        TATGATACAT GTACAAATGA ACATCTTTGA GCAAGGAATC CCCGTTTG-A ATGATTCACG
SH72        TATGATACAT GTACAAATGA ACATCTTTGA GCAAGGAATC CCCGTTTG-A ATGATTCACG
SH111       TATGATACAT GTMCAAATGA ACATCTTTGG GCAAGGAATC CCCGTTTG-A ATGATTCCCG
SH88        TATGATACAT GTACAAATGA ACATCTTTGA GCAAGGAATC CCCGTTTG-A ATGATTCACC
SH77        TATGATACAT GTACAAATGA AC-TCTTTGA GCAAGGAATC CCCGTTTGGA ATGATTCCCG
SH52        TATGATACAT GTACAAATGA ACATCTTTGA GCAAGGAATC CCCGTTTG-A ATGATTCCCG
NPB1820     TATGATACAT GTACAAATGA ACATCTTTGA GCAAGGAATC CCCGTTTG-A ATGATTCACG
AW20        TATGATACAT GTACAAATGA ACATCTTTGA GCAAGGAATC CCCGTRTG-A ATGATTCACG

                                                                          660
SH44_Oju    ATCAATATCT GAAATTTACA AAGTTTTTCT TTTGACAAAT T????????? ??????????
SH50        ATCAATATCT GAAATTTACA AAGTTTTTCT TTTGACAAGT TATAGGACCT GGATGAGGCT
SR178       ATCAATATCT GAAATTTACA AAGTTTTTCT TTTGACAAGT TATAGGACCT GGATGAGGCT
NPB1831     ATCAATATCT GAAATTTACA AAGTTTTTCT TTTGACAA?? ?????????? ??????????
SH40        ATCAATATCT GAAATTTACA AAGTTTTTCT TTTGACAAAT TATAGGACCT GGATGAGGCT
SH57        ATCAATATCT GAAATTTACA AAGTTTTTCT TTKGACAAAT TATAGGAC?? ??????????
SH58        ATCAATATCT GAAATTTACA AAGTTTTTCT TTTGACAAAT TATAGGACCT GGATGAGGCT
MP452       ATCAATATCT GAAATTTACA AAGTTTTTCT TTTGACAAAT TATAGGACCT GGATGAGGCT
CP491       ATCAATATCT GAAATTTACA AAGTTTTTCT TTTGACAAAT TATAGGACCT GGATGAGG??
SR173       ATCAATATCT GAAATTTACA AAGTTTTTCT TTTGACAAAT TATAGGACCT GGATGAGGCT
SH75        ATCAATATCT GAAATTTACA AAGTTTTTCT TTTGACAAAT TATAGGACCT GGATGAGGCT
SH72        ATCAATATCT GAAATTTACA AAGTTTTTCT TTTGACAAGT TATAGGACCT G?????????
SH111       CTCAATAT?? ?????????? ?????????? ?????????? ?????????? ??????????
SH88        ATCAATATCT GAAATTTACA AAGTTTTTCT TTTGACAAAT TATAGGACCT GGA???????
SH77        ATCAATATCT GAATTTT??? ?????????? ?????????? ?????????? ??????????
SH52        ATCAATATCT GAAATTTACA AAGTTTTTCT TTTGACAAAT TATAGGACCT GGATGAGGCT
NPB1820     ATCAATATCT GAAATTTACA AAGTTTTTCT TTTGACAAAT TATAGGACCT GGAGGAGGC?
AW20        ATCAATATCT GAAATTTACA AAGTTTTTCT TTTGACAAAT TATAGGACCT GGATGAGGCT


SH44_Oju    ?????????? ?????????? ????
SH50        TTGTA-CAAT TGACATAGAC CCCC
SR178       TTGTAGCAAT TGACATAGAC CCCC
NPB1831     ?????????? ?????????? ????
SH40        TTGTAGCAAT TGACATAGAC CCCC
SH57        ?????????? ?????????? ????
SH58        ?????????? ?????????? ????
MP452       TTGTA-CAAT TGACATAGAC CCCC
CP491       ?????????? ?????????? ????
SR173       TTGTAGCAAT TGACATAGAC CCCC
SH75        TTGTA-CAAT TGACATAGAC CCCC
SH72        ?????????? ?????????? ????
SH111       ?????????? ?????????? ????
SH88        ?????????? ?????????? ????
SH77        ?????????? ?????????? ????
SH52        TTGTA-CAAT TGACATAGAC CCC?
NPB1820     ?????????? ?????????? ????
AW20        TTGTA-CAAT TGACATAGAC CCCC
```

psbA-trnH

```
                                                                               60
SH44_Oju     CGTAATGCTC ATAATTTCCC TCTAGACTTA GCTGCTATTG AAGCTCCATC TACAAATGGA
SH50         CGTAATGCTC ATAATTTCCC TCTAGACTTA GCTGCTATTG AAGCTCCATC TACAAATGGA
SR178        CGTAATGCTC ATAATTTCCC TCTAGACTTA GCTGCTATTG AAGCTCCATC TACAAATGGA
NPB1831      ?GTAATGCTC ATAATTTCCC TCTAGACTTA GCTGCTATTG AAGCTCCATC TACAAATGGA
SH40         ?????????? ?????????? ?????????? ?????????? ?????CCATC TACAAATGGA
SH57         ?????????? ?????????? ?????????? ?????????G AAGCTCCATC TACAAATGGA
SH58         CGTAATGCTC ATAATTTCCC TCTAGACTTA GCTGCTATTG AAGCTCCATC TACAAATGGA
MP452        ?GTAATGCTC ATAATTTCCC TCTAGACTTA GCTGCTATTG AAGTTCCATC TATAAATGGA
CP491        ?????????? ?????????? ?????????? ?????????? ?AGCTCCATC TACAAATGGA
SR173        ?????????? ?TAATTTCCC TCTAGACTTA GCTGCTATTG AAGCTCCATC TACAAATGGA
SH75         ?????????? ?????????C TCTAGACTTA TCTGCTATTG GAGCTCCATC TACAA-TGGA
SH72         ?GTAATGCTC ATAATTTCCC TCTAGACTTA GCTGCTATTG AAGCTCCATC TACAAATGGA
SH111        CGTAATGCTC ATAATTTCCC TCTAGACTTA GCTGCTATTG AAGCTCCATC TACAAATGGA
SH88         ??TAATGCTC ATAATTTCCC TCTAGACTTA GCTGCTATTG AAGCTCCATC TACAAATGGA
SH77         ?????????? ?????????? ????GACTTA GCTGCTATTG AAGCTCCATC TACAAATGGA
SH52         ?????????? ?????????? ?????????? ?CTGCTATTG AAGCTCCATC TACAAATGGA
NPB1820      ?????????? ?????????? ???AGACTTA GCTGCTATTG AAGCTCCATC TACAAATGGA
AW20         CGTAATGCTC ATAATT-CCC TCTAGACTTA GCTGCTATTG AAGCTCCATC TACAAATGGA

                                                                              120
SH44_Oju     TAAGACTTTG GTCTGATTGT ATAGGAGTTT TTGAACTAAA AAAGGAGCAA CAGCTTCCCT
SH50         TAAGACTTTG GTCTGATTGT ATAGGAGTTT TTGAACTAAA AAAGGAGCAA CAGCTTCCCT
SR178        TAAGACTTTG GTCTGATTGT ATAGGAGTTT TTGAACTAAA AAAGGAGCAA CAGCTTCCCT
NPB1831      TAAGACTTTG GTCTGATTGT ATAGGAGTTT TTGAACTAAA AAAGGAGCAA CAGCTTCCCT
SH40         TAAGACTTTG GTCTGATTGT ATAGGAGTTT TTGAACTAAA AAAGGAGCAA CAGCTTCCCT
SH57         TAAGACTTTG GTCTGATTGT ATAGGAGTTT TTGAACTAAA AAAGGAGCAA CAGCTTCCCT
SH58         TAAGACTTTG GTCTGATTGT ATAGGAGTTT TTGAACTAAA AAAGGAGCAA CAGCTTCCCT
MP452        TAAGACTTTG GTCTGATTGT ATAGGAGTTT TTGAACTAAA AAAGGAGCAA CAGCTTCCCT
CP491        TAAGACTTTG GTCTGATTGT ATAGGAGTTT TTGAACTAAA AAAGGAGCAA CAGCTTCCCT
SR173        TAAGACTTTG GTCTGATTGT ATAGGAGTTT TTGAACTAAA AAAGGAGCAA CAGCTTCCCT
SH75         TAAGACTTTG GTCTGATTGT ATAGGAGTTT TTGAACTAAA AAAGGAGCAA CAGCTTCCCT
SH72         TAAGACTTTG GTCTGATTGT ATAGGAGTTT TTGAACTAAA AAAGGAGCAA CAGCTTCCCT
SH111        TAAGACTTTG GTCTGATTGT ATAGGAGTTT TTGAACTAAA AAAGGAGCAA CAGCTTCCCT
SH88         TAAGACTTTG GTCTGATTGT ATAGGAGTTT TTGAACTAAA AAAGGAGCAA CAGCTTCCCT
SH77         TAAGACTTTG GTCTGATTGT ATAGGAGTTT TTGAACTAAA AAAGGAGCAA CAGCTTCCCT
SH52         TAAGACTTTG GTCTGATTGT ATAGGAGTTT TTGAACTAAA AAAGGAGCAA CAGCTTCCCT
NPB1820      TAAGACTTTG GTCTGATTGT ATAGGAGTTT TTGAACTAAA AAAGGAGCAA CAGCTTCCCT
AW20         TAAGACTTTG GTCTGATTGT ATAGGAGTTT TTGAACTAAA AAAGGAGCAA CAGCTTCCCT

                                                                              180
SH44_Oju     CTTGTTTTAT CAAGAGGGCG GTATTGCTCC TTTTTTTATT TAGTACTATT TGCCTTACAT
SH50         CTTGTTTTAT CAAGAGGGCG TTATTGCTCC TTTTTTTATT TAGTACTATT TGCCTTACAT
SR178        CTTGTTTTAT CAAGAGGGCG TTATTGCTCC TTTTTTTATT TAGTACTATT TGCCTTACAT
NPB1831      CTTGTTTTAT CAAGAGGGCG TTATTGCTCC TTTTTTTATT TAGTACTATT TGCCTTACAT
SH40         CTTGTTTTAT CAAGAGGGCG TTATTGCTCC TTTTTTTATT TAGTACTATT TGCCTTACAT
SH57         CTTGTTTTAT CAAGAGGGCG TTATTGCTCC TTTTTTTATT TAGTACTATT TGCCTTACAT
SH58         CTTGTTTTAT CAAGAGGGCG TTATTGCTCC TTTTTTTATT TAGTACTATT TGCCTTACAT
MP452        CTTGTTTTAT CAAGAGGGCG TTATTGCTCC TTTTTTTATT TAGTACTATT TGCCTTACAT
CP491        CTTGTTTTAT CAAGAGGGCG TTATTGCTCC TTTTTTTATT TAGTACTATT TGCCTTACAT
SR173        CTTGTTTTAT CAAGAGGGCG TTATTGCTCC TTTTTTTATT TAGTACTATT TGCCTTACAT
SH75         CTTGTTTTAT CAAGAGGGCG TTATTGCTCC TTTTTTTATT TAGTACTATT TGCCTTACAT
SH72         CTTGTTTTAT CAAGAGGGCG TTATTGCTCC TTTTTTTATT TAGTACTATT TGCCTTACAT
SH111        CTTGTTTTAT CAAGAGGGCG TTATTGCTCC TTTTTTTATT TAGTACTATT TGCCTTACAT
SH88         CTTGTTTTAT CAAGAGGGCG TTATTGCTCC TTTTTTTATT TAGTACTATT TGCCTTACAT
SH77         CTTGTTTTAT CAAGAGGGCG TTATTGCTCC TTTTTTTATT TAGTACTATT TGCCTTACAT
SH52         CTTGTTTTAT CAAGAGGGCG TTATTGCTCC TTTTTTTATT TAGTACTATT TGCCTTACAT
NPB1820      CTTGTTTTAT CAAGAGGGCG TTATTGCTCC TTTTTTTATT TAGTACTATT TGCCTTACAT
AW20         CTTGTTTTAT CAAGAGGGCG TTATTGCTCC TTTTTTTATT TAGTACTATT TGCCTTACAT
```

```
                                                                          240
SH44_Oju    AGTTTCTTTA AAAATAAAAA GAATATGGGC TTTTTCTAGT TTGGTTCGAT TCGCATGTTT
SH50        AGTTTCTTTA AAACTAAAAA GAATATGGGC TTTTTCTAGT TTGGTTCGAT TCGCATGTTT
SR178       AGTTTCTTTA AAAATAAAAA GAATATGGGC TTTTTCTAGT TTGGTTCGAT TCGCATGTTT
NPB1831     AGTTTCTTTA AAAATAAAAA GAATATGGGC TTTTTCTAGT TTGGTTCGAT TCGCATGTTT
SH40        AGTTTCTTTA AAAATAAAAA GAATATGGGC TTTTTCTAGT TTGGTTCGAT TCGCATGTTT
SH57        AGTTTCTTTA AAACTAAAAA GAATATGGGC TTTTTCTAGT TTGGTTCGAT TCGCATGTTT
SH58        AGTTTCTTTA AAACTAAAAA GAATATGGGC TTTTTCTAGT TTGGTTCGAT TCGCATGTTT
MP452       AGTTTCTTTA AAACTAAAAA GAATATGGGC TTTTTCTAGT TTGGTTCGAT TCGCATGTTT
CP491       AGTTTCTTTA AAAATAAAAA GAATATGGGC TTTTTCTAGT TTGGTTCGAT TCGCATGTTT
SR173       AGTTTCTTTA AAAATAAAAA GAATATGGGC TTTTTCTAGT TTGGTTCGAT TCGCATGTTT
SH75        AGTTTCTTTA AAAATAAAAA GAATATGGGC TTTTTCTAGT TTGGTTCGAT TCGCATGTTT
SH72        AGTTTCTTTA AAACTAAAAA GAATATGGGC TTTTTCTAGT TTGGTTCGAT TCGCATGTTT
SH111       AGTTTCTTTA AAAATAAAAA GAATATGGGC TTTTTCTAGT TTGGTTCGAT TCGCATGTTT
SH88        AGTTTCTTTA AAACTAAAAA GAATATGGGC TTTTTCTAGT TTGGTTCGAT TCGCATGTTT
SH77        AGTTTCTTTA AAAATAAAAA GAATATGGGC TTTTTCTAGT TTGGTTCGAT TCGCATGTTT
SH52        AGTTTCTTTA AAACTAAAAA GAATATGGGC TTTTTCTAGT TTGGTTCGAT TCGCATGTTT
NPB1820     AGTTTCTTTA AAACTAAAAA GAATATGGGC TCTTTCTAGT TTGGTTCGAT TCGCATGTTT
AW20        AGTTTCTTTA AAAATAAAAA GAATATGGGC TTTTTCTAGT TTGGTTCGAT TCGCATGTTT

                                                                          300
SH44_Oju    TCTCTTTGTA TTCATTTATA TTATAGGTAT AGGTTTCTAT ATCTTTTTCC CAATCTTTTA
SH50        TCTCTTTGTG TTCATTTATA TTATAGGTAT AGGTTTCTAT ATCTTTTTCC CAATCTTTTA
SR178       TCTCTTTGTA TTCATTTATA TTATAGGTAT AGGTTTCTAT ATCTTTTTCC CAATCTTTTA
NPB1831     TCTCTTTGTA TTCATTTATA TTATAGGTAT AGGTTTCTAT ATCTTTTTCC CAATCTTTTA
SH40        TCTCTTTGTA TTCATTTATA TTATAGGTAT AGGTTTCTAT ATCTTTTTCC CAATCTTTTA
SH57        TCTCTTTGTA TTCATTTATA TTATAGGTAT AGGTTTCTAT ATCTTTTTCC CAATCTTTTA
SH58        TCTCTTTGTG TTCATTTATA TTATAGGTAT AGGTTTCTAT ATCTTTTTCC CAATCTTTTA
MP452       TCTCTTTGTA TTCATTTATA TTATAGGTAT AGGTTTCTAT ATCTTTTTCC CAATTTTTTA
CP491       TCTCTTTGTA TTCATTTATA TTATAGGTAT AGGTTTCTAT ATCTTTTTCC CAATCTTTTA
SR173       TCTCTTTGTA TTCATTTATA TTATAGGTAT AGGTTTCTAT ATCTTTTTCC CAATCTTTTA
SH75        TCTCTTTGTA TTCATTTATA TTATAGGTAT AGGTTTCTAT ATCTTTTTCC CAATCTTTTA
SH72        TCTCTTTGTA TTCATTTATA TTATAGGTAT AGGTTTCTAT ATCTTTTTCC CAATCTTTTA
SH111       TCTCTTTGTA TTCATTTATA TTATAGGTAT AGGTTTCTAT ATCTTTTTCC CAATCTTTTA
SH88        TCTCTTTGTG TTCATTTATA TTATAGGTAT AGGTTTCTAT ATCTTTTTCC CAATCTTTTA
SH77        TCTCTTTGTA TTCATTTATA TTATAGGTAT AGGTTTCTAT ATCTTTTTCC CAATCTTTTA
SH52        TCTCTTTGTG TTCATTTATA TTATAGGTAT AGGTTTCTAT ATCTTTTTCC CAATCTTTTA
NPB1820     TCTCTTTGTA TTCATTTATA TTATAGGTAT AGGTTTCTAT ATCTTTTTCC CAATCTTTTA
AW20        TCTCTTTGTA TTCATTTATA TTATAGGTAT AGGTTTCTAT ATCTTTTTCC CAATCTTTTA

                                                                          360
SH44_Oju    TGAAGTTTTA TTTCCAATTC AATTTCATTC AATTTCAACT GAAAATAGAT AAAAATGAAA
SH50        TCAAGTTTTA TTTCCAATTC AATTTCA--- -------ACT GAAAATAGAT AAAAATGAAA
SR178       TGAAGTTTTA TTTCCAATTC AATTTCATTC AATTTCAACT GAAAATAGAT AAAAATGAAA
NPB1831     TGAAGTTTTA TTTCCAATTC AATTTCATTC AATTTCAACT GAAAATAGAT AAAAATGAAA
SH40        TGAAGTTTTA TTTCCAATTC AATTTCATTC AATTTCAACT GAAAATAGAT AAAAATGAAA
SH57        TCAAGTTTTA TTTCCAATTC AATTTCA--- -------ACT GAAAATAGAT AAAAATGAAA
SH58        TCAAGTTTTA TTTCCAATTC AATTTCA--- -------ACT GAAAATAGAT AAAAATGAAA
MP452       TCAAGTTTTA TTTCCAATTC AATTTCA--- -------ACT GAAAATAGAT AAAAATGAAA
CP491       TGAAGTTTTA TTTCCAATTC AATTTCATTC AATTTCAACT GAAAATAGAT AAAAATGAAA
SR173       TGAAGTTTTA TTTCCAATTC AATTTCATTC AATTTCAACT GAAAATAGAT AAAAATGAAA
SH75        TGAAGTTTTA TTTCCAATTC AATTTCATTC AATTTCAACT GAAAATAGAT AAAAATGAAA
SH72        TCAAGTTTTA TTTCCAATTC AATTTCA--- -------ACT GAAAATAGAT AAAAATGAAA
SH111       TGAAGTTTTA TTTCCAATTC AATTTCATTC AATTTCAACT GAAAATAGAT AAAAATGAAA
SH88        TCAAGTTTTA TTTCCAATTC AATTTCA--- -------ACT GAAAATAGAT AAAAATGAAA
SH77        TGAAGTTTTA TTTCCAATTC AATTTCATTC AATTTCAACT GAAAATAGAT AAAAATGAAA
SH52        TCAAGTTTTA TTTCCAATTC AATTTCA--- -------ACT GAAAATAGAT AAAAATGAAA
NPB1820     TCAAGTTTTA TTTCCAATTC AATTTCA--- -------ACT GAAAATAGAT AAAAATGAAA
AW20        TGAAGTTTTA TTTCCAATTC AATTTCATTC AATTTCAACT GAAAATAGAT AAAAATGAAA
```

```
                                                                          420
SH44_Oju    ATTTTGCTTA TTTATTACTT T--------- -----AA-AT ACTAAATAAG AGATTTCAGA
SH50        ATTTTGCTTA TTTATTACTT T--------- -----AATTT CCTAAATAAG AGATTTCAGA
SR178       ATTTTGCTTA TTTATTACTT T--------- -----AA-AT ACTAAATAAG AGATTTCAGA
NPB1831     ATTTTGCTTA TTTATTACTT T--------- -----AA-AT ACTAAATAAG AGATTTCAGA
SH40        ATTTTGCTTA TTTATTACTT T--------- -----AA-AT ACTAAATAAG AGATTTCAGA
SH57        ATTTTGCTTA TTTATTACTT T--------- -----AATTT CCTAAATAAG AGATTTCAGA
SH58        ATTTTGCTTA TTTATTACTT T--------- -----AATTT CCTAAATAAG AGATTTCAGA
MP452       ATTTTGCTTA TTTATTACTT G--------- -----AATTT CCTAAATAAG AGATTTCAGA
CP491       ATTTTGCTTA TTTATTACTT T--------- -----AA-AT ACTAAATAAG AGATTTCAGA
SR173       ATTTTGCTTA TTTATTACTT T--------- -----AA-AT ACTAAATAAG AGATTTCAGA
SH75        ATTTTGCTTA TTTATTACTT TATATTTATT ACTTTAA-AT ACTAAATAAG AGATTTCAGA
SH72        ATTTTGCTTA TTTATTACTT T--------- -----AATTT CCTAAATAAG AGATTTCAGA
SH111       ATTTTGCTTA TTTATTACTT T--------- -----AA-AT ACTAAATAAG AGATTTCAGA
SH88        ATTTTGCTTA TTTATTACTT T--------- -----AATTT CCTAAATAAG AGATTTCAGA
SH77        ATTTTGCTTA TTTATTACTT TATATTTATT ACTTTAA-AT ACTAAATAAG AGATTTCAGA
SH52        ATTTTGCTTA TTTATTACTT T--------- -----AATTT CCTAAATAAG AGATTTCAGA
NPB1820     ATTTTGCTTA TTTATTACTT T--------- -----AATTT CCTAAATAAG AGATTTCAGA
AW20        ATTTTGCTTA TTTATTACTT TATATTTATT ACTTTAA-AT ACTAAATAAG AGATTTCAGA

                                                                          480
SH44_Oju    AAGAAGAGAG AAATAATATG CTTTTTTATA TTAATGGAAA AATATTGATA ATACTGGATA
SH50        AATAAGAAAG AAATAATATG CTTTTTTATA TTAATGGAAA ----TTGATA ATACTGGATA
SR178       AATAAGAGAG AAATAATATG CTTTTTTATA TTAATGGAAA AATATTGATA ATACTGGATA
NPB1831     AATAAGAGAG AAATAATATG CTTTTTTATA TTAATGGAAA AATATTGATA ATACTGGATA
SH40        AATAAGAGAG AAATAATATG CTTTTTTATA TTAATGGAAA AATATTGATA ATACTGGATA
SH57        AATAAGAAAG AAATAATATG CTTTTTTATA TTAATGGAAA ----TTGATA ATACTGGATA
SH58        AATAAGAAAG AAATAATATG CTTTTTTATA TTAATGGAAA ----TTGATA ATACTGGATA
MP452       AATAAGAAAG AAATAATATG GTTTTTTATA TTAATGGAAA ----TTGATA ATACTGGATA
CP491       AATAAGAGAG AAATAATATG CTTTTTTATA TTAATGGAAA AATATTGATA ATACTGGATA
SR173       AATAAGAGAG AAATAATATG CTTTTTTATA TTAATGGAAA AATATTGATA ATACTGGATA
SH75        AATAAGAGAG AAATAATATG CTTTTTTATA TTAATGGAAA A-TATTGATA ATACTGGATA
SH72        AATAAGAAAG AAATAATATG CTTTTTTATA TTAATGGAAA ----TTGATA ATACTGGATA
SH111       AATAAGAGAG AAATAATATG CTTTTTTATA TTAATGGAAA AATATTGATA ATACTGGATA
SH88        AATAAGAAAG AAATAATATG CTTTTTTATA TTAATGGAAA ----TTGATA ATACTGGATA
SH77        AATAAGAGAG AAATAATATG CTTTTTTATA TTAATGTAAA AATATTGATA ATACTGGATA
SH52        AATAAGAAAG AAATAATATG CTTTTTTATA TTAATGGAAA ----TTGATA ATACTGGATA
NPB1820     AATAAGAAAG AAATAATATG CTTTTTTCTA TTAATGGAAA ----TTGATA ATACTGGATA
AW20        AATAAGAGAG AAATAATATG CTTTTTTATA TTAATGTAAA AATATTGATA ATACTGGATA

                                                                          540
SH44_Oju    ATAGTAGAAA TAGAAATAAT ACTAGATAAT ACTAGATAAT AGTAGAGGGG CGGATGTAGC
SH50        ATAGTAGAAA TA------AT ACTAGATAAT ACTAGATAAT AGTAGAGGGG CGGATGTAGC
SR178       ATAGTAGAAA TAGAAATAAT ACTAGATAAT ACTAGATAAT AGTAGAGGGG CGG???????
NPB1831     ATAGTAGAAA TAGAAATAAT ACTAGATAAT ACTAGATAAT AGTAGAGGGG CGGATGTAGC
SH40        ATAGTAGAAA TAGAAATAAT ACTAGATAAT ACTAGATAAT AGTAGAGGGG CGGATGTAGC
SH57        ATAGTAGAAA TA------AT ACTAGATAAT ACTAGATAAT AGTAGAGGGG CGGATGTAGC
SH58        ATAGTAGAAA TA------AT ACTAGATAAT ACTAGATAAT AGTAGAGGGG CGGATGTAGC
MP452       ATAGTAGAAA TA------AT ACTAGATAAT ACTAGATAAT AGTAGAGGGG CGGATGTAGC
CP491       ATAGTAGAAA TAGAAATAAT ACTAGATAAT ACTAGATAAT AGTAGAGGGG CGGA??????
SR173       ATAGTAGAAA TAGAAATAAT ACTAGATAAT ACTAGATAAT AGTAGAGGGG CGGATGTAGC
SH75        ATAGTAGAAA TAGAAATAAT ACTAGATAAT ACTAGATAAT AGTATAGGGG CGGATGTAGC
SH72        ATAGTAGAAA TA------AT ACTAGATAAT ACTAGATAAT AGTAGAGGGG CGGATGTAGC
SH111       ATAGTAGAAA TAGAAATAAT ACTAGATAAT ACTAGATAAT AGTAGAGGGG CGGATGTAGC
SH88        ATAGTAGAAA TA------AT ACTAGATAAT ACTAGATAAT AGTAGAGGGG CGGATGTAGC
SH77        ATAGTAGAAA TAGAAATAAT ACTAGATAAT GCTAGATAAT AGTATAGGGG CGGAGG????
SH52        ATAGTAGAAA TA------AT ACTAGATAAT ACTAGATAAT AGTAGAGGGG CGGATGTAGC
NPB1820     ATAGTAGAAA TA------AT ACTAGATAAT ACTAGATAAT AGTAGAGGGG CGGATGTAGC
AW20        ATAGTAGAAA TAGAAATAAT ACTAGATAAT ACTAGATAAT AGTATAGGGG CGGATGTAGC
```

```
SH44_Oju     CAAGTGGATC AAGGCAGTGG ATTTGTGA
SH50         CAAGTGGATC AAGGCAGTGG ATTTGT??
SR178        ?????????? ?????????? ????????
NPB1831      CAAGTGGATC AAGGCAGTGG ATTTGTGA
SH40         CAAG?????? ?????????? ????????
SH57         CAAG?????? ?????????? ????????
SH58         CAAGTGGATC AAGGCAGTGG ATTTGTGA
MP452        CAAGTGGATC AAGGCAGTGG ATTTGTGA
CP491        ?????????? ?????????? ????????
SR173        CAAGTGGATC AAGGCAGTGG ATTTGTGA
SH75         CAAG?????? ?????????? ????????
SH72         CA-GTGGATC AAGGCAGTGG ATTTGTGA
SH111        CAAGTGGATC AAGGCAGTGG ATTTGTG?
SH88         CA-GTGGATC AAGGCAGTGG ATTTGTGA
SH77         ?????????? ?????????? ????????
SH52         CAAGTGGATC AAGGCAGTGG ATTTGTGA
NPB1820      CAAGTGGATC AAGGCAGTGG ATTTGTGA
AW20         CAAGTGGATC AAGGCAGTGG ATTTGTGA
```

ITS ("Complete" + "AW20" and "SH75" for the Limited ITS data set)

```
                                                                          60
cstf          TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
NPB1831       TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
NPB1832       ?????????? ??TCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
pibotdt       ?GGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
pimatt        TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
pithomb       TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
CP480         ?????????? ??TCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
f_inc         TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
CP485         TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
CP491         TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
BSR2          TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
BSR1          ?????????? ?????TAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
TD4442        ?GGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
SH59          TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
SH72          TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
SH86          TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
SH90          TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
mp_452        ????GAAAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
SH100         TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
SH112         ?????????? ??????AACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
NPB1879       TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
SH40          TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
SH42          TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
SH44          TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
SH49          TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
SH50          TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
NPB1830       ?????????? ??TCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
CH43          TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
SH51          TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
SH54          TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
SH55          TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
SH87          TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
CH57          TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
SH58          TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
SH62          TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
SH65          TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
NPB1836       TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
SH102         TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
SH103         TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
SH52          TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
SH53          TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
SH66          TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
SH68          TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
SH70          TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
SH105         TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
SH73          TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
SH91          TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
SH94          TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
SH104         TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
NPB1820       TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
NPB1815       TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
NPB1818       TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
SR181         TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
SR173         TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
SR180         ?GGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
SR184         TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
SR188         ?GGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
SR189         ?GGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
NPB1888       TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
NPB1833       TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
SH71          TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
SH97          TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
SH111         TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
SH120         TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
NPB1877       TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
NPB1880       TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
NPB1881       TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
NPB1887       TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
PE2           TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
R_KOS         TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
SR178         ?GGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
SR179         ?????????? ???CGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
SH60          TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
NPB1882       TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
```

```
SH77           AGGAAGGAGA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
SH79           AGGAAGGAGA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CGTCGTCGC-
SH78           ?????????? ????GTAACA ACAATTCCGT AGGAGAACCT CCGGAAGGAT CATCATCGAA
AW20           AGGAAGGAGA AGTCGTAACA AGGTTTCCGT AGGTGAACCT GCGGAAGGAT CATTGTCGAA
SH75           TGGAAGTAAA AGTCGTAACA AGGTTTCCGT AGGTGAGCCT GCGGAAGGAT CATTGTCGAA

                                                                             120
cstf           CCCTGCATAG CAAAACGACC CGCGAACATG TACTAACAAC CGGGCGTCAT GGGGATGGGG
NPB1831        CCCTGCATAG CAAAACGACC CGCGAACATG TACTAACAAC CGGGCGTCAT GGGGATGGGG
NPB1832        CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
pibotdt        CCCTGCATAG CAAAACGACC CGCGAACATG TACTAACAAC CGGGCGTCAT GGGGATGGGG
pimatt         CCCTGCATAG CAAAACGACC CGCGAACATG TACTAACAAC CGGGCGTCAT GGGGATGGGG
pithomb        CCCTGCATAG CAAAACGACC CGCGAACATG TACTAACAAC CGGGCGTCAT GGGGATGGGG
CP480          CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
f_inc          CCCTGCATAG CARAACGACC CGCGAACATG TACTAACAAC CGGGCGTCAT GGGGATGGGG
CP485          CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
CP491          CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
BSR2           CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
BSR1           CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
TD4442         CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
SH59           CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
SH72           CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
SH86           CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
SH90           CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
mp_452         CCCTGCAAAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGTGATGGG-
SH100          CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
SH112          CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
NPB1879        CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
SH40           CCCTGCATAG CARAACGACC CGCGAACATG TACTAACAAC CGGGCGTCAT GGGGATGGGG
SH42           CCCTGCATAG CARAACGACC CGCGAACATG TACTAACAAC CGGGCGTCAT GGGGATGGGG
SH44           CCCTGCATAG CAGAACGACC CGCGAACATG TACCAACAAC CGGGCGACGC GGGGATGGGG
SH49           CCCTGCATAG CAAAACGACC CGCGAACATG TACTAACAAC CGGGCGTCAT GGGGATGGGG
SH50           CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
NPB1830        CCCTGCATAG CAGAACGACC CGCGAACATG TACCAACAAC CGGGCGACGC GGGGATGGGG
CH43           CCCTGCATAG CAAAACGACC CGCGAACATG TACTAACAAC CGGGCGTCAT GGGGATGGGG
SH51           CCCTGCATAG CAAAACGACC CGCGAACATG TACTAACAAC CGGGCGTCAT GGGGATGGGG
SH54           CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
SH55           CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
SH87           CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
CH57           CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
SH58           CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
SH62           CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
SH65           CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
NPB1836        CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
SH102          CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
SH103          CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
SH52           CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
SH53           CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
SH66           CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
SH68           CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
SH70           CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
SH105          CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
SH73           CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
SH91           CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
SH94           CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
SH104          CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
NPB1820        CCCTGCATAG CAGAACGACC CGCGAACATG TACTGACAAC CGGGCGTCAT GGGGATGGGG
NPB1815        CCCTGCATAG CAGAACGACC CGCGAACATG TACTGACAAC CGGGCGTCGT GGGGATGGGG
NPB1818        CCCTGCATAG CAGAACGACC CGCGAACATG TACTGACAAC CGGGCGTCRT GGGGATGGGG
SR181          CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
SR173          CCCTGCATAG CAAAACGACC CGCGAACATG TACTAACAAC CGGGCGTCAT GGGGATGGGG
SR180          CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
SR184          CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
SR188          CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
SR189          CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
NPB1888        CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
NPB1833        CCCTGCAWAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
SH71           CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
SH97           CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
SH111          CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCAT GGGGATGGGG
SH120          CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
NPB1877        CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
NPB1880        CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
NPB1881        CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
NPB1887        CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
PE2            CCCTGCATAG CAGAACGACC CGCGAACATG TACTTACAAC CGGGCGTCGT GGGGATGGGG
R_KOS          CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCRT GGGGATGGGG
```

```
SR178         CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
SR179         CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
SH60          CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
NPB1882       CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG
SH77          CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGT??? ??????????
SH79          CCCTGCATAG CAGAACGAGC C------NTC ATCTAGCAAC CGGG-GTCTT GGGGATGGTG
SH78          CCCCCCATAA CAGAASAACC CGCGAACATC AACTAACAAC CGGGCGTCGT GGGGAWGGGG
AW20          CCCTGAATAG CAGAACGACA CGCGAACATG TACTAACAAC CGGGC????? ??????????
SH75          CCCTGCATAG CAGAACGACC CGCGAACATG TACTAACAAC CGGGCGTCGT GGGGATGGGG

                                                                            180
cstf          CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
NPB1831       CA---CTT-- -GTCCTGTCC CTGTGGYGCC TCGTTGGTGT GCGTGTCTGG TGGCCGCTTC
NPB1832       CA---CTT-- -GTCCTGTCC CTGTGGYGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
pibotdt       CA---CTT-- -GTCCTGTCC CTGTGGYGCC TCGTTGGTGT GCGTGTCTGG TGGCCGCTTC
pimatt        CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
pithomb       CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
CP480         CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
f_inc         CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTT
CP485         CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
CP491         CA---C---- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
BSR2          CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
BSR1          CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
TD4442        CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
SH59          CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
SH72          CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
SH86          CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
SH90          CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
mp_452        -A---CTT-- -GTCCTGTCC CTGTGGTGCC TCGTCGGTGT GCGTGTTTGG TGGCCGCCTT
SH100         CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
SH112         CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
NPB1879       CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
SH40          CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
SH42          CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
SH44          CA---C---- -GTCCTGTCC CCGTGGTGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTT
SH49          CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
SH50          CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
NPB1830       CA---CTT-- -GTCCTGTCC CCGTGGTGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTT
CH43          CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
SH51          CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
SH54          CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
SH55          CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
SH87          CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
CH57          CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
SH58          CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
SH62          CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
SH65          CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
NPB1836       CA---CTT-- -GTCCTGTCC CTGTGGCKCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
SH102         CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
SH103         CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
SH52          CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
SH53          CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
SH66          CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
SH68          CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
SH70          CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
SH105         CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
SH73          CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
SH91          CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
SH94          CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
SH104         CA---C---- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
NPB1820       CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGR TGGCCGCTTT
NPB1815       CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTT
NPB1818       CA---C---- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTT
SR181         CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
SR173         CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
SR180         CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
SR184         CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
SR188         CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
SR189         CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
NPB1888       CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
NPB1833       CG---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTTTGG TGRCCGCTTY
SH71          CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
SH97          CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
SH111         CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
SH120         CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
NPB1877       CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
NPB1880       CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
```

```
NPB1881       CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
NPB1887       CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
PE2           CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
R_KOS         CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
SR178         CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
SR179         CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
SH60          CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
NPB1882       CA---CTT-- -GTCCTGTCC CTGTGGCGCC TCGTCGGTGT GCGTGTCTGG TGGCCGCTTC
SH77          ????????GT AGTCCTGTCC TTGTGGCGCC TCGTCGYTGT GCGTGTCTGG TGGCCGCTTC
SH79          GTGCGCTTGT CCTCGTGTCC TTGTGGCGCC TCGTCGYTGT GCGTGTTTGG TGGCCGCTTC
SH78          CA---CCAGT CGTCCTGTCC TTGTGGCGCC TCGTCGATGT GCGTGTTTGG TGGCCGCTTC
AW20          ?????????? ?????????? ?????????? ?????????? ?????????? ??????????
SH75          CA---CTTGT CGTCCTGTCC TTGTGGCGCC TCGTCGATGT GCGTGTTTGG TGGCCGCTTC

                                                                             240
cstf          CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA TGTGCCAAGG
NPB1831       CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA TGTGCCAAGG
NPB1832       CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA TGTGCCAAGG
pibotdt       CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA TGTGCCAAGG
pimatt        CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA TGTGCCAAGG
pithomb       CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA TGTGCCAAGG
CP480         CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA TGTGCCAAGG
f_inc         CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA TGTGCCAAGG
CP485         CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA TGTGCCAAGG
CP491         CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA TGTGCCAAGG
BSR2          CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA CGTGCCAAGG
BSR1          CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA CGTGCCAAGG
TD4442        CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA TGTGCCAAGG
SH59          CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA YGTGCCAAGG
SH72          CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA YGTGCCAAGG
SH86          CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA CGTGCCAAGG
SH90          CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA YGTGCCAAGG
mp_452        TGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACGGCA CGTGCCAAGG
SH100         CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA CGTGCCAAGG
SH112         YGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA TGTGCCAAGG
NPB1879       CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA CGTGCCAAGG
SH40          CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA TGTGCCAAGG
SH42          CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA TGTGCCAAGG
SH44          GGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACGGCA CGTGCCAAGG
SH49          CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA TGTGCCAAGG
SH50          TGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA TGTGCCAAGG
NPB1830       GGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACGGCA CGTGCCAAGG
CH43          CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACARCA TGTGCCAAGG
SH51          CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA TGTGCCAAGG
SH54          CGGGGCCTCG AGCATGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA CGTGCCAAGG
SH55          CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA CGTGCCAAGG
SH87          CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA CGTGCCAAGG
CH57          CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA CGTGCCAAGG
SH58          CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA CGTGCCAAGG
SH62          CGGGGCCTCA AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA CGTGCCAAGG
SH65          CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA CGTGCCAAGG
NPB1836       CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA TGTGCCAAGG
SH102         CGGGGCCTCG AGMACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA CGTGCCAAGG
SH103         CGGGGCCTCG AGMACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA CGTGCCAAGG
SH52          TGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA TGTGCCAAGG
SH53          TGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA TGTGCCAAGG
SH66          CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA CGTGCCAAGG
SH68          CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA CGTGCCAAGG
SH70          CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA CGTGCCAAGG
SH105         CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA CGTGCCAAGG
SH73          CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA CGTGCCAAGG
SH91          CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA CGTGCCAAGG
SH94          CGGGGCCTCG AGAACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA CGTGCCAAGG
SH104         CGGGGCCTCG AGAACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA CGTGCCAAGG
NPB1820       CGGGGCCTCG AGCACGTCTC ATCGGCACAA CAACAAACCC CGGCACGGCA CGTGCCAAGG
NPB1815       CGGGGCCTCG AGCACGTCTC ATCGGCACAA CAACAAACCC CGGCACGGCA CGTGCCAAGG
NPB1818       CGGGGCCTCG AGCACGTCTC ATCGGCACAA CAACAAACCC CGGCACGGCA CGTGCCAAGG
SR181         TGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA TGTGCCAAGG
SR173         CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA TGTGCCAAGG
SR180         CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA CGTGCCAAGG
SR184         TGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA TGTGCCAAGG
SR188         CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA YGTGCCAAGG
SR189         CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA CGTGCCAAGG
NPB1888       CGGGGCCTCG AGCAYGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA CGTGCCAAGG
NPB1833       TGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA YGTGCCAAGG
SH71          CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA CGTGCCAAGG
SH97          CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA YGTGCCAAGG
```

```
SH111        CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA TGTGCCAAGG
SH120        CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA TGTGCCAAGG
NPB1877      CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA CGTGCCAAGG
NPB1880      CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA CGTGCCAAGG
NPB1881      CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA CGTGCCAAGG
NPB1887      CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA CGTGCCAAGG
PE2          CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA TGTGCCAAGG
R_KOS        YGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA TGTGCCAAGG
SR178        GGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA CGTGCCAAGG
SR179        YGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA YGTGCCAAGG
SH60         CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA TGTGCCAAGG
NPB1882      CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA CGTGCCAAGG
SH77         CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA TGTGCCAAGG
SH79         CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA TGTGCCAAGG
SH78         CGGGGCCTCG ASCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA TGTGCCAAGG
AW20         ?????????? ?????????? ?????????? ?????????? ?????????? ??????????
SH75         CGGGGCCTCG AGCACGTCGC ATCGGCACAA CAACAAACCC CGGCACAACA TGTGCCAAGG

                                                                          300
cstf         AAAACTAAAC TTAAG-AAGG GCTCGTAACA CGACGCCCCG TTCGCGGGGT GCTCGTGTGG
NPB1831      AAAACTAAAC TTAAG-AAGG GCTCGTAACA CGACGCCCCG TTCGCGGGGT GCTCGTGTGG
NPB1832      AAAACTAAAC TTAAG-AAGG GCTCGTAACA CGACGCCCCG TTCGCGGGGT GCTCGTGTGG
pibotdt      AAAACTAAAC TTAAG-AAGG GCTCGTAACA CGACGCCCCG TTCGCGGGGT GCTCGTGTGG
pimatt       AAAACTAAAC TTAAG-AAGG GCTCGTAACA CGACGCCCCG TTCGCGGGGT GCTCGTGTGG
pithomb      AAAACTAAAC TTAAG-AAGG GCTCGTAACA CGACGCCCCG TTCGCGGGGT GCTCGTGTGG
CP480        AAAACTAAAC TTAAG-AAGG GCTCGTAACA CGACGCCCCG TTCGCGGGGT GCTCGTGTGG
f_inc        AAAACTAAAC TTAAG-AAGG GCTCGTAACA CGACGCCCCG TTCGCGGGGT GCTCGTGTGG
CP485        AAAACTAAAC TTAAG-AAGG GCTCGTAACA CGACGCCCCG TTCGCGGGGT GCTCGTGTGG
CP491        AAAACTAAAC TTAAG-AAGG GCTCGTAACA CGACGCCCCG TTCGCGGGGT GCTCGTGTGG
BSR2         AAAACTAAAC TTAAG-AAGG GCTCGTAGCA CGACGCCCCG TTCGCGGTGT GCTCGTGTGG
BSR1         AAAACTAAAC TTAAG-AAGG GCTCGTAGCA CGACGCCCCG TTCGCGGTGT GCTCGTGTGG
TD4442       AAAACTAAAC TTAAG-AAGG GCTCGTAACA CGACGCCCCG TTCGCGGGGT GCTCGTGTGG
SH59         AAAACTAAAC TTAAG-AAGG GCTCGTAGCA CGACGCCCCG TTCGCGGTGT GCTCGTGTGG
SH72         AAAACTAAAC TTAAG-AAGG GCTCGTAGCA CGACGCCCCG TTCGCGGTGT GCTCGTGTGG
SH86         AAAACTAAAC TTAAG-AAGG GCTCGTAGCA CGACGACCCG TTCGCGGTGT GCTCGTGTGG
SH90         AAAACTAAAC TTAAG-AAGG GCTCGTAGCA CGACGCCCCG TTCGCGGTGT GCTCGTGTGG
mp_452       AAAACTAAAC TTAAG-AAGG GCTCGTAACA CGACGCCCCG TTCGCGGTGT GCTCGTGTGG
SH100        AAAACTAAAC TTAAG-AAGG GCTCGTAGCA CGACGCCCCG TTCGCGGTGT GCTCGTGTGG
SH112        AAAACTAAAC TTAAG-AAGG GCTCGTAACA CGACGCCCCG TTCGCGGKGT GCTCGTGTGG
NPB1879      AAAACTAAAC TTAAG-AAGG GCTCGTAGCA CGACGMCCCG TTCGCGGTGT GCTCGTGYGG
SH40         AAAACTAAAC TTAAG-AAGG GCTCGTAACA RGACGCCCCG TTCGCGGGGT GCTCGTGTGG
SH42         AAAACTAAAC TTAAG-AAGG GCTCGTAACA CGACGCCCCG TTCGCGGGGT GCTCGTGTGG
SH44         AAAACTAAAC TTAAGGAAGG GCTCGTAACA CGACGCCCCG TTCGCGGTGT GCTCGTGTGG
SH49         AAAACTAAAC TTAAG-AAGG GCTCGTAACA CGACGCCCCG TTCGCGGGGT GCTCGTGTGG
SH50         AAAACTAAAC TTAAG-AAGG GCTCGTAACA CGACGCCCCG TTCGCGGGGT GCTCGTGTGG
NPB1830      AAAACTAAAC TTAAGGAAGG GCTCGTAACA CGACGCCCCG TTCGCGGTGT GCTCGTGTGG
CH43         AAAACTAAAC TTAAG-AAGG GCYCGTAACA CGACGCCCCG TTCGCGGGGT GNTCGTGTGG
SH51         AAAACTAAAC TTAAG-AAGG GCTCGTAACA CGACGCCCCG TTCGCGGGGT GCTCGTGTGG
SH54         AAAACTAAAC TTAAG-AAGG GCTCGTAGCA CGACGCCCCG TTCGCGGTGT GCTCGTGCGG
SH55         AAAACTAAAC TTAAG-AAGG GCTCGTAGCA CGACGCCCCG TTCGCGGTGT GCTCGTGCGG
SH87         AAAACTAAAC TTAAG-AAGG GCTCGTAGCA CGACGACCCG TTCGCGGTGT GCTCGTGTGG
CH57         AAAACTAAAC TTAAG-AAGG GCTCGTAGCA CGACGCCCCG TTCGCGGTGT GCTCGTGCGG
SH58         AAAACTAAAC TTAAG-AAGG GCTCGTAGCA CGACGCCCCG TTCGCGGTGT GCTCGTGRGG
SH62         AAAACTAAAC TTAAG-AAGG GCTCGTAGCA CGACGACCCG TTCGCGGTGT GCTCGTGTGG
SH65         AAAACTAAAC TTAAG-AAGG GCTCGTAGCA CGACGCCCCG TTCGCGGTGT GCTCGTGTGG
NPB1836      AAAACTAAAC TTAAG-AAGG GCTCGTAACA CGACGCCCCG TTCGCGGGGT GCTCGTGTGG
SH102        AAAACTAAAC TTAAG-AAGG GCTCGTAGCA CGACGCCCCG TTCGCGGTGT GCTCGTGYGG
SH103        AAAACTAAAC TTAAG-AAGG GCTCGTAGCA CGACGCCCCG TTCGCGGTGT GCTCGTGYGG
SH52         AAAACTAAAC TTAAG-AAGG GCTCGTAACA CGACGCCCCG TTCGCGGGGT GCTCGTGTGG
SH53         AAAACTAAAC TTAAG-AAGG GCTCGTAACA CGACGCCCCG TTCGCGGGGT GCTCGTGTGG
SH66         AAAACTAAAC TTAAG-AAGG GCTCGTAGCA CGACGCCCCG TTCGCGGTGT GCTCGTGTGG
SH68         AAAACTAAAC TTAAG-AAGG GCTCGTAGCA CGACGCCCCG TTCGCGGTGT GGTCGTGTGG
SH70         AAAACTAAAC TTAAG-AAGG GCTCGTAGCA CGACGCCCCG TTCGCGGTGT GCTCGTGTGG
SH105        AAAACTAAAC TTAAG-AAGG GCTCGTAGCA CGACGCCCCG TTCGCGGTGT GCTCGTGYGG
SH73         AAAACTAAAC TTAAG-AAGG GCTCGTAGCA CGACGCCCCG TTCGCGGTGT GCTCGTGTGG
SH91         AAAACTAAAC TTAAG-AAGG GCTCGTAGCA CGACGCCCCG TTCGCGGTGT GCTCGTGTGG
SH94         AAAACTAAAC TTAAG-AAGG GCTCGTAGCA CGACGCCCCG TTCGCGGTGT GCTCGTGCGG
SH104        AAAACTAAAC TTAAG-AAGG GCTCGTAGCA CGACGCCCCG TTCGCGGTGT GCTCGTGYGG
NPB1820      AAAACTAAAC TTAAG-AAGG GCTCGTAACA CGACGCCCCG TTCGCGGTGT GCTCGTGTGG
NPB1815      AAAACTAAAC TTAAG-AAGG GCTCGTAACA CGACGCCCCG TTCGCGGTGT GCTCGTGTGG
NPB1818      AAAACTAAAC TTAAG-AAGG GCTCGTAACA CGACGCCCCG TTCGCGGTGT GCTCGTGTGG
SR181        AAAACTAAAC TTAAG-AAGG GCTCGTAACA CGACGCCCCG TTCGCGGGGT GCTCGTGTGG
SR173        AAAACTAAAC TTAAG-AAGG GCTCGTAACA CGACGCCCCG TTCGCGGGGT GCTCGTGTGG
SR180        AAAACTAAAC TTAAG-AAGG GCTCGTAGCA CGACGCCCCG TTCGCGGTGT GCTCGTGTGG
SR184        AAAACTAAAC TTAAG-AAGG GCTCGTAACA CGACGCCCCG TTCGCGGGGT GCTCGTGTGG
SR188        AAAACTAAAC TTAAG-AAGG GCTCGTAGCA CGACGCCCCG TTCGCGGKGT GCTCGTGTGG
SR189        AAAACTAAAC TTAAG-AAGG GCTCGTAGCA CGACGCCCCG TTCGCGGTGT GCTCGTGTGG
```

```
NPB1888     AAAACTAAAC TTAAG-AAGG GCTCGTAGCA CGACGCCCCG TTCGCGGTGT GCTCGTGYGG
NPB1833     AAAACTAAAC TTAAG-AAGG GCYCGTAACA CGACGCCCCG TTCGCGGKGT GYTCGTGTGG
SH71        AAAACTAAAC TTAAG-AAGG GCTCGTAGCA CGACGCCCCG TTCGCGGTGT GCTCGTGTGG
SH97        AAAACTAAAC TTAAG-AAGG GCTCGTAGCA CGACGCCCCG TTCGCGGTGT GCTCGTGTGG
SH111       AAAACTAAAC TTAAG-AAGG GCTCGTAACA CGACGCCCCG TTCGCGGGGT KCTCGTGTGG
SH120       AAAACTAAAC TTAAG-AAGG GCTCGTAACA CGACGCCCCG TTCGCGGGGT GCTCGTGTGG
NPB1877     AAAACTAAAC TTAAG-AAGG GCTCGTAGCA CGACGCCCCG TTCGCGGTGT GCTCGTGTGG
NPB1880     AAAACTAAAC TTAAG-AAGG GCTCGTAGCA CGACGMCCCG TTCGCGGTGT GCTCGTGTGG
NPB1881     AAAACTAAAC TTAAG-AAGG GCTCGTAGCA CGACGACCCG TTCGCGGTGT GCTCGTGTGG
NPB1887     AAAACTAAAC TTAAG-AAGG GCTCGTAGCA CGACGACCCG TTCGCGGTGT GCTCGTGTGG
PE2         AAAACTAAAC TTAAG-AAGG GCTCGTAACA CGACGCCCCG TTCGCGGGGT GCTCGTGTGG
R_KOS       AAAACTAAAC TTAAG-AAGG GCTCGTAACA YGACGCCCCG TTCGCGGGGT GCTCGTGTGG
SR178       AAAACTAAAC TTAAG-AAGG GCTCGTAGCA CGACGCCCCG TTCGCGGTGT GCTCGTGTGG
SR179       AAAACTAAAC TTAAG-AAGG GCTCGTARCA CGACGCCCCG TTCGCGGGGT GCTCGTGTGG
SH60        AAAACTAAAC TTAAG-AAGG GCTCGTAGCA CGACGCCCCG TTCGCGGTGT GCTCGTGTGG
NPB1882     AAAACTAAAC TTAAG-AAGG GCTCGTAGCA CGACGCCCCG TTCGCGGTGT GCTCGTGTGG
SH77        AAAACTAAAC TTAAG-AAGG GCTCGTAACA CGACGCCCCG TTCGCGGG-T GCTCGTGTGG
SH79        AAAACTAAAC TTAAG-AAGG GCTCGTAACA CGACGCCCCG TTCGCGGG-T GCTCGTGTGG
SH78        AAAACTAAAC TTAAG-AAGG GCTCGTAACA CGACGCCCCG TTCGCGGG-T GCTCGTGTGG
AW20        ?????????? ?????????? ?????????? ?????????? ?????????? ??????????
SH75        AAAACTAAAC TTAAG-AAGG GCTCGTAACA CGACGCCCCG TTCGCGGG-T GCTCGTGTGG

                                                                         360
cstf        CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
NPB1831     CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
NPB1832     CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
pibotdt     CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
pimatt      CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
pithomb     CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
CP480       CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
f_inc       CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
CP485       CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
CP491       CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
BSR2        CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
BSR1        CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
TD4442      CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
SH59        CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
SH72        CGTGTCCTCT TTGTAATCAC AAACGACTCT CGGCAAKGGA TATCTAGGNT CNCGCATCGA
SH86        CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
SH90        CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
mp_452      CGTGGCCTCT TTGTAATCAC AAATGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
SH100       CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
SH112       CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
NPB1879     CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CYCGCATCGA
SH40        CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
SH42        CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
SH44        CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
SH49        CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
SH50        CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
NPB1830     CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
CH43        CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
SH51        CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
SH54        CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
SH55        CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
SH87        CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
CH57        CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
SH58        CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
SH62        CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
SH65        CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
NPB1836     CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
SH102       CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
SH103       CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
SH52        CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
SH53        CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
SH66        CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
SH68        CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
SH70        CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
SH105       CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
SH73        CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
SH91        CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
SH94        CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
SH104       CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
NPB1820     CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
NPB1815     CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
NPB1818     CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
SR181       CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
SR173       CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
```

```
SR180          CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
SR184          CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
SR188          CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
SR189          CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
NPB1888        CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
NPB1833        CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
SH71           CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
SH97           CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
SH111          CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
SH120          CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
NPB1877        CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
NPB1880        CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
NPB1881        CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
NPB1887        CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
PE2            CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
R_KOS          CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
SR178          CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
SR179          CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
SH60           CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
NPB1882        CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
SH77           CGTGGCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
SH79           CGTGCCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
SH78           CGTGCCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA
AW20           ?????????? ?????????? ?????????? ?????????? ?????????? ??????????
SH75           CGTGCCCTCT TTGTAATCAC AAACGACTCT CGGCAACGGA TATCTCGGCT CACGCATCGA

                                                                             420
cstf           TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
NPB1831        TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
NPB1832        TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
pibotdt        TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
pimatt         TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
pithomb        TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
CP480          TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
f_inc          TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
CP485          TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
CP491          TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
BSR2           TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
BSR1           TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
TD4442         TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
SH59           TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
SH72           TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
SH86           TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
SH90           TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
mp_452         TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
SH100          TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
SH112          TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
NPB1879        TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
SH40           TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
SH42           TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
SH44           TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
SH49           TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
SH50           TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
NPB1830        TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
CH43           TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
SH51           TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
SH54           TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
SH55           TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
SH87           TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
CH57           TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
SH58           TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
SH62           TGAAGAACGT AGCAWAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
SH65           TGAAGAACGT AGCAAAATGC GATAGTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
NPB1836        TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
SH102          TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
SH103          TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
SH52           TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
SH53           TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
SH66           TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
SH68           TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
SH70           TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
SH105          TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
SH73           TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
SH91           TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
SH94           TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
SH104          TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
NPB1820        TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
```

```
NPB1815       TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
NPB1818       TGAAGA-CGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
SR181         TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
SR173         TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
SR180         TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
SR184         TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
SR188         TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
SR189         TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
NPB1888       TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
NPB1833       TGAAGAACG- AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
SH71          TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
SH97          TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
SH111         TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
SH120         TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
NPB1877       TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
NPB1880       TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
NPB1881       TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
NPB1887       TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
PE2           TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
R_KOS         TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
SR178         TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
SR179         TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
SH60          TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
NPB1882       TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
SH77          TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
SH79          TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
SH78          TGAAGAACGT AGCAAAATGC GATACTTGGT GTGAATTGCA GAATCCCGTG AACCATCGAG
AW20          ?????????? ?????????? ?????????? ?????????? ?????????? ??????????
SH75          TGAAGAACNT AGC??????? ?ATACTTGGT GTGAATTGCA GAATCCCGSG AACCATCGAG

                                                                            480
cstf          TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
NPB1831       TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
NPB1832       TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TTGGCCGAGG GCACGTCTGC C-TGGGCGTC
pibotdt       TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
pimatt        TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
pithomb       TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
CP480         TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
f_inc         TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
CP485         TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
CP491         TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
BSR2          TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
BSR1          TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
TD4442        TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
SH59          TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGYGTC
SH72          TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
SH86          TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
SH90          TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
mp_452        TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
SH100         TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
SH112         TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
NPB1879       TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCTGAGG GCACGTCTGC C-TGGGCGTC
SH40          TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCAGGTCTGC C-TGGGCGTC
SH42          TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
SH44          TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
SH49          TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
SH50          TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
NPB1830       TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
CH43          TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TYGGCCGAGG GCACGTCTGC C-TGGGCGTC
SH51          TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
SH54          TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
SH55          TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
SH87          TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
CH57          TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
SH58          TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
SH62          TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
SH65          TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
NPB1836       TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
SH102         TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
SH103         TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
SH52          TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
SH53          TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
SH66          TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
SH68          TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
SH70          TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCAGCCGAGG GCACGTCTGC C-TGGGCGTC
SH105         TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
SH73          TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
```

```
SH91          TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
SH94          TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
SH104         TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
NPB1820       TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCYGAGG GCACGTCTGC C-TGGGCGTC
NPB1815       TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCTGAGG GCACGTCTGC C-TGGGCGTC
NPB1818       TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCTGAGG GCACGTCTGC C-TGGGCGTC
SR181         TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
SR173         TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
SR180         TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
SR184         TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
SR188         TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
SR189         TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
NPB1888       TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCYGAGG GCACGTCTGC C-TGGGCGTC
NPB1833       TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TYGGCCGAGG GCACGTCTGC C-TGGGCGTC
SH71          TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
SH97          TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
SH111         TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
SH120         TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
NPB1877       TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
NPB1880       TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCYGAGG GCACGTCTGC C-TGGGCGTC
NPB1881       TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCYGAGG GCACGTCTGC C-TGGGCGTC
NPB1887       TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
PE2           TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
R_KOS         TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
SR178         TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
SR179         TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
SH60          TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
NPB1882       TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCAGCCGAGG GCACGTCTGC C-TGGGCGTC
SH77          TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC CCTGGGCGTC
SH79          TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
SH78          TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC C-TGGGCGTC
AW20          ?????????? ?????????? ?????????? ?????????? ???????TGC C-TGGGCGTC
SH75          TTTTTGAACG CAAGTTGCGC CCGAAGCCTT TCGGCCGAGG GCACGTCTGC CCTGGGCGTC

                                                                             540
cstf          ACGCATCGCG TCGCCCCCC- ACCACCCATC CCATACGGTG ATGCATGGTA CGGGGGCGGA
NPB1831       ACGCATCGCG TCGCCCCCC- ACCACCCATC CCATACGGTG ATGCATGGTA CGGGGGCGGA
NPB1832       ACGCATCGCG TCGCCCCCC- ACCAYCCATC CCATAMGGTG ATGCATGGTA CGGGGGCGGA
pibotdt       ACGCATCGCG TCGCCCCCC- ACCACCCATC CCATACGGTG ATGCATGGTA CGGGGGCGGA
pimatt        ACGCATCGCG TCGCCCCCC- ACCACCCATC CCATACGGTG ATGCATGGTA CGGGGGCGGA
pithomb       ACGCATCGCG TCGCCCCCC- ACCACCCATC CCATACGGTG ATGCATGGTA CGGGGGCGGA
CP480         ACGCATCGCG TCGCCCCCC- ACCACCCATC CCATACGGTG ATGCATGGTA CGGGGGCGGA
f_inc         ACGCATCGCG TCGCCCCCC- ACCACCCATC CCATACGGTG ATGCATGGTA CGGGGGCGGA
CP485         ACGCATCGCG TCGCCCCCC- ACCACCCATC CCATACGGTG ATGCATGGTA CGGGGGCGGA
CP491         ACGCATCGCG TCGCCCCCC- ACCACCCATC CCATACGGTG ATGCATGGTA CGGGGGCGGA
BSR2          ACGCATCGCG TCGCCCCCC- ACCACSCATC CC-------- ATGCATGGTA CGGGGGCGGA
BSR1          ACGCATCGCG TCGCCCCCC- ACCACGCATC CC-------- ATGCATGGTA CGGGGGCGGA
TD4442        ACGCATCGCG TCGCCCCCC- ACCACCCATC CCATACGGTG ATGCATGGTA CGGGGGCGGA
SH59          ACGCATCGCG TCGCCCCCC- TCCACGCATC CC-------- ATGCATGGTA CGGGGGCGGA
SH72          ACGCATCGCG TCGCCCCCC- TCCACGCATC CC-------- ATGCATGGTA CGGGGGCGGA
SH86          ACGCATCGCG TCGCCCCCC- ACCACGCATC CC-------- ATGCATGGTA CGGGGGCGGA
SH90          ACGCATCGCG TCGCCCCCC- TCCACGCATC CC-------- ATGCATGGTA CGGGGGCGGA
mp_452        ACGCATCGCG TCGCCCCCC- ACCACCCATC CCCTAAGGTG ATGCATGGTA -GGGGGCGGA
SH100         ACGCATCGCG TCGCCCCCC- ACCACGCATC CC-------- ATGCATGGTA CGGGGGCGGA
SH112         ACGCATCGCG TCGCCCCC?? ?????????? ?????????? ?????????? ??????????
NPB1879       ACGCATCGCG TCGCCCCCC- ACCACGCATC CC-------- ATGCATGGTA CGGGGGCGGA
SH40          ACGCATCGCG TCGCCCCCC- ACCACCCATC CCATACGGTG ATGCATGGTA CGGGGGCGGA
SH42          ACGCATCGCG TCGCCCCCC- ACCACCCATC CCATACGGTG ATGCATGGTA CGGGGGCGGA
SH44          ACGCATCGCG TCGCCCCCC- ACCACCCATC CCCTACGGTG ATGCATGGTA CGGGGGCGGA
SH49          TCGCATCGCG TCGCCCCCC- ACCACCCATC CCATACGGTG ATGCATGGTA CGGGGGCGGA
SH50          ACGCATCGCG TCGCCCCCCC ACCACCCATC CCATACGGTG ATGCATGGTA CGGGGGCGGA
NPB1830       ACGCATCGCG TCGCCCCCC- ACCACCCATC CCCTACGGTG ATGCATGGTA CGGGGGCGGA
CH43          ACGCATCGCG TCGCCCCCC- ACCAYCCATC CCMTAMGGTG ATGCATGGTA CGGGGGCGGA
SH51          ACGCATCGCG TCGCCCCCC- ACCACCCTTC CCATACGGTG ATGCATGGTA CGGGGGCGGA
SH54          ACGCATCGCG TCGCCCCCC- ACCACGCATC CC-------- ATGCATGGTA CGGGGGCGGA
SH55          ACGCATCGCG TCGCCCCCC- ACCACGCATC CC-------- ATGCATGGTA CGGGGGCGGA
SH87          ACGCATCGCG TCGCCCCCC- ACCACGCATC CC-------- ATGCATGGTA CGGGGGCGGA
CH57          ACGCATCGCG TCGCCCCCC- ACCACGCATC CC-------- ATGCATGGTA CGGGGGCGGA
SH58          ACGCATCGCG TCGCCCCCC- ACCACGCATC CC-------- ATGCATGGTA CGGGGGCGGA
SH62          ACGCATCGCG TCGCCCCCC- ACCACGCATC CC-------- ATGCATGGTA CGGGGGCGGA
SH65          ACGCATCGCG TCGCCCCCC- ACCACGCATC CC-------- ATGCATGGTA CGGGGGCGGA
NPB1836       ACGCATCGCG TCGCCCCCC- ACCACCCATC CCATACGGTG ATGCATGGTA CGGGGGCGGA
SH102         ACGCATCGCG TCGCCCCCC- ACCACGCATC CC-------- ATGCATGGTA CGGGGGCGGA
SH103         ACGCATCGCG TCGCCCCCC- ACCACGCATC CC-------- ATGCATGGTA CGGGGGCGGA
SH52          ACGCATCGCG TCGCCCCCCC ACCACCCATC CCATACAGTG ATGCATGGTA CGGGGGCGGA
SH53          ACGCATCGCG TCGCCCCCCC ACCACCCATC CCATACGGTG ATGCATGGTA CGGGGGCGGA
SH66          ACGCATCGCG TCGCCCCCC- ACCACGCATC CC-------- ATGCATGGTA CGGGGGCGGA
```

```
SH68        ACGCATCGCG TCGCCCCCC- ACCACGCATC CC-------- ATGCATGGTA CGGGGGCGGA
SH70        ACGCATCGCG TCGCCCCCC- ACCACGCATC CC-------- ATGCATGGTA CGGGGGCGGA
SH105       ACGCATCGCG TCGCCCCCC- ACCACGCATC CC-------- ATGCATGGTA CGGGGGCGGA
SH73        ACGCATCGCG TCGCCCCCC- ACCACGCATC CC-------- ATGCATGGTA CGGGGGCGGA
SH91        ACGCATCGCG TCGCCCCCC- ACCACGCATC CC-------- ATGCATGGTA CGGGGGCGGA
SH94        ACGCATCGCG TCGCCCCCC- ACCACGCATC CC-------- ATGCATGGTA CGGGGGCGGA
SH104       ACGCATCGCG TCGCCCCCC- ACCACGCATC CC-------- ATGCATGGTA CGGGGGCGGA
NPB1820     ACGCATCGCG TCGCCCCCC- ACCACCCATC CCCTACGGTG AGGCATGGTA CGGGGGCGGA
NPB1815     ACGCATCGCG TCGCCCCCC- ACCACCCATC CCCTACGGTG AGGCATGGTA CGGGGGCGGA
NPB1818     ACGCATCGCG TCGCCCCCC- ACCACCCATC CCCTACGGTG AGGCATGGTA CGGGGGCGGA
SR181       ACGCATCGCG TCGCCCCCCC ACCACCCATC CCATACAGTG ATGCATGGTA CGGGGGCGGA
SR173       ACGCATCGCG TCGCCCCCC- ACCACCCATC CCATACGGTG ATGCATGGTA CGGGGGCGGA
SR180       ACGCATCGCG TCGCCCCCC- ACCACGCATC CC-------- ATGCATGGTA CGGGGGCGGA
SR184       ACGCATCGCG TCGCCCCCCC ACCACCCATC CCATACAGTG ATGCATGGTA CGGGGGCGGA
SR188       ACGCATCGCG TCGCCCCCC- ACCACGCATC CC-------- ATGCATGGTA CGGGGGCGGA
SR189       ACGCATCGCG TCGCCCCCC- ACCACGCATC CC-------- ATGCATGGTA CGGGGGCGGA
NPB1888     ACGCATCGCG TCGCCCCCC- ACCACGCATC CC-------- ATGCATGGTA CGGGGGCGGA
NPB1833     ACGCATCGCG TCGCCCCCC- ACCAYCCATC CCMTAMGGKG ATGCATGGTA CGGGGGCGGA
SH71        ACGCATCGCG TCGCCCCCC- ACCACGCATC CC-------- ATGCATGGTA CGGGGGCGGA
SH97        ACGCATCGCG TCGCCCCCC- ACCACGCATC CC-------- ATGCATGGTA CGGGGGCGGA
SH111       ACGCATCGCG TCGCCCCCC- ACCACCCATC CCATACGGGG ATGCATGGTA CGGGGGCGGA
SH120       ACGCATCGCG TCGCCCCCC- ACCACCCATC CCATACGGTG ATGCATGGTA CGGGGGCGGA
NPB1877     ACGCATCGCG TCGCCCCCC- ACCACGCATC CC-------- ATGCATGGTA CGGGGGCGGA
NPB1880     ACGCATCGCG TCGCCCCCC- ACCACGCATC CC-------- ATGCATGGTA CGGGGGCGGA
NPB1881     ACGCATCGCG TCGCCCCCC- ACCACGCATC CC-------- ATGCATGGTA CGGGGGCGGA
NPB1887     ACGCATCGCG TCGCCCCCC- ACCACGCATC CC-------- ATGCATGGTA CGGGGGCGGA
PE2         ACGCATCGCG TCGCCCCCC- ACCACCCATC CCATACGGTG ATGCATGGTA CGGGGGCGGA
R_KOS       ACGCATCGCG TCGCCCCCC- ACCACCCATC CCATACGGTG ATGCATGGTA CGGGGGCGGA
SR178       ACGCATCGCG TCGCCCCCC- ACCACGCATC CC-------- ATGCATGGTA CGGGGGCGGA
SR179       ACGCATCGCG TCGCCCCCCC ACAACCCATC CCATACAGTG ATGCATGGTA CGGGGGCGGA
SH60        ACGCATCGCG TCCGCCCCCC ACCACGCATC CC-------- ATGCATGGTA CGGGGGCGGA
NPB1882     ACGCATCGCG TCGCCCCCC- ACCACGCATC CC-------- ATGCATGGTA CGGGGGCGGA
SH77        ACGCATCGC- TCGCCCCCC- ACCACCCATC CCTTACGGTG ATGCATGGTA CGGGGGCGGA
SH79        ACGCATCGCG TCGCCCCCC- ACCACCCATC CCATACGGTG ATGCATGGTA CGGGGGCGGA
SH78        ACGCATCGCG TCGCCCCCC- ACCACCCATC CCATACGGTG ATGCATGGTA CGGGGGCGGA
AW20        ACGCATCGCG TCGCCCCCC- ACCACCCATC CCTTACTGTG ATGCATGGTA CGGGGGCGGA
SH75        ACGCATCGCG TCGCCCCCC- ACCACCCATC CCATACGGTG ATGCATGGTA CGGGGGSGGA

                                                                          600
cstf        -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
NPB1831     -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
NPB1832     -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
pibotdt     -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
pimatt      -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
pithomb     -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
CP480       -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
f_inc       AGACTGGCCT CCCGTTCCCA CGGTGCGGCT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
CP485       -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
CP491       -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
BSR2        -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
BSR1        -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
TD4442      -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
SH59        -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
SH72        -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
SH86        -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
SH90        -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
mp_452      -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
SH100       -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
SH112       ?????????? ?????????? ?????????? ?????????? ?????????? ??????????
NPB1879     -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
SH40        -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
SH42        -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
SH44        -GACTGGCCT CCCGTTCCCC CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
SH49        -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
SH50        -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
NPB1830     -GACTGGCCT CCCGTTCCCG CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
CH43        -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
SH51        -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
SH54        -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
SH55        -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
SH87        -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
CH57        -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
SH58        -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
SH62        -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
SH65        -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
NPB1836     -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
SH102       -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
```

```
SH103         -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
SH52          -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
SH53          -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
SH66          -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
SH68          -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
SH70          -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
SH105         -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
SH73          -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
SH91          -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
SH94          -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
SH104         -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
NPB1820       -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
NPB1815       -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAAKC GGAGTCCCCT TCGGTGGACG
NPB1818       -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAAKC GGAGTCCCCT TCGGTGGACG
SR181         -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
SR173         -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
SR180         -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
SR184         -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
SR188         -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
SR189         -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
NPB1888       -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
NPB1833       -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
SH71          -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
SH97          -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
SH111         -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
SH120         -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
NPB1877       -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
NPB1880       -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
NPB1881       -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
NPB1887       -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
PE2           -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
R_KOS         -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
SR178         -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
SR179         -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTG????
SH60          -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
NPB1882       -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
SH77          -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
SH79          -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
SH78          -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
AW20          -GACTGGCCT CCCGTTCCCA CGGTGCGGTT GGCCAAAATC GGAGTCCCCT TCGGTGGACG
SH75          -GACKGGCCT CCCGTTCCCA CGGKGSGGTK GGCCAAAATC GGAGTCCCCT TCGGTGGACG

                                                                           660
cstf          CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
NPB1831       CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
NPB1832       CACGACTAGT GGTGGTTGRC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
pibotdt       CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
pimatt        CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
pithomb       CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
CP480         CACGACTAGT GGTGGTTGGC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
f_inc         CACGACTAGT GGTGGTTGAC AAAAACCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
CP485         CACGACTAGT GGTGGTTGGC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
CP491         CACGACTAGT GGTGGTTGGC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
BSR2          CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTT AGCCGCGAGG
BSR1          CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTY AGCCGCGAGG
TD4442        CACGACTAGT GGTGGTTGGC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
SH59          CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
SH72          CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
SH86          CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGTGG
SH90          CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
mp_452        CACGACTAGT GGTGGTTGAC ATAA-CCCTC GTCTCGTGTC GTGTGTCGTC ATCCGCAAGG
SH100         CACGACTAGT GGTGGTTGGC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
SH112         ?????????? ?????????? ?????????? ?????????? ?????????? ??????????
NPB1879       CACGACTAGT GGTGGTTGAC AAGG-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
SH40          CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
SH42          CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
SH44          CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
SH49          CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
SH50          CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
NPB1830       CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
CH43          CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTYGTGTC GTGTGTCGTC AGCCGCGAGG
SH51          CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
SH54          CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
SH55          CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
SH87          CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGTGG
CH57          CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
SH58          CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
```

```
SH62       CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGTGG
SH65       CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
NPB1836    CACGACTAGT GGTGGTTGRC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
SH102      CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
SH103      CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
SH52       CACGGCTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
SH53       CACGGCTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
SH66       CACGACTAGT GGTGGTTGAC AAAA-CCCTC YTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
SH68       CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGYGTCGTC AGCCGCGAGG
SH70       CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
SH105      CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
SH73       CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
SH91       CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
SH94       CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
SH104      CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
NPB1820    CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCATGTC GTGTGTCGTC AGCCGCGAGG
NPB1815    CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCATGTC GTGTGTCGTC AGCCGCGAGG
NPB1818    CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCATGTC GTGTGTCGTC AGCCGCGAGG
SR181      CACGGCTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
SR173      CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
SR180      CACGACTAGT GGTGGTTGAC AAAA-CCCTC KTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
SR184      CACGGCTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
SR188      CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
SR189      CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTT AGCCGCGAGG
NPB1888    CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGWGG
NPB1833    CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTYGTGTC GTGTGTCGTS AGCCGCGAGG
SH71       CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
SH97       CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
SH111      CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCNTC AGCCGC????
SH120      CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
NPB1877    CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
NPB1880    CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGWGG
NPB1881    CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGTGG
NPB1887    CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGTGG
PE2        CACGACTAGT GGTGGTTGGC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
R_KOS      CAYGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
SR178      CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
SR179      ?????????? ?????????? ?????????? ?????????? ?????????? ??????????
SH60       CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
NPB1882    CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
SH77       CACGACTAGT TGTGGTTGAC AAAA-CCCTG GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
SH79       CACGACTAGT GGTGGTTGAC AAAA-CCCTS GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
SH78       CACGACTAGT GGTGGTTGAC AAAA-CCCTS GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
AW20       CACGACTAGT GGTGGTTGAC AAAA-CCCTC GTCTCGTGTC GTGTGTCGTC AGCCGCGAGG
SH75       CACGAC???? ?????????? ?????????? ?????????? ?????????? ??????????

                                                                         720
cstf       GAAGACCTCT TCAAAGACCC CAA-CGCGTC -GTCTCGCGA CGATGCTT-C GACCTTGA-C
NPB1831    GAAGACCTCT TCAAAGACCC CAA-CGCRTC -GTCTCGCGA CGATGCTT-C GACCTTGA-C
NPB1832    GAAGACCTCT TCAAAGACCC CAA-CGCRTC -GTCTCGCGA CGATGCTT-C GACCGCGA-C
pibotdt    GAAGACCTCT TCAAAGACCC CAA-CGCGTC -GTCTCGCGA CGATGCTT-C GACCTTGA-C
pimatt     GAAGACCTCT TCAAAGACCC CAA-CGCGTC -GTCTCGCGA CGATGCTT-C GACCTTGA-C
pithomb    GAAGACCTCT TCAAAGACCC CAA-CGCGTC -GTCTCGCGA CGATGCTT-C GACCTTGA-C
CP480      GAAGACCTCT TCAAAGACCC CAA-CGCATC -GTCTCGCGA CGATGCTT-C GACCGCGA-C
f_inc      GAAGACCTCT TCAAAGACCC CAA-CGCGTC ?????????? ?????????? ??????????
CP485      GAAGACCTCT TCAAAGACCC CAA-CGCATC -GTCTCGCGA CGATGCTT-C GACCGCGA-C
CP491      GAAGACCTCT TCAAAGACCC CAA-CGCATC -GTCTCGCGA CGATGCTT-C GRCCGCAA-C
BSR2       GAAGACCTCT TCAAAGACCC CAA??????? ?????????? ?????????? ??????????
BSR1       GAAGACCTCT TCAAAGACCC CAA-CGCGTC -GTCTCGCGA CGATGCTT-C GACCGCGA-C
TD4442     GAAGACCTCT TCAAAGACCC CAA-CGCATC -GTCTCGCGA CGATGCTT-C GACCGCGA-C
SH59       GA???????? ?????????? ?????????? ?????????? ?????????? ??????????
SH72       GAAGACCTCT TCAAAGACCC CAA-CGCGTC -GTCTC???? ?????????? ??????????
SH86       GAAGACCTCT TCAAAGACCC CAA-CGCGTC -GTCTCGCGA CGATGCTT-C GACCGCG???
SH90       GAAGACCTCT TCAAAGACCC CAA-CGCGTC -GTCTCGCGA ?????????? ??????????
mp_452     GAAGACCTCT TCAAAGACCC CAA-CGCGTC -GTCTCGCGA CGAAGCTT-C GACCGCGA-C
SH100      GAAGACCTCT TCAAAGACCC CAAACGCGTC -GTCTCGCGA CGATGCT??? ??????????
SH112      ?????????? ?????????? ?????????? ?????????? ?????????? ??????????
NPB1879    GAAGACCTCT TCAAAGACC? ?????????? ?????????? ?????????? ??????????
SH40       GAAGACCTCT TCAAAGACCC CAA-CGCGTC -GTCTCGCGA CGATGCTT-C GACCTTGA-C
SH42       GAAGACCTCT TCAAAGACCC CAA-CGCGTC -GTCTCGCGA CGATGCTT-C GACCTTGA-C
SH44       GAAGACCTCT TCAAAGACCC CAA-CGCGTC -GTCTCGCGA CGATGCTT-C GACCGCGA-C
SH49       GAAGACCTCT TCAAAGACCC CAA-CGCGTC -GTCTCGCGA CGATGCTT-C GACCTTGA-C
SH50       GAAGACCTCT TCAAAGACCC CAA-CGCGTC -GTCTYGCGA CGATGCTT-C GACCGCGA-C
NPB1830    GAAGACCTCT TCAAAGACCC CAA-CGCGTC -GTCTCGCGA CGATGCTT-C GACCGCGA-C
CH43       GAAGACCTCT TCAAAGACCC CAA-CGYGTC -GTCTYGCGA CGATGCTT-C GACCTTGA-C
SH51       GAAGACCTCT TCAAAGACCC CAA-CGCGTC -GTCTCGCGA CGATGCTT-C GACCTTGA-C
SH54       GAAGACCTCT TCAAAGACCC MAA-CGCGTC -GTCTCGCGA CGATGCTT-C GACCGCGA-C
```

```
SH55       GAAGACCTCT TCAAAGACCC AAA-CGCGTC -GTCTCGCGA CGATGCTT-C GACCGCGA-C
SH87       GAAGACCTCT TCAAAGACCC CAA-CGCGTC -GTCTCGCGA CGATGCTT-C GACCGCGA-C
CH57       GAAGACCTCT TCAAAGACCC MAA-CGCGTC -GTCTCGCGA CGATGCTT-C GACCGCGA-C
SH58       GAAGACCTCT TCAAAGACCC CAA-CGCGTC -GTCTCGCGA CGATGCTT-C GACCGCGA-C
SH62       GAAGACCTCT TCAAAGACCC CAA-CGCGTC -GTCTCGCGA CGATGCTT-C GACCGCGA-C
SH65       GAAGACCTCT TCAAAGACCC CAA-CGCGTC -GTCTCGCGA CGATGCTT-C GACCGCGA-C
NPB1836    GAAGACCTCT TCAAAGACCC CAA-CGCKTC -GTCTCGCGA CGATGCTT-C GACCGCGA-C
SH102      GAAGACCTCT TCAAAGACCC MAA-CGCGTC -GTCTCGCGA CGATGCTT-C GACCGCGA-C
SH103      GAAGACCTCT TCAAAGACCC MAA-CGCGTC -GTCTCGCGA CGATGCTT-C GACCGCGA-C
SH52       GAAGACCTCT TCAAAGACCC CAA-CGCGTC ?????????? ?????????? ??????????
SH53       GAAGACCTCT TCAAAGACCC CAA-CGCGTC AGTCTCGCGA CGATGCTT-C GACCGCGA-C
SH66       GAAGACCTCT TCAAAGACCC CAA-CGCGTC -GTCTCGCGA CGATGCTT-C GACCGCGA-C
SH68       GAAGACCTCT TCAAAGACCC CAA-CGCGTC -GTCTCGCGA CGATGCTT-C GACCGCGA-C
SH70       GAAGACCTCT TCAAAGACCC CAA-CGCGTC -GTCTCGCGA CGATGCTT-C GACCGCGA-C
SH105      GAAGACCTCT TCAAAGACCC MAA-CGCGTC -GTCTCGCGA CGATGCTT-C GACCGCGA-C
SH73       GAAGACCTCT TCAAAGACCC CAA-CGCGTC -GTCTCGCGA CGATGCTT-C GACCGCGA-C
SH91       GAAGACCTCT TCAAAGACCC CAA-CGCGTC -GTCTCGCGA CGATGCTT-C GACCGCGA-C
SH94       GAAGACCTCT TCAAAGACCC AAA-CGCGTC -GTCTCGCGA CGATGCTT-C GACCGCGA-C
SH104      GAAGACCTCT TCAAAGACCC AAA-CGCGTC -GTCTCGCGA CGATGCTT-C GACCGCGA-C
NPB1820    GAAGGCCTCT TCAAAGACCC CAA-YGCGTC -GTCTTGCGA CGATGCTT-C GRCCGCGA-C
NPB1815    GAAGGCCTCT TCAAAGACCC CAA-TGCGTC -GTCTTGCGA CGATGCTT-C GACCGCGA-C
NPB1818    GAAGGCCTCT TCAAAGACCC CAA-YGCGTC -GTCTTGCGA CGATGCTT-C GACCGCGA-C
SR181      GAAGACCTCT TCAAAGACCC CAA-CGCGTC -GTCTCGCGA CAATGCTT-C GACCGC????
SR173      GAAGACCTCT TCAAAGACCC CAA-CGCGTC -GTCTCGCGA CGATGCTT-C GACCTTGA-C
SR180      GAAGACCTCT TCAAAGACCC CAA-CGCGTC -GTCTCGCGA CGATGCTT-C GACCGCGA-C
SR184      GAAGACCTCT TCAAAGACCC CAA-CGCGTC -GTCTCGCGA CGATGCTT-C GACCGCGA-C
SR188      GAAGACCTCT TCAAAGACCC CAA-CGCGTC -GTCTCGCGA CGATGCTT-C GACCGCGA-C
SR189      GAAGACCTCT TCAAAGACCC CAA-CGCGTC -GTCTCGCGA CCATGCTT-C GA????????
NPB1888    GAAGACCTCT TCAAAGACCC MAA-CGCGTC -GTCTCGCGA CGATGCTT-C GACCGCGA-C
NPB1833    GAAGACCTCT TCAAAGACCC CAA-CGYGTC -GTCTYGCGA CGATGCTT-C GACCGCGA-C
SH71       GAAGACCTCT TCAAAGACCC CAA-CGCGTC -GTCTCGCGA CGATGCTT-C GACCGCGA-C
SH97       GAAGACCTCT TCAAAGACCC CAA-CGCGTC -GTCTCGCGA CGATGCTT-C GACCGCGA-C
SH111      ?????????? ?????????? ?????????? ?????????? ?????????? ??????????
SH120      GAAGACCTCT TCAAAGACCC CAA-CGCATC -GTCTCGCGA CGATGCTT-C GACCGCGA-C
NPB1877    GAAGACCTCT TCAAAGACCC MAA-CGCGTC -GTCTCGCGA CGATGCTT-C GACCGCGA-C
NPB1880    GAAGACCTCT TCAAAGACCC CAA-CGCGTC -GTCTCGCGA CGATGCTT-C GACCGCGA-C
NPB1881    GAAGACCTCT TCAAAGACCC CAA-CGCGTC -GTCTCGCGA CGATGCTT-C GACCGCGA-C
NPB1887    GAAGACCTCT TCAAAGACCC CAA-CGCGTC -GTCTCGCGA CGATGCTT-C GACCGCGA-C
PE2        GAAGACCTCT TCAAAGACCC CAA-CGCRTC -GTCTCGCGA CGATGCTT-C GACCGCGA-C
R_KOS      GAAGACCTCT TCAAAGACCC CAA-CGCGTC -GTCTCGCGA CGATGCTT-C GACCKYGA-C
SR178      GAAGACCTCT TCAAAGACCC CAA-CGCGTC -GTCTCGCGA CKATGCTTTC GACCGCTA-C
SR179      ?????????? ?????????? ?????????? ?????????? ?????????? ??????????
SH60       GAAGACCTCT TCAAAGACCC CAA-CGCGTC ?????????? ?????????? ??????????
NPB1882    GAAGACCTCT TCAAAGACCC CAA-CGCGTC -GTCTCGCGA CGATGCTT-C GACCGCGA-C
SH77       GAAGACCTCT TCAAAGACCC CAAGCGCGTC -GTCTCGCGA CGATGCTT-C GACCGCGAGC
SH79       GAAGACCTCT TCAAAGACCC CAAGCGCGTC -GTCTCGCGA CGATGCTT-C GACCGCGAGC
SH78       GAAGACCTCT TCAAAGACCC CAAGCGCGTC -GTCTCGCGA CGATGCTT-C GACCGCGAGC
AW20       GAAGACCTCT TCAAAGACCC CAA-CGCGTC -GTCTCGCGA CGATGCTT-C GACCGCGA-C
SH75       ?????????? ?????????? ?????????? ?????????? ?????????? ??????????


cstf       CCCAGGTCAG GCGGGACTAC CCGC-TGAGT TTAA
NPB1831    CCCAGGTCAG GCGGGACTAC CCGC-TGAGT TTAA
NPB1832    CCCAGGTCAG GCGGGACTAC CCGC-TGAGT TTAA
pibotdt    CCCAGGTCAG GCGGGACTAC CCGC-TGAGT TTAA
pimatt     CCCAGGTCAG GCGGGACTAC CCGC-TGAGT TTAA
pithomb    CCCAG-TCAG GCGGGACTAC CCGC-TGAGT TTAA
CP480      CCCAGGTCAG GCGGGACTAC CCGC-TGAGT TTAA
f_inc      ?????????? ?????????? ?????????? ????
CP485      CCCAGGTCAG GCGGGACTAC CCGC-TGAGT TTAA
CP491      CCCAGGTCAG GCGGGACTAC CCGC-TGA?? ????
BSR2       ?????????? ?????????? ?????????? ????
BSR1       CCCAGGTCAG GCGGGACTAC CCGC-TGAGT TTAA
TD4442     CCCAGGTCAG GCGGGACTAC CCGC-TGAGT TTAA
SH59       ?????????? ?????????? ?????????? ????
SH72       ?????????? ?????????? ?????????? ????
SH86       ?????????? ?????????? ?????????? ????
SH90       ?????????? ?????????? ?????????? ????
mp_452     CCCAGGTCAG GCGGGACTAC CCGCCTGAGT TTA?
SH100      ?????????? ?????????? ?????????? ????
SH112      ?????????? ?????????? ?????????? ????
NPB1879    ?????????? ?????????? ?????????? ????
SH40       CCCAGGTCAG G????????? ?????????? ????
SH42       CCCAGGTCAG GCGGGACTAC CCGC-TGAGT TTA?
SH44       CCCAGGTCAG GCGGGACTAC CCGC-TGAGT TTAA
SH49       CCCAGGTCAG GCGGGACTAC CCGC-TGAGT TTAA
SH50       CCCAGGTCAG GCGGGACTAC CCGC-TGAGT TTAA
```

```
NPB1830    CCCAGATCAG GCGGGACTAC CCGC-TGAGT TTAA
CH43       CCCAGGTCAG GCGGGACTAC CCGC-TGAGT TTAA
SH51       CCCAGGTCAG GCGGGACTAC CCGC-TGAGT TTAA
SH54       CCCAGGTCAG GCGGGACTAC CCGC-TGAGT TTAA
SH55       CCCAGGTCAG GCGGGACTAC CCGC-TGAGT TTAA
SH87       CCCAGGTCAG ACGGGACTAC CCGC-TGAGT TTAA
CH57       CCCAGGTCAG GCGGGACTAC CCGC-TGAGT TTAA
SH58       CCCAGGTCAG ACGGGACTAC CCGC-TGAGT TTAA
SH62       CCCAGGTCAG ACGGGACTAC CCGC-TGAGT TTAA
SH65       CCCAGGTCAG ACGGGACTAC CCGC-TGAGT TTAA
NPB1836    CCCAGGTCAG GCGGGACTAC CCGC-TGAGT TTAA
SH102      CCCAGGTCAG RCGGGACTAC CCGC-TGAGT TTAA
SH103      CCCAGGTCAG ACGGGACTAC CCGC-TGAGT TTAA
SH52       ?????????? ?????????? ?????????? ????
SH53       CCCAGGTCAG GCGGGACTAC CCGC-TGAKT TTAA
SH66       CCCAGGTCAG RCGGGACTAC CCGC-TGAGT TTAA
SH68       CCCAGGTCAG ACGGGACTAC CCGC-TGAGT TTAA
SH70       CCCAGGTCAG ACGGGACTAC CCGC-TGAGT TTAA
SH105      CCCAGGTCAG ACGGGACTAC CCGC-TGAGT TTAA
SH73       CCCAGGTCAG ACGGGACTAC CCGC-TGAGT TTAA
SH91       CCCAGGTCAG ACGGGACTAC CCGC-TGAGT TTAA
SH94       CCCAGGTCAG GCGGGACTAC CCGC-TGAGT TTAA
SH104      CCCAGGTCAG RCGGGACTAC CCGC-TGAGT TTAA
NPB1820    CCCAGGTCAG GCGGGACTAC CCGC?????? ????
NPB1815    CCCAGGTCAG GCGGGACTAC CCGC-TGAGT TTAA
NPB1818    CCCAGGTCAG GCGGGACTAC CCGC-TGAGT TTAA
SR181      ?????????? ?????????? ?????????? ????
SR173      CCCAGGTCAG GCGGGACTAC CCGC-TGAGT TTAA
SR180      CCCAGGTCAG RCGGGACTAC CCGC-TGAGT TTAA
SR184      CCCAGGTCAG GCGGGACTAC CCGC-TGAGT TTAA
SR188      CCCAGG???? ?????????? ?????????? ????
SR189      ?????????? ?????????? ?????????? ????
NPB1888    CCCAGGTCAG GCGGGACTAC CCGC-TGAGT TTAA
NPB1833    CCCAGGTCAG GCGGGACTAC CCGC-TGAGT TTAA
SH71       CCCAGGTCAG ACGGGACTAC CCGC-TGAGT TTAG
SH97       CCCAGGTCAG GCGGGACTAC CCGC-TGAGT TTAA
SH111      ?????????? ?????????? ?????????? ????
SH120      CCCAGGTCAG GCGGGAATAC CGCT-T-AGT TTAG
NPB1877    CCCAGGTCAG RCGGGACTAC CCGC-TGAGT TTAA
NPB1880    CCCAGGTCAG ACGGGACTAC CCGC-TGAGT TTAA
NPB1881    CCCAGGTCAG ACGGGACTAC CCGC-TGAGT TTAA
NPB1887    CCCAGGTCAG ACGGGACTAC CCGC-TGAGT TTAA
PE2        CCCAGGTCAG GCGGGACTAC CCGC-TGAGT TTAA
R_KOS      CCCAGGTCAG GCGGGACTAC CCGC-TGAGT TTAA
SR178      CCCAGGTCAG GCGGGACTAC CCGC-TGAGT TTAT
SR179      ?????????? ?????????? ?????????? ????
SH60       ?????????? ?????????? ?????????? ????
NPB1882    CCCAGGTCAG ACGGGACTAC CCGC-TGAGT TTAA
SH77       CCCAGTTCAG GCGGGACTAC CCGC-T???? ????
SH79       CCCAGTTCAG GCGGGAC-AC CCGC-TG??? ????
SH78       CCCAGGTC?? ?????????? ?????????? ????
AW20       CCCAGGTCA? ?????????? ?????????? ????
SH75       ?????????? ?????????? ?????????? ????
```

rps16

```
                                                                                  60
pi_173      ????????? ?GCTCATCGG GATAGATGTA GATGAAAAGA ACCCCCCCTC CCTAGAACCG
Oju_1830    GATCGATGAC GGCTCATCGG GATAGATGTA GATGAAAAGA ACCCCCCCTC CCTAGAACCG
rot_491     GATCGATGAC GGCTCATCGG GATAGATGTA GATGAAAAGA ACCCCCCCTC CCTAGAACCG
flo_bsrl    GATCGATGAC GGCTCATCGG GATAGATGTA GATGAAAAGA ACCCCCCCTC CCTAGAACCG

                                                                                 120
pi_173      TATAGGAAGT TTTCTCCTCG TACGGCTCGA GAAAAAATGA TTCGAAGTTT TGTCTATGGA
Oju_1830    TATAGGAAGT TTTCTCCTCG TACGGCTCGA GAAAAAATGA TTCGAAGTTT TGTCTATGGA
rot_491     TATAGGAAGT TTTCTCCTCG TACGGCTCGA GAAAAAATGA TTCGAAGTTT TGTCTATGGA
flo_bsrl    TATAGGAAGT TTTCTCCTCG TACGGCTCGA GAAAAAATGA TTCGAAGTTT TGTCTATGGA

                                                                                 180
pi_173      TAAAATTAGA ATAAATAGGA AAGTAATCCG TAAAATAAAT TAGTCTATAA TTTAACTCAT
Oju_1830    TAAAATTAGA ATAAATAGGA AAGTAATCCG TAAAATAAAT TAGTCTATAA TTTAACTCAT
rot_491     TAAAATTAGA ATAAATAGGA AAGTAATCCG TAAAATAAAT TAGTCTATAA TTTAACTCAT
flo_bsrl    TAAAATTAGA ATAAATAGGA AAGTAATCCG TAAAATAAAT TAGTCTATAA TTTAACTCAT

                                                                                 240
pi_173      AGTCATTTTT ATTTAGCCTC CTTACTGAAA AATAAAAAAT CATTTGTACT CAGAACTCAA
Oju_1830    AGTCATTTTT ATTTAGCCTC CTTACTGAAA AATAAAAAAT CATTTGTACT CAGAACTCAA
rot_491     AGTCATTTTT ATTTAGCCTC CTTACTGAAA AATAAAAAAT CATTTGTACT CAGAACTCAA
flo_bsrl    AGTCATTTTT ATTTAGCCTC CTTACTGAAA AATAAAAAAT CATTTGTACT CAGAACTCAA

                                                                                 300
pi_173      GTTCAATAAT TCTCAAAAAT CTTAAAGATT TTTCTTTCAG ATTTTTTTTG AGTGGTCTTT
Oju_1830    GTTCAATAAT TCTCAAAAAT CTTAAAGATT TTTCTTTCAG ATTTTTTTTG AGTGGTCTTT
rot_491     GTTCAATAAT TCTCAAAAAT CTTAAAGATT TTTCTTTCAG ATTTTTTTTG AGTGGTCTTT
flo_bsrl    GTTCAATAAT TCTCAAAAAT CTTAAAGATT TTTCTTTCAG ATTTTTTTTG AGTGGTCTTT

                                                                                 360
pi_173      AACTCACCCT TTTTGTCTCG TTTAAAATCT ATTTGGATTC TTTATTCGGA TCTGTGAGAC
Oju_1830    AACTCACCCT TTTTGTCTCG TTTAAAATCT ATTTGGATTC TTTATTCGGA TCTGTGAGAC
rot_491     AACTCACCCT TTTTGTCTCG TTTAAAATCT ATTTGGATTC TTTATTCGGA TCTGTGAGAC
flo_bsrl    AACTCACCCT TTTTGTCTCG TTTAAAATCT ATTTGGATTC TTTATTCGGA TCTGTGAGCC

                                                                                 420
pi_173      AATTGAAGGC TGTTTTCTTG TTCTGGGATC CTTTATCTTT GTTTTAAATC ATTGGGGTTA
Oju_1830    AATTGAAGGC TGTTTTCTTG TTCTGGGATC CTTTTTCTTT GTTTTAAATC ATTGGGGTTA
rot_491     AATTGAAGGC TGTTTTCTTG TTCTGGGATC CTTTATCTTT GTTTTAAATC ATTGGGGTTA
flo_bsrl    AATTGAAGGC TGTTTTCTTG TTCTGGGATC CTTTATCTTT GTTTTAAATC ATTGGGGTTA

                                                                                 480
pi_173      GACATTACTT CGGTGCTTCT TAATCCTTTC AAAAGGTAGC AACATACCCC CTTTGTGATT
Oju_1830    GACATTACTT CGGTGCTTCT TAATCCTTTC AAAAGGTAGC AACATACCCC CTTTGTGATT
rot_491     GACATTACTT CGGTGCTTCT TAATCCTTTC AAAAGGTAGC AACATACCCC CTTTGTGATT
flo_bsrl    GACATTACTT CGGTGCTTCT TAATCCTTTC AAAAGGTAGC AACATACCCC CTTTGTGATT

pi_173      TCTTTCTATC AAAGAATCAT ACCGACGGGT GATTCGTGCG CGA
Oju_1830    TCTTTCTATC AAAGAATCAT ACCGACGGGT GATTCGTGCG CGA
rot_491     TCTTTCTATC AAAGAATCAT ACCGACGGGT GATTCGTGCG CGA
flo_bsrl    TCTTTCTATC AAAGAATCAT ACCGACGGGT GATTCGTGCG CGA
```

trnS-trnfM

```
                                                                       60
can_1820    ?????????? ?????????? ?????????? ?????????? ???AACTAGT AATTAGTTAA
rot_491     GCCTCGCATA CTTTTTGTGA AGGATTGTAT GTTTTACTCC GGTAACTAGT AATTAGTTAA
pis_40      ?????????? ?TTTTTGTGA AGGATTGTAT GTTTTACTCC GGTAACTAGT AATTAGTTAA
inc_75      GCCTCGCATA CTTTTTGTGA AGGATTGTAT GTTTTACTCC GGTAACTGGT AATTAGTTAA

                                                                      120
can_1820    TGCTTTTTCT TTTGGGGTAG AAGACAAAGA GGTGAAGAAT AGAATCACTA CACTATGACG
rot_491     CGCTTTTTCT TTTGGGGTAG AAGACAAAGA GGTGAAGAAT AGAATCACTA CACTATGACG
pis_40      TGCTTTTTCT TTTGGGGTAG AAGACAAAGA GGTGAAGAAT AGAATCACTA CACTATGACG
inc_75      TGCTTTTTCT TTTGGGGTAG AAGACAAAGA GGTGAAGAAT AGAATCACTA CACTATGACG

                                                                      180
can_1820    GCTAACTATA CCAAATCCTT TATAGTTTTA CATTTTATTA GATATGAAAT AAATGACTTC
rot_491     GCTAACTATA CCAAATCCTT TATCGTTTTA CATTTTATTA GATATGAAAT AAATGACTTC
pis_40      GCTAACTATA CCAAATCCTT TATCGTTTTA CATTTTATTA GATATGAAAT AAATGACTTC
inc_75      GCTAACTATA CCAAATCCTT TATCGTTTTA CATTTTATTA GATATGAAAT AAATGACTTC

                                                                      240
can_1820    ATTTTGGGCG GATAGCGGGA ATCGAACCCG CGTCTTCTCC TTGGCAAAGA GAAATTTTAC
rot_491     ATTTTGGGCG GATAGCGGGA ATCGAACCCG CGTCTTCTCC TTGGCAAAGA GAAATTTTAC
pis_40      ATTTTGGGCG GATAGCGGGA ATCGAACCCG CGTCTTCTCC TTGGCAAAGA GAAATTTTAC
inc_75      ATTTTGGGCG GATAGCGGGA ATCGAACCCG CGTCTTCTCC TTGGCAAAGA GAAATTTTAC

                                                                      300
can_1820    CATTCGACTA TACCCGCTTT TTGTTTCGTT CTTGATACAC AATATAATAT GTCTATAATA
rot_491     CATTCGACTA TMTCCGCTTT TTGTTTCGTT CTTGATACAC AATATAATAT GTCTATAATA
pis_40      CATTCGACTA TATCCGCTTT TTGTTTCGTT CTTGATACAC AATATAATAT GTCTATAATA
inc_75      CATTCGACTA TMTCCGCTTT TTGTTTCGTT CTTGATACAC AATATAATAT GTGTATAATA

                                                                      360
can_1820    CACAATATGT CCACACATAT ATTATATATA TTTCTGGAGC ATATTTGTGC AATGCTGGAC
rot_491     CACAATATGT CCACACATAT ATTATATATA TTTCTGGAGC ATATTTGTGC AATGCTGGGC
pis_40      CACAATATGT CCACACATAT ATTATATATA TTTCTGGAGC ATATTTGTGC AATGCTGGGC
inc_75      CACAATATGT CCACACATAT ATTATATATA TGTCTGGAGC ATATTTGTGC AATGCTGGGC

                                                                      420
can_1820    CGGAYACTCT CTTCAGATTG AGCCCAAATT TTTTATTAAT TTGAATTTCA TATTCATTTT
rot_491     CGGAYACTCT CTTCAGATTG AGCCCAAATT TTTTATTAAT TTGAATTTCA TATTCATTTT
pis_40      CGGAYACTCT CTTCAGATTG AGCCCAAATT TTTTATTAAT TTGAATTTCA TATTCATTTT
inc_75      CGGAYACTCT CTTCAGATTG AGCCCAAATT TTTTATTAAT TTGAATTTCA TATTCATTTT

                                                                      480
can_1820    TTTCAAGAAC AAGAAGTTTG ACCCCCTTCT AATTTTTCAT TTTTTTTCTT ATATTTATTT
rot_491     TTTCAAGAAC AAGAAGTTTG ACCCCCTTCT AATTTTTCAT TTTTTTTCTT ATATTTATTT
pis_40      TTTCAAGAAC AAGAAGTTTG ACCCCCTTCT AATTTTTCAT TTTTTTTCTT AAATTTATTT
inc_75      TTTCAAGAAC AAGAAGTTTG ACCCCCTTCT AATTTTTCAT TTTTTTTCTT AAATTTATTT

can_1820    GGATTTTGGG GACTTATATT GACTGTGA
rot_491     GGATTTTGGG GACTTATATT GACTGTGA
pis_40      GGATTTTGGG GACTTATATT GACTGTGA
inc_75      GGATTTTGGG GACTTATATT GACTGTGA
```

Appendix 4: Fluorescence transients: formulae for calculations.

$V_J = (F_J - F_0)/(F_M - F_0)$

$M_0 = 4(F_{300\mu s}) - F_0)(F_M - F_0)$

Specific energy fluxes expressed per reaction centre (RC)

$ABS/RC = (M_0/V_J)/[1-(F_0/F_M)]$

$TR_0/RC = (M_0/V_J) = (ABS/RC)\ \varphi_{P0}$

$DI_0/RC = (ABS/RC) - (TR_0/RC)$

$ET_0/RC = (M_0/V_J)(1-V_J) = (TR_0/RC)\ \psi_0$

Yields (or flux ratios)

$\varphi_{P0} = [1-(F_0/F_M)] = (TR_0/RC)(ABS/RC) = TR_0/ABS$

$\varphi_{E0} = [1-(F_0/F_M)](1-V_J) = (ET_0/RC)/(ABS/RC) = ET_0/ABS$

$\psi_0 = 1-V_J = (ET_0/RC)(TR_0/RC) = (ET_0/TR_0)$

Performance index

$PI_{ABS} = (RC/ABS)\ (\varphi_{P0}/\ 1-\varphi_{P0})\ (\psi_0/1-\psi_0)$

Appendix 5: Table of Fluorescence transients values.

Locality	F_M	F_0	F(300 µs)	F_J	F_I	Area
Joubertina	3372 ± 383	608 ± 82	1350 ± 209	1943 ± 276	2899 ± 374	41883 ± 10294
Tsitsikama	3656 ± 228	497 ± 45	1518 ± 174	2166 ± 156	3117 ± 160	52146 ± 14231
St Francis	3586 ± 205	512 ± 47	1268 ± 119	1875 ± 155	2968 ± 211	55257 ± 9074