Phylogenetic relationships within the subtribe Chironiinae (Gentianaceae), with a particular focus on the evolutionary trends and biogeographic patterns of the genus *Centaurium* Hill s. l.

Guilhem Mansion

****phD Thesis ****
November 2001
Université de Neuchâtel - Institut de Botanique
Laboratoire de Botanique Evolutive (Prof. P. Küpfer)
02 november 2001

A mes filles, Anaïs & Amélie.

A mes parents.

Acknowledgements

Sans le concours de nombreuses personnes qui, durant ces dernières années, m'ont apporté leur soutien et leur aide, le présent travail ne serait qu'ébauche…

Je tiens donc à remercier tout particulièrement :

A

***Alessia Guggisberg**, qui m'a soutenu et supporté dans les moments les difficiles et particulièrement éprouvants, que sont la rédaction et la mise en page d'une thèse. Merci pour ton lumineux sourire!

***Anouk Béguin**, qui m'a rendu de nombreux services de part sa compétence, son savoir technique et sa grande sympathie.

B

***Bob**, autrement dit **Roberto Guadagnuolo**, collègue de bureau mais surtout un grand ami qui, durant toutes ces années, a sû m'écouter, voire me supporter quotidiennement (et réciproquement!). Sans toi, ces moments n'auraient pas eu la même force! Alors grazie mile y scuzi per la musica!

***Bret**, plus connu sous le pseudonyme de Prof. **François Bretagnolle**, grand dans l'amitié et la générosité! Many thanks pour les longues discussions philosophiques, quelquefois plus matérielles.., autour d'un café (en poudre, uniquement!) ou autre breuvage dyonisiaque. Merci pour tes conseils, pour le temps passé à la relecture de ce papier et pour ta présence dans ce jury.

C

***Calamus** (nom vernaculaire: **François Calame**), qui a fait preuve de beaucoup de compréhension et de bon goût, en allant jusqu'à apprécier certains de mes style musicaux! Merci également pour ta bonne humeur journalière et tes blagues homériques.

***Cindy Nussbaumer**, dont le rire et le dynamisme ont laissé une empreinte indélibile dans ce laboratoire. Merci pour toute l'aide que tu as pû m'apporter durant ces années.

***Charles Jollès**, un grand amateur de gentianes (surtout la jaune!), à qui je dédie tout particulièrement ce travail. Ton dynamisme et ta passion ont été, assurement, des catalyseurs très efficaces dans l'accomplissement de ce travail.

D

***Daniel Oppizi,** avec qui j'ai eu beaucoup de plaisir à partager les premiers moments de ma thèse, dans une excellente ambiance!

E

*L'incontournable **Ernest Fortis** sur qui j'ai toujours pu compter. Merci pour ton professionnalisme sans faille, ta capacité d'écoute et surtout ton grand cœur!

F

***François Felber**, dont la compétence n'a d'égale que la grande modestie. Merci pour tous tes conseils pertinents et efficaces.

K

***Kirsten Wolff**, pour m'avoir transmis une partie de ses immenses connaissances. Ton rire restera à jamais gravé dans les mémoires de ceux qui ont eu la chance de te côtoyer!

L

***Lena Struwe**, THE specialist of the Gentianaceae family, for her participation in the jury. Many thanks for your help, readings and interesting comments of the present dissertation.

*__Louis__ __&__ __Nicole__ __Zeltner__, qui furent mes fidèles compagnons de voyage au cours de nombreuses "expéditions *Centaurium*". Merci pour l'aide immense, le dévouement sans pareil et les réunions de travail à la Saignotte (Jura Neuchâlelois), autant de facteurs qui ont contribué à la réussite (j'espère!) de ce travail.

*__Louis__ __Renaud__, avec qui j'ai eu beaucoup de plaisir à partager les nombreuses discussions et sorties dans lesquelles il m'entraînait souvent!!

P

__*Philippe__ __Küpfer__, qui a su me donner la chance de pratiquer quotidiennement la Botanique. Merci pour ton enthousiasme passionné, ta sympathie sans bornes, et toutes les randonnées alpiennes où j'ai eu la chance de profiter de ton immense savoir.

__*Pierre__ __Taberlet__, qui m'a fait le grand honneur d'accepter de juger ce travail. Merci pour les remarques pertinentes et tous les conseils prodigués.

*Toute l'équipe de __Phanérogamie__ (aujourd'hui Botanique Evolutive), passée et présente, avec qui j'ai partagé de grands moments. Mes meilleurs souvenirs vont à Louis Azevedo, Frédéric Bilet, Ariane Calliau, Martin Callmander, Philippe Chassot, Luigi D'Andrea, Gabrielle Ducommun, Adrienne Godiot, Jason Grant, Ali Mezouar, Christian Monnerat, Louis Nusbaumer, Catherine Perret, Janine Piguet, Ariane Pittet, Michal Ronikier, François Rion, Dessislava Savova-Bianchi, Nicola Schoenenberger, Olivier Stauffer, Sarah-Lan Stieffel… et tous ceux que ma mémoire défaillante oublie!

* Mes meilleures pensées aux amis rencontrés durant mes années "__Pharmaciennes"__, Olivier & Florence Marty, Joël & Nelly Pistre, Michel & Caroline Martinez, Jean-Yves Fournier, Sylvie Chambert, Sophie Barthe et Stéphanie Prunac-Mansion… et tous les autres.

R

*Merci au Professeur __Raphaël__ __Tabbachi__, pour le temps consacré au jugement de ce travail qui, pour certaines raisons temporelles, n'a pas eu l'ampleur phytochimique initialement souhaitée.

S

*Mes pensées les plus amicales vont à __Siwert__ __Nilsson__, pour la réalisation des microphotographies de grains de pollen, et surtout pour sa gentillesse sans bornes.

Y

*Many thanks to __Yong-Ming__ __Yuan__, THE gentian's specialist, for all his comments, judgements and advices on the present work.

CONTENTS

Introduction

Summary

One of the aims of the phylogenetic reconstruction is to provide natural classification by inferring relationships between taxa, but also to understand the history of a group and its evolution through space and time. Presently, the species occurring on Earth are the results of diverse and numerous selective strength and constraints inducing reproductive isolation. The genetic divergence may be either progressive, with the occurrence of physical barriers, or abrupt with sudden chromosome changes. In this context, the genus *Centaurium* may be considered as an ideal group to underlie some evolutionary trends such as allopatric speciation with geographical isolation or polyploidy effects, known to provide abrupt speciation, via hybridization. This genus comprises annuals or biennials species, often forming polyploid complexes and mainly distributed in two distinct geographical areas of the Northern Hemisphere. In the Mediterranean region and Europe, diploid to polyploid taxa are represented, whereas only polyploids are known to occur in the western North America.

At the beginning of the present study, the systematics of *Centaurium* was greatly uncertain and consequently, many questions remained unresolved such as the relationships between the respective groups and their biogeographical history. Moreover, the importance of polyploidy processes and chromosome rearrangements as evolutionary trends has been questioned.

For morphological or karyological data sets often generate a weak number of informative characters, DNA sequences of either maternal inheritance (*trn*L intron and *trn*L-F spacer) or biparental origin (ITS and 5S-NTS), have been used to reconstruct molecular phylogenies of *Centaurium*.

A first study performed on the North American species of *Centaurium* (Chapter 1) revealed the occurrence of three well-supported groups in this New World clade. Each of them occupies a particular geographic region and share similar chromosome number. In this area, important geological events, such as deserts and mountains formation in the Miocene or Quaternary ice ages, are known to have influenced plant evolution and repartition, and may have modulated the present allopatric distribution of the North American centauries. Diploid species have not been encountered in this zone and dysploidy, or sometimes allopolyploidy, appears to be the main trend of chromosome evolution in this group.

Similar analyses achieved on the Mediterranean group of *Centaurium* (Chapter 2) have depicted four clades characterized by the presence of diploid and polyploid taxa. A large sampling of more than 1400 populations indicates that diploid species are generally confined to restricted areas corresponding to Pleistocene glacial refugia, whereas their relative polyploids are distributed in peripheral zones. Moreover, one clade comprising the majority of the European species reveals weak interspecific sequences divergence contrarily to the

morphological differentiation. Environmental pressures may have been important on this group and recent Quaternary glaciations have, in all likelihood, contributed to the present distribution and genetic differentiation of the Eurasian *Centaurium*. In this group, unlike the North American one, polyploidy appears to be the main trend of chromosome evolution with frequent allopolyploidy and occasional dysploidy.

Such difference in evolutionary trends and chromosome behavior have been reinforced by the evidence of the polyphyly of *Centaurium* revealed by a large data set of ITS and *trn*LF sequences (Chapter 3), and further confirmed with 5S-NTS analyses (Chapter 4). Moreover, *Centaurium* appears to be an assemblage of four independent lineages having phylogenetic affinities with other genera of the subtribe Chironiinae. For example, in the ITS cladograms, section *Spicaria* appears to be closer to the genus *Exaculum* and section *Gyrandra* shares affinities with *Sabatia*, whereas the American and Eurasian clades are separated by a deep phylogenetic gap.

The polyphyly of *Centaurium* not only complicates the systematic of the whole subtribe, with few morphological synapomorphies available to support molecular clades, but also have important consequences on the evolutionary pattern of the group.

In a biogeographic point of view, several episodes of vicariance and intercontinental dispersal, with frequent back migrations, may be proposed from a Mediterranean center of origin (*Blackstonia*, Eurasian *Centaurium*, *Cicendia*, *Exaculum*, *Ixanthus* or section *Spicaria*) to North America (American *Centaurium*, *Cicendia*, *Eustoma*, section *Gyrandra* or *Sabatia*), South Africa (*Chironia, Orphium*) or Australia (section *Spicaria*).

The chromosome evolution also has to accommodate the whole subtribe diversity. Moreover, chromosome numbers has not yet been determinate for several taxa such as *Chironia* or *Orphium*; a general pattern seems difficult to trace. Nevertheless, on the base of molecular or biogeographical evidences, some karyological schemes may be drawn, with a basal number of n = 10 and involving both dysploidy and polyploidy processes. The latter appear to be of particular importance in *Centaurium* s. l. evolution. Nevertheless, other patterns including gene introgression or homoploid hybridization may be of particular importance, as suggested by a punctual analysis of the *C. muehlenbergii* complex (Chapter 5).

Finally, the use of molecular tools to reconstruct the phylogeny of a particularly difficult group of plants, the genus *Centaurium*, has revealed different evolutionary scenarios along with particular biogeographic patterns, in four particular groups of divergent origins.

Major trends in plant evolution

Species concepts and speciation processes

Phylogeny may be regarded as the evolutionary history of a particular group of organisms, depicted on a general tree of life, with the existing taxa present at the tips of the outermost branches. One of the major aims of the phylogenetic reconstruction is to provide the circumscription of natural groups of species and, subsequently to infer patterns in plant evolution.

Thus, the species concept, which is one of the more fascinating and debated theories in biology, had to accommodate a temporal scale. Historically, this definition has changed and is still changing. First descriptions dealt with fixed and invariable entities ("Typological Species Concept"), deriving in a "Morphological" concept, measuring the degree of total resemblance or divergence between taxa, to finally produce the "Biological Species Concept" (Dobzhansky, 1937) refined by Mayr (1963): "species are groups of actually or potentially interbreeding natural populations that are reproductively isolated from other such groups". Latter, Wiley (1981) proposed a modified version of the "Evolutionary Species Concept" (Simpson, 1951) and described ancestor-descendant lineage maintaining their identity in a unique evolutionary history. In this view, the species appears to be "the best bridge principle between evolutionary processes and patterns of descent". Lastly, the Phylogenetic Species Concept (Cracraft, 1983) proposed the clade as the smallest diagnosable cluster".

All these species definitions have respective advantages and limitations, but the evolutionary concept appears to be more compatible with various models of speciation.

In all the cases, the main condition for speciation to occur is the reproductive isolation between the different groups of populations composing the respective species. Several prezygotic and postzygotic barriers have been proposed (review in Avise, 1994).

Another important point concerns the origins of the species. Several modes of speciation are now accepted and mainly differ in the number of taxa produced. The phyletic evolution (or anagenesis) proceeds by a directional change within a single lineage, over a long period of time and without multiplication. The fusion of species may occur when isolating mechanisms are completely broken, resulting in hybrid swarms or populations (i.e., populations consisting entirely of hybrids). More important is the cladogenesis or branching of the phylogenetic tree by speciation. This last phenomenon results in a multiplication of the species and may be subdivided under

two headings, namely gradual and abrupt speciation. This separation is only convenient and the respective processes mentioned may rely on similar processes such as hybridization or introgression.

The gradual speciation involves a progressive genetic divergence between populations until reproductive isolation occurs, generally with or without geographic isolation (allopatry to sympatry).

On the other hand, abrupt speciation or saltation occurs via chromosomal changes such as polyploidy or chromosome rearrangements, producing offspring isolation (Grant, 1981).

Speciation by hybridization and introgression

Hybridization process is relatively common in plants but gamete fusion of divergent species generally fail to produce fertile offspring due to the lack of homology between different genomes, resulting in an insufficient pairing of chromosomes at the meiosis. Nevertheless, a small percentage of unreduced eggs or pollen may be sometimes produced, together with unbalanced haploid meiotic products. Moreover, alternative processes may restore hybrid fertility and further produce speciation. Polyploidy and chromosome rearrangements can produce sudden speciation when combined with hybridization process, whereas allopatric populations may introgress in hybrid zones and produce gradual speciation (Grant, 1981).

In the context of allopolyploidy process, hybridization between two taxa can represent a process of speciation (Thompson & Lumaret, 1992). In such well-documented pattern of abrupt speciation (Levin, 1993), reproductive isolation arises as an instantaneous by-product of hybrid genome duplication. On the other hand, fertile and stabilized hybrid may be produced, without changes in chromosome number. Such homoploid speciation, unlike allopolyploidy, is not instantaneous, and several models have been proposed to test hybrid segregant success (Rieseberg *et al.*, 2000).

Along with hybridization, the role of introgression or interspecific gene exchanges, in plant evolution has been extensively debated (Stebbins, 1950; Grant, 1981; Rieseberg and Soltis, 1991). Whereas different types have been described (localized and dispersed introgressions, ancient hybridization) and unequivocally detected by several molecular markers (Arnold, 1992, 1997; Liston and Kadereit, 1994; Hardig *et al.*, 2000), the frequency of occurrence, extent, and evolutionary significance of introgression is still unclear (Rieseberg and Soltis, 1991; Abbott, 1992; Rieseberg and Wendel, 1993; Lowe and Abbott, 2000). The difficulty to document natural hybridization or introgression lies on the fact that there are multiple explanations for morphological intermediacy, molecular additivity or phylogenetic incongruence (Rieseberg *et al.*, 2000).

Polyploidy and abrupt speciation

Polyploidy (i.e. the presence of more than two genomes per cell) is a speciation process widely active in the plant kingdom evolution. Estimations of plant species that have experienced at least one polyploidization event vary from 50% to 70% (Averett, 1980; Grant, 1981). Hence, polyploidy is now considered as one of the major evolutionary trends in plant speciation (Thompson and Lumaret, 1992; Soltis and Soltis, 1993; Bretagnolle and Thompson, 1995; Ramsey and Schemske, 1998; Petit *et al.*, 1999).

Several processes result in the emergence of polyploids depending on the degree of genetic differentiation between parental species (Bretagnolle and Thompson, 1995; Ramsey and Schemske, 1998). First, polyploids may results from hybridization mechanisms between two genetically distinct diploid taxa. This fact characterizes allopolyploidy and involves basically a hybridization process in addition to a polyploidization event. Second, polyploids may result within a single diploid species, without hybridization event, and are called autopolyploids (Stebbins, 1947; Lewis, 1980). In many cases, distinction between allo- and autopolyploids is not so clear, and mainly depends of the degree of genome homology. Indeed, these two phenomena represent ends of a continuum between duplicate sets of genetically homologous and non-homologous chromosomes. Segmental allopolyploidy has been defined to describe homeologous duplicated genomes and appears to be indistinguishable with the concept of autopolyploidy. Moreover, the recent concept of recurrent and multiple origins of many polyploid species (Doyle *et al.*, 1999; Soltis and Soltis, 1999, 2000) tend to support a polyphyletic origin for most of them and argue for a weak occurrence of autopolyploids.

Allopatry (geographic speciation) and gradual speciation

Contrarily to the rapid and abrupt process of speciation via hybridization and polyploidy, other mechanisms may invoke gradual genetic divergence between species population until reproductive isolation occurs. Several modes of speciation have been proposed, depending on the degree of geographic isolation. On one extreme, the absence of physical barriers with genetic divergence and reproductive isolation in the same parental populations, or in contiguous but slightly different environments, characterizes the processes of sympatric and parapatric speciation, respectively. On the other extreme, the full spatial isolation between populations of the same ancestral species may progressively create isolating mechanisms and produce new species. This pattern is known as allopatric or geographic speciation (Mayr, 1963). A single population is first divided in two or several gene pools by geological

or ecological processes, such as mountain uplifts, glacier formations or vegetation changes. The selective pressures then become different in the respective new areas, increasing the continuous evolutionary divergence and resulting in genotype / phenotype differentiation between each population. Finally, the establishment of complete isolating mechanisms occurs, and the resultant species may either evolve independently or come again into contact after geographic barriers fall. In the last case, character displacement accentuating differences between species in areas of sympatry, or competitive exclusion with extinction or extirpation of one species, may occur. Alternatively, when reproductive and geographic isolation are incomplete, the new species may come into secondary contact with each other along a common border and produce a hybrid zone. In such active sites of evolutionary changes, hybridization and introgression patterns can (1) increase genetic diversity within species, (2) induce the breakdown or the reinforcement of reproductive barriers, and (3) lead to the emergence of a new species (Abbott, 1992; Arnold, 1992; Rieseberg, 2000). Particular analyses of diploid - polyploid contact zones have allowed a better understanding of the dynamic and evolution of polyploid complexes (Petit *et al.*, 1999).

Historical biogeography and phylogeography: genes through space and time

The availability of DNA sequences data gave good opportunities to assess the biogeographic repartition of different populations across a species range (Avise, 1979). Therefore, it was conceivable to describe gene repartition in space, but also in time because mutational sequence changes accumulate over time, following or not a molecular clock. This spatio-temporal scale was the base of a new biogeographic discipline, namely the phylogeography that may be defined as "the principles and processes governing the geographic distributions of genealogical lineages" (Avise *et al.*, 1987). Phylogeography may be considered as a critical juncture among macroevolutionary disciplines comprising historical biogeography or paleontology, and population genetics (Avise, 2000). This discipline is of particular interest to detect present geographic subdivision, premating barriers or hybrid zones within or between species and to understand range changes through time (Hewitt, 1988).

One of the most fascinating consequence of phylogeographic studies, in parallel with the advances in paleoclimatology, is the actual comprehension of the Quaternary ice ages history in Europe and North America, and its repercussion on the current genetic diversity and species repartition (Hewitt, 2000). From 2.4 mya to until 0.9 mya, the ice sheets advanced and receded following periodic rhythms though to be in phase with the Milankowitch cycles (Hays *et al.* 1976). These severe climatic oscillations produced great changes in species distribution, some of them being extinct to a large part of their range, others dispersed to new places or surviving in refugia and then

expanding again. In these refugia, genomes may have diverged by repeated allopatry, protected by hybrid zones, and speciation may have occurred over many ice ages (Hewitt, 1996). The multiple episodes of genetic differentiation in refugia and the loss of alleles and homozygosity in the re-colonizant populations are thought to have played a major role in current genetic diversity of species occurring in Europe or in North America (Hewitt, 2000). Despite their similar size and latitude, these regions differ significantly in geographical features and geological barrier orientation. Nevertheless, most plant or animal species in the Pacific Northwest show a "southern richness and a northern purity", and most of the phylogeographic studies conclude in patterns similarity between northwestern regions of North America and Europe (Hewitt, 1996).

The genus Centaurium Hill: a useful model to appraise trends in plant evolution

Occurrence of polyploid complexes

The genus *Centaurium* Hill (Gentianaceae) appears to be a useful model to appraise the process and pattern of polyploid species evolution in Europe or in North America. In the first area, some complexes comprising diploid and tetraploid (sometime hexaploid) taxa have been described and their distribution is well known (Zeltner, 1970). For the New World species, no diploid were hitherto described and the occurrence of high and mature polyploid complexes has been proposed (Broome, 1973). The present repartition of different kind of polyploid complexes, in areas where glaciations are known to have drastically modulated plant distribution, may be the consequence of different patterns in glacier advances and retreats. On the other hand, older paleoclimatic or geological events may have contributed to the current genus repartition.

Occurrence of hybridization or introgression patterns

In *Centaurium*, many variable forms are supposed to be of hybrid or introgressive origin (Melderis, 1931; Melderis, 1972), but only a few case cases have been documented, on the base of intermediate characters (Ubsdell, 1976a, b, 1979; Zeltner, 1978). In the British Isles, *C. intermedium* appears to be an allohexaploid taxon resulting from a cross between two ecological-isolated parents, namely *C. erythraea* subsp. *erythraea* and *C. littorale* subsp. *littorale* (both tetraploid with 2n = 4x = 40). The low fertility of F1' hybrids acts as a post zygotic isolating mechanisms and hybridization does not normally extend beyond the F1 generation. Nevertheless, in some sympatric populations caused by man's habitat perturbation, backcrosses with *C. erythraea* may occur resulting in the production *C. intermedium* (Ubsdell, 1979).

Another striking case has been reported in Majorca (Spain) where diploid species *C. maritimum* (yellow corollas) and *C. tenuiflorum* subsp. *acutiflorum* (pink corollas) normally occur. In some places, one can encounter tetraploid taxa with salmon, pink or yellow corollas, suggesting hybridization with Mendelian inherited character (Zeltner, 1978).

Such case of putative allopolyploidization needs to be confirmed with molecular studies based on DNA markers of either maternal (cpDNA) or biparental (nrDNA) inheritance. On the other hand, some questions concerning the evolutionary trends encountered in *Centaurium* need to be addressed. For example, what is the importance of allopolyploidy and autopolyploidy phenomena in *Centaurium*? Is there some pattern of repartition and evolution of the polyploid complexes? Does homoploid speciation occur? Finally, how frequent, if detectable, is introgression?

The genus *Centaurium* Hill: current knowledge and interpretations

Generic circumscription, systematic position and infrageneric classification

The genus *Centaurium* Hill is a moderate sized group of the Gentianaceae family, with presently 50-60 species recorded. This genus, commonly called centaury, comprises short-lived herbaceous species, mostly annuals and biennials (rarely perennials), glabrous or sometime papillose on leaves and calyx margins. Inflorescence are represented by lax to compact cymes, with simple or compound, dichasial (or monochasial) branching, sometimes becoming uniflorate by reduction. These cymes can be either racemose, paniculate or corymbiform. The flowers, pentamerous (often tetramerous in species with short corollas), are either pedicellate or sessile. The main character for the genus delimitation is the complete coiling of the anthers after pollen release: it goes from an only half-twisted to up to four gyres, depending of the stamen length. (This character occurs to a lesser extent in other genera such as *Orphium* or *Chironia*). The style (1 to 25 mm) is generally undivided or slightly bifid, and many stigmata shapes can be encountered, such as ligulate, rhomboid, cuneate, reniform to flabelliform or subcapitate. The taxonomic importance of this character has been suggested (Grisebach, 1839; Broome, 1973) and needs further investigations. Capsules are narrowly cylindrical to oblate and may contain 50 to 700 minute seeds, (light) dark brown to black, borne on parietal, intruding placenta.

The name *Centaurium*, as a taxonomic entity, appeared for the first time during the 16th century (Dodoens, 1583) and was cited by Bauhin (1623), Tournefort (1700) or Ray (1724). Nevertheless, the term *Erythraea* proposed by

Borckhausen (1796) has been widely used till the middle 20th century, before its logical illegitimate, following the priority rule (Robyns, 1954). Indeed, the genus was first described by Hill in 1752 and typified later (Gillett 1963), on the base of *C. littorale* (= *C. vulgare*).

The exact number of species in the genus is difficult to estimate because of penury of sufficient taxonomic characters available to clearly separate the taxa. A high clinal variation in gross morphology, often depending of the ecological conditions, is frequently observed. In addition, natural hybridization has sometimes been reported and tends to obscure taxon delimitation (Ubsdell, 1976, a, b, 1979; Zeltner, 1970, 1978). This implies difficulties in the definition of the species, abundance of confusing synonyms and nomenclatural divergences.

First studies mainly underlie the difficulty of the species circumscription due to a high number of transition forms (Hegi, 1927; Lemee, 1931; Melderis, 1931; de Litardiere, 1948 and Robyns, 1954). Grisebach (1839), in the first monograph treating the whole genus, recognized 18 species of *Erythraea*. Gilg (1895), in his synopsis of the Gentian family numbered 29 species of *Centaurium*. Others rough estimations have been given in some floras (Standley and Williams, 1969), but no complete study covering the whole genus diversity has been yet undertaken. Consequently, the systematic of the genus is presently unresolved or imprecise in many parts.

Grisebach (1839) initially subdivided the genus for *Erythraea* in four sections (*Euerythraea*, *Trichostylus*, *Spicaria* and *Xanthaea*), and further considered a neighbor genus *Gyrandra*, comprising two Mexican species. Gray (1878) lowered *Gyrandra* to sectional rank, whereas Gilg (1895) included it in *Euerythraea*. Latter, Ronniger (1916) clustered European centauries in six sections (*Parviflorae*, *Centauria*, *Linariaefoliae*, *Caespitosae*, *Spicaria* and *Xanthaea*). Melderis (1931) was the first to create subdivisions in section *Euerythraea*. Presently, the Eurasian species are grouped in four main sections (Table 1).

On the other hand, American sections are not yet delimited. Section *Trichostylus* was first maintained for some American taxa (Melderis, 1931) while Broome (1973) included most of the North-American species in the section *Centaurium* and conserved section *Gyrandra* for five Mexican centauries (*C. chironioides*, *C. tenuifolium*, *C. madrense*, *C. pauciflorum* and *C. brachycalyx*).

Section	Subsection	Zeltner's treatment (1970, 1985)	Melderis' treatment (1931, 1972)	2n
Spicaria Griseb		*C. spicatum* L.	*C. spicatum* L.	22
Xanthaea Reichb.		*C. maritimum* (L.) Fritch	*C. maritimum* (L.) Fritch	20
Caespitosa Ronn.		*C. scilloides* (L.) Druce	*C. scilloides* (L.) Druce	20
Centaurium	*Parviflora* (Ronn.) Melderis	*C. tenuiflorum* (Hoffm. & Link) Fritsch	C. *tenuiflorum* (Hoffm. & Link) Fritsch	40
		---subsp. *tenuiflorum*	---subsp. *tenuiflorum*	
		---subsp. *acutiflorum* (Schott.) Zeltner	---subsp. *acutiflorum* (Schott.) Zeltner	20
		C. pulchellum (Swartz) Druce	*C. pulchellum* (Swartz) Druce	36
		C. mairei Zeltner	**	54
	Vulgaria Melderis	*C. barrelieri* (Duf.) F. Q. & Rothm.	*C. linariifolium* (Lam.) G. Beck	20
		C. gypsicola (B. & R.) Ronn.	*C. triphyllum* (Schmidt) Melderis	20
		C. favargeri Zeltner	*C. favargeri* Zeltner	20
		C. chloodes (Brot.) Samp.	*C. chloodes* (Brot.) Samp.	40
		C. vulgare Rafn.	*C. littorale* (Turner) Gilmour	40
		---subsp. *vulgare*	---subsp. *littorale*	40
		---subsp. *uliginosum* (W. & K.) Soo	---subsp. uliginosum (W. & K.) Melderis	40
	Centaurium	*C. majus* (Hoffm. & Link) Zeltner	*C. erythraea* Rafn.	
		---subsp. *rhodense* (B. & R.) Zeltner	---subsp. *rhodense* (B. & R.) Melderis	40
		---subsp. *majus*	---subsp. *majus* (Hoffm. & Link) Melderis	
		-------var. *majus*		20
		---subsp. *majus*		
		-------var. *suffruticosum* (Salzm.) Zeltner	*C. suffruticosum* (Griseb.) Ronn	20
		C. minus Gars.	*C. erythraea* Rafn.	
		---subsp. *minus*	---subsp. *erythraea*	
		------var. *minus*		40
		------var. *capitatum*		
		---subsp. *turcicum* (Velen.) Zeltner	---subsp. *turcicum* (Velen.) Melderis	60
		---subsp. *rumelicum* (Velen.) Zeltner	---subsp. *rumelicum* (Velen.) Melderis	20
		---subsp. *bernardii* (Maire & Sauv.) Zeltner	**	20

Table 1. Sectional classification of the European species of *Centaurium* following Melderis (1931, 1972) and Zeltner (1970, 1985).

Geographic repartition and ecology

The genus *Centaurium* is of temperate distribution and mainly occurs in the Old World and New World regions of the Northern Hemisphere.

In Eurasia, most of the species are centered in the Mediterranean basin, and diploid taxa generally occupy restricted areas (e. g. *C. maritimum*, *C. favargeri* or *C. majus*). Polyploids mainly irradiate to northern or eastern parts of Europe (*C. erythraea*, *C. littorale*) and may be encountered in India (*C. centaurioides* subsp. *centaurioides*). Only a few are confined to the Atlantic coasts (*C. scilloides* and *C. chloodes*). The *C. pulchellum* complex (including tetraploid and hexaploid taxa) possesses the widest distribution, ranging from the Finnish coasts to Northern Africa and throughout Asia to the eastern part of China.

New World taxa occur in the western part of United States (mainly in California and Texas) and in Mexico. Around 40 taxa, all polyploids, have been described in this area. Few species appear to be native from the Southern Hemisphere such as *C. cachanlahuen*, *C. lomae* or *C. ameghinoi* present in South America. Other species, morphologically close to the Mediterranean taxon *C. spicatum*, occur in Hawaiian Islands (*C. sebaeoides*) or in Australia (*C. clementii*).

Centaurium species may be encountered in various habitats, e.g. along roadside, on stream banks, in fields and pastures, in open forests or in oasis. In these respective stations, two ecological factors such as the degree of moisture and the level of insolation appear to be necessary for *Centaurium* germination and growth. *Centauries* are dominantly found in plains, at level field, except in Mexico, where most of the species inhabit the pine-oak woodland communities, from 1500 to 3000 meters above sea level. In California, *C. venustum* subsp. *abramsii* may also be found in the alpine meadows of the Sierra Nevada Mountains whereas subsp. *venustum* generally occurs near the sea.

Except some punctual detailed studies (Freijsen, 1967), a little is presently known concerning the ecology of *Centaurium*. Most of the species are calcicolous, some are acidiphilous (*C. scilloides*) and a few are restricted to gypseous soils (*C. gypsicola* and *C. maryannum*). Nevertheless, edaphic conditions appear to be of little importance compared to the moisture and insolation of the station. For example, *Centaurium exaltatum* occurs on alkaline soil but is high calcareous tolerant; meanwhile, it does not support shade and never occurs in dry localities. Lastly, another important parameter for *Centaurium* presence in a particular place seems to be the relative lack of competition with other species. Therefore, many centauries may be encountered in disturbed and open places, and sometime on volcanic or clay soils such as the rare *C. sebaeoides* (Medeiros, *et al.* 2000).

The Eurasian species of Centaurium: description and main characteristics

Presently, fourteen species are described in Flora Europaea (Melderis, 1972). Some of them present well-marked morphological characters and are relatively easy to identify. *Centaurium maritimum* is immediately recognizable with its yellow corolla and bifid style, and usually occurs near the sea. The variety *Shuttleworthianum* slightly differs from the type by the basal insertion of the stamen (vs. insertion in the upper part of the tube). *Centaurium scilloides*, with oval leaves on many decumbent and non-flowering shoots, is neighbor to subsection Vulgaria grouping taxa with relatively showy flowers and linear leaves (Zeltner, 1970). The tetraploid species *C. chloodes* and *C. littorale* (subsp. *littorale* and subsp. *uliginosum*) mainly differ by the caespitose aspect and are thought to derive from a diploid set comprising *C. favargeri*, *C. gypsicola* (= *C. triphyllum*) and *C. barrelieri* (= *C. linariifolium*). Doubtful species such as *C. rigualii* or C. *barrelieroides*, collected from only one or few location, are presently treated in synonymy with *C. barrelieri* (= *C. linariifolium* or *C. quadrangularis* subsp *barrelieri*) (Gonzalez, 1979; Bayer and Gonzalez, 1991).

Centaurium spicatum possesses spiciform-cymed inflorescence and is frequent around the Mediterranean basin and in Pacific Islands (Japan, Australia, Hawaii..), as well. Some interesting varieties have been described in Portugal and in Australia (Fernandes, 1965; Adams, 1996).

The remaining European species mainly belong to the *C. erythraea* complex comprising numerous taxa, which name and systematic rank diverge according to the different authors. Two main rough groups can be described, depending on the flower size. The tetraploid *C. erythraea* s. s., along with diploid subspecies *rumelicum* and *bernardii* (not described in Flora Europeae) and the hexaploid *C. erythraea* subsp. *turcicum*, generally possess small flowers (i.e. less than 1 cm long). On the other hand, the *C. majus* group (subsp. *majus* and subsp. *rhodense*) comprises taxa with large corollas (diameter > 1.5 cm). Nevertheless, many intermediate forms have been described and a clinal variation undoubtedly exists between these respective sets.

The *C. pulchellum* group comprises widespread taxa dichotomously branched from below the middle of the stem and having different chromosome valence (from diploid to hexaploid one). *Centaurium tenuiflorum* is morphologically close to this group and the diploid cytotype has sometimes been considered as a putative genitor of *C. pulchellum* (Zeltner, 1970).

Lastly, some Eurasian species are not described in Flora Europeae, either because the do not belong to the range covered the study (*C. centaurioides* subsp *centaurioides* and subsp. *malzacianum* - *C. pulchellum* var. *altaicum*) or for unclear reasons (*C. erythraea* subsp. *bernardii*, *C. serpentinicola* or *C. somedanum*).

The North American species of Centaurium: description and main characteristics

North American centauries comprise about 30 taxa located in three main geographic areas. In California, one can encounter two different sets of species based on the flower size (Munz and Keck, 1959). A first cluster comprises relatively big-flowered taxa (diameter > 1.5 cm), such as *C. venustum* with 2 subspecies described from the north of the Sierra Nevada Mountains to Baja-California, *C. trichanthum*, growing preferably in coastal and salt-marsh habitats of the Northern California, and *C. namophilum* var. *namophilum*, restricted to the Death Valley region (Reveal *et al.*, 1974).

Other species mainly have small flowers. *Centaurium muehlenbergii* s.l. (known under different synonyms such as *C. floribundum*, *C. curvistamineum* or *C. tenuiflorum*) is supposed to be native to California and is confined to Western North America. The desert Centaury (*C. exaltatum* or *C. namophilum* var. *nevadense)* is characterized by long pedicels and peduncles and extends from Baja-California to western USA and Canada. One close related species, *C. davyi* (*C. exaltatum* subsp. *davyi*) appears to be endemic to the coastal range of California.

Texan species are showy and often form important populations along roadsides. *Centaurium calycosum* s.l. is very abundant in the center of the state, *C. texense* in the eastern part and *C. beyrichii* in the North. Some taxa are restricted in very limited area (e. g. *C. breviflorum*, *C. glanduliferum* or *C. maryannum*) while *C. arizonicum*, encountered from Arizona to Chihuahua region, is morphologically close to *C. multicaule.* A last species, *C. nudicaule* reaches the northern limits of its repartition in Texas (Guadalupe Mountains, Hudspeth Co).

Lastly, 16 annual species of *Centaurium* are native to Mexico (Broome, 1973). They inhabit mostly pine-oak woodland community, which occurs from the sea level to more than 2500 m.

Within section *Gyrandra*, all the five species but *C. brachycalyx* show large flowers (diameter: 1.5 - 2 cm). *Centaurium gentryi*, known from only one population, can be added to this set (Broome, 1977).

The remaining species generally possess small flowers. Two Mexican endemic *C. strictum* and *C. wigginsii* bear similarities in vegetative characters (pattern of branching, leaf shape). *Centaurium quitense*, widespread in Central America, is morphologically variable but may be easily recognized by the openness of flowers branching. Close related taxa appear to be *C. pusillum* and *C. nudicaule*, a taxon characterized by a great phenotypic variation. Relationships have been also suggested between *C. setaceum* and *C. martinii* (= *C. phenax*), based on chromosomal grounds (Broome, 1978).

Finally, *C. capense* appears to be endemic of the Cape region of Baja California, but has been often confounded with *C. exaltatum* or *C. nudicaule*, two small-flowered sympatric species. Therefore, the repartition of *C. capense* is not really well established and this taxon may be more frequent than hitherto supposed.

Finally, a little is known concerning the South American species such as *C. ameghinoi*, *C. cachanlahuen*, or *C. lomae* whereas some European native species, namely *C. erythraea*, *C. pulchellum* or *C. spicatum* are often present on the shores of the American continent.

Morphology

Whereas many morphological studies have been proposed for some major groups of Gentianaceae such as *Gentiana* (e, g., Miege and Wüest, 1984; Ho and Liu, 1990; Yuan, 1993; Ho *et al.*, 1996) or Lisianthoid species (Struwe *et al.*, 1997; Struwe and Albert, 1998a, b; Systma, 1988), a few if not treatments have been done on *Centaurium*. For example, early works dealt with anatomy of some particular species (Perrot, 1898) or with morphogenesis of inflorescence and flowers (Robyns, 1954). A phenetic approach has been undertaken on some North American species of *Centaurium* but no systematics conclusions emerged of this study (Dunn, 1967).

One explanation of the lack of taxonomic treatment in *Centaurium* may be that taxon delimitation is always difficult due to a frequent clinal morphological variation, resulting from ecological conditions or possible introgression. On the other hand, the profusion of synonyms and disputed binomials, along with the description of many dubious taxa, is a priori discouraging.

Cytotaxonomy

The soar of karyological and cytological techniques was of first importance in the systematic comprehension of the genus. First approximate chromosome numbers were established on Eurasian species (Wulff, 1937; Rork, 1949; de Mesquita Rodrigues, 1953; Khoshoo and Khushu, 1966). Later, Zeltner gave one of the most valuable contributions towards the understanding of relationships between European species (Zeltner, 1963, 1970, 1978, 1980, 1985, 1987, 1990 and 1991).

The New World taxon set received poor attention and was mainly investigated by Broome (1976). These previous chromosome data showed a great range of numbers.

Several counts, including 2n = 18, 20, 22, 34, 36, 40, 42, 44, 54, 56, 60, 72, 74, 80, 82 and 84, have been reported in the genus (Figure 1). Some of them seem to be restricted to well-delimited groups while other assemblages show greater variation. In the first case, one can mention section *Xanthaea* (2n = 20) or *Spicaria* (2n = 22). Among the variable taxa, some show different ploidy levels, e. g. the *C. pulchellum* complex (2n = 18 / 36 or 54) or *C. erythraea* sensu lato (2n = 20 / 40 / 60), whereas others appear to be more heterogeneous as shown with the Texan assemblage, with n = 40, 42, 82 and 84 (Broome, 1976).

Generally, x = 10 was suggested to be a primary basic number in the genus from which derived x = 9 and x = 11. So dysploidy was thought to be one of the early trends in *Centaurium* evolution (Zeltner, 1970). Polyploidization then occurs in separate lineage giving tetraploid to hexaploid taxa in Eurasia and even higher polyploids (octoploids) in North America (Broome, 1973).

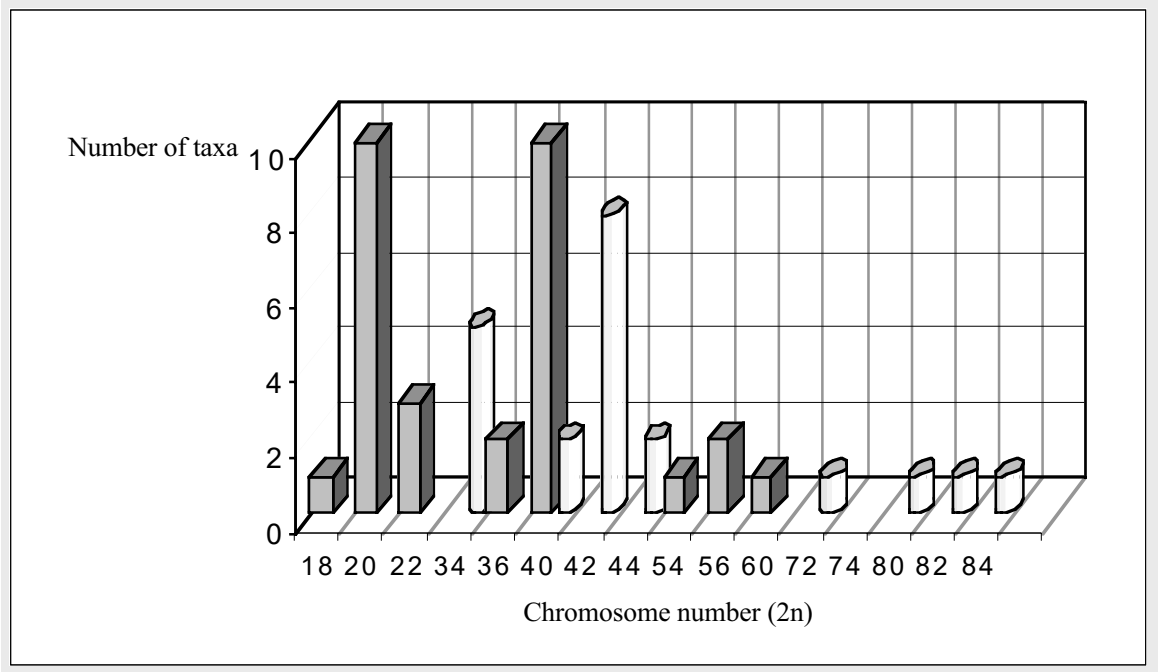


Figure 1. Number of taxa (literature data) per chromosome number in Old World (grey) and New World (white) species of *Centaurium*.

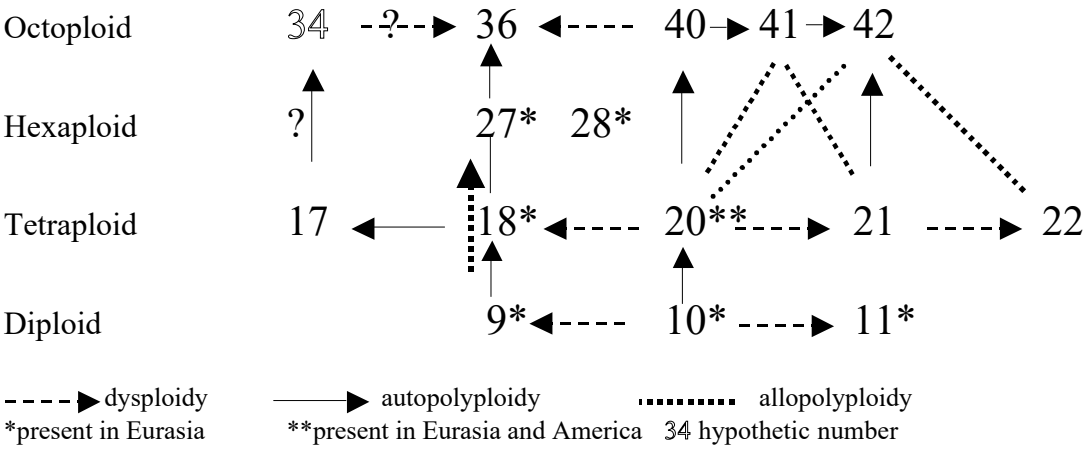


Figure 2. Hypothetical chromosome evolution in *Centaurium*, modified from Broome (1973) to accommodate the Eurasian diversity of the genus (Zeltner, 1970 and after).

As shown in Figure 2, the knowledge of the chromosome number is of main importance in *Centaurium*, for most of the systematics conclusions hitherto proposed are based on this character (Zeltner, 1970, Broome, 1973). Nevertheless, if the chromosome number knowledge is a useful systematic tool for species determination or ploidy level evaluation, the global similarity between taxa, in the absence of other data, may be homoplasic. Thus, the proposal of new systematic combinations or taxa, on the base of karyological data alone is highly critical. In *Gentiana*, for example, several monobasic and non-natural genera (e. g. *Ciminalis*) or other combinations have been proposed (Löve and Löve, 1972; Omer and Qaiser, 1992).

In the light of previous hypotheses concerning karyological evolution in *Centaurium*, the following questions need to be faced in the present investigation: (1) what is the range of chromosome number variation within the genus? (2) are diploid taxa definitively absent on the American continent? (3) does the striking difference reported between New World and Old World *Centaurium* chromosome numbers reflect a common evolutionary history or independent scenario? (4) is x = 10 the basic ancestral number in *Centaurium?* (5) to what extent are basic numbers useful for taxonomic delimitation? (6) are there polarities in chromosome number variation? and, (7) what are the main patterns of chromosome evolution within the genus?

These questions require accurate investigations of the groups comprising many divergence and variations in chromosome numbers, and obviously those hitherto poorly studied. Consequently, (1) the chromosomal diversity of the New World taxa need to be fully checked, (2) peripheral areas encompassing the Mediterranean basin are of great importance in the comprehension of patterns of chromosome evolution (polyploid complexes), and (3) Southern Hemisphere taxa have to be investigated in order to cover the whole extent of chromosome variation.

Phytochemistry

Most of the members of the Gentianaceae family are characterized by the occurrence, in their aerial parts and roots, of particularly secondary compounds, namely xanthones and secoiridoids (Box 1). The former give yellow dyes and have interesting pharmacological activities (mainly versus depression) while secoiridoids are responsible of the bitterness of many of the gentians, and are used for their digestive properties. More than 70 xanthones aglycones have been described in the Gentianaceae and some of them have been detected in seven other plant families. Polysaccharides of xanthones, and particularly xanthones-O-glycosides, are specific and characteristic of the Gentianaceae (Schaufelberger, 1986). Moreover, the concomitant presence of secoiridoids and xanthones appears to be synapomorphic in some groups of the Gentianaceae. Thus, these chemical compounds have been regarded as good taxonomic markers and may be valuable tools at intergeneric level (Massias *et al.*, 1982) or below (Hostettmann-Kaldas and Jacot-Guillarmod, 1978).

In the Chironieae, some species of *Centaurium* (e. g. *C. erythraea*, *C. pulchellum*, *C. spicatum* or *C. cachanlahuen*), *Canscora* (*C. decussata*) or *Chironia* (*C. baccifera*, *C. krebsyi*), have been widely investigated for their chemical constitution and medicinal activity (Van der Sluis and Labadie, 1981a, b and c; Wolfender and Hostettmann, 1992, 1993). Nevertheless, a few concern the taxonomic utility of these particular secondary metabolites in *Centaurium* (Lebreton and Dangy-Caye, 1973). The use of both xanthone and secoiridoids for

systematic purposes has been attempted on nine European and two American *Centaurium* species (Van der Sluis, 1985).

BOX 1. PHYTOCHEMICAL ASPECTS OF THE GENTIANACEAE / CHIRONIEAE

Chemical compounds found in the Gentianaceae--The main secondary metabolites found in the Gentianaceae can be roughly divided in 3 classes [(a)] (Wolfender, 1993).

* ***Polyphenols***: Anthocyanidins (5) / benzophenones (1) / Flavanes (3) / Flavanones (2) / Flavones (54) / Flavonols (7) / Xanthones (134) / Tetrahydroxyxanthones (1)

* ***Terpenes***: Monoterpenes (>130) / Secoiridoids (18) / Iridoids (8) pentacyclic triterpenes (9) / Sterols (7)

* **Other** Alkaloids (4 - may be artifacts) / Aryl-6-pyrones-2 (3) / Carotenoids (?) / Phtalids (1)

[(a)] The numerically more important classes are underlined, and the number of compounds presently known in the Gentianaceae are mentioned in parentheses.

Xanthones: structure and characteristics--Xanthones (or 9H-Xanthene-9-on) are pigments deriving from a dibenzo-γ-pyrone structure: they generally produce yellow dyes (from the Greek, *xanthos* = yellow). The molecules encountered in higher plants and ferns are probably biosynthesized via the shikimate pathway (Sultanbawa, 1980), whereas those found in fungi and lichens are of polyketide origin (Turner, 1971). The dibenzo-γ-pyrone skeleton can be esterified, producing C- or O-glycosides. Xanthones C-glycosides are widely encountered in higher plants and ferns, and mangiferin, a tetra-oxygenated xanthone-C-glycoside, is one of the best-known examples. On the other hand, most of xanthone-O-glycosides and xanthone aglycones found in Angiosperms have been isolated in the Gentianaceae. Nevertheless, they occur in seven other families as well. The Clusiaceae (Guttiferae) is one of the most important, with as many xanthones as in the Gentianaceae, but with different kind of substitution, such as isoprenoyls and geranoyls (Sordat, 1992). The same type can be found in the Moraceae (Sultanbawa, 1980). Within the Polygalaceae, ca. 40 xanthones have been isolated (Sultanbawa, 1980).

Chemotaxonomic importance-of the xanthone oxidation patterns- in the Gentianaceae---Only some positions can be substituted, by hydroxyl or methoxyl radicals, on the dibenzo-γ-pyrone structure, producing different oxidation patterns. In the Gentianaceae, xanthones are tri-, tetra-, penta-, and hexa-oxygeneted and more than 20 patterns have been described (Wolfender, 1993). These polyphenols are not distributed equally in the different genera of the family (Table 1.1) and may be view as potential taxonomic markers.

Pharmacological properties of xanthones-O-glycosides---Some xanthones isolated in the Gentianaceae (mainly the 1,3,5,8 aglycones) are known to have stimulant properties in the central nervous system (CNS) (Hostettmann and Wagner, 1977). These compounds may act as monoamine oxidase inhibitors (MAO A and sometimes of MAO B). This enzyme decreases the intersynaptic concentration of dopamine or serotonine and creates depressant effects. MAO A inhibition observed on ca. 16 xanthones of the Gentianaceae (Schaufelber and Hostettmann, 1988) increases intersynaptic concentration in nervous-mediators and thus products CNS stimulation.

Secoiridoids: structure and characteristics---Secoiridoids glycosides are terpenes derived biosynthetically from an iridoid glycoside (Inouye *et al.*, 1976). They have been classified into four groups (sweroside type, morroniside type, oleuropeine type, and alkaloidal glycosides). Only the sweroside and morroniside types occur in the Gentianaceae. The most important molecules of the first type are gentiopicroside (specific of the Gentianaceae), sweroside, swertiamarin or amaroswerine, one of the bitterest compounds hitherto known (it remains bitter after dilutions of 1 / 58.000.000). The sweroside type is frequent in the Gentianales (Gentianaceae, including *Anthocleista* previously placed in the Loganiaceae and Apocynaceae) and in the Menyanthaceae. The morroniside has been found only in *Gentiana thunbergii* (Inouye and Nakamura, 1971). The alkaloid type has been reported in the Loganiaceae and Apocynaceae and the oleuropein type in the Oleaceae (Jensen *et al.*, 1975)

SUBTRIBE	SPECIES	Xanthone 3	Xanthone 4	Xanthone 5	Xanthone 6	Xanthone C-glycosyl	Secoiridoid
Chironiinae	*Blackstonia perfoliata*	0	0	0	0	no	yes
	Centaurium cachanlahuen	0	1/3/5/8 1/3/7/8		1/2/3/4/6/8		
	C. chloodes	0	1/3/5/8 1/3/7/8		1/2/3/4/6/8		
	C. erythraea	0	1/2/3/5 1/3/4/5 1/3/5/8 1/3/7/8	1/3/4/6/8	1/2/3/4/6/8		yes
	C. linariifolium	0	1/3/5/6 1/3/7/8	1/3/4/6/8	1/2/3/4/6/8		
	C. littorale	0	1/3/5/8		1/2/3/4/6/8		yes
	C. majus	0			1/2/3/4/6/8		
	C. maritimum	0	0		0		0
	C. pulchellum	0	1/3/5/8 1/3/7/8		1/2/3/4/6/8		
	C. quitense	0			1/2/3/4/6/8		
	C. scilloides	0	1/3/5/8 1/3/7/8		1/2/3/4/6/8		
	C. spicatum	0	1/3/5/8		1/2/3/4/6/8		yes
	C. tenuiflorum	0	1/3/5/8 1/3/7/8		1/2/3/4/6/8	no	
	Chironia baccifera	1/3/5	1/3/7/8 1/3/5/8		1/3/5/6/7/8	no	yes
	C. krebsii	1/3/5 1/3/7	1/3/5/6 1/3/7/8	1/3/5/6/8	1/3/5/6/7/8	no	yes
	C. palustris	1/3/5 1/3/7	1/3/5/6 1/3/5/8 1/3/7/8	1/3/5/6/8	1/3/5/6/7/8	no	yes
	C. purpurascens	1/3/5	1/3/5/6 1/3/7/8	1/3/5/6/8	1/3/5/6/7/8	no	yes
	Cicendia filiformis	0	0	0	0	no	yes
	Eustoma grandiflora	1/3/5 1/3/7	1/3/7/8	1/3/4/7/8	1/2/3/4/6/8		
	Ixanthus viscosus			1/2/3/6/8			
	Orphium frutescens	yes	yes	yes	yes		
	Sabatia angularis	0	0	0	0	0	0
	S. elliotti	0	0	0	0	0	0
Canscorinae	*Canscora decussata*	1/3/5	1/3/5/6 1/3/7/8	1/2/3/6/8		mangiferin	
	C. decurrens		yes				
	C. diffusa		yes				
	Hoppea dichotoma	1/3/5	1/3/5/6	1/3/5/6/7		homomangiferin	
	Hoppea fastigiata		1/3/5/7				

Table 1.1 Repartition of xanthones and secoiridoids within the different genera of the tribe Chironieae (modified from Van der Sluis, 1985)

Preliminary phylogenies and intergeneric classification

Before the soar of the molecular phylogeny, taxonomists used to combine large sets of characters resulting from several approaches such as macro- and micromorphology, palynology, karyology or phytochemistry, in order to obtain natural classifications. Such traditional or phenetic approaches often create conflicts or controversies depending on the importance or weight one can accord to a particular character in a systematic treatment.

Rare attempts to provide phenetic or cladistic treatment of the genus *Centaurium* can be found in the literature. As an example, the combination of karyological and phytochemical data has been performed on a major part of the European species (Van der Sluis, 1985). Nevertheless, such reconstruction mainly relied on subjective assumptions and a-posteriori choice of selected characters (Figure 3).

Whereas the genus *Centaurium* received poor contributions, more important studies have been undertaken either on the subtribe Erythraeinae (presently, Chironiinae), or on the whole family. One of the goals of these approaches was to provide a new classification of the Gentianaceae. Indeed, the last major one, proposed a century ago (Gilg, 1895), reveals many artificial assemblages and therefore becomes unsatisfactory. One can mention here some approaches combining morphological (Figure 4) or phytochemical (Figure 5) characters (Meszaros, 1994; Meszaros *et al.*, 1996)

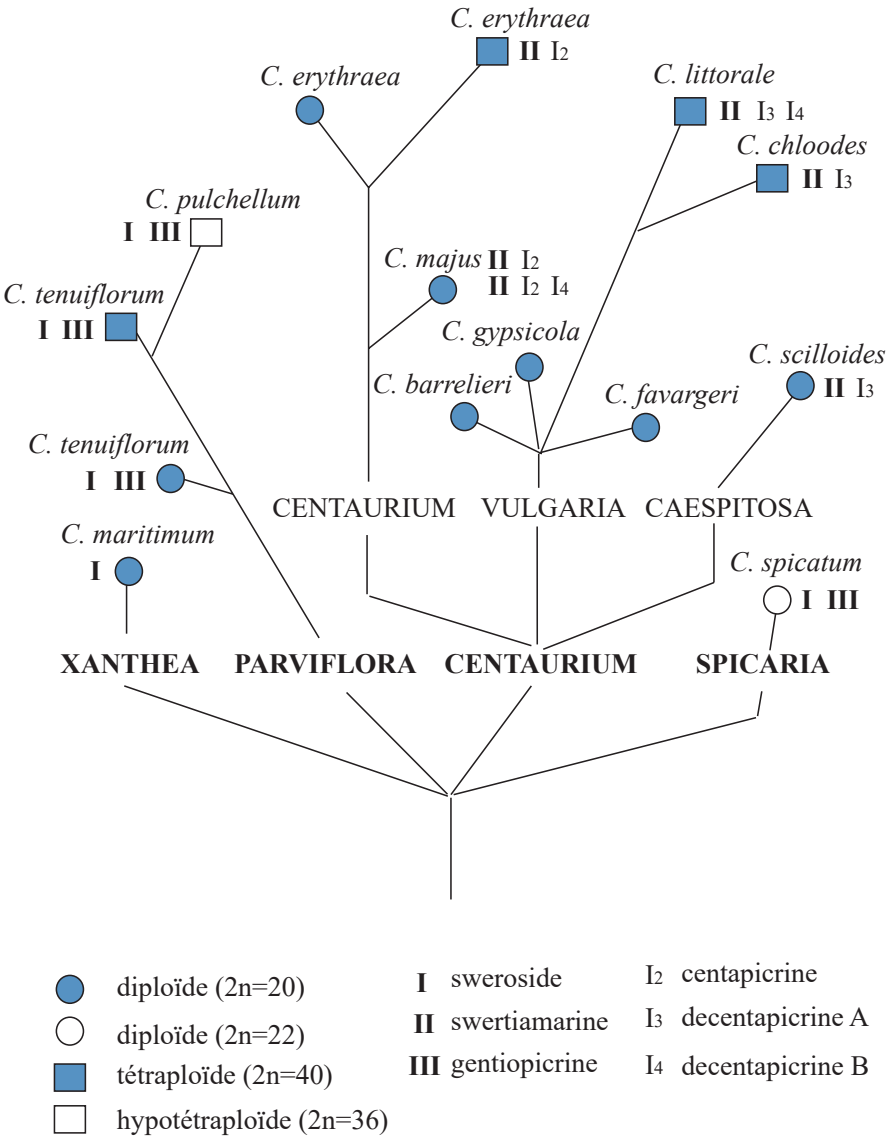


Figure 3. Mapping of secoiridoid presence on an hypothetical phylogenetic cladogram of some European *Centaurium* species, based on chromosome grounds (modified from Van der Sluis, 1985)

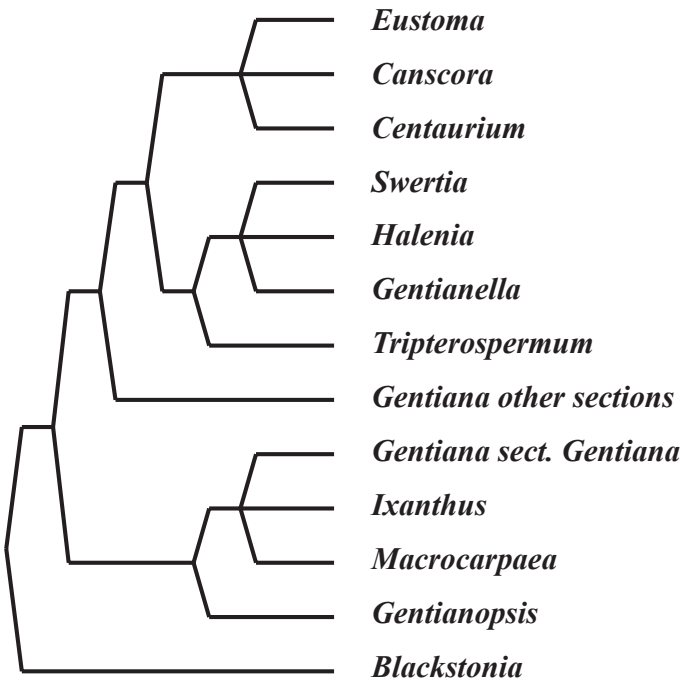


Figure 4. Phytochemical consensus cladogram of some representative genera of the Gentianaceae (modified from Meszaros, 1994)

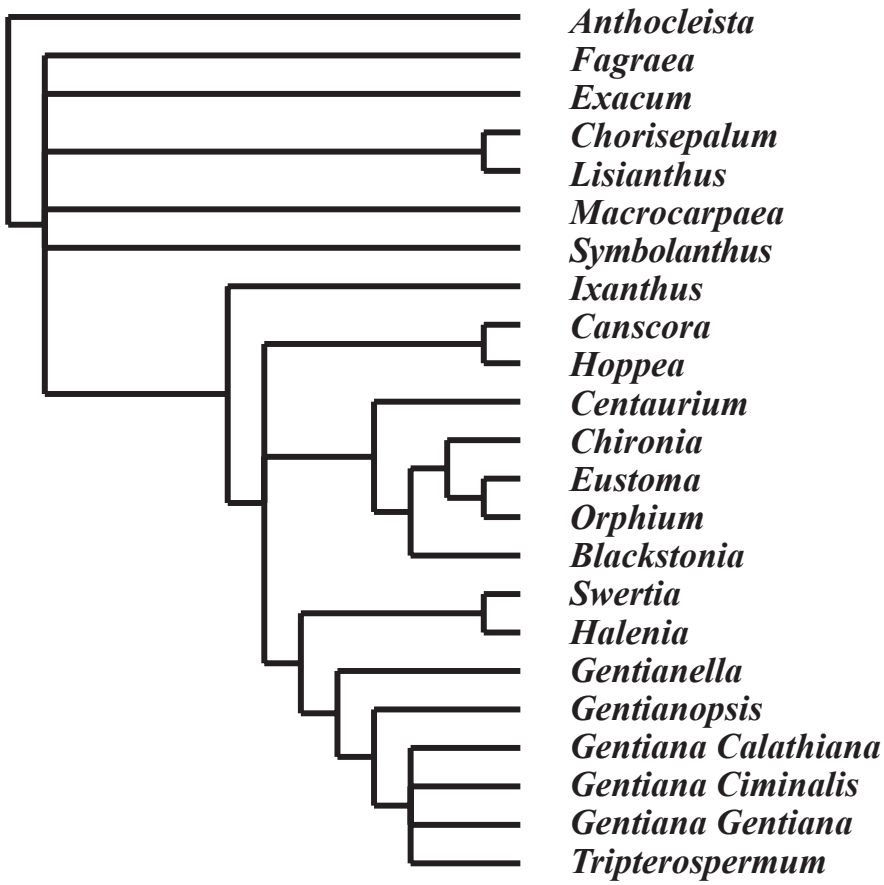


Figure 5. Morphological consensus cladogram of some representative genera of the Gentianaceae (modified from Meszaros *et al*., 1996)

Molecular markers: powerful tools to assess plant phylogeny

The last two decades have been the cradle of one of the most exciting and revolutionary progress in biology with the birth a new discipline, the molecular systematics. The application of nucleic acid data to resolve plants relationships has revolutionized the traditional classifications, bringing answers to hitherto unresolved questions, opening doors to non prospected and fascinating fields of investigations and, as a matter of fact, creating new problems. Presently, we know that molecular phylogenies, first considered as a panacea, do not necessary represent the true phylogeny of a lineage (i.e. the species tree). Inferred phylogenetic trees are only hypotheses dealing with the evolution of some genes in investigated species (the gene tree). Notwithstanding, this Pandora's box has brought a new and healthy breath in Systematics producing a plethora of studies and the first molecular phylogenetic classification in the plant kingdom (APG, 1998).

Which DNA for which purpose?

Two primary classes of DNA data have been used to assess phylogenetic reconstruction in plants: the nuclear ribosomal DNA (nrDNA) and the chloroplast DNA (cpDNA).

The nuclear genes that code for the rRNA are repeated thousand of times in plant genomes and may comprise as much as 10 % of the total plant DNA. Ribosomal DNA is arranged in tandem repeats in one or a few chromosomal loci.

Despite the important size of the nuclear genome, only a very restricted part has been used for phylogenetic purposes, such as particular spacers of ribosomal multigene family. Nowadays, the phylogenetic utility of the Internal Transcribed Spacer region at different taxonomic levels suffers no controversy. Hundred of studies based on ITS sequence data have been performed during the last decade and only few report the lack of phylogenetic utility for this marker. In the Gentianaceae, ITS appeared to be a useful tool at both intergeneric and interspecific levels (Yuan and Küpfer, 1995; Yuan *et al*., 1996; Hungerer and Kadereit, 1998; Thiv *et al*., 1999). On the other hand, sequences of the 5S Non Transcribed Spacer (5S-NTS), although of fast evolution rate and occurring in many tandem multiple copies in the genome, have received little attention (Udovic *et al.*, 1995; Ansari *et al.*, 1999; Crisp *et al.*, 1999; Ladiges *et al.*, 1999; Adams *et al.*, 2000; Baker *et al.*, 2000; Persson, 2000). The main reason for this relative lack of interest is the potential problem with paralogy. In other terms copy sequenced from one taxon may not be homologous with copy from another taxon. Within a locus,

concerted evolution homogenizes the repeated units only weakly and episodically, and such homogenization does not necessary occur between loci (Kellogs *et al.*, 1996).

BOX 2. Phylogenetic utility of the nrDNA genome

Organization of the different gene families---The nrDNA gene family comprises very conserved regions (i.e. 18S and 26S genes) as well as more rapidly evolving segments (IGS, ITS or 5S-NTS). These nrDNA cistrons are located in the NOR of several different chromosomes. The 5S-gene family is composed by the 5S gene (~120 bp) and the 5S non-transcribed spacer (5S-NTS: ~300-400 bp). The 18S-5.8S-25S region is a part of the 45S nrDNA repeat and consists of transcription units separated each other's by an intergenic spacer (IGS). The sequential arrangement of the transcription unit is the following: (1) the external transcribed spacer (ETS), (2) the 18S gene, (3) the internal transcribed spacer 1 (ITS1), (4) the 5.8 S gene, (5) the internal transcribed spacer 2 (ITS2) and (6) the 25S gene (Figure 2.1).

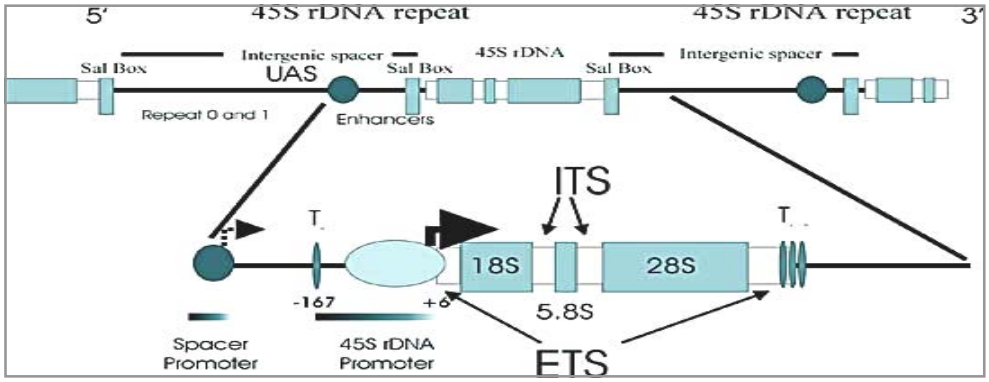


Figure 2.1. General schematic diagram of nrDNA repeats in plants (from: http://www.bioscience.org)

Function of the nrDNA---In plants, each ribosome is composed by a large (60S) and a small (40S) subunit containing RNA and proteins. Whereas these corpuscles occur in chloroplast, mitochondria, and cytoplasm, and may therefore be produced by three potential genomic sources, only nuclear ribosomal DNA has received notable attention. Each nuclear gene families encode for the rRNA molecules constituting the ribosome. The 28S, 5.8S and 5S genes code for rRNA and proteins forming the large subunit, whereas the 18 S gene is involved in the 40S subunit formation (Figure 2.2). ITS1 and ITS2 regions appear to play a role in the maturation of nuclear rRNAs, bringing the large and small subunits into close proximity into a processing model (Baldwin *et al*, 1995).

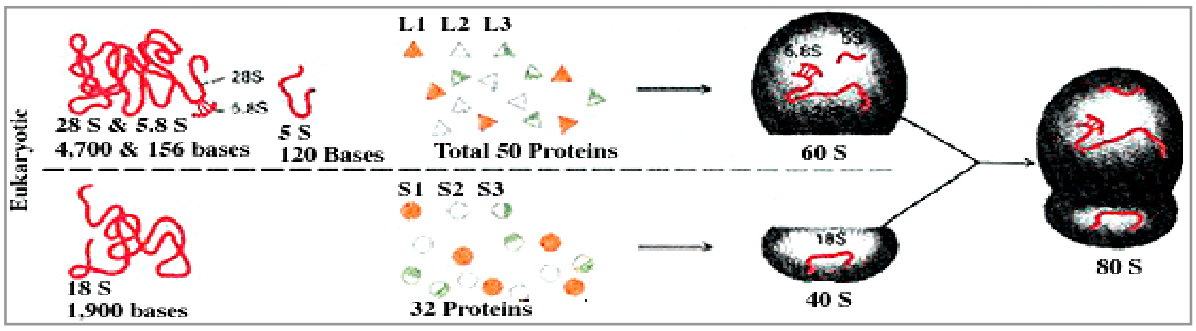


Figure 2.2. Constitution of plant ribosome (from:http://ntri.tamuk.edu)

***Concerted evolution in the nrDNA---**Within an organism, sequence homogeneity of cistrons is extremely important and has been referred to concerted evolution or inter-locus homogenization (Zimmer *et al.*, 1980). Different mechanisms have been proposed for this concerted evolution including unequal crossing-over or exchanges, transposition, gene conversion or slippage (Arnheim *et al.*, 1980).

***Phylogenetic utility of ITS and 5S-NTS sequences (Table 2.1)---**Since the last decade, sequences of the ITS region have been extensively used to infer phylogeny at several taxonomic levels in diverse organisms, including

angiosperms (e. g. Baldwin, 1992), algae (e. g. Bakker *et al.*, 1995), fungi (e. g.Vilgalys and Sun, 1994) or insects (e.g. Campbell *et al.*, 1993). Among the advantages of the ITS region for phylogenetic purposes, one can note the presence of conserved sequences pattern, a relative facility in amplification (due to small size, high copy number and applicability of universal primer) and a good amount of sequence divergence. On the other hand, ITS sequences and can be used to document reticulate evolution (Sang *et al.*, 1997). In allopolyploid species, for example, mechanisms of concerted evolution may be more or less complete, giving either heterozygote sites or entirely homogenized sequence of one parent.
The 5S non-transcribed spacer (5S-NTS) appears to be more variable than the ITS region. The 5S-NTS sequences have a good phylogenetic potential at interspecific and intergeneric levels. However, the lack of concerted evolution among repeat units and areas make this region difficult to isolate and to sequence. Successive PCR amplifications or cloning methods are often necessary.

Characteristics / region	**ITS1 + ITS2**	**5.8 S gene**	**5S-NTS**	**5S gene**
Length range (Angiosperms)	600-700 bp	160 bp	350-400 (or more) bp	
Rate of Concerted evolution	strong	*	weak	*
Taxonomic level of utility	Species / Genus	Order to Phylum	(Population)/Species/Genus	*
Paralogy	yes	*	yes	*
Possible contamination	Alga e / Fungi	*	Alga e / Fungi / Animal	*

Table 2.1. Main characteristics of the principal nrDNA regions used for phylogenetic reconstruction

Along with nuclear DNA, chloroplast DNA has been widely used in plant systematics (BOX 3). The literature abounds of publications treating different aspects of molecular structure, genomic organization, rates of molecular evolution and utility in phylogenetic reconstruction (Palmer, 1987, 1995; Clegg *et al.*, 1994; Olmstead and Palmer, 1994; Hilu and Liang, 1997). In phylogeny, DNA variation may be screened by two major approaches such as (1) the mapping of the entire genome using RFLP and AFLP methods or (2) direct sequencing of a particular region delimited by selected primers. The first approach covers a wide range of taxonomic levels from infraspecific (Schmidt and Jensen, 2000; Sunnucks, 2000), interspecific or higher (Wolff and Schaal, 1992; Okada *et al.*, 1997; Vijverberg *et al.*, 1999). On the other hand, cpDNA sequencing have been performed on many species and genera of seed plants and complete chloroplast genomes are presently available for some of them: 20 complete plant genomes have been sequenced in 2001. The different genes and non-coding fragments suitable for phylogenetic studies and their respective rate of evolution have been extensively reviewed (Olmstead and Palmer, 1994). While intergenic spacers are known to give higher resolution than coding genes at lower taxonomic levels, they usually evolve slower than ITS or 5S-NTS sequences data (Sanderson, 1998). Generally, cpDNA sequences fail to resolve low-taxonomic levels phylogenies. Nevertheless, a combined utilization with nrDNA regions often help to interpret incongruence in phylogenetic reconstruction, due to hybridization or chloroplast capture (Rieseberg and Soltis, 1991; Soltis and Kuzoff, 1995; (Sang, Crawford *et al.* 1997). On the other hand, the maternal inheritance of cpDNA fragments in Angiosperms, is useful to identify mother parent of putative hybrids (Wendel *et al.*, 1995).

BOX 3. Phylogenetic utility of the chloroplast genome

The use of chloroplast DNA (cpDNA) for phylogenetic reconstruction has been so important and extensive, with an exponential explosion in data accumulation, that this molecule may be viewed as a major tool in the rise of molecular systematics. A lot of reviews have been performed within the last decade (e. g., Downie and Palmer, 1987; Clegg *et al.*, 1994); therefore, only the main characteristics of cpDNA are presented.

Structure and organization of the cpDNA genome---The plant chloroplast genome is a circular molecule (120 - 200 kb), characterized by two inverted repeated segments that separate the remainder of the molecule into a small and a large single-copy region, SSC and LSC respectively. Complete DNA sequences are presently available for ca. 20 organisms (Table 3.1)

Organism	Phylum	Accessions	Reference
Arabidopsis thaliana	Embryophyta / Brassicaceae	AP000423	Sato *et al.*, 1999
Chlorella vulgaris	Chlorophyta	AB001684	Wakasugi *et al.*, 1996
Cyanidium caldarium	Rhodophyta	AF022186	Vogel *et al*., 1996
Cyanophora paradoxa	Glaucocystophyceae	U30821	Stirewalt *et al.*, 1995
Epifagus virginiana	Embryophyta / Fagaceae	M81884	Wolfe *et al.*, 1992
Euglena gracilis	Euglenozoa	X70810	Hallick *et al.*, 1993
Guillardia theta	Cryptophyta	AF041468	Douglas and Penny, 1999
Lotus japonicus	Embryophyta / Fabaceae	NC002694	Kato *et al*, 2001
Marchantia polymorpha	Streptophyta	X04465	Ohyama *et al.*, 1986
Mesostigma viride	Chlorophyta	AF166114	Lemieux *et al.*, 2000
Nephroselmis olivacea	Chlorophyta	NC000927	Turmel *et al.*, 1999
Nicotania tabacum	Embryophyta / Solanaceae	Z00044	Shinozaki *et al.*, 1986
Odontella sinensis	Stramenopiles	Z67753	Kowallik *et al.*, 1995
Oryza sativa	Embryophyta / Poaceae	X15901	Hiratsuka *et al.*, 1989
Pinus thunbergii	Embryophyta / Pinaceae	D17510	Wakasugi *et al.*, 1994
Porphyra purpurea	Rhodophyta	U38804	Reith and Munholland, 1994
Spinacia oleracea	Embryophyta / Chenopodiaceae	AJ400848	Schmitz-Linneweber *et al.*, 2001
Toxoplasma gondii	Alveolata	U87145	Kissinger *et al*., *
Zea mays	Embryophyta / Poaceae	X86563	Maier *et al.*, 1995

Table 3.1. Non-exhaustive list of complete chloroplast DNA genome sequences presently available (2001)

Properties of cpDNA in phylogenetic reconstruction---The advantages of cpDNA in evolutionary reconstruction studies include: (1) small sized molecule, making the entire genome relatively easy to examine; (2) quantitatively abundant and easy to amplify; (3) conservative in structure, most of the genes being essentially single-copy (Palmer, 1995); (4) relatively conservative in evolution with non-coding spacer and intron evolving at higher rates than coding genes (Downie and Palmer, 1992). The latter property may be regarded as either positive, with the wide range of taxonomic levels which can be screened with cpDNA variation (Soltis and Soltis, 1998), or negative with generally failure to resolve relationships at species level or below. A second disadvantage of cpDNA utilization for phylogeny is the potential occurrence of chloroplast transfer from one species to another (introgression), creating bias in inferred trees (Rieseberg and Soltis, 1991). But if detected, this chloroplast capture may be very informative about evolutionary processes (Wendel and Doyle, 1998).

Different approaches and cpDNA regions used for phylogenetic purposes---Four mains approaches are used to infer plant relationships, namely restriction sites analysis (population level analyses), structural changes such as inversions, deletions of genes and intron (genus level or higher), PCR-based techniques such as cpDNA ST-SSR (Powell *et al.*, 1995) and DNA sequencing (Chase *et al.*, 1993). Many cpDNA regions have been used for phylogeny estimations (Table 3.2), from *rbc*L gene to a plethora of "new" alternatives recently proposed.

Locus	Length range (bp)	Occurrence of indels (size)	Taxonomic level of utility
*rbc*L	1428 - 1431 - 1434	NO	Genus to Subclass
*atp*B	1428 - 1431 - 1434	NO	Genus to Subclass
*atp*B - *rbc*L IGR	~900	frequent (50 to 204 bp)	Species - Genus
*mat*K	~1550	weak (3 to 9 bp)	Species to Order
*ndh*F	2233 (tobacco)	between bp1443 and bp1697	Genus to Order
16S rRNA	~1600	?	Order to Phylum
*trn*L intron	350 - 600	frequent	Species to Family
*trn*L-F IGS	120 - 350	frequent	Species to Family
cpITS2 and cpITS3	~500		Order to Phylum

Table 3.2. Main characteristics of some DNA regions frequently used for phylogenetic purposes. Shaded areas represent sequences used in the present study.

Phylogenetic inferences: concepts and methods

The development of Polymerase Chain Reaction (PCR) and automatic thermo-cyclers, was one of the major revolutions in molecular systematics, allowing investigations in DNA variations on a large scale. Different PCR-based techniques, such as direct sequencing or DNA fingerprinting methods (microsatellites, Amplified Fragment Length Polymorphism, Random Amplified Length Polymorphism…), have spawned increasingly large data sets available for diverse molecular studies.

Alongside the advances in biotechnology, the sophistication of phylogenetic analyses has grown rapidly, assessing phylogenetic accuracy (Hillis, 1995), testing different hypotheses or new evolutionary models (e. g., Felsenstein, 1981; Hasegawa *et al.*, 1985) or showing important progression in the study of characters evolution (Maddison and Maddison, 1997).

Presently, phylogenetic inferences are always based on two principal assumptions (homology and independence of the characters) and use optimally criteria, such as parsimony, distance or probabilistic methods. The search of the different phylogenetic trees is performed with independent algorithms, and only the trees that fit the optimally criteria are kept. Then, different methods (bootstrapping, decay index) are applied to evaluate the confidence of trees (Box 4).

BOX 4: Phylogenetic Reconstruction

The main goal of molecular phylogeny is to convert the information contained in DNA sequences into inferred evolutionary trees. To do so, one must first considerate the kind of data to be used, then create a matrix of aligned characters and finally input that matrix into a method of phylogenetic reconstruction.

The data--In ***distance characters*** based methods, aligned sequences are first converted into a pairwise distance matrix, whereas ***discrete characters*** methods consider each nucleotide site directly. Distances or characters matrix are then input in a tree building method.

The tree building methods--A plethora of approaches are presently available and can be divided in two main categories, depending on how they handle the data. The ***clustering methods*** follow an algorithm (a number of computation steps) to produce a tree. A starting tree of three sequences is first constructed, and then remaining sequences are successively added, depending on the kind of distances chosen. The ***search methods*** use optimality criteria to choose among all of the possible trees obtained for a given number of sequences. Trees with best scores under a given criterion are those that better fit the data. In the *Maximum parsimony methods* the best trees are those with the minimum number of steps (i.e. assuming the minimum level of homoplasy) whereas the *Maximum likelihood methods* incorporate explicit models of sequence evolution and compute the likelihood to obtain observed data under the model chosen.

Advantages and limits of the tree-building methods--Five desirable properties have been suggested to compare the numerous tree-building methods (Penny et al., 1992). The ***efficiency*** is the speed of a given method, i.e. the time in which a computer program can find a tree. The ***power*** is the amount of data (nucleotide sites sequenced) needed to obtain correct results. A phylogenetic method that converges to the true tree as we add more sequences data is ***consistent***. The sensitivity to violation of an underlying model of evolution measures the ***robustness*** of a given approach. Lastly, the ***falsibiality*** of a method is its capacity to tell us if its assumptions are violated or not.

Tree search methods--For a given phylogenetic method, many inferred trees may be produced. Several kind of searches may be used in order to find the best tree for the optimally criterion chosen. The ***exhaustive search*** checks all the possible fully resolved trees in order to choose the best one. Unfortunately this method is only feasible for small data sets, for the number of possible tree quickly become enormous with the taxa increase. The ***branch and bound search*** saves effort by only checking trees that are likely to be the best (and not all the trees). A reference tree is first constructed and is set at the best score (upper bound in parsimony). Then taxa are added and tree scores are measured. At each step, bad trees (compared to the reference) are discarded. Although this approach requires checking fewer trees than the exhaustive one, it is still time consuming. Therefore, for big data sets, only inexact ***heuristic search*** can be used. To avoid to find locally optimal sets of trees instead of the global one, two strategies are frequently used, namely the *Stepwise addition* and the *Branch swapping*.

Tree statistics and confidence—A phylogenetic tree either correctly depict historical relationships, or it does not! So, one can ask how much confidence can be ascribed from a particular cladogram? There are two ways to estimate branch or clade support in a phylogenetic tree. The first way is to calculate several statistical parameters that can give the scores and confidence limits of the inferred trees; a second possibility is the use resampling methods.

Statistical parameters for tree confidence---Some statistical parameters are of current use to evaluate branch support on a cladogram (Table 4.1).

Optimum Criterion	Parameter	Formula	Description
Maximum Parsimony	tree length (L)	*	Number of steps or character state changes in the fully resolved tree
	Consistency Index (C)	m / s	m = minimum amount of change possible for a character s = actual number of steps in the tree
	Homoplasy Index (H)	1 - C	*
	Retention Index (R)	(g - s) / (g - m)	g = maximum number of steps on the tree
Maximum Likelihood	Likelihood score	*	depends on the model parameters
Distance	Minimum Evolution Score	*	*

Table 4.1. Different statistical parameters used to evaluate branch supports or tree scores under several criterions.

Resampling methods---These methods estimate the variance of the sampling distribution by repeatedly resampling data from the original matrix. The use of these techniques in phylogenetic reconstruction was performed early by Felsenstein (1985). The bootstrap and the Jackknife are the more widely resampling techniques used so far. They mainly differ by the way of resampling. In the bootstrap, data are sampled randomly and with replacement, from the original matrix, until a new matrix with the same number of data is obtained. The jackknife method resamples the original matrix but drops a certain proportion of data at a time, and computes the estimate of the remaining data. In both methods, a certain number of replicates are performed, and branch supports will be given by the percentage of clade presence in the respective matrices.

Example: a bootstrap value of 85, for a particular clade, indicates that this group appeared 85 times within 100 replicates.

Bremer support or Decay Index---Another way to test the robustness of a clade is to determine whether this group occurs in other trees, almost equally short. The number of extra steps required to collapse this clade is called the decay index or Bremer support (Bremer, 1998).

Example: Let's take a most parsimonious tree of length L with a clade A. Let's considerate other trees of length L+1 and L+2. Let's perform the strict consensus trees of L and L+1 and L and L+2, sc1 and sc2, respectively. If the clade A appears on sc1 but not on sc2, the Bremer support of this clade is 1.

BOX 5: Evolutionary models - Molecular clock

The concept of molecular clock was early proposed, based on the observation that various protein or DNA sequences may evolve at a constant rate over time, and thus may provide useful information for dating past evolutionary events (Zuckerkandl and Pauling, 1965). Presently, there are indisputable evidences that many genes and proteins do not evolve a constant rate through time (Avise, 1994). A rate is a change in some quantity in an interval of time, and may be either absolute (with a time scale), or relative (comparison in the same time interval) (Sanderson, 1998). These rates vary among regions of genes and proteins for diverse reasons (genetic code constraints, protein structure and function, expression). Nevertheless, the molecular rate constancy continues to be view as a reasonable model, with respect to the neutral theory (Wray *et al.*, 1996), and has led to many estimation of time divergence, without any attempt to test for a molecular clock behavior and thus to check the validity of the underlying model.

Evolutionary models--In order to understand the details of molecular variation, some models have been proposed for DNA and amino-acid sequences, but also for restriction sites, allozymes, or insertion-deletion events (Swofford *et al.*, 1996).
All the models of sequence evolution rely on a Poisson process, describing discrete changes in continuous time. Moreover, the substitution process encountered in these models may be regarded as a Markov chain. Considering 3 time points in evolution (t1 < t2 < t3), the Markovian model assumes that the status of the nucleotide site at t3 depends only on the status at t2, but not the status at t1, if the status of the site at t2 is known. The characteristics of some proposed models are described in Table 5.1. In each of them, one can assume some sites to be invariant, while allowing the remaining variable sites to also exhibit among-site rate heterogeneity (measured by the gamma shape parameter).

Model	Base frequencies	Substitution rate	Reference
Jukes – Cantor (JC)	Equal	Equal	Jukes & Cantor, 1969
Kimura 2-parameters (K2P)	Equal	Transitions ≠ Transversions	Kimura, 1969
Felsenstein (F81)	Unequal	Equal	Felsenstein, 1981
Hasegawa *et al.* (HKY 85)	Unequal	Unequal	Hasegawa *et al*., 1985
General reversible (REV)	Unequal	Unequal for each substitution	Yang *et al.*, 1994

Table 5.1. Comparisons between five current models used to estimate the nucleotide substitutions among pairs of DNA sequences. The REV model encompasses all the remaining ones.

Testing the molecular clock---One of the most exciting challenges of the phylogenetic approach is to estimate time divergence between the respective lineages to better understand their evolutionary history through space and time. Therefore, the assumption of rate constancy must be preliminary tested, and many methods have been proposed.

Relative Rate Tests--These widely used tests compare two lineages A and B, sharing a common ancestor, with a more distantly related outgroup taxon (O = reference). If rate are constants over time, the accumulated changes from the common ancestor to A and B, respectively, are expected to be identical (except for difference due to sampling error). Relative Rate Tests measure the difference Δ between O and each of the ingroup taxa A and B, which is expected to be zero under rate constancy (Figure 5.1).

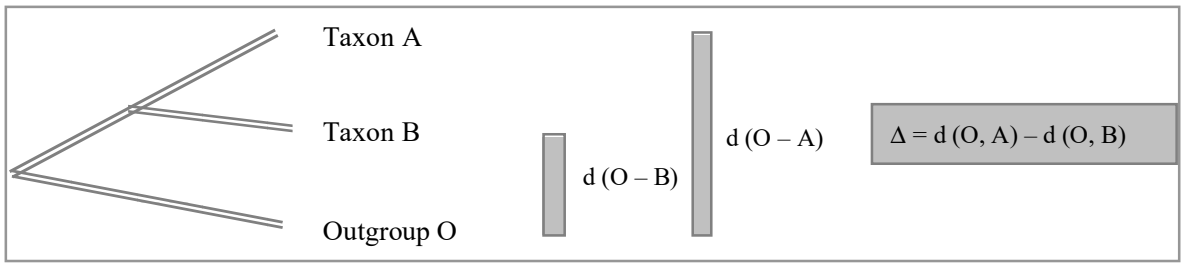


Figure 5.1. Relative Rate Test between two taxa A and B and a reference O. Δ is the statistic tested against the expectation of zero when rates are constants

These tests generally apply pairwise distance (e. g., Kimura 2-parameters model) as a measure of divergence between the lineages and use the same model to predict the variance (Wu and Li, 1985). A standard two-tailed test is then applied ($z = \Delta / \sqrt{(var\Delta)}$), which is expected to be distributed with mean zero and variance one.

Available Software: RRTree (Robinson *et al.*, 1998)

***Likelihood ratio test*---In this method, the entire tree in considered, instead of multiple tests at each node, and thus the global hypothesis of rate constancy is tested (Felsenstein, 1988, 1993). The likelihood of a given tree is first calculated assuming a molecular clock (L_0: null hypothesis), then calculated without rate constancy (L_1: alternative hypothesis), and a ratio R is finally calculated:

$R = -2\ln(L_0 / L_1)$ or $R = -2(\ln L_0 - \ln L_1)$.

This ratio is assumed to follow a chi squared distribution with N - 2 degree of freedom. The latter is given by the difference in the number of free parameters in the two models.
Any model of sequence evolution can be chosen to perform these tests, and the best model is selected using the LRT method (comparing the likelihood of one model over the likelihood of an alternative model).

Example from Chapter 1. Analyses of a reduced ITS data sets (twenty-two representative taxa of the North American *Centaurium*) were performed with Heuristic Search, under the Maximum likelihood criterion, using the HKY85 model of sequence evolution. Model parameters (base frequencies, transition over transversion ratio, proportion of invariant sites, gamma shape) were first estimated in Paup 4.0b3 (Swofford, 1999), using the Neighbor Joining tree. Then, the respective likelihood of the tree constrained by a molecular clock ($\ln L_0$), or not ($\ln L_1$) was computed, and the ratio was calculated:
$R = -2 \ln (L_0 / L_1) = 22$.
The Chi-square value critical value at $\alpha = 0.05$ is 31, with n = 20 (22 - 2) degrees of freedom.
There is no significant difference between the respective assumptions and the null hypothesis of a clock like model should not be rejected.

Available Software: Paup 4.0 (Swofford, 1999); Phylip (Felsenstein, 1993)

Previous molecular phylogeny in the Gentianaceae

One of the first molecular investigations made on the Gentianaceae was performed on the tropical genus *Lisianthus* using restriction fragment analysis of cpDNA and sequences of nrDNA (Systma & Schaal, 1985, 1990). Later, the phylogeny of European gentians was inferred using non-coding sequences of cpDNA, namely *trn*L intron and *trn*L-*trn*F spacer (Gielly and Taberlet, 1996; Gielly *et al.*, 1996). The *trn*L intron, along with the matK gene, has then provided the first intergeneric classification of the family (Struwe *et al*, 1998; Thiv *et al.*, 1999b). On the other hand, the use of ITS data allowed larger investigations on *Gentiana* at both infra- and intergeneric level, including many Asiatic species and close relative genera (Yuan and Küpfer, 1995; Yuan *et al.*, 1996). One may cite other punctual studies on *Gentiana* at the sectional level, for example *Chondrophyllae* (Yuan and Küpfer, 1997) or *Ciminalis* (Hungerer and Kadereit, 1998). Recently, the pantropical tribe Chironieae has received attention (Thiv *et al*, 1999a).

Based on *trn*L intron, *mat*K and ITS sequences, a new tribal and subtribal classification is presently proposed (Struwe *et al.*, 200*). Subtribe Erythraeinae and Chironiinae sensu Gilg (1895), along with several other genera such as *Ixanthus* or *Eustoma*, have been blended in one tribe Chironieae further subdivided (Chironiinae, Canscorinae and Coutoubeinae). On the other hand, tropical Erythraeinae (9 genera) are included in a new subtribe (Faroinae / Potalieae).

BOX 6: Evolution of the intergeneric classification of *Centaurium*

Family	**GENTIANACEAE**					
Tribe	**Gentianaeae**					**Helieae**
Subtribe	**Erythraeinae (19/19)**		**Chironiinae (3/3)**	**Gentianinae (1/17)**	**Tachiinae (2/9)**	**(5/15)**
	Bartonia	*Hoppea*	*Chironia*	*Ixanthus*	*Eustoma*	*Coutoubea*
	Bisgoeppertia	*Lapithea* (1)*	*Gentianothamnus*		*Hockinia*	*Deianera*
	Blackstonia	*Microcala* (2)*	*Orphium*			*Schultesia*
	Canscora	*Neurotheca*				*Symphyllophyton*
	Centaurium	*Obolaria*				*Xestaea*
	Cicendia	*Oreonesion*				
	Curtia	*Sabatia*				
	Enicostema	*Schinziella*				
	Faroa	*Tapeinostemon*				
	Geniostemon					

Table 6.1: Partial Classification of the Gentianaceae, comprising *Centaurium* and more or less closely relative genera (Gilg, 1895). Shaded genera are part of the subtribe Chironiinae as defined in the present work. Numbers in brackets represent the proportion of genera for each subtribe. (*) Non-valid names, (1) included in *Sabatia*, (2) included in *Cicendia*

Family	**GENTIANACEAE**					
Tribe	**Chironieae**			**Gentianeae**	**Saccifolieae**	**Potalieae**
Subtribe	**Chironiinae (12/12)**	**Canscorinae (5/5)**	**Coutoubeinae (5/5)**	**Swertiinae (2/14)**	**(2/5)**	**Faroinae (9/9)**
	Bisgoeppertia	*Canscora*	*Coutoubea*	*Bartonia*	*Curtia*	*Congolanthus**
	Blackstonia	*Cracosna**	*Deianera*	*Obolaria*	*Tapeinostemon*	*Djaloniella**
	Centaurium	*Hoppea*	*Schultesia*		*Hockinia*	*Enicostema*
	Chironia	*Microrphium**	*Symphyllophyton*			*Faroa*
	Cicendia	*Schinziella*	*Xestaea**			*Karina**
	Eustoma					*Neurotheca*
	*Exaculum**					*Oreonesion*
	Geniostemon					*Pycnosphaera**
	Ixanthus					*Urogentias**
	Orphium					
	Sabatia					
	*Zygostigma**					

Table 6.2: Partial Classification of the Gentianaceae, based on molecular characters, according to Struwe *et al.* (200*). Shaded genera are part of the subtribe Chironiinae as defined in the present work. Numbers in brackets represent the proportion of genera for each subtribe. (*) Genera non-present in Gilg's treatment.

The present dissertation will focus on the phylogenetic relationships of the core genus *Centaurium* of the subtribe Chironiinae, at both interspecific and intergeneric levels.

Account of the premises and aims of the present study

What was the context of the systematic knowledge on *Centaurium* and relatives at the beginning of the present investigation? On one hand, one part of the genus, the "Eurasian group", had been (and is still!) extensively studied by one of our collaborator, and a systematic treatment based on karyological and biogeographical data was available, but not yet complete. On the other hand, the New World group lacked such taxonomic treatment and many results (such as chromosome counts) were based on a very restricted set of studies. Moreover, some of them needed either confirmation or deeper investigations.

Therefore, the first logical step in this thesis was to increase our knowledge and comprehension of the North-American *Centaurium* that was, so far, the most obscure group for us. Three field expeditions, in the Western part of the United States and Mexico areas, were organized to obtain, if not an exhaustive, at least a very representative sampling of populations for further karyological and molecular investigations.

Then, in a second part of this work, we focused on the Eurasian species in order to check previous systematics treatment and hypotheses with molecular tools. Existing karyological data resulting for more than 30 years of field expedition (Zeltner, op. cit.) were completed with some expeditions in the peripheral parts of the genus repartition area (e. g. Sultanate of Oman or Australia).

Lastly, a global investigation of the genus *Centaurium* was performed to understand the evolutionary affinities of the two major groups hitherto delimited and to understand biogeographic patterns. Concomitantly, the place of *Centaurium* within the subtribe Chironieae and subsequently intergeneric relationships were inferred using four molecular data sets.

In this context, many aspects concerning the natural evolution of *Centaurium* were envisaged:

1. The phylogenetic relationships within and between the American and Eurasian groups of *Centaurium,* with perspectives of further systematic treatment.

2. The intergeneric affinities between the genera constituting the subtribe Chironiinae, with a particular focus on characters evolution and biogeographic patterns.

3. The chromosomal variability and karyological evolution in *Centaurium* and closed relatives.

4. The importance of polyploidy and hybridization processes in this plant group.

Main results and discussion

Phylogenetic reconstruction and biogeographical implications

***The North American species of *Centaurium* form a strongly supported clade, well separated from the Old World group and from two taxa of the section *Gyrandra* (*C. tenuifolium* and *C. brachycalyx*)**. Within the New World group, three main geographical groups can be drawn, based on molecular and karyological data sets (Chapter 1). The first group comprises species having a haploid chromosome number of n = 17 (n = 20), and mainly occurring in California. The second assemblage, with n = 20 (n = 21), may be found in Texas, whereas most of the Mexican taxa (n = 21 or n = 22) constitute a third clade. The chromosome numbers recorded in the American group are generally less heterogeneous than previously proposed (Broome, 1976) and dysploidy appears to be the main trend of chromosome evolution (17 < 20 > 21 > 22). Hybridization phenomena occur, generally obscuring taxon delimitation and sometime providing speciation process. A new allopolyploid taxon, *C. sp.nov.1* (n = 37), seems to have originated via hybridization between *C. exaltatum* (n = 20) and *C. venustum* (n = 17). The diversification of the North-American *Centaurium* clade likely occurred in the early Pliocene (ca. 6 mya). Geological events such as deserts formation (Chihuahua, Mojave or Sonora) or mountains orogenies (Sierra Nevada and Cascade Ranges) have created physical barriers that today separate three probably allopatric groups.

***The Eurasian species may be divided in 5 clades, with a well-separated section *Spicaria* (Chapter 2)**. Phylogenetic reconstruction do not always supports the existing classification, and some groups such as subsections *Vulgaria* and *Parviflora* appear to be polyphyletic. On the other hand, subsection *Centaurium* forms a fully unresolved clade, comprising *C. erythraea* sensu lato and *C. muehlenbergii*, native from California. This group may have recently evolved under strong environmental pressures. Within the Eurasian *Centaurium* assemblage, polyploidy seems to be a major force of speciation, for diploid to tetra- or hexaploid taxa are found in the respective clades. Dysploidy also occurs, with a descending pattern from n = 10 (basal chromosome number) to n = 9, in the *C. pulchellum* complex. Allopolyploidy with hybridization processes may be proposed for several taxa origin such as *C. bianoris*, with *C. tenuiflorum* as parent donor and *C. maritimum* as recipient or *C. malzaciannum* (= *C. maritimum* x *C. pulchellum*). Finally, the present geographic repartition of the Eurasian species of *Centaurium* may be seen as the consequence of major geological events such as the Messinian salinity crisis or more recently, the Quaternary ice ages. Nowadays, diploid taxa generally occur in restricted areas

corresponding to some ancestral glacial refugia, whereas their polyploid relatives occupy a more peripheral and northern repartition, suggesting possible patterns of post-glacial recolonization.

***Phylogenetic studies based on four data sets clearly demonstrate the polyphyly of the genus *Centaurium* (Chapter 3, 4)**. All the different cladograms support the segregation of *Centaurium* in several independent clades. Trees inferred from the ITS data set depict a derivative position for the New-World *Centaurium* species, close to a (*Sabatia* - *Eustoma* - section *Gyrandra*) sister clade and near section *Spicaria*. Lastly, the Eurasian clade occupies a basal position close to *Chironia* and *Orphium* genera. The 5S-NTS region confirms the deep separation between the Old World and New World *Centaurium*, as well. Notwithstanding, respective trees obtained with these sequences of nrDNA showed much incongruence in topology. Finally, two regions of the cpDNA, namely the *trn*L intron and *trn*L-F IGS, also supports the subdivision of *Centaurium* in four separate groups, but otherwise failed to satisfactory resolve interspecific relationships within each ones. In all the cases, molecular data conflict with the apparent morphological homogeneity occurring in *Centaurium* s. l. and may suggest strong environmental pressures in the respective groups. A major consequence of the weak clade support by classical data results in important difficulties to provide a strong systematic treatment of the entire group. Few characters appear to be synapomorphic within the subtribe, and in *Centaurium*, only the degree of style division or the stigmata shape tend to confirm molecular subdivisions.

Phylogenetic utility of the different DNA regions investigated

***Degree of resolution and scale of taxa divergence**

Four different regions of DNA of plastid origin (*trn*L intron and *trn*LF IGS) or of nuclear one (ITS and 5S-NTS) have been used to produce phylogenies within the Chironiinae subtribe.

Because the chloroplast genome is inherited as a unit and is not subject to recombination (Olmstead and Palmer, 1994), *trn*L and *trn*F regions have been associated in a common data set, in order to increase phylogenetic signal. Meanwhile, the latter remained weak, with few nucleotide substitutions (generally less than 10 % of informative characters), and occurrence of informative indels. These combined regions provide more or less important resolution at generic or sectional levels. On the other hand, topology conflicts with nrDNA sequences provide evidences for reticulate evolution or other biological phenomena.

The use of ITS sequences appeared to be very helpful to resolve phylogenetic relationships within the Chironiinae subtribe, at both intergeneric and interspecific levels. However, the phylogenetic utilization became

limited at the infra-specific level. On the other hand, 5S-NTS region appeared to evolve faster than ITS but was of difficult use at both inter- and infrageneric levels, with a high rate of sequence divergence between the different genera and many difficult-to-align sequence regions (Chapter 4). Presently, the choice of the 5S-spacer region appears questionable to assess interspecific relationships within the Chironiinae and should better fit lower phylogenetic levels.

***Sources of incongruence between ITS and 5S-NTS reconstruction (Chapter 4).**

Congruence tests performed between the respective data set topologies indicated high significant differences. The first cause of conflict appeared to be the high level of variation within the 5S data set compared to the scale of taxa divergence. The high frequency of point mutations along with an elevated occurrence of indels, made the alignments and subsequent analyses problematic, resulting in conflicting phylogenetic reconstructions with ITS data. On the other hand, PCR amplification of the 5S spacer generally produced more than 2 fragments, indicating potential occurrence of paralogy in this region. The presence of two loci arising from gene duplication in a same individual genome, has been described as a potential source of incongruence between molecular data sets (Doyle, 1992; Sanderson and Doyle, 1992).

Furthermore, interlocus concerted evolution (Arnheim, 1983), is generally complete in the ITS region whereas weak and partial homogenization has been documented in 5S sequences, not only between loci (Sanderson and Doyle, 1992), but also within a locus (Kellogs *et al.*, 1996; Cronn *et al.*, 1996). Lastly, hybridization may further complicate phylogenetic reconstruction and therefore increase the existing incongruence.

Evolutionary trends in Centaurium s. l.

***Polyploidy and hybridization**

The present investigations clearly show that polyploidy process generally occurs within several clades of *Centaurium*. Within the Old World clade, three ploidy levels can be encountered, suggesting the possibility of autopolyploidy or segmental allopolyploidy, depending on the level of genome divergence (Bretagnolle *et al.*, 1998). This may be the case for diploid and tetraploid populations of *C. tenuiflorum* or *C. majus*. On the other hand, allopolyploidy appears to be more common, either occurring between close related genomes (e. g. hexaploid species *C. mairei* and diploid / tetraploid taxa of *C. pulchellum*) or between divergent ones giving new allopolyploid taxa (e. g., *C. malzacianum* [n = 28] with *C. maritimum* [n = 10] and *C. pulchellum* [n = 18] or *C.*

sp.nov.1 [n = 37] with *C. exaltatum* [n = 20] and *C. venustum* [n = 17]). Other putative cases of allopolyploid species (e. g. *C. somedanum* or *C. serpentinicola*) have been suggested in the present study and need further confirmation.

Moreover, diploid species along with tetraploid relatives are present only in Eurasian *Centaurium* and in section *Spicaria* and autopolyploidy processes may be expected in these clades. Allopolyploidy is mainly encountered in the Old World *Centaurium* and occasionally in the New World clade, where dysploidy appears to be more frequent.

From a biogeographic point of view, diploid populations of *Centaurium* s. s. seem to occupy some ancestral glacial refugia proposed for other groups of plants (Huntley and Birks, 1983), whereas polyploid taxa clearly show a peripheral repartition (Chapter 2).

In some particular groups, a great morphological variation may be found, without changes in chromosome valence suggesting either occurrence of homoploid hybridization or more likely, introgression. RAPD analyses performed on several North American and Australian populations of *C. muehlenbergii* have shown diverse degree of introgression patterns with introduced populations of *C. erythraea*, a taxon of Mediterranean origin (Chapter 5).

***Chromosome evolution**

With the current demonstration of *Centaurium* polyphyly, some karyological hypotheses invoking the possibility of an early Tertiary migration of diploid *Centaurium*, from the Mediterranean to the Americas, followed by further episodes of dysploidy and polyploidy (Zeltner, 1970; Broome, 1973), have to be revised. Presently, chromosome evolution in *Centaurium* not only involves four separate clades but other genera of the Chironiinae as well, and thus several independent scenarios must be envisaged. Nevertheless, the knowledge of the chromosome number is currently not available for all the taxa of the subtribe and only fragmental conclusions can be addressed.

The basal chromosome number in this group appears to be n = 10, present in basal lineage such as *Blackstonia* or the Eurasian *Centaurium*, and to a lesser extant in *Exaculum*. This number may then evolve either by polyploidy, with n = 10 → 20 → 30 (or 31) in Eurasian *Centaurium* (or in *Blackstonia* and paleopolyploid *Ixanthus*) or by dysploidy pattern in the *Centaurium pulchellum* group (n = 10 → n = 9). Whereas no general pattern may be proposed due to some missing information, dysploidy appeared to be relatively frequent at diploid and polyploid levels in the Chironiinae, but with no particular polarity. Both descending and ascending dysploidy can be

observed in the different clades. Thus, chromosomal evolution in *Centaurium* and relative taxa involves separate episodes of dysploidy and polyploidy, along with hybridization process obscuring the general pattern of chromosome changes. Following the evolutionary trends depicted by the ITS trees, a scheme of karyotype evolution can be suggested (Figure 6).

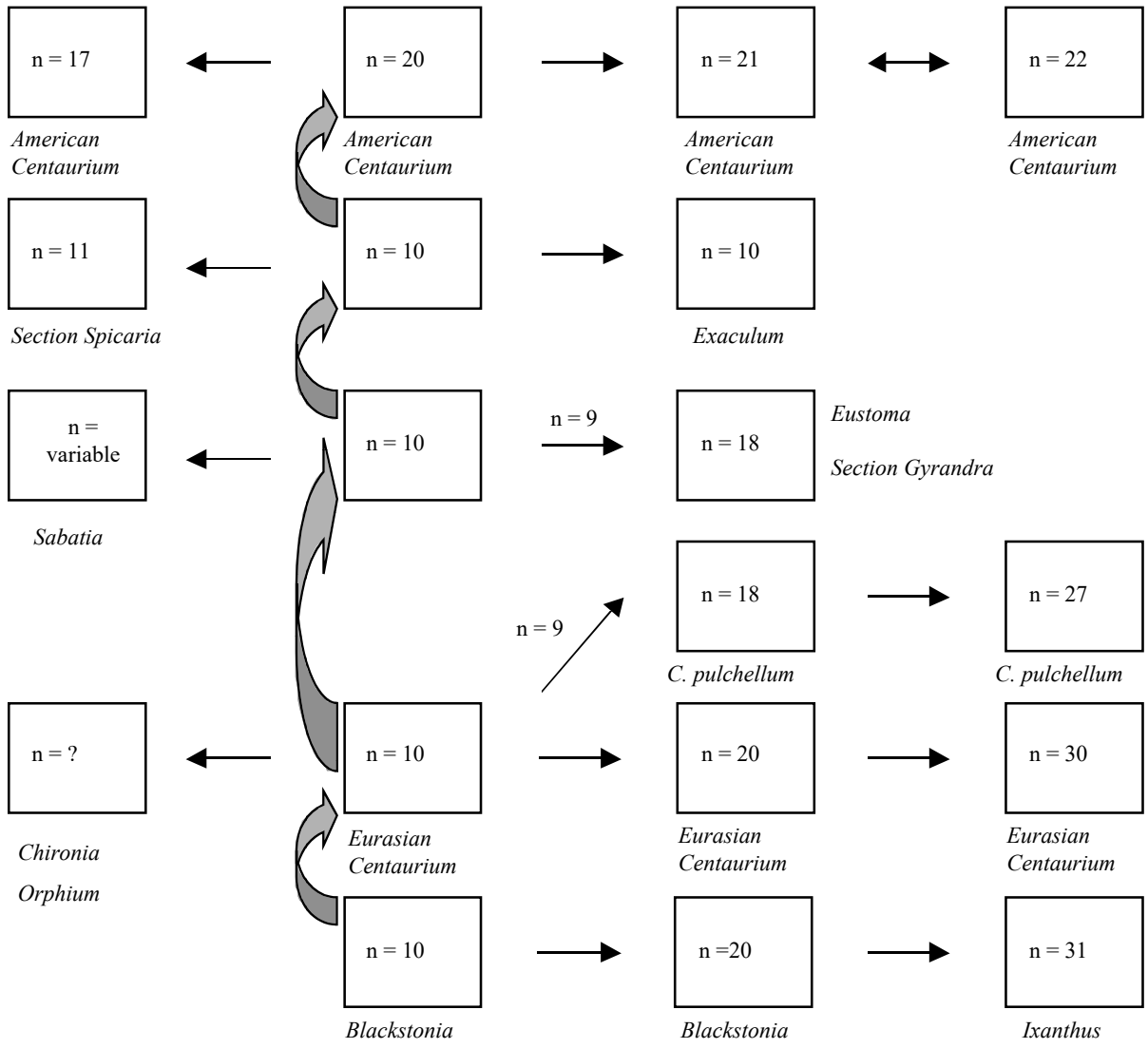


Figure 6. Hypothetical karyological evolution within the subtribe Chironiinae. Curved arrows indicate phylogenetic patterns described on the ITS strict consensus tree (Chapter 3). Black arrows denote either dysploidy or polyploidy patterns.

References

ABBOTT, R. J. 1992. Plant invasions, interspecific hybridization and the evolution of new plant taxa. *Trends in Ecology and Evolution* 7 (12) : 401-405.

ADAMS, L. G. 1996. *Gentianaceae*. in *Flora of Australia* 28 : 73-103. A. Wilson Editor. CSIRO. Melbourne.

ADAMS, S., I. J. LEITCH, M. D. BENNETT, M. W. CHASE, and A. R. LEITCH. 2000. Ribosomal DNA evolution and phylogeny in *Aloe* (Asphodelaceae). *American Journal of Botany* 87 (11) : 1578-1583.

ANSARI, H. A., N. W. ELLISON, S. M. READER, E. D. BADAEVA, B. FRIEBE*, et al.* 1999. Molecular cytogenetic organization of 5S and 18S-26S rDNA loci in white clover (*Trifolium repens* L.) and related species. *Annals of Botany* 183 : 199-206.

APG (Angiosperm Phylogenetic Group). 1998. An ordinal classification for the families of flowering plants. *Annals of the Missouri Botanical Garden* 85 : 531-553.

ARNHEIM, N. D. 1983. Concerted evolution of multigene family. in *Evolution of genes and proteins*. M. Nei and R. K. Koehn Editors. Sinauer Associates. Sunderland.

ARNHEIM, N. M., R. KRYSTAL, R. SCHMICKEL, G. WILSON, O. RYDER*, et al.* 1980. Molecular evidences for genetic exchanges among ribosomal genes of nonhomologous chromosomes in man and apes. *Proceedings of the National Academy of Science USA* 77 : 7323-7327.

ARNOLD, M. L. 1992. Natural hybridization as an evolutionary process. *Annual Review of Ecology and Systematic* 23 : 237-261.

ARNOLD, M. L. 1997. *Natural hybridization and evolution*. Oxford University Press. Oxford

AVERETT, J. H. 1980. Polyploidy in plant taxa: summary. in *Polyploidy, biological relevance* **:** 262-273. Lewis Editors. Plenum Press. New York.

AVISE, J. C. 1994. *Molecular markers, natural history and evolution*. Chapman & Hall. New York.

AVISE, J. C. 2000. *Phylogeography - The history and formation of species*. Harvard University Press. Cambridge.

AVISE, J. C., J. ARNOLD, R. M. BALL, E. BERMINGHAM, T. LAMB*, et al.* 1987. Intraspecific phylogeography: the mitochondrial bridge between population genetics and systematics. *Annual Review of Ecology and Systematic* 18 : 489-522.

BAKER, W. J., T. A. HEDDERSON, and J. DRANSFIELD. 2000. Molecular phylogenetics of *Calamus* (Palmae) and related rattan genera based on the 5S nrDNA spacer sequence data. *Molecular Phylogenetics and Evolution* 14 (2) : 218-231.

BAKKER, F. T., J. L. OLSEN, and W. T. STAM. 1995. Evolution of the nuclear rDNA ITS sequences in the *Cladophora albida* / *sericea* clade (Chlorophyta). *Journal of Molecular Evolution* 40 : 640-651.

BALDWIN, B. G. 1992. Phylogenetic utility of the Internal Transcribed Spacer of ribosomal DNA in plants: an example from the Compositae. *Molecular Phylogenetics and Evolution* 1 : 3-16.

BAUHIN, G. 1623. Pinax theatri botanici, sive index in Theophrasti, Dioscoridis, Plinii et Botanicorum qui a saeculo scripserunt. *Basiliae.* 278.

BAYER, E. and G. L. GONZALEZ. 1991. *Centaurium barrelieroides* Pau y *C. rigualii* Esteve (Gentianaceae), ¿dos endemicos mediterraneos de area muy limitada? *Anales Jardin Botanico de Madrid.* 49 (1) : 57-65.

BORCKHAUSEN, M. B. 1796. Uber Linne's gattung *Gentiana*. *Arch. I.* 1 : 29-30.

BREMER, K. 1988. The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. *Evolution* 42 : 795-803.

BRETAGNOLLE, F., F. FELBER, F. G. CALAME, and P. KÜPFER. 1998. La polyploidie chez les plantes. *Botanica Helvetica* 108 : 5-37.

BRETAGNOLLE, F. and J. D. THOMPSON. 1995. Gametes with somatic chromosome number: their mechanisms of formation and role in the evolution of autopolyploid plants. *The New Phytologist* 129 : 1-22.

BROOME, C. R. 1973. Systematics of *Centaurium* (Gentianaceae) of Mexico and Central America. Thesis. Duke University. Duke

BROOME, R. C. 1976. The Central American species of *Centaurium* (Gentianaceae). *Brittonia* 28 : 413-426.

BROOME, R. C. 1977. Four new species of *Centaurium* from Mexico. *Madroño* 24 : 237-244.

BROOME, R. C. 1978. Chromosome numbers and meiosis in North and Central American species of *Centaurium* (Gentianaceae). *Systematic Botany* 3 : 299-312.

CAMPBELL, B. C., J. D. STEFFEN - CAMPBELL, and J. H. WERREN. 1993. Phylogeny of *Nasonia* species complex (Hymenoptera: Pteromalidae) inferred from the Internal Transcribed Spacer. *Insect Molecular Biology* 2 : 225-237.

CHASE, M. W., D. E. SOLTIS, R. G. OLMSTEAD, D. MORGAN, D. H. LES*, et al.* 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid *rbc*L. *Annals of the Missouri Botanical Garden* 80 : 528-580.

CLEGG, M. T., B. S. GAUT, G. H. LEARN, and B. R. MORTON. 1994. Rates and pattern of chloroplast DNA evolution. *Proceedings of the National Academy of Science USA* 91 : 6795-6801.

CRACRAFT, J. 1983. Species concepts and speciation analysis. in *Current Ornithology*. R. F. Johnston Editor. Plenum Press. New York.

CRISP, M. D., R. APPELS, S. M. SMITH, and W. M. S. KEYS. 1999. Phylogenetic evaluation of 5S ribosomal RNA gene and spacer in the *Callistachys* group (Fabaceae: Mirbalieae). *Plant Systematic and Evolution* 218 : 33-42.

CRONN, R. C., X. ZHAO, H. PATERSON, and J. F. WENDEL. 1996. Polymorphism and concerted evolution in a multigene family: 5S ribosomal DNA in polyploids and allopolyploids cottons. *Journal of Molecular Evolution* 42 : 685-705.

De LITARDIERE, R. 1948. Nouvelles contribution à l'étude de la flore de Corse. *Candollea* 11 (7) : 213.

De MESQUITA RODRIGUES, J. E. 1953. Contribuiçao para o conhecimento cariologico das halofitas e psamofitas litorais. *Coïmbra* 110.

DODOENS, R. 1583. Stirpium historiae pemptades sex, sive libri XXX. *Antwerpiae* 333.

DOUGLAS, S.E., S. L. PENNY. 1999. The plastid genome of the cryptophyte alga, *Guillardia theta*: complete sequence and conserved synteny groups confirm its common ancestry with red algae. *Journal of Molecular Evolution.* 48(2) : 236-44

DOWNIE, S. R. and J. D. PALMER. 1992. Use of chloroplast DNA rearrangements in reconstructing plant phylogeny. in *Molecular Systematics of Plants* : 14-35. D. E. Soltis, P. S. Soltis and J. J. Doyle Editors. Chapman & Hall. New York.

DOYLE, J. J. 1992. Gene trees and species trees: molecular systematics as one character taxonomy. *Systematic Botany* 17 : 144-163.

DOYLE, J. J., J. L. DOYLE, and H. D. BROWN. 1999. Origins, colonization, and lineage recombination in a widespread perennial soybean polyploid complex. *Proceedings of the National Academy of Science USA* 96 : 10741-10745.

DOBZHANSKY, T. 1937. *Genetics and the origin of the species*. Columbia University Press. New York.

DUNN, R. A. 1967. A revision of the genus *Centaurium* of continental United States. Thesis. Catholic University of America. Washington D. C.

FELSENSTEIN, J. 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. *Journal of Molecular Evolution* 17 : 368-376.

FELSENSTEIN, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 17 : 223-263.

FELSENSTEIN, J. 1988. Phylogenies from molecular sequences: inference and reliability. *Annual Review of Genetics* 22 : 521-565.

FELSENSTEIN, J. 1993. *PHYLIP: Phylogenetic Inference Package*. University of Washington. Seattle.

FERNANDES, R. 1965. Duas variedades novas de *Centaurium spicatum* (L.) Fritsch. *Boletim da Sociedade Broteriana* 31 : 15-29.

FREIJSEN, A. H. J. (1967). A field study of the ecology of *Centaurium vulgare* Rafn. Thesis. Tilburg.

GIELLY, L. and P. TABERLET. 1996. A phylogeny of the European gentians inferred from chloroplast *trn*L (UAA) intron sequences. *Botanical Journal of the Linnean Society* 120 (1) : 57-75.

GIELLY, L., Y.-M. YUAN, P. KÜPFER, and P. TABARLET. 1996. Phylogenetic use of noncoding regions in the genus *Gentiana* L: Chloroplast *trn*L (UAA) intron versus nuclear ribosomal internal transcribed spacer sequences. *Molecular Phylogenetics and Evolution* 5 (3) : 460-466.

GILG, E. 1895. Gentianaceae. in *Die Naturlichen Pflanzenfamilien*. 4**:** 50-108. A. Engler and K. Prantl Editors

GILLETT, J. M. 1963. *The gentians of Canada and Greenland*. Research Branch, Canada Department of Agriculture. Ottawa.

GONZALEZ, G. L. 1979. Sobre el *Centaurium linariifolium* (Lam.) G. Beck. *Anales del Jardin Botanico de Madrid* 36 : 123-128.

GRANT, V. 1981. *Plant speciation*. Columbia University press. New York.

GRANT, V. 1998. Primary classification and phylogeny of the Polemoniaceae, with comments on molecular cladistics. *American Journal of Botany* 85 (6) : 741-752.

GRAY, A. 1878. Gamopetalae after Compositae. in *Synoptical flora of North America 2*. Ivison, Blakeman and Taylor Editors. New York.

GRISEBACH, A. H. R. 1839. *Genera et Species Gentianearum adjectis observationibus quibusdam phytogeographycis* : 116-125. J. G. Cotta. Stuttgart et Tübingen.

HALLICK, R. B., L. HONG, R. G. DRAGER, M. FAVREAU, A. MONFORT*, et al.* 1993. Complete sequence of *Euglena gracilis* chloroplast DNA. *Nucleic Acids Research* 21 : 3537-3544.

HARDIG, T. M., S. J. BRUNSFELD, R. S. FRITZ, M. MORGAN, and C. M. ORIENS. 2000. Morphological and molecular evidence for hybridization and introgression in a willow (*Salix*) hybrid zone. *Molecular Ecology* 9 : 9-24.

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

HAYS, J. D., J. IMBRIE, and N. H. SCHAKLETON. 1976. Variations in the Earth's orbit: pacemaker of the ice ages. *Science* 194 : 1121-1132.

HEGI, G. 1966. Gentianaceae. In : *Illustrierte Flora von Mittel-Europa*. 5 (3) : 1953-2047. Carl Hanser Verlag. München.

HEWITT, G. 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* 58 : 247-276.

HEWITT, G. 2000. The genetic legacy of the Quaternary ice ages. *Nature* 405 : 907-913.

HILLIS, D. M. 1995. Approaches for assessing phylogenetic accuracy. *Systematic Biology* 44 : 3 -16.

HILU, K. W. and H. LIANG. 1997. The *mat*K gene: sequence variation and application in plant systematics. *American Journal of Botany* 84 (6) : 830-839.

HIRATSUKA, J., H. SHIMADA, R. WHITTIER, T. ISHIBASHI, M. SAKAMOTO*, et al.* 1989. The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes account for a major plastid DNA inversion during the evolution of cereals. *Molecular and General Genetics* 217 : 185-194.

HO, T.-N. and S.-W. LIU. 1990. The infrageneric classification of *Gentiana* (Gentianaceae). *Bulletin British Museum National History (Botany)* 20 (2) : 169-192.

HO, T.-N., S.-W. LIU, and X.-F. LU. 1996. A phylogenetic analysis of *Gentiana* (Gentianaceae). *Acta Phytotaxonomica Sinica* 34 (5) : 505-530.

HOSTETTMANN, K. and H. WAGNER. 1977. Xanthone glycosides. *Phytochemistry* 16 : 821-829.

HOSTETTMANN-KALDAS, M. and A. JACOT-GUILLARMOD. 1978. Xanthones and C-glucosides flavoniques du genre *Gentiana* (sous-genre *Gentianella*). *Phytochemistry* 17 : 2083-2086.

HUNGERER, K. B. and J. W. KADEREIT. 1998. The phylogeny and biogeography of *Gentiana* L. sect. *Ciminalis* (Adans.) Dumort. : A Historical interpretation of distribution ranges in the European high mountains. *Perspectives in Plant Ecology, Evolution and Systematic* 1 (1) : 121-135.

HUNTLEY, B. and H. J. B. BIRKS. 1983. *An atlas for past and present pollen maps for Europe : 0 - 13 000 years ago*. Cambridge University Press. Cambridge.

INOUYE, H. and Y. NAKAMURA. 1971. XVI. Occurrence of secoiridoid glucosides in gentianaceous plants especially in genera *Gentiana* and *Swertia*. in *Studies on monoterpene glucosides and related natural products* 91 (7) : 755-759. Y. Zasshi Editor

INOUYE, H., K. UOBE, M. HIRAI, Y. MASADA, and K. HASHIMOTO. 1976. Studies on monoterpene glucosides and related natural products XXXI. Gas chromatography - mass spectrometry on iridoid and secoiridoid glucosides. *Journal of Chromatography* 118 : 201-216.

JENSEN, S. R., B. J. NIELSEN, and R. DAHLGREN. 1975. Iridoid compounds, their occurrence and systematic importance in the Angiosperms. *Botaniska Notiser* 128 : 148-180.

JUKES, T. H. AND C. R. CANTOR. 1969. Evolution of protein molecules. in *Mammalian protein metabolism.* : 21-132. H. N. Munros Editor. Academic Press. New York.

KATO T, KANEKO T, SATO S, NAKAMURA Y, TABATA S. 2000. Complete structure of the chloroplast genome of a legume, *Lotus japonicus*. *DNA Research.* 7(6) : 323-30

KELLOGS, E. A., R. APPELS, and R. MASON-GAMER. 1996. When genes tell different stories: incongruent gene trees for diploid genera of Triticeae (Gramineae). *Systematic botany* 21 : 231-247.

KHOSHOO, T. N. and C. L. KHUSHU. 1966. Biosystematics of Indian plants. II. The problem of *Centaurium pulchellum* complex. *Proceedings of Indian Academy of Sciences* 63 : 152-160.

KISSINGER, J .C., R. G.DONALD, A. L. MOULTON, R. GUTELL, D. P. AIELLO, N. KOHLER, S., C. F. DELWICHE, P. W. DENNY, L. G. TILNEY, P. WEBSTER, R. J. WILSON, J. D. PALMER, D. S. ROOS. 1997. A plastid of probable green algal origin in Apicomplexan parasites. *Science*. 275 (5305) : 1422

LADIGES, P. Y., G. I. McFADDEN, N. MIDDLETON, D. A. ORLOVICH, N. TRELOAR*, et al.* 1999. Phylogeny of *Melaleuca*, *Callistemon* and related genera of *Beaufortia* suballiance (Myrtaceae) based on 5S and ITS1 spacer regions of the nrDNA. *Cladistics* 15 : 151-172.

LEBRETON, P. and P. DANGY-CAYE. 1973. Contribution biochimique à l'étude taxonomique des Gentianacées. *Plantes Médicinales et Phytothérapie* 7 : 87-94.

LEMEE, A. 1931. *Dictionnaire descriptif et synonymique des genres de plantes phanérogames*. Paris

LEMIEUX, C., C. OTIS, M. TURMEL. M. 2000. Ancestral chloroplast genome in *Mesostigma viride* reveals an early branch of green plant evolution. *Nature*. 403 (6770) : 649-52

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

LEWIS, W. H. 1980. Polyploidy in angiosperms: dicotyledons. in *Polyploidy: biological relevance* : 241-268. W. H. Lewis Editor. Plenum Press. New York.

LISTON, A. and J. W. KADEREIT. 1994. Chloroplast DNA evidence for introgression and long distance dispersal in the desert annual *Senecio flavus* (Asteraceae). *Plant Systematic and Evolution* 197 : 33-41.

LÖVE, A. and D. LÖVE. 1972. *Favargera* and *Gentianodes*, two new genera of alpine Gentianaceae. *Botaniska Notiser* 125 : 255-258.

LOWE, A. J. and R. J. ABBOTT. 2000. Routes of origin of two recently evolved hybrid taxa: *Senecio vulgaris* var. *hibernicus* Ans York radiate groundsel (Asteraceae). *American Journal of Botany* 87 (8).

MADDISON, W. P. R. and D. R. MADDISON. 1997. *MacClade 3.07*. Sinauer Associates. Sunderland.

MAIER, R. M., K. NECKERMANN, G. L. IGLOI, and H. KÖSSEL. 1995. Complete sequence of maize chloroplast genome: gene content, hotspot of divergence and fine tuning of genetic information by transcript editing. *Journal of Molecular Biology* 251 : 614-628.

MASSIAS, M., J. CARBONNIER, and D. MOLHO. 1982. Chemotaxonomy of *Gentianopsis*: xanthones, C-glycosylflavonoids and carbohydrates. *Biochemical Systematics and Ecology* 10 (4) : 319-327.

MAYR, E. 1963. *Animal species and evolution*. Harvard University Press. Cambridge.

MEDEIROS, A. C., C. G. CHIMERA, L. L. LOOPE, S. M. JOE, and P. D. KRUSHELNYCKY. 2000. Notes on status and ecology of the endangered Hawaiian annual `Awiwi, *Centaurium sebaeoides* (Gentianaceae). *Pacific Science.* 54 (4) : 417.

MELDERIS, A. 1931. Genetical and taxonomical studies in the genus *Erythraea* Rich. *Acta Horti. bot. Univ. Latv.* 6 **:** 123-256.

MELDERIS, A. 1972. Taxonomic studies on the European species of the genus *Centaurium* Hill. *Botanical Journal of the Linnean Society* 65 (2) : 224-250.

MESZAROS, S. 1994. Evolutionary significance of xanthones in Gentianaceae: a reappraisal. *Biochemical Systematics and Ecology* 22 (1) : 85-94.

MESZAROS, S., J. de LAET, and E. SMETS. 1996. Phylogeny of temperate Gentianaceae: a morphological approach. *Systematic Botany* 21 (2) : 153-168.

MIEGE, J. and J. WÜEST. 1984. Les surfaces tégumentaires des graines de *Gentiana* et G*entianella* vues au microcope électronique à balayage. *Botanica Helvetica* 94 (1) : 41-59.

MUNZ, P. A. and D. D. KECK. 1959. *A California Flora*. University of California Press. Beckerley

OHYAMA, K. H., H. FUKUZAWA, T. KOHCHI, H. SHIRAI, T. SANO*, et al.* 1986. Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. *Nature* 322 : 572-574.

OKADA, M., R. WHITKUS, and T. K. LOWREY. 1997. Genetics of adaptive radiation in Hawaiian and Cook islands species of *Tetramolopium* (Asteraceae; Astereae). I. Nuclear RFLP maker diversity. *American Journal of Botany* 84 (9) : 1236-1246.

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

OMER, S. and M. QAISER. 1992. Generic limits of *Gentiana* (Gentianaceae) and related genera in Pakistan and adjoining areas along with a new genus *Kurraniana*. *Pakistan Journal of Botany* 24 : 95-106.

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

PALMER, J. D. 1995. Rubisco rules fall; gene tranfer triumphs. *BioEssays* 17 : 1005-1008.

PERROT, M. E. 1898. Anatomie comparée des Gentianacées. *Annales des Sciences Naturelles de Botanique* VII : 107-292.

PERSSON, C. 2000. Phylogeny of the tropical *Alibertia* group (Rubiaceae), with emphasis on the genus *Alibertia*, inferred from ITS and 5S ribosomal DNA. *American Journal of Botany* 87 (7) : 1018-1028.

PETIT, C., F. BRETAGNOLLE, and F. FELBER. 1999. Evolutionary consequences of diploid-polyploid zones in wild species. *Trends in Ecology and Evolution* 14 (8) : 306-311.

POWELL, W., M. MORGANTE, R. McDEWITT, G. G. VENDRAMIN, and J. A. RAFALSKI. 1995. Polymorphic simple sequence repeat regions in chloroplast genomes: applications to the population genetics of pines. *Proceedings of the National Academy of Science USA* 92 : 7759-7763.

RAMSEY, J. and D. W. SCHEMSKE. 1998. Pathways, mechanisms, and rate of polyploid formation in flowering plants. *Annual Review of Ecology and Systematic* 29 : 467-501.

RAY, J. 1724. Synopsis methodica stirpium britannicarum, in qua tum notae generum charasteristicae traduntur, tum species singulae brevinter describuntur. *Londini* III : 286.

REITH, M. and J. MUNHOLLAND. 1994. Complete nucleotide sequence of the *Porphyra purpurea* chloroplast genome. *Plant Molecular Biology Reporter* 13 : 333-335.

REVEAL, J. L., C. R. BROOME, and J. C. BEATLEY. 1974. A new *Centaurium* from the Death Valley region of Nevada and California. *Bulletin of the Torrey Botanic Club* 100 : 353-356.

RIESEBERG, L. 2000. Crossing relationships among ancient and experimental sunflower hybrid lineages. *Evolution* 54 (3) : 859-865.

RIESEBERG, L. H. and D. E. SOLTIS. 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary Trends in Plants* 5 (1) : 65-83.

RIESEBERG, L. and J. F. WENDEL. 1993. Introgression and its consequences in plants. in *Hybrid zone and the evolutionary process* : 70-109. R. G. Harrisson Editors. Oxford University Press. Oxford.

RIESEBERG, L. H., S. J. E. BAIRD, and K. A. GARDNER. 2000. Hybridization, introgression and linkage evolution. *Plant Molecular Biology.* 42 : 205-224.

ROBINSON, M., M. GOUY, C. GAUTIER, and D. MOUCHIROUD. 1998. Sensitivity of the relative-rate test to taxonomic sampling. *Molecular Biology an Evolution* 15 (9) : 1091-1098.

ROBYNS, A. 1954. Essai d'étude systématique et écologique des *Centaurium* de Belgique. *Bulletin du Jardin Botanique de Bruxelle* 24 : 349-398.

RONNIGER, K. 1916. *Centaurium* (*Erythraea*). *Mitteilungen des Naturwissenschaftlichen Vereines für Steiermark* 52 : 312-321.

RORK, C. L. 1949. Cytological studies in the Gentianaceae. *American Journal of Botany* 36 : 687-701.

SANDERSON, M. J. 1998. Estimating rate and time in molecular phylogenies: beyond the molecular clock? in *Molecular systematics of plants 2. DNA sequencing* : 242-264. D. E. Soltis, P. M. Soltis and J. J. Doyle Editors. Kluwer Academic. Boston.

SANDERSON, M. J. and J. J. DOYLE. 1992. Reconstruction of organismal and gene phylogenies from data on multigene families: concerted evolution, homoplasy, and confidence. *Systematic Biology* 41 : 4-17.

SANG, T., D. J. CRAWFORD, and T. F. STUESSY. 1997. Documentation of reticulate evolution in peonies (*Paeonia*) using Internal Transcribed Spacer of nuclear ribosomal DNA: Implications for biogeography and concerted evolution. *Proceedings of the National Academy of Science USA* 92 : 6813-6817.

SANG, T., M. J. DONOGHUE, and D. ZHANG. 1997. Evolution of Alcohol Deshydrogenase genes in Peonies (*Paeonia*) : Phylogenetic relationships of putative non-hybrid species. *Molecular Biology an Evolution* 14 (10) : 994-1007.

SATO, S., Y. NAKAMURA, T. KANEKO, E. ASAMIZU, S. TABATA. 1999. Complete structure of the chloroplast genome of *Arabidopsis thaliana*. DNA Research. 6 (5) : 283-90

SCHAUFELBER, D. and K. HOSTETTMANN. 1988. Chemistry and pharmacology of *Gentiana lactea. Planta Medica* 54 : 219-221.

SCHAUFELBERGER, D. 1986. Contribution à l'étude phytochimique des Gentianacées. Thesis. Université de Lausanne. Lausanne.

SCHMIDT, K. and K. JENSEN. 2000. Genetic structure and AFLP variation of remnant populations in the rare plant *Pedicularis palustris* (Scrophulariaceae) and its relation to population size and reproductive component. *American Journal of Botany* 87 (5) : 678-689.

SCHMITZ-LINNEWEBER, C., R. M. MAIER, J. P. ALCARAZ, A. COTTET , R. G. HERRMANN, R. MACHE. 2001. The plastid chromosome of spinach (*Spinacia oleracea*): complete nucleotide sequence and gene organization. Plant Molecular Biology. 45 (3) : 307-15

SHINOZAKI, K., M. OHME, M. TANAKA, T. WAKAZUGI, N. HAYASHIDA*, et al.* 1986. The complete nucleotide sequence of tobacco chloroplast genome: its gene organization and expression. *EMBO Journal* 5 : 2043-2049.

SIMPSON, G. G. 1951. The species concept. *Evolution* 5 : 285-298.

SOLTIS, D. E. and P. S. SOLTIS. 1993. Molecular data and the dynamic nature of polyploidy. *Critical Reviews in Plant Science* 12 : 243-273.

SOLTIS, D. E. and R. K. KUZOFF. 1995. Discordance between nuclear and chloroplast phylogenies in the *Heuchera* group (Saxifragaceae). *Evolution* 49 (4) : 727-742.

SOLTIS, D. E. and P. S. SOLTIS. 1998. Choosing an approach and an appropriate gene for phylogenetic analysis. in *Molecular systematics of plants 2. DNA sequencing* : 1-42. D. E. Soltis, P. M. Soltis and J. J. Doyle Editors. Kluwer Academic. Boston.

SOLTIS, D. E. and P. S. SOLTIS. 1999. Polyploidy: recurrent formation and genome evolution. *Trends in Ecology and Evolution* 14 (9) : 348-352.

SOLTIS, P. S. and D. E. SOLTIS. 2000. The role of genetic and genomic attributes in the success of polyploids. *Proceedings of the National Academy of Science USA.* 97 (13) : 7051-7057.

SORDAT, D. 1992. Etude des xanthones du *Garcinia gerrardii* et du *Garcinia livingstonei* (Guttiferae) et de *Polygala niykensis* (Polygalaceae). Thesis. Université de Lausanne. Lausanne.

STANDLEY, P. C. and L. O. WILLIAMS. 1969. Flora of Guatemala. *Feldiana: Bot.* VIII (4) : 263-474.

STEBBINS, G. 1947. Types of polyploids: their classification and significance. *Advances in Genetics* 1 : 403-429.

STEBBINS, G. L. 1950. *Variation and Evolution in Plants*. Columbia University Press. New York.

STIREWALT, V. L., C. B. MICHALOWSKI, W. LOFFELHARDT, H. J. BOHNERT, and D. B. BRYANT. 1995. Nucleotide sequence of the cyanelle genome from *Cyanophora paradoxa*. *Plant Molecular Biology Reporter* 13 : 327-332.

STRUWE, L., M. THIV, J. W. KADEREIT, A. S.-R. PEPPER, T. J. MOTLEY*, et al.* 1998. *Saccifolium* (Saccifoliaceae), an endemic of Sierra de la Neblina of the Brazilian-Venezuelan border, is related to the temperate-alpine lineage of Gentianaceae. *Harvard Paper in Botany.* 3 (2) : 199-214.

STRUWE, L., J. W. KADEREIT, J. KLACKENBERG, S. NILSSON*, et al.* 200*. Systematics, character evolution, and biogeography of Gentianaceae, including a new tribal and subtribal classification. in *Gentianaceae - Systematics and natural history*. L. Struwe and V. Albert Editors. Cambridge University Press. Cambridge.

SULTANBAWA, M. U. 1980. Xanthonoids of tropical plants. *Tetrahedron* 36 : 1465-1506.

SUNNUCKS, P. 2000. Efficient genetic markers for population biology. *Trends in Ecology and Evolution* 15 (5) : 199-203.

SWOFFORD, D. L. 1999. *PAUP* Phylogenetic Analysis Using Parsimony (*and other methods) Version 4.* Sinauer Associates. Sunderland.

SWOFFORD, D. L., G. K. OLSEN, P. J. WADDELL, and D. M. HILLIS. 1996. Phylogeny reconstruction. in *Molecular Systematics, second edition*. D. M. Hillis, C. Moritz and B. K. Mable Editors. Sinauer Associates. Sunderland.

SYSTMA, K. J. 1988. Taxonomic revision of the Central American *Lisianthus skinneri* species complex (Gentianaceae). *Annals of the Missouri Botanical Garden* 75 : 1587-1602.

THIV, M., L. STRUWE, and J. W. KADEREIT. 1999a. The phylogenetic relationships and evolution of the Canarian laurel forest endemic *Ixanthus viscosus* (Aiton) Griseb. (Gentianaceae) : evidence from *mat*K and ITS sequences, and floral morphology and anatomy. *Plant Systematics and Evolution* 218 : 299-317.

THIV, M., L. STRUWE, V. A. ALBERT, & J. W. KADEREIT. 1999b. The phylogenetic relationships of *Saccifolium bandeirae* Maguire & Pires (Gentianaceae) reconsidered. Harvard Papers in Botany 4 : 519-526.

THOMPSON, J. D. and R. LUMARET. 1992. The evolutionary dynamics of polyploid plants: origin, establishment and persistence. *Trends in Ecology and Evolution* 7 (9) : 302-307.

TOURNEFORT, J. P. D. 1700. Institutiones rei herbarieae. *Parisiis* 122.

TURMEL M, OTIS C, LEMIEUX C. 1999. The complete chloroplast DNA sequence of the green alga *Nephroselmis olivacea*: insights into the architecture of ancestral chloroplast genomes. *Proceedings of the National Academy of Science USA.* 96 (18) : 10248-53

TURNER, W. B. 1971. *Fungal metabolites.* Academic Press. New York.

UBSDELL, R. A. E. 1976a. Studies and evolution in *Centaurium erythraea* Rafn. and *Centaurium littorale* (D. Turner) in the British Isles; taxonomic and biometrical studies. *Watsonia* 11 : 7-31.

UBSDELL, R. A. E. 1976b. Studies and evolution in *Centaurium erythraea* Rafn. and *Centaurium littorale* (D. Turner) in the British Isles; Cytology. *Watsonia* 11 : 33-43.

UBSDELL, R. A. E. 1979. Studies and evolution in *Centaurium erythraea* Rafn. and *Centaurium littorale* (D. Turner) in the British Isles; breeding systems, floral biology and general discussion. *Watsonia* 12 : 225-232.

UDOVIC, F., G. MCFADDEN, and P. Y. LADIGES. 1995. Phylogeny of *Eucalyptus* and *Angophora* based on 5S rDNA spacer sequence data. *Molecular Phylogenetics and Evolution* 4 (3) : 247-256.

VAN DER SLUIS, W. G. 1985. Chemotaxonomical investigations of the genera *Blackstonia* and *Centaurium* (Gentianaceae). *Plant Systematic and Evolution* 149 : 253-286.

VAN DER SLUIS, W. G. and R. P. LABADIE. 1981. Secoiridoids and xanthones in the genus *Centaurium* part II: secoiridoid glucosides in *Centaurium spicatum*.. *Planta Medica* 41 : 221-231.

VAN DER SLUIS, W. G. and R. P. LABADIE. 1981. Secoiridoids and xanthones in the genus *Centaurium.* Part III: Decentapicrin A, B, C, new M-hydroxybenzoyl esters of sweroside from *Centaurium littorale*. *Planta Medica.* 41 : 150-160.

VIJVERBERG, K., T. H. M. MES, and K. BACHMANN. 1999. Chloroplast DNA evidence for the evolution of *Microseris* (Asteraceae) in Australia and New Zealand after long-distance dispersal from western North America. *American Journal of Botany* 86 (10) : 1448-1463.

VILGALYS, R. and B. L. SUN. 1994. Ancient and recent pattern of geographic speciation in the oyster mushroom *Pleurotus* revealed by phylogenetic analysis of ribosomal DNA sequences. *Proceedings of the National Academy of Science USA* 91 : 4599-4603.

VOGEL, H., S. FISCHER, K. A.VALENTIN. 1996. Model for the evolution of the plastid sec apparatus inferred from secY gene phylogeny. Plant Molecular Biology. 32 (4) : 685-92

WAKASUGI, J., S. I. TSUDZUKI, M. SHIBATA, and M. SUGIURA. 1996. A physical map and clone bank of the black pine (*Pinus thunbergii*) chloroplast genome. *Plant Molecular Biology Reporter* 12 : 227-241.

WAKASUGI, T., J. TSUDZUKI, S. ITO, K. NAKASHIMA, T. TSUDZUKI*, et al.* 1994. Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of black pine *Pinus thunbergii*. *Proceedings of the National Academy of Science USA* 91 : 9794-9798.

WENDEL, J. F. and J. J. DOYLE. 1998. Phylogenetic incongruence: window into genome history and molecular evolution. in *Molecular systematics of plants 2. DNA sequencin*g : 265-296. D. E. Soltis, P. M. Soltis and J. J. Doyle Editors. Kluwer Academic. Boston.

WENDEL, J. F., A. SCHNABEL, and T. SEELANAN. 1995. Bidirectionnal interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). *Proceedings of the National Academy of Science USA* 92 : 280-284.

WILEY, E. O. 1981. *Phylogenetics*. J. Wiley and sons Editors. New York.

WOLFE, K. H., C. W. MORDEN, and J. D. PALMER. 1992. Function and evolution of a minimal plastid genome from a non-photosynthetic parasitic plant. *Proceedings of the National Academy of Science USA* 89 : 10648-10652.

WOLFENDER, J. L. (1993). Investigation phytochimique et analyse par chromatographie liquide - spectrométrie de masse de quatre espèces du genre *Chironia* (Gentianaceae). Thesis. Université de Lausanne. Lausanne.

WOLFENDER, J.-L. and K. HOSTETTMANN. 1992. Search for xanthones in *Chironia* species: a rapid method for the screening of crude extracts. *Planta Medica* 58 (1) : A673-A674.

WOLFENDER, J.-L. and K. HOSTETTMANN. 1993. Liquid chromatographic-UV detection and liquid chromatographic-thermospray mass spectometric analysis of *Chironia* (Gentianaceae) species. A rapid method for the screening of polyphenols in crude plant extracts. *Journal of Chromatography* 647 : 191-202.

WOLFF, K. and B. SCHAAL. 1992. Chloroplast DNA variation within and among five *Plantago* species. *Journal of Evolutionary Biology* 5 : 325-344.

WRAY, G. A., J. S. LEVINTON, and L. H. SHAPIRO. 1996. Molecular evidence for deep Precambrian divergences among metazoan phyla. *Science* 274 : 568-573.

WU, C.-I. and W.-H. LI. 1985. Evidence for higher rates of nucleotide substitution in rodents than in man. *Proceedings of the National Academy of Science USA* 82 : 1741-1745.

WULFF, H. D. 1937. Karyologische Untersuchungen an der Halophytenflora Schleswig-Holsteins. *Jahrbüchen für Wissens Chaftliche Botanik* 84 : 820.

YANG, Z., N. GOLDMAN, AND A. E. FRIDAY. 1994. Comparison of models for nucleotide substitution used in maximum likelihood phylogenetic estimation. *Molecular Biology an Evolution* 11: 316-324

YUAN, Y.-M. 1993. Seedcoat micromorphology and its systematic implications for Gentianaceae of Western China. *Botanica Helvetica* 103 : 73-82.

YUAN, Y.-M. and P. KÜPFER. 1995. Molecular phylogenetics of the subtribe Gentianinae (Gentianaceae) inferred from the sequences of the Internal Transcribed Spacer (ITS) of nuclear ribosomal DNA. *Plant Systematic and Evolution* 196 : 206-226.

YUAN, Y.-M. and P. KÜPFER. 1997. The monophyly and rapid evolution of *Gentiana* Sect. *Chondrophyllae* Bunge s.l. (Gentianaceae) : evidence from the nucleotide sequences of the Internal Transcribed Spacer of nuclear ribosomal DNA. *Botanical Journal of the Linnean Society* 123 : 25-43.

YUAN, Y.-M., P. KÜPFER, and J. J. DOYLE. 1996. Infrageneric phylogeny of the genus *Gentiana* (Gentianaceae) inferred from nucleotide sequences of the Internal Transcribed Spacer (ITS) of nuclear ribosomal DNA. *American Journal of Botany* 83 (5) : 641-652.

ZELTNER, L. 1963. Recherches sur quelques taxa méditerranéens du genre *Centaurium* Hill (Gentianaceae). *Bulletin de la Société Neuchâteloise des Sciences Naturelles* 86 : 93-100.

ZELTNER, L. 1970. Recherches de biosystématique sur les genres *Blackstonia* Huds. et *Centaurium* Hill (Gentianaceae). *Bulletin de la Société Neuchâteloise des Sciences Naturelles* 93 : 1-164.

ZELTNER, L. 1978. Notes de cytotaxonomie sur les genres *Blackstonia* Huds. et *Centaurium* Hill en Crête. *Bulletin de la Société Neuchâteloise des Sciences Naturelles* 86 : 93-100.

ZELTNER, L. 1978. Recherches sur le *Centaurium bianoris* (Sennen) Sennen. *Revue de Biologie et Ecologie Mediterranéenne* 5 (1) : 51-58.

ZELTNER, L. 1980. Contribution à la cytotaxonomie en Iran de trois espèces du genre *Centaurium* Hill. *Revue de Biologie et Ecologie Mediterranéenne* 7 (1) : 57-62.

ZELTNER, L. 1985. Etude cytotaxonomique et cytogéographique du *Centaurium pulchellum* (Swartz) Druce, sensu lato. *Botanica Helvetica* 95 : 47-57.

ZELTNER, L. 1987. Contribution a l'étude cytotaxonomique et cytogéographique des genres *Blackstonia* Huds. et *Centaurium* Hill en URSS. *Bulletin de la Société Neuchâteloise des Sciences Naturelles* 110 : 17-35.

ZELTNER, L. 1990. Sur quelques *Centaurium* Hill (Gentianaceae) d'Egypte, de la Péninsule Arabique et des Indes. *Bulletin de la Société Neuchâteloise des Sciences Naturelles* 113 : 255-270.

ZELTNER, L. 1991. Contribution à l' étude cytogéographique des genres *Blackstonia* Huds. et *Centaurium* Hill (Gentianaceae) en Turquie, à Rhodes et à Chypre. *Bulletin de la Société Neuchâteloise des Sciences Naturelles* 114

ZIMMER, E. A., S. L. MARTIN, S. M. BEVERLEY, Y. W. KAN, and A. C. WILSON. 1980. Rapid duplication and loss of gene coding for the chains of hemoglobin. *Proceedings of the National Academy of Science USA* 77 : 2158-2162.

ZUCKERKANDL, E. and L. PAULING. 1965. Evolutionary divergence and convergence in proteins. in *Evolving genes and proteins*. V. Bryson and H. J. Vogel Editors. Academic Press. New York.

Chapter 1

Phylogenetic relationships between North-American species of *Centaurium* (Gentianaceae – Chironiinae) inferred from molecular and karyological data

Phylogenetic relationships between North-American species of *Centaurium* (Gentianaceae - Chironiinae) inferred from molecular and karyological data

ABSTRACT

North-American species of *Centaurium*, also known as centaury, consist of 34 species, mainly distributed in the western part of the United States and Mexico. Chromosome counts of 139 populations (28 species), native to the Americas, revealed some incongruence with previous surveys and tend to support a particular origin of the American set. The latter was then submitted to molecular analyses, along with European naturalized species and their native relatives. The monophyly of the New World group, except for two species, was confirmed using phylogenetic reconstructions based on DNA sequences from both nuclear ribosomal DNA (ITS) and chloroplast DNA (*trn*L intron and *trn*L-*trn*F intergenic spacer regions). Inferred trees were similar in topology and defined strongly supported clades, which were congruent with the chromosomal groups. Nevertheless, in each well-delimited set, morphological variability was high at the population level, mainly due to ecological conditions and, occasionally, to hybridization. Karyological and molecular data sets show three main biogeographic patterns for the American species of *Centaurium*: two clades were centered in the Texas region (n = 20 and n = 21) and Mexico (n = 21 and n = 22), respectively, whereas an unresolved assemblage (n = 17) occurred in California. Under the assumption of a molecular clock, the group showed a considerable diversification in the early Pliocene (ca. 6 mya), with a basal Californian group from which arose the Texan and Mexican clades. Geological events such as desert formation or mountain orogenies have created insuperable barriers that today separate the three major groups.

INTRODUCTION

Within the subtribe *Chironiinae* of the Gentianaceae - *Chironieae*, *Centaurium* (~50-60 species) is the largest genus (Struwe *et al*., 200*). This genus, commonly called centaury, comprises mostly short-lived (annuals or biennials) herbaceous species and is characterized by the coiling of the anthers after pollen release: it goes from only half-twisted to up to four gyres, depending of the anther length. The circumscription of different *Centaurium* taxa is difficult because morphological characters discriminate the species poorly. Moreover, a strong morphological plasticity occurs, depending of the ecological conditions, and finally, natural hybridization has sometimes been reported, obscuring species delimitation (Ubsdell, 1976, a, b, 1979; Zeltner, 1970, 1978). This confusion has led to the naming of many taxa *Centaurium* (114 epithets are listed in Index Kewensis) and some nomenclatural differences between authors who have contributed to the systematics of the genus (e. g., Grisebach, 1839; Gilg, 1895, Hegi, 1927; Melderis, 1931, Robyns, 1954, and Zeltner, 1970). The taxa are widely distributed in wet and unstable habitats, where the competition is weak, and can be encountered along roadsides, on stream banks, in fields and pastures or in open forest, in most continental and insular parts of the northern Hemisphere. Two main regions of diversification are recognized in the Old World and New World, respectively. Old World species are centered in the Mediterranean basin and radiate in a northeasterly manner: in this region, diploid and polyploid species (ca. 20) occur. All the New World species (~34 spp.) are polyploids and occur in California, Texas and Mexico. However, despite a series of works, the systematics of the North American taxa is still confused (Dunn, 1967; Reveal *et al.*, 1974; Broome, 1976, 1978, 1981; Turner, 1993). Broome (1973) included most of the North-American species in the European section *Centaurium* except for a set of five Mexican species forming section *Gyrandra*. Today, Old and New World species of *Centaurium* compose two distinct groups, each of them showing a high level of diversification. American centauries seem to form well-marked geographic taxa, and their chromosome numbers are different from the Eurasian species, mainly in the occurrence of dysploid and high polyploid chromosome numbers previously reported (Broome, 1976; Turner, 1993). Therefore, karyological variation needs to be fully examined in New World taxa since (1) morphologically closely related taxa, in some cases, exhibit different chromosome numbers (Broome, 1976), (2) distant taxa may share the same number, and (3) some of the New World species have not yet been karyologically investigated. Moreover, the large amount of morphological plasticity shared by Mexican and Texan species lead to potential taxonomic confusion (Broome, 1976; Turner, 1993).

The aim of this study was first, to fully examine the variation of chromosome numbers within North American *Centaurium*. To do so, a large set of field-collected populations was investigated for chromosome number.

The other goal of the present paper was to reconstruct the phylogeny of the North American *Centaurium* using molecular markers. Sequences of the ITS1 and ITS2 regions of the 18S-25S nrDNA might provide satisfying information to resolve phylogenetic relationships. These genes have been used in the Gentianaceae at the intergeneric and infrageneric level (Yuan and Küpfer, 1995; Yuan *et al.*, 1996; Hungerer and Kadereit, 1998; Thiv *et al.*, 1999). The combined sequences of the *trn*L (UAA) intron and *trn*L - *trn*F (GAA) spacer were used in this study (1) to interpret incongruence in phylogenetic reconstruction, due to hybridization or chloroplast capture (Rieseberg and Soltis, 1991; Soltis and Kuzoff, 1995; Sang *et al.*, 1997a, b), and (2) to identify maternal parent of putative hybrids (Wendel *et al.*, 1995). Finally, molecular inferences and karyological data were combined to infer phylogeography of this group.

MATERIAL AND METHODS

Taxon sampling---*Centaurium* species were collected in the western part of the USA (Arizona, California, Nevada, New Mexico, Texas and Utah) and in Mexico in 1996 to 1999. Voucher specimens were deposited in the herbarium of the University of Neuchâtel, Switzerland (NEU). We were able to obtain most of the representative native taxa of this group, covering ca. 95% of the species of *Centaurium* native to the United States, and ca 80% for Mexico. We have also included introduced Eurasian species (*C. erythraea*, *C. pulchellum*, *C. spicatum*) in the analyses (table 1). The choice of the outgroup was difficult. Previous analyses strongly suggested that *Centaurium*, *Orphium*, *Chironia*, *Eustoma* and *Sabatia* form a clade, close to a *Blackstonia* - *Ixanthus* clade (Thiv *et al*., 1999). At the moment, relationships within the entire genus *Centaurium* and between close relatives (*Sabatia* and *Chironia*) are not fully resolved (Chapter 3). This is why *Blackstonia*, belonging to the sister clade, has been selected as the outgroup in the phylogenetic trees.

Karyological study---Chromosome numbers were obtained from the analysis of pollen mother cells during meiosis (tetrads), or from mitosis in ovary tissues. Floral buds were fixed directly in the field with a freshly prepared solution containing absolute alcohol and glacial acetic acid, in a final 3:1 concentration. When fresh buds were not available, chromosome numbers were obtained on the mitosis of root tips from seeds germinated on a wet filter paper in Petri dishes. In that case, root tips were pre-treated with a saturated aqueous solution of α-

bromonaphtalene (1.5 hours), and fixed for at least 24 hours with Carnoy 3:1. The different tissues were then stained with carmin for 1 hour, gently heated for 2 minutes and squashed in acetic carmin. Observations and photographs were performed with a Leitz Aristoplan microscope coupled with a Leitz Orthomat E system for microphotographs. For each population, chromosomes were independently counted by the two authors, one of them being uninformed about the identity of the preparation (blind count). At least ten different unambiguous and congruent observations were performed separately to identify the chromosome number of a population.

DNA extraction, amplification and sequencing---Eighty-five accessions of 32 *Centaurium* species were sequenced for the combined ITS1 and ITS2 regions and 56 for the cpDNA region. For most taxa, fresh leaves collected in natural populations of North America and Mexico were dried with silica gel; in some cases, herbarium samples were used. Total DNA extraction was performed from leaf tissues using either the CTAB method (Doyle and Doyle, 1987) with an additional cleaning step using ammonium acetate 7.5 M, or using a DNeasy® kit (Qiagen, Basel). Double stranded nrDNA and cpDNA were amplified by symmetric PCR using a Perkin-Elmer or a Biometra thermo-cycler. The conditions parameters were as follow: 1 cycle of 3 minutes at 94°C linked to 30 cycles of 10 sec at 94°C, 20 sec at 50-55°C, 1 min 30 at 72°C and then 4 min at 72°C to complete primer extension. Primer ITS1 (GGA AGT AGA AGT CGT AAC AAG G) and ITS2 (TCC TCC GCT TAT TGA TAT GC), respectively binding to the 3' end of the 18S rRNA gene and 5' end of the 26S rRNA gene (White *et al.*, 1990) were optimally used at a final concentration of 0.2 pmol/µl. In the same conditions, primers *trn*L'c' (CGA AAT CGG TAG ACG CTA CG) and *trn*L'd' (GGG GAT AGA GGG ACT TGA AC) were used to amplify the chloroplast *trn*L (UAA) intron; for *trn*L - *trn*F spacer, primers *trn*L'f'(ATT TGA ACT GGT GAC ACG AG) and *trn*L'e' (GGT TCA AGT CCC TCT ATC CC) were used (Taberlet *et al.*, 1991). PCR products were first checked on a 1% agarose gel stained with ethidium bromide (in order to evaluate the quality and quantity of the amplified templates) and then purified using the QIAquick PCR purification kit (Qiagen). For difficult specimens (e.g. herbarium), a lower annealing temperature (50°C) was used, which often amplify fungi DNA, providing multiple ITS fragments. The latter were separated directly on the gel, by cutting and purifying the desired region with the QIAEX II Gel Extraction Kit (Qiagen). When this separation was not possible, PCR products were cloned as follows: PCR products were ligated with plasmids (pGEM(r)T vector) and transformed into *Escherichia coli* competent cells, then selected on LB plates containing ampicillin and X-gal (pGem(r)T and pGem(r)T easy Vector Systems, Promega). Sixteen white *E. coli* clones (in which ligation was successful) were picked and cultured on LB containing ampicillin broth, for isolating plasmids.

Purified PCR products or plasmids were automatically sequenced on an Applied Biosystems 310 DNA sequencer, using the dideoxy chain termination technique and a BigDye Terminator Cycle Sequencing Ready Reaction Kit (Perkin Elmer). A 5 µl reaction was performed with 2µl of premix (containing the 4 labeled terminators, the deoxynucleotide triphosphates, the AmpliTaq DNA, $MgCl_2$ and Tris-HCl buffer), 0.2 µl of primer (10 ng/µl) and 2.8 µl of amplified DNA (about 10 ng/µl). The cycle sequencing program was performed on a thermo-cycler (Biometra or Perkin Elmer) : 30 cycles of 10 sec at 96°C followed by 4 min at 60°C.

Phylogenetic analysis---All data sets were analyzed with PAUP* 3.1.1 (Swofford, 1993) and the PAUP* 4.0 beta 3 version (Swofford, 1999) using Fitch parsimony methods (i.e. assuming unordered, equally weighted, character states). DNA sequences were aligned using the Clustal program (Higgins *et al.*, 1996) with manual modification to optimize the alignment. Ambiguous regions (i.e. where many different alignments were possible) were generally excluded from the analysis. However, an analysis with all the characters was also performed, in order to quantify the effect of removing ambiguous regions. Heuristic search, with TBR branch swapping, MULPARS and ACCTRAN options ON, was employed using the simple sequence addition options (Steepest Descent On). In all the analysis, characters states were weighted equally and unambiguous gaps were either deleted, treated as missing data or as new state (fifth base). In order to test the robustness of individual branches, bootstrap values (Felsenstein, 1985) and decay indices (Bremer, 1988) were calculated. Bootstrap analyses consist of 100 replicates with resampling, using Heuristic Search, simple taxon addition and TBR branch swapping, saving 100 trees per replicate. All clades with at least 70 % bootstrap values could be considerate as well supported (Hillis & Bull, 1993). The Bremer's support (or decay index) invoking topological constraints, was calculated using Heuristic Search, 100 random sequence addition (Maxtree limit reset to 100) and TBR branch swapping, and was performed with TreeRot version 2 (Sorenson, 1999). The Maxtree limit restriction imposed on both bootstrap (BS) and decay index (DI) may not supply the best estimate of branch support. However, different tests (not shown) performed with no Maxtree limits (reaching limits of computer memory) do not substantially change BS or DI values. The amount of phylogenetic signal was reflected by the classical descriptive statistics (consistency index C, retention index R and derived rescaled index RC) along with the skewness test, by calculating the g1 value (Hillis and Huelsenbeck, 1992). The latter is obtained by computing the tree length distribution of 10000 random trees using the Random Trees option in PAUP*. The g1 statistic evaluates the amount of phylogenetic signal in the data set, in comparison of random noise. Separate analyses were first performed on the ITS matrix (85 sequences of ITS and ITS2) and on the *trn*L-F data set (56

sequences). In order to combine data in a single ITS-*trn*L-F analysis, respective matrices were reduced to 51 sequences. The partition homogeneity test (PHT), was used to assess whether or not the respective data sets were only partitions of a same large assemblage (Farris *et al.*, 1995). A likelihood-ratio test was performed, under the HKY85 model of molecular evolution, to test for a molecular clock within the North American *Centaurium* (Felsenstein, 1981; Hasegawa *et al.*, 1985; Goldman, 1993; Yang *et al.*, 1995; Huelsenbeck and Crandall, 1997). The analysis was performed on a reduced subset of 22 representative species (chosen from the best supported clades on the combined tree) for the ITS data set alone, because it was found to contain more phylogenetic information.

RESULTS

Karyological study---A total of 139 populations was analyzed, corresponding to ca. 28 species. No diploid or highly autopolyploid taxa were detected in our investigations. The present study reveals striking differences with the literature, mainly in the occurrence of previously reported high chromosome numbers. Of the 28 species surveyed, 14 have either newly reported (9) or different (5) chromosome numbers (Table 2). Haploid complement mainly ranges from n=17 to n=22; however, three taxa have n=36 or n=37 (Fig. 7). The 2n=34 assemblage comprises morphologically similar species, with generally big flowers, and occurring in California (*C. trichanthum*, *C. namophilum* var. *namophilum*, *C. namophilum* var. *nevadense*, and *C. venustum* subsp. *venustum*). *Centaurium capense*, an endemic from the Cape Province of Baja California, is the only American taxon hitherto known to have a chromosome count of n=18 chromosomes. The n=20 set is frequent in the Texan populations (6 species) but can also be encountered in three Californian (or locally introduced) taxa. *Centaurium exaltatum* accessions, collected on the Pacific range, have been found to be tetraploid and not octoploid (n=40), as previously reported (Broome, 1978). Two hypertetraploid numbers (2n=42 and 2n=44) can be found in Mexican species. They generally characterize well-defined species, which are relatively easy to recognize in the field (e. g. *C madrense*, *C. martinii* or *C. wigginsii*). On the other hand, the case of *C. quitense* and *C. nudicaule* is problematic as different populations of these taxa show either n=21 or n=22 chromosomes. Until now, no morphological or biogeographical variants have been described for these species. The highest numbers (2n=72) were reported for two different species: *C. tenuifolium* and *C. brachycalyx*. This number, which may be viewed as a derivative of x=18, supports a particular position for these taxa. Lastly, 2n=74 has been reported in one

taxon, previously described as *C. exaltatum* (n=20) or *C. namophilum* var. *nevadense* (n=17). This additive number supports allopolyploid processes in this group.

Characteristics of the ITS region in *Centaurium*---The main characteristic of the ITS region, excluding the 5.8S cistron, are shown in Table 3. Total length of the ITS sequences ranged from 454 to 464 base pairs (bp) with an average GC content of 61.4 %. The ITS1 and ITS2 spacers are about the same length (~230 bp) and have a similar GC content (58 % - 62.9 % vs. 59.6 % - 65.2 %, respectively). The introduction of 75 indels was necessary in the multiple alignment of 85 sequences, producing a matrix of 479 characters of which 179 positions were parsimony informative (37.4%). The deletion of 17 ambiguous characters due to difficulties in alignment resulted in a final matrix of 462 characters containing 65 indels (coded as presence / absence characters) and 170 informative characters.

Characteristics of the *trn*L-F region in *Centaurium*---Sequences of the *trn*L intron and *trn*L-F spacer were similar in length (391-416 bp and 368-394 bp) and mean GC content (36.6% and 33.3%), respectively. For phylogenetic analyses, 51 ambiguous characters were excluded from the matrix and informative indels were either coded (presence / absence character) or not. In the first case, the matrix was 764 bp, with 24 indels (3.1%) and 45 parsimony informative characters (5.9%), and the non-coded matrix contained 793 bp with 58 indels (7.3%) and 42 informative characters (5.3%).

Characteristics of the combined ITS-*trn*L-F region in *Centaurium*---Because the respective data sets were not incongruent to each other (PHT: P = 0.3), we combined 51 sequences in two matrices. The coded matrix was 1226 bp, with 80 indels and 206 informative characters, whereas the non-coded one was 1255 bp with 114 indels and 203 informative characters.

Phylogenetic reconstruction of *Centaurium*---In the separate and combined analysis, the g1 statistic test, including all the characters, show relatively high negative values (Table 3). These respective values of 10000 random trees indicate significant skewness (g1 ranging from - 0.56 to - 0.79) of the length distribution rather than a normal random distribution: in the latter case, g1= - 0.09 for 250 variable positions and 25 or more taxa, with $P<0.01$ (Hillis and Huelsenbeck, 1992).

ITS analysis--Heuristic searches performed on the ITS matrix, excluding 17 ambiguous sites and treating gaps as missing data, reached computer memory limits, saving 9000 MP trees with 448 steps, C = 0.659 and R = 0.929. The strict consensus of these trees is shown in Figure 1. When gaps were treated as new state (fifth base), heuristic searches generated 1367 most parsimonious trees (L = 556, C = 0.640 and R = 0.916). Strict consensus trees from different analyses had quite the same topology, with several well-supported clades. The most basal clade in *Centaurium* (clade A) comprises all the Eurasian taxa along with two American relatives, *C. muehlenbergii* and *C. capense*. The highly supported clade B (BS=100 - DI=8) is composed of three accessions of two Mexican species: *C. tenuifolium* and *C. brachycalyx.* A third group (clade C) is highly supported (BS=100 - DI=12), and comprises different accessions of *C. spicatum*. The present study does not support the inclusion of the European section *Spicaria* in the Old World clade. Lastly, the remaining American species occurring in Western United States and Mexico, form a monophyletic assemblage (clade D: BS=80 - DI=3). The Californian group is polytomized and mainly comprises species with 2n=34 chromosomes, along with *C. exaltatum* sensu lato (i.e. populations with 2n=40 and 2n=74). A second well-supported American clade (BS=82 and DI=3) is composed of eight species, six collected in central Texas (n = 20) and two from New Mexico and Arizona (n = 21). Finally, Mexican species form a moderately resolved clade (BS=72 - DI=1) with three main subdivisions. Morphologically well-marked species such as *C. madrense*, *C. setaceum*, *C. martinii* and *C. nudicaule* (n=21), recognizable by the degree of inflorescence branching, form a first group (BS=69 - DI=1). *Centaurium strictum*, *C. wigginsii* and *C. pusillum* (n=22) are positioned in a second clade (BS=96 - DI=4), whereas several accessions of *C. quitense* (n=21 or n=22) compose a well-marked set (BS=98 - DI=4). Other analyses performed on the ITS matrix, but keeping all the characters or treating gaps as a new state, depicted nearly the same clade with similar branch supports (not shown).

*trn*L-F analysis--Heuristic searches performed on the entire *trn*L-F data set reached computer limits and gave 35900 shortest trees of 113 steps with similar topologies (C = 0.841, R = 0.923). The resolution of the strict consensus tree is relatively weak compared to ITS, but the main biogeographical groups found with the ITS matrix are supported by the *trn*L-F analysis. Bootstrap and decay index values ranged from 51 to 100 % and from 1 to 3, respectively, the higher scores confirming the clades found in the ITS analysis. Few incongruence were found with ITS data, except for the position of the (*C. brachycalyx* - *C. tenuifolium*) clade as the sister group of *Blackstonia*, in the majority rule consensus tree (not resolved in the strict consensus tree).

Combined analysis--Both ITS and *trn*L-F data sets do not show significant differences in their respective topologies; furthermore, PHT value does not contradict their phylogenetic congruence. Heuristic searches, with

simple sequence addition, gaps treated as missing data, and informative indels either coded or not (collapse option OFF), resulted in 972 most parsimonious trees of 529 and 523 steps, respectively. The strict consensus tree is well resolved, except for the Californian group, and shows clades supported by bootstrap values ranging from 62 to 100 % (Fig. 3). The general topology of the combined tree is close to the ITS one, with only a few modifications: *C. maryannum* (n=21) is much closer to *C. multicaule* (n=21) than to *C. arizonicum* (n=20) as suggested by ITS data alone. Generally, bootstrap values and decay index of the combined analysis are higher than those obtained for each data set alone.

Molecular clock hypothesis--Using the log-likelihood ratio test suggests that the null hypothesis of a clock like model should not be rejected (likelihood score ~ 22 and chi-square Critical Value = 31 with df = 20). Therefore, the branch lengths of inferred phylogenetic trees can be considered to be approximately proportional to time. For the fossils evidences are rare for annuals plants, and hitherto non-reported in *Centaurium*, molecular clock calibration was performed with particular geological events (see below).

DISCUSSION

Chromosome evolution---Geographical groups in American centauries generally display a uniform chromosome number: the n=17 set is present in California, the n=20 in Texas, while n=21 and n=22 characterize the Mexican clade. When we mapped these number on the strict consensus tree of the combined analysis (Fig. 3), we obtained a pattern of evolution from n=17 to n=20 in the California range and from n=21 to n=20 and n=22 in the Texan and Mexican region, revealing several episodes of dysploidy in this group. ACCTRAN optimization favors the n=20 state to be ancestral, since it appears in the basal Californian set within the ingroup (*C. exaltatum*), and in the Eurasian species. Mechanisms of chromosome breakage and rearrangement have occurred during speciation in the American species. Dysploidy could be the result of Robertsonian fusion in the n=17 group since the chromosomes appear to be larger in this group than in the other groups (personal observations). However, this suggestion needs to be fully investigated with future DNA content analysis and chromosome banding. New World taxa karyotypes, on the other hand, do not show either marked differences in chromosome size or meiotic anomalies. No multivalent was observed in prophase, and regular pairing along with a good segregation at the anaphase seems to be the rule in the whole ingroup. Sometimes, aggregations occur but this is probably the consequence of the strong chromosome interactions observed for all the New World species. Our observations of

chromosome meiotic behavior suggest that American centauries have attained a relative stability in their rearranged karyotype.

The chromosome aggregation was the main difficulty in chromosome counts and may explain some discordance between our results and the literature, particularly for the Texan and Mexican species. Nevertheless, it fails to resolve the problem of high polyploids previously described (e.g. n=40 in *C. exaltatum*, n=41 in *C. beyrichii* and n=42 in *C. calycosum*). In the first case, the n=40 is close to the n=37 we have reported for *C. sp.nov.1* and minor chromosome aggregation may reflect this difference. Other high numbers may be viewed either as erroneous reports or as small octoploid populations of *C. beyrichii* and *C. calycosum*, along with very narrow hybrid zones we failed to detect.

Phylogenetic inferences

Eurasian native species--Molecular approaches, based on nuclear and chloroplast DNA, suggest an early separation between the Eurasian set selected for the present study and the American clade. This supports previous hypotheses based on morphological and karyological data (Broome, 1973; Zeltner, 1970). According to these authors, the Eurasian and American groups have evolved separately for much of their history. Putative diploid species of *Centaurium* were supposed be present in North America during the early Tertiary, as a part of the madrean Tertiary migration to California from the Mediterranean (Zeltner, 1970). A great divergence then occurred in the New World species, implying polyploidy and dysploidy phenomena, and resulting in a new group which, presently, differs from the ancestral Mediterranean lineage by a "constellation of characters" (Broome, 1973: *op cit*).

The Eurasian clade A comprises a few species introduced to the Americas along with two taxa hitherto considered as having a New World origin: *C. muehlenbergii* and *C. capense*. The former species has been subject to many systematic controversies and a more detailed study is presented in Chapter 5. Divergences mainly rely on the holotype description and on early confusions with some European species (e. g. *C. tenuiflorum* and *C. pulchellum*), leading to erroneous concepts of *C. muehlenbergii* with many synonyms (Abrams, 1951; Holmes and Wivagg, 1996; Pringle, unpublished data). Present molecular data support the inclusion of *C. muehlenbergii* in the section *Centaurium* (close to *C. erythraea*) and reject the synonymy with *C. tenuiflorum*. The other American species, *C. capense*, belongs to the *C. pulchellum* group, and may have been introduced in the Americas. An examination of herbarium samples would be necessary to evaluate the age and history of such possible introduction. Morphological characters (e. g. general outlook, particular division of the style found in

Eurasian species), chromosome number identity (n = 18; 2n = 36), and similar flowering period (June to August, instead of December-March for the remaining Mexican species) support the present phylogenetic conclusions. Nevertheless, *C. capense* appears to be a well-differentiated species and today inhabits different areas of Baja California, notably the Cape region (Johnston, 1924). The origin of this taxon remains obscure and two hypotheses are possible. First, *C. capense* could have resulted from migration of a *C. pulchellum* ancestor, from Europe to different zones of the American continent, via the Bering land bridge. The occurrence of *C. pulchellum* in northeastern China supports this hypothesis. On the other hand, a long distance dispersal of this taxon, from Eurasia to different regions of the Americas, cannot be rejected. The examination of different herbarium specimens, collected in West Indies and South America (e.g., Bermuda, 1854, Holton. 1744-73 (K); Bermuda, 1872. 1744-19 (K); Peru, near Lima, 1862, Nation. 1744-73 (K)), confirms the presence of *C. pulchellum* (often called *C. chilense* or *C. lomae*) in several separated areas of the Americas. This particular repartition better fits with the long-range dispersal scenario. Moreover, the mainly coastal repartition of *C. pulchellum* in North America (Holmes and Wivagg, 1996) supports this view . Since the native American origin of *C. capense* and *C. muehlenbergii* (see Chapter 5) is presently subject to caution, we propose to exclude them from the American group that, until now, refers to clade D only.

Section *Spicaria*--The section *Spicaria* falls between the Eurasian (clade A) and the American (clade D) groups in the ITS analysis (Fig. 1). These spicate-flowered species are centered in the Mediterranean basin and Russia, but can be also found in Japan, Australia and Hawaii to Eastern Island. Molecular, morphological (spike-like cyme inflorescence, subsessile flowers and subcapitate stigmata) and karyological (n=11) evidences support the particular identity of this widespread section. A detailed study of this group is under way, revealing the occurrence of tetraploids (2n=44) in the Southern Hemisphere (Chapter 2). In North America, the diploid variety of *C. spicatum* has been introduced, and appears to be well established only in Nantucket, Massachusetts (Pringle, personal communication).

Section *Gyrandra*--One of the goals of this study was to infer the relationships between all *Centaurium* species present on the North American continent. Karyological investigations and molecular data support the exclusion of this particular set of Mexican species (*C. brachycalyx* and *C. tenuifolium*), from the New World clade. A chromosome number of 2n=72 and morphological characters such as (1) the large-sized nearly rotate corolla (not in *C. brachycalyx*); (2) an elliptic to oval capsule; (3) stems and leaves minutely papillose, and (4) outer corolla lobes viscid to touch, characterize these taxa. The systematic position and rank of section *Gyrandra* will not be

discussed here, but these preliminary results clearly support its exclusion of the New World *Centaurium* clade (Chapter 3).

Californian group--Despite a common haploid chromosome number (n=17) and a similar gross morphology (e. g. generally large corollas with a white throat), the Californian assemblage might not form a clade. This group is unresolved on the strict consensus tree, but each separated subset of Californian taxa is well supported (Fig. 3). *Centaurium trichanthum* and *C. namophilum* are localized in restricted areas, in NW and SE directions of the Sierra Nevada Mountains axis, and form a basal clade. Molecular and karyological data help to discriminate between *C. namophilum* var. *nevadense* (n=17) and *C. exaltatum* (n=20), and support the occurrence of a new allopolyploid species, *C. sp.nov.1* (n=37). Its systematic status will be further discussed in a latter paper (Mansion, unpublished). In the *trn*L + *trn*LF spacer analysis, *C. exaltatum* and *C. sp.nov.1* form a clade, which comes close to *C. venustum* subsp. *venustum* in ITS analyses. The latter possesses larger flowers and ranges from Cismontane southern California to Baja California, where *C. sp.nov.* generally occurs. The last taxon, *C. venustum* subsp. *abramsii* replaces subsp. *venustum* in open meadows of the Sierra Nevada Mountains. Present phylogenetic analyses argue in favor of species rank for this taxon. Moreover, previous experimental crosses failed between subspecies *venustum* and *abramsii* suggesting reproductive isolation between these two taxa (Broome, 1973). Presently, we propose to use the binomial *C. abramsii* (Munz) Mansion instead of *C. venustum* (Gray) B.L. Robins. subsp. *abramsii* Munz

Texan clade--Combined analyses (Fig. 3) tend to support descending dysploidy, from n=21 (*C. multicaule* and *C. maryannum*) to n=20 chromosomes (all the remaining species). Texan and Mexican species share a common ancestor with 2n=42, from which derives *C. multicaule*, present in both areas, and *C. maryannum*, isolated in restrictive areas of southern New Mexico. Hence, chromosome loss occurred in the morphologically close *C. arizonicum* and in all the taxa present in the Edward Plateau region of Texas (n=20). *Centaurium beyrichii* is located from the northern part of the Texan state to Oklahoma and is easily recognized by several morphological features such as vegetative stolons produced from true roots, anthers not forming a cluster at the opposite side of the deflected style, and a subcapitate stigma. This attractive species, forming pink densely-flowered bouquets along Texan roadsides, seems to be relatively ancient regarding the Texan clade topology, and thus cannot be viewed here as a recent allopolyploid (Broome, 1973). (Previously known as a variety of *C. beyrichii)*, *Centaurium glanduliferum*, a rare endemic of the Trans Pecos region, often bears minute glands on the stem, leaves and calyx (a plastic character we did not observed on individuals cultivated in the greenhouse). Our study supports the species rank of this taxon, following Turner (1993). The clade consisting of *C. breviflorum*, *C.*

texense and *C. calycosum* is not fully resolved in the ITS analysis (BS=67 - DI-1), and these three entities are not easily distinguishable in the field. Flower size and leaf shape, which are the main distinctive characters between these species, are generally of low taxonomic values in *Centaurium*, because of frequent clinal variation. However, each one occurs in a distinctive ecogeographic setting (Turner, 1993), and a subspecific rank may possibly fit these taxa. *Centaurium breviflorum* is found in southern Texas (sandy soils), *C. texense* in the eastern part of the Edward Plateau (calcareous seeps) and *C. calycosum* in the more western portion of this Plateau (rocky soils). Because of the difficulty to discriminate between the varieties previously described, we prefer to consider *C. calycosum* as a heterogeneous assemblage of variable populations (Turner, 1993).

Mexican clade--The Mexican clade can be roughly divided in three groups. A first group (BS=69 - DI=1) contains four species, namely *C. madrense, C. martinii*, *C. setaceum*, with n = 21 chromosomes, and *C. nudicaule* (n = 21 and n = 22). *Centaurium madrense*, the unique Mexican species with showy corolla, appears to be a large-flowered version of *C. setaceum*, a taxon recognizable by its very thin leaves. *Centaurium martinii* also has very narrow leaves appressed to the stem, giving the plant a leafless aspect. This species shows an unusual disjunction in its distribution for populations can be found in the Transverse Volcanic Belt of Mexico (between Guadalajara and Mexico) and in Honduras, but not in the intervening area despite the presence of suitable habitats. The fourth species of the group, *C. nudicaule*, is generally characterized by the presence of a basal rosette and a slender, unique stem and peduncle. However, another accession of *C. nudicaule*, comprising stout-ramified plants, shows n = 22 chromosomes and forms a basal taxon in this clade. An early study on Baja Californian *Centaurium* species discriminated between slender and stout individuals of *C. nudicaule* (Brandegee, 1891). The former, found from south of Arizona to Mexico, is similar to the type of the species (Arizona, Streams of the Santa Catalina Mountains, 18 may 14, C. G. Pringle. 8358/112 (G)) and therefore, must be called *C. nudicaule*. The systematics of the new taxon (n=22), temporally called *C. sp.nov.2*, will be further examined.

A second well resolved clade (n=22) contains three species, namely *C. strictum*, *C. wigginsii* and *C. pusillum*. The first two taxa are morphologically closely related, and only distinguishable by the pattern of the branching (branching strict for *C. strictum* vs. divaricate for *C. wigginsii*) and inflorescence (conspicuously leafy in *C. strictum* and with few bracts in *C. wigginsii*). The two taxa may have certainly diverged by geographic isolation, for they presently occur in allopatric areas, with a more northern repartition for *C. wigginsii* (Sinaloa and Nayarit) and a southern one for *C. strictum* (Oaxaca to Mexico). Our karyological data do not support an aneuploid increase in *C. wigginsii* as previously proposed (Broome, 1973). Lastly, *C. pusillum* is basal in this clade and

shares great morphological features with *C. quitense.* Both taxa mainly differ by the small size, the basal branching, and a possible basal rosette in *C. pusillum*.

Nevertheless, the different accessions of *C. quitense* form a third and well supported clade (BS=98 - DI=4). As mentioned above, two distinct chromosome numbers occur in *C. quitense*, and populations with n=21 form a well-supported group (BS=86 - DI=2). However, populations with n=22 chromosomes were mainly collected in the southern part of Mexico, whereas the n=21 populations seem to occur in a more northern range (Fig. 6, map B). These preliminary data support further segregation in the *C. quitense* complex.

Hybridization in *Centaurium*---Natural hybridization in *Centaurium* was described for some Eurasian species (Druce, 1928; Jonker, 1950; Zeltner, 1970; Ubsdeel, 1976a, b), and also seems to occur in the New World. Yet, the term hybrid has often been employed to describe intermediate or variable forms that may have only been the result of clinal variation. Fixed genomic allopolyploids may be recognized by their additive chromosome number, if parental sets are numerically different. In the other case (identity of the putative parental chromosome number), distinction between allo- and autopolyploidy is not easy. Moreover, the difficulty of homoploid hybrid identification in the field lies on the fact that there are multiple explanations for morphological variation, molecular additivity or phylogenetic incongruence (Rieseberg *et al*., 2000).

Chromosome counts of *C. sp.nov.1* (n=37) indicate an allopolyploid origin between putative parents with n=20 and n=17 chromosomes. *C. sp.nov.1* is common from Baja California to Nevada and Utah while morphologically close relatives, namely *C. exaltatum* (n=20) or *C. namophilum* var. *nevadense* (n=17) occur in more restricted and peripheral geographic areas (Fig. 6 - Map A). Chloroplast phylogenetic analyses support *C. exaltatum* as the mother parent (chloroplast capture of *C. exaltatum* genome in *C. sp.nov.1*) while ITS tree group the putative hybrid with *C. exaltatum* and *C. venustum* (n=17). *C. venustum* has showy flowers and sometimes occurs in sympatry with *C. sp.nov.1,* especially in Baja California where very variable forms of the two taxa may be found. This suggests an episode of hybridization between these two species resulting in a widespread allopolyploid with a chromosome number of n=37, close to the previous count of n=40 reported for *C. exaltatum* sensu lato. Another similar case of genomic allopolyploidy occurs in Arabia, where *C. malzacianum* (n=28) falls near *C. maritimum* (n=10) and *C. pulchellum* (n=18) in phylogenetic trees based on different data sets (Chapter 2). Along with allopolyploidy, homoploid hybrid speciation and introgression may explain a part of the high variability in *Centaurium*. For example, *C. muehlenbergii* has been subject to many unsatisfactory taxonomic classifications and is now recognized as a variable entity by most authors. The different forms described show morphological

and molecular resemblance to *C. erythraea*, a species recently introduced in California. Nevertheless, in the field, some populations of *C. muehlenbergii* are indistinct from the Mediterranean tetraploid *C. tenuiflorum*. Human introduction of *C. tenuiflorum* and further natural hybridization with *C. erythraea* may have been possible in California. Once the hybrid species *C. muehlenbergii* becomes established and expands its range, it may come back into contact and merge with *C. erythraea* due to asymmetric gene flow, resulting in a wide range of morphological forms for the *C. muehlenbergii* complex. The same phenomenon seems to occur in Australia, in places where the introduced species *C. tenuiflorum* and *C. erythraea* are sympatric. Different herbarium sheets described as *C. erythraea* X *C. tenuiflorum* look like *C. muehlenbergii* and preliminary ITS and 5S-NTS sequences obtained from Australian *C. tenuiflorum* and *C. erythraea* are identical to the *C. muehlenbergii* ones (Mansion, unpublished). Moreover, artificial crosses between the different forms generally yield vigorous hybrids, intermediate in vegetative and floral morphology (Broome, 1973).

Other examples of variable populations are found in Texas with *C. calycosum*. From a systematic point of view, the taxon may comprise several forms, as shown by the resolution of the clade in the ITS tree topology. Moreover, *C. calycosum*, with *C. texense* or *C. arizonicum*, produce sympatric populations with numerous ambiguous forms that may be the result of homoploid or introgressive hybridization. Such a variation exists in *Sabatia*, a genus with mainly proterandrous species, where introgression has been reported (Bell and Lester, 1978). In Texan *Centaurium*, each individual bears numerous showy corollas surmounted by the anthers that engender a high degree of pollen donation, so favoring allogamy (either geitonogamy or xenogamy). In Mexican species, the latter phenomenon seems to be uncommon, due to the very small size of the flowers with stamens not exceeding the corolla tube. Sympatric populations often contain three to four different species and intermediate forms are infrequent. Only a few number of apparently natural hybrids, such as *C. strictum* x *quitense* and *C. nudicaule* x *wigginsii*, have been mentioned in the literature (Broome, 1973). In the documented cases of hybridization cited above, no nucleotide additivity was observed in ITS sequences of putative hybrids. So, concerted evolution, via gene conversion or unequal crossing-over, seems to operate in the genus, given the high homogeneity of ITS sequence within a species. The rate of concerted evolution may be fast in *Centaurium* since species are annuals or biennials, with a short generation time, and higher rates of sequence evolution may be expected (Gaut *et al.*, 1992). Vegetative reproduction that may prolong generation time significantly, and so favoring maintenance of both parent ITS sequences in the hybrid (Coen *et al.*, 1982), does not occur in the genus. As a result, as shown for *C. sp.nov.1* or *C. muehlenbergii*, homogenization of different parent genomes may take place in hybrids, and consequently, only one parental genome has been detected (Wendel *et al.*, 1995).

Biogeography---The present study reveals three major groups located in three distinct geographic areas, and together forming a monophyletic North-American assemblage. Combined analyses show a basal Californian set of polyphyletic origin from which diverge two clades respectively centered in Texas and in Mexico. In California, geographical barriers such as Sierra Nevada and deserts may have divided the widespread range of a common ancestor. Speciation could subsequently have then occurred by vicariance, as shown by the isolated areas of basal taxa (*C. trichanthum*, *C. namophilum*), on both western and eastern parts of the Sierra Nevada mountains. Other geoclimatic events, such as the Wisconsin glaciation, have directly determined refugia and pattern of migration, for most of plant and animal species during the Pleistocene (Brouillet and Whetstone, 1993). This major glacial episode should mainly explain the actual repartition of *Centaurium* species in western North America and particularly in Texas where some taxa seem to occur in a particular ecogeographic area (see above).

As shown with a likelihood ratio test, a molecular clock hypothesis could not be rejected, meaning that branch lengths of the phylogenetic inferred trees could be proportional to time. The use of a clock assumption to estimate time divergence between lineages is often subject to controversies (Wen, 1999). The major obstacle from molecular rate estimation is the lack of accurate calibration for evolutionary rates of genes in different lineage. For ITS, different substitution rates have been published and vary from 4.5 $x10^{-10}$ substitutions per site per year in the Winteraceae (Suh *et al.*, 1993) to 1.6 - 2.2 x 10^{-8} for perennial gentians of *Gentiana* section *Ciminalis* (Hungerer and Kadereit, 1998). Sometimes, congruent estimates are obtained from two different markers, e.g., in *Liriodendron* and *Magnolia* (Magnoliaceae) or *Gleditsia* (Leguminosae) (Parks and Wendel, 1990; Qiu *et al.*, 1995; Sewell *et al.*, 1996; Schnabel and Wendel, 1998). One way to infer time divergence is to use higher and lower known rates and to check the compatibility of the results obtained with some geological events or fossil records. Unfortunately, fossil records are often very rare for annual species and absent in *Centaurium*. The way to roughly calibrate a molecular clock is to associate one phylogenetic node of a tree with a well-dated geoclimatic or geological event. Hence, the rate of mutation can be calculated by dividing the pairwise distance between the taxa by twice the estimated time divergence (Schnabel and Wendel, 1998). With our data, we have calibrated vicariance speciation rates using the closely related but disjunct *C. trichanthum* and *C. namophilum*, and the maximum uplifts of the Cascade and Sierra Nevada Ranges (ca. two million years ago). For a time divergence of 2 mya, we obtain an ITS mutation rate of 5.5 x 10^{-9} mutations per site and per year which is slower than in perennial *Gentiana*. The phylogram obtained using the maximum likelihood criterion, under the assumption of a molecular clock (Fig. 4), then permits us to estimate the separation between the (*C. trichanthum* - *C.*

namophilum) basal clade and the remaining American taxa as 7-8 mya. Following this, the (*C. venustum* - *C. exaltatum* - *C. abramsii*) group and Texan and Mexican ones seem to have diverged around 6 mya. This is congruent with an alternative-calibrating event under the molecular clock hypothesis. During the late Miocene and Pliocene (from 6 to 2 mya), widespread arid habitats developed into the modern deserts, Chihuahua, Mojave and Sonora, respectively (Morafka *et al.*, 1992). Nowadays, these unfavorable zones for annual species, clearly separate the three North American groups of *Centaurium*. This time estimation likewise fits with the hypothesis of an early uplift, during late Miocene and early Pliocene, of the Mexican Sierras that today shelter some *Centaurium* species such as *C. madrense*, *C. quitense* or *C. strictum* (Delgadillo, 1971). No diploid native species have been found in the New World, and some authors have previously concluded that the American centauries have a derived position compared to the Eurasian species (Zeltner, 1970; Broome, 1973). On the molecular clock tree, a separation between a reduced set of Old World *Centaurium* and the New World group can be estimated to have occurred around 14 mya. The late Miocene is one of the five major periods of exchanges between Asia and North America (Tiffney, 1985a, b). By that time, fossil evidence tends to support cooling of the Bering region (due to southward migrating air masses), and subtropical vegetation that might have included genera like *Liriodendron* (Magnoliaceae) or *Liquidambar* (Hamamelidaceae), could not have survived (Parks and Wendel, 1990; Hoey and Parks, 1991). On the other hand, the Bering Land Bridge was suitable for the exchange of temperate deciduous plants until about 3.5 mya (Wen, 1999), and annual species such as *Centaurium* may have migrated this way. Today, the genus reaches its northern limit under the 60°N parallel in Europe (*C. littorale* and *C. pulchellum* in Finland) and so migration at the Beringian latitudes, i.e. 69° to 75°N, may have been possible under the late Miocene bioclimatical conditions. In addition, the existence of a presumed Aleutian Land Bridge in the south of the Bering Land Bridge, during the Tertiary, may have favored plant migration. So, alternative hypotheses, such as California colonization from an ancestral Eurasian species or vicariant speciation event during the middle Miocene, may be equally proposed. Nevertheless, each of them is based on the assumption of the monophyly of *Centaurium*. In this study, only the (*C. pulchellum* - *C. capense*) set may be seen as a candidate for the Madrean-Tethys hypothesis (Axelrod, 1975) which predicts: (1) the monophyly of the group under investigation and (2) a divergence time, estimated from molecular data, earlier than Paleogene/Neogene boundary, i.e. more than 25 Mya (Fritsch, 1996). Anyway, even if the *C. pulchellum* group is a clade and forms an intercontinental sister species with *C. capense*, divergence time roughly estimated with a molecular clock do not agree with the second aspect of this hypothesis.

CONCLUSION

The monophyly of the New World species *Centaurium* is strongly supported by two different and independent molecular data sets, namely the Internal Transcribed Spacer of nrDNA, and trnLF region of cpDNA. Furthermore, the phylogenetic reconstruction presented here, depict four clades. Most of the Eurasian species, along with *C. muehlenbergii* and *C. capense*, previously considered as native of the Americas, form a basal assemblage on the ITS tree. The New World clade occupies a derived and terminal position on the cladogram, whereas the Mediterranean section *Spicaria*, along with section *Gyrandra* of Mexican origin, form two intermediate clades.

Therefore, the phylogenetic reconstruction of this group does not agree with the previous systematic placement of the American species in the European section *Centaurium* (Gilg, 1885; Broome, 1973). Moreover, the different chromosome numbers found in the respective Eurasian and American clades, argue for a distinct evolutionary history of these groups. Within the New World *Centaurium*, karyological and biogeographical data depict three major sets, namely the Californian group (n = 17, n = 20), the Texan group (n = 20, n = 21) and the Mexican one (n = 21, n = 22). These three assemblages are congruent with those deriving from the phylogenetic trees. Under the assumption of a molecular clock, and further calibration with geological events, an evolutionary scenario of this group may be proposed. The common ancestor of the North American centauries was probably present in California by the end of the Miocene (ca 7 mya). Early divergence gave the first Californian lineage (*C. trichanthum – C. namophilum*). Then, ca. 6 mya ago, the development of the modern deserts (Chihuahua, Mojave and Sonoran) lead to an important diversification in three separate areas, giving a new Californian group (e. g. *C. venustum* and *C. abramsii*), along with the present Texan and Mexican ones.

In the light of the present study, the North American clade must be considered as a distinct section or subgenus, further subdivided in three sets, under the assumption of a monophyletic genus *Centaurium*. Studies are presently being undertaken to test this hypothesis, and it necessitates a global analysis including many other closely related genera.

REFERENCES

ABRAMS, L. R. 1951. *An illustrated flora of the Pacific States, Washington, Oregon, and California*. Stanford University Press, Stanford.

ADAMS, L. G. 1996. *Gentianaceae*. in *Flora of Australia*, vol. 28 : 73-103. A. Wilson, Editor. CSIRO, Melbourne.

AXELROD, D. I. 1975. Evolution and biogeography of Madrean-Tethyan sclerophyll vegetation. *Annals of the Missouri Botanical Garden* 62 : 280-334.

BELL, N. B., and L. J. LESTER. 1978. Genetic and morphological detection of introgression in a clinal population of *Sabatia* section *Campestria* (Gentianaceae). *Systematic Botany* 3 (1) : 87-104.

BRANDEGEE, T. S. 1891. Flora of the Cape region of Baja California. *Proceedings of the California Academy of Science* 2 (3) : 108-182.

BREMER, K. 1988. The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. *Evolution* 42 : 795-803.

BROOME, R. C. 1973. Systematics of *Centaurium* (Gentianaceae) of Mexico and Central America. Thesis. Duke University. Duke

-----. 1976. The central American species of C*entaurium* (Gentianaceae). *Brittonia* 28 : 413-426.

-----. 1978. Chromosome numbers and meiosis in North and Central American species of C*entaurium* (Gentianaceae). *Systematic Botany* 3 : 299-312.

-----. 1981. A new variety of *Centaurium namophilum* (Gentianaceae) from the great basin. *Great basin naturalist* 41 (2) : 192-197.

BROUILLET, L., and D. WHETSTONE. 1993. Climate and physiography of North America, north of Mexico. In *Flora of North America, north of Mexico*, vol. 1. Oxford University Press. New York.

CHASE, M. W., C. M. MORTON, and J. A. KALLUNKI. 1999. Phylogenetic relationships of Rutaceae : a cladistic analysis of the subfamilies using evidence from *rbc*L and *atp*B sequence variation. *American Journal of Botany* 86 (8) : 1191-1199.

COEN, E., T. STRACHAN, and G. A. DOVER. 1982. Dynamics of concerted evolution of ribosomal DNA and histone gene families in the *Melanogaster* species subgroup of *Drosophila. Journal of Molecular Biology* 158 : 17-35.

DELGADILLO, M. C. 1971. Phytogeographic studies on alpine mosses of Mexico. *Bryologist* 74 : 331-346.

DOYLE, J. J., and J. L. DOYLE. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical bulletin* 19 : 11-15.

DRUCE, G. C. 1928. British Plant List. *Arbroath* 78

DUNN, R. A., 1967. A revision of the genus *Centaurium* of continental United States. Thesis. Catholic University of America. Washington D. C.

FARRIS, J. S., M. KALLERSJÖ, A. G. KLUGE, AND C. BULT. 1995. Testing significance of incongruence.

Cladistics 10 : 315-319.

FELSENSTEIN, J. 1981. Evolutionary trees from DNA sequences : A maximum likelihood approach. *Journal of Molecular Evolution* 17 : 368-376.

-----. 1985. Confidence limits on phylogenies : an approach using the bootstrap. *Evolution* 17 : 223-263.

FERNANDES, R. 1965. Duas variedades novas de *Centaurium spicatum* (L.) Fritsch. *Boletim da Sociedade Broteriana* 31 : 15-29.

FRITSCH, P. 1996. Isozyme analysis of intercontinental disjuncts within *Styrax* (Styracaceae) : implication for the Madrean-Tethyan hypothesis. *American Journal of Botany* 83 (3) : 342-355.

GAUT, B. S., S. V. MUSE, W. D. CLARK, and M. T. CLEGG. 1992. Relative rates of nucleotide substitution at the *rbc*L locus of monocotyledonous plants. *Journal of Molecular Evolution* 35 : 292-303.

GILG, E. 1895. Gentianaceae. in *Die Naturlichen Pflanzenfamilien*. 4**:** 50-108. A. Engler and K. Prantl Editors

GOLDMAN, N. 1993. Statistical tests of models of DNA substitution. *Journal of Molecular Evolution* 36 : 182-198.

GRISEBACH, A. H. R. 1839. *Genera et Species Gentianearum adjectis observationibus quibusdam phytogeographycis* : 116-125. J. G. Cotta. Stuttgart et Tübingen.

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

HEGI, G. 1966. Gentianaceae. In : *Illustrierte Flora von Mittel-Europa*. 5 (3) : 1953-2047. Carl Hanser Verlag. München.

HIGGINS, D. G., J. D. THOMPSON, and T. J. GIBSON. 1996. Using CLUSTAL for multiple sequence alignments. *Methods in Enzymology* 266 : 383-402.

HILLIS, D. M., and J. P. HUELSENBECK. 1992. Signal, noise and reliability in molecular phylogenetic analyses. *Journal of Heredity* 83 : 189-195.

HILLIS, D. M., and J. J. BULL. 1993. An empirical method of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Systematics Biology* 42 : 182-192

HOEY, M. T., and C. R. PARKS. 1991. Isozyme divergence between eastern Asian, North American and Turkish species of *Liquidambar* (Hamamelidaceae). *American Journal of Botany* 78 : 938-947.

HOLMES, W. C., and D. E. WIVAGG. 1996. Identification and distribution of *C. muhlenbergii* (Griseb.) Piper and *C. pulchellum* (Sw.) Druce (Gentianaceae) in Louisiana, Mississippi, and Texas. *Phytologia* 80 : 23-29.

HUELSENBECK, J. P., and K. A. CRANDALL. 1997. Phylogeny estimation and hypothesis testing using maximum likelihood. *Annual Review of Ecology and Systematics* 28 : 437-466.

HUNGERER, K. B., and J. W. KADEREIT. 1998. The phylogeny and biogeography of *Gentiana* L. sect. *Ciminalis* (Adans.) Dumort. : a historical interpretation of distribution ranges in the European high mountains. *Perspectives in Plant Ecology, Evolution and Systematic* 1 (1) : 121-135.

JOHNSTON, I. M. 1924. Expedition of the California Academy of Sciences to the Gulf of California in 1921. *Proceedings of the California Academy of Science* 4 (2) : 951-1218.

JONKER, F. P. 1950. Revisie van de Nederlandse Gentianaceae : I. *Centaurium* Hill. *Nederllandsch Kruidkundig Archief* 57 : 170-198.

MELDERIS, A. 1931. Genetical and taxonomical studies in the genus *Erythraea* Rich. *Acta Horti Botanici Universitatis Latviensis.* 6 : 123-256

MORAFKA, D. J., G. A. ADEST, and L. M. REYES. 1992. Differentiation of the North American deserts : a phylogenetic evaluation of a vicariance model, in *Biogeography of Mesoamerica* : 195-226. Tulasne University. New Orleans

PARKS, C. R., and J. F. WENDEL. 1990. Molecular divergence between Asian and North American species of *Liriodendron* (Magnoliaceae) with implications for interpretation of fossil floras. *American Journal of Botany* 77 (10) : 1243-1256.

QIU, Y.-L., M. W. CHASE, and C. R. PARKS. 1995. A chloroplast DNA phylogenetic study of the eastern Asia-eastern North America disjunct in section *Rytidospermum* of *Magnolia* (Magnoliaceae). *American Journal of Botany* 82 : 1582-1588.

REVEAL, J. L., C. R. BROOME, and J. C. BEATLEY. 1974. A new *Centaurium* from the Death Valley region of Nevada and California. *Bulletin of the Torrey Botanic Club* 100 : 353-356.

RIESEBERG, L. H., and D. E. SOLTIS. 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary trends in plants* 5 (1) : 65-83.

ROBYNS, A. 1954. Essai d'étude systématique et écologique des *Centaurium* de Belgique. *Bulletin du Jardin Botanique de Bruxelle* 24 : 349-398.

SANG, T., D. J. CRAWFORD, and T. F. STUESSY. 1997a. Documentation of reticulate evolution in peonies (*Paeonia*) using internal transcribed spacer of nuclear ribosomal DNA : Implications for biogeography and concerted evolution. *Proceedings of the National Academy of Science USA* 92 : 6813-6817.

-----, M. J. DONOGHUE, and D. ZHANG. 1997b. Evolution of Alcohol Deshydrogenase genes in Peonies (*Paeonia*) : Phylogenetic relationships of putative non-hybrid species. *Molecular Biology an Evolution* 14 (10) : 994-1007.

SCHNABEL, A., and J. F. WENDEL. 1998. Biogeography of *Gleditsia* (Leguminosae) based on ndhF and rpl16 chloroplast gene sequences. *American Journal of Botany* 85 (2) : 1753-1765.

SEWELL, M. L., C. R. PARKS, and M. W. CHASE. 1996. Intraspecific chloroplast DNA variation and biogeography of North American *Liriodendron* L (Magnoliaceae). *Evolution* 50 (3) : 1147-1154.

SOLTIS, D. E., and R. K. KUZOFF. 1995. Discordance between nuclear and chloroplast phylogenies in the *Heuchera* group (Saxifragaceae). *Evolution* 49 (4) : 727-742.

SORENSON, M. D. 1999. *TreeRot version 2*. Boston University, Boston.

SUH, Y., L. B. THIEN, H. E. REEVE, and E. Z. ZIMMER. 1993. Molecular evolution and phylogenetic implications of internal transcribed spacer sequences of ribosomal DNA in Winteraceae. *American Journal of Botany* 80 : 1042-1055.

STRUWE, L., J. W. KADEREIT, J. KLACKENBERG, J. S. NILSSON, M. THIV, K. B. VON HAGEN, and V. A. ALBERT, 200*. Systematics, character evolution, and biogeography of the Gentianaceae, including a new tribal and subtribal classification. in *Gentianaceae - systematics and natural history*. L. Struwe and V. A. Albert Editors. Cambridge University Press. Cambridge

SWOFFORD, D. L. 1993. "PAUP : Phylogenetic Analysis Using Parsimony" Version 3.1.1. Illinois Natural History Survey. Champaign

-----. 1999. "PAUP : Phylogenetic Analysis Using Parsimony" Version 4.0. Sinauer. Sunderland

TABERLET, P., L. GIELLY, G. PAUTOU, and J. BOUVET. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology* 17 : 1105-1109.

THIV, M., L. STRUWE, and J. W. KADEREIT. 1999. The phylogenetic relationships and evolution on the Canarian laurel forest endemic *Ixanthus viscosus* (Aiton) Griseb. (Gentianaceae) : evidence from *mat*K and ITS sequences, and floral morphology and anatomy. *Plant Systematics and Evolution* 218 : 299-317.

TIFFNEY, B. H. 1985a. The Eocene North Atlantic land bridge : its importance in tertiary and modern phytogeography of the Northern hemisphere. *Journal of the Arnold Arboretum* 66 : 243-273.

-----. 1985b. Perspectives on the origin of the floristic similarity between Eastern Asia and Eastern North America. *Journal of the Arnold Arboretum* 66 : 73-94.

TURNER, B. L. 1993. The Texas species of *Centaurium* (Gentianaceae). *Phytologia* 75 (3) : 259-275.

UBSDELL, R. A. E. 1976a. Studies and evolution in *Centaurium erythraea* Rafn. and *Centaurium littorale* (D. Turner) in the British Isles; taxonomic and biometrical studies. *Watsonia* 11 : 7-31.

-----. 1976b. Studies and evolution in *Centaurium erythraea* Rafn. and *Centaurium littorale* (D. Turner) in the British Isles; Cytology. *Watsonia* 11 : 33-43.

-----. 1979. Studies and evolution in *Centaurium erythraea* Rafn. and *Centaurium littorale* (D. Turner) in the British Isles; breeding systems, floral biology and general discussion. *Watsonia* 12 : 225-232.

VIJVERBERG, K., T. H. M. MES, and K. BACHMANN. 1999. Chloroplast DNA evidence for the evolution of *Microseris* (Asteraceae) in Australia and New Zealand after long-distance dispersal from western North America. *American Journal of Botany* 86 (10) : 1448-1463.

WEN, J. 1999. Evolution of eastern Asian and eastern North American disjunct distributions in flowering plants. *Annual Revue of Ecology and Systematics* 30 : 421-455.

WENDEL, J. F., A. SCHNABEL, and T. SEELANAN. 1995. Bi-directional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). *Proceedings of the National Academy of Science USA* 92 : 280-284.

WHITE, T. J., T. BRUNS, S. LEE, and J. TAYLOR. 1990 Amplification and direct sequencing of fungal ribosomal DNA for phylogenetics. in *PCR protocols : a guide to methods and applications*. M. A. Innis *et al.* Editors. Academic Press. San Diego.

YANG, Z., N. GOLDMAN, and A. E. FRIDAY. 1995. Maximum likelihood trees from DNA sequences : a peculiar statistical estimation problem. *Systematic Biology* 44 : 384-399.

YUAN, Y.-M., and P. KÜPFER. 1995. Molecular phylogenetics of the subtribe Gentianinae (Gentianaceae) inferred from the sequences of the internal transcribed spacer (ITS) of nuclear ribosomal DNA. *Plant Systematic and Evolution* 196 : 206-226.

YUAN, Y.-M., P. KÜPFER, and J. J. DOYLE. 1996. Infrageneric phylogeny of the genus *Gentiana* (Gentianaceae) inferred from nucleotide sequences of the internal transcribed spacer (ITS) of nuclear ribosomal DNA. *American Journal of Botany* 83 (5) : 641-652.

ZELTNER, L. 1970. Recherche de biosystématique sur les genres *Blackstonia* Huds. et *Centaurium* Hill (Gentianaceae). *Bulletin de la Societé Neuchâteloise des Sciences Naturelles* 93 : 1-164.

ZELTNER, L. 1978. Recherches sur le *Centaurium bianoris* (Sennen) Sennen. *Revue de Biologie et Ecologie Méditerranéenne* 5 (1) : 51-58

Table 1. Accession data for taxa sampled for karyological study and phylogenetic analyses of the internal transcribed spacer (ITS1 and ITS2) of nrDNA and *trn*L-F region (*trn*L intron and *trn*L-F spacer) of cpDNA. Taxa marked with an asterisk are not included in the molecular analyses. All the chromosome numbers have been established in the present study except # (Zeltner, 1970 or unpublished data) and ◊ (Broome, 1978).

TAXON	Accession	n	2n	GenBank ITS1	GenBank ITS2	GenBank *trn*L	GenBank *trn*F	Source
B. perfoliata	MG98712			AY047793	AY047878	AF402198	AF402254	France - 43.47 N / 04.40 W
C. abramsii	MG990903	17◊	34◊	AY047712	AY047797	AF402145	AF402201	California - 37.43.45 N / 119.38.47 W
C. altaicum	MG98810	18#	36#	AY047788	AY047873			China - 42.22.49 N / 86.19.56 W
C. arizonicum	MG97726	20	40	AY047724	AY047809	AF402155	AF402211	Texas - 30.03.89 N / 99.33.02 W
C. arizonicum	MG97729	20	40	AY047725	AY047810			Texas - 30.00.21 N / 100.02.94 W
*C. arizonicum**	970606-9	20	40					Texas - 30.04.02 N / 99.43.19 W
*C. arizonicum**	970606-10	20	40					Texas - 30.03.54 N / 99.54.66 W
*C. arizonicum**	970607-2	20	40					Texas - 29.50.51 N /100.01.58 W
*C. arizonicum**	970607-3a	20	40					Texas - 29.25.05 N / 100.03.66 W
*C. arizonicum**	970608-2	20						Texas - 29.27.28 N / 100.56.24 W
*C. arizonicum**	970608-5a	20						Texas - 30.10.80 N / 102.24.11 W
*C. arizonicum**	970609-1a	20	40					Texas - 30.29.08 N / 102.31.96 W
C. beyrichii	MG96522	20	40	AY047733	AY047818			Texas - 30.02.33 N / 99.08.66 W
C. beyrichii	MG96526	20		AY047732	AY047817	AF402160	AF402216	Texas - 29.49.96 N / 98.09.65 W
*C. beyrichii**	960625-3	20	40					Texas - 30.20.41 N / 98.47.36 W
*C. beyrichii**	960625-4	20	40					Texas - 30.45.23 N / 98.12.16 W
*C. beyrichii**	970605-5a	20						Texas - 29.40.38 N / 98.50.08 W
*C. beyrichii**	970606-6	20	40					Texas - 30.02.33 N / 99.08.66 W
*C. beyrichii**	970606-7	20						Texas - 30.03.58 N / 99.27.71 W
C. brachycalyx	MG990205			AY047771	AY047856	AF402184	AF402240	Mexico - 16.46.04 N / 92.35.32 W
C. brachycalyx	MG990239	36	72	AY047770	AY047855	AF402185	AF402241	Mexico - 19.32.46 N / 100.22.40 W
C. breviflorum	MG97735	20	40	AY047736	AY047821	AF402162	AF402218	Texas - 29.15.76 N / 100.45.39 W
C. calycosum	MG96519	20	40	AY047740	AY047825	AF402164	AF402220	Texas - 29.42.26 N / 101.21.69 W
C. calycosum	MG96520	20	40	AY047737	AY047822			Texas - 29.43.64 N / 100.47.87 W
C. calycosum	MG96521	20	40	AY047739	AY047824	AF402163	AF402219	Texas - 30.04.33 N / 99.41.08 W
C. calycosum	MG97733	20		AY047738	AY047823			Texas - 29.25.05 N / 100.03.66 W
C. calycosum	MG97738	20	40	AY047741	AY047826			Texas - 29.28.46 N / 100.59.20 W
C. calycosum	MG98710					AF402165	AF402221	Texas

Table 1. continued

*C. calycosum**	960605-2	20	40					Texas - 30.05.73 N / 99.04.89 W
*C. calycosum**	970605-3b	20						Texas - 29.37.18 N / 98.43.77 W
*C. calycosum**	970605-5b	20	40					Texas - 29.40.38 N / 98.50.08 W
*C. calycosum**	970605-6	20						Texas - 29.44.58 N / 98.58.16 W
*C. calycosum**	970605-7	20						Texas - 29.43.28 N / 99.07.60 W
*C. calycosum**	970606-1a	20	40					Texas - 29.38.07 N / 99.11.06 W
*C. calycosum**	970606-2	20						Texas - 29.45.44 N / 99.10.72 W
*C. calycosum**	970606-3	20						Texas - 29.49.28 N / 99.15.61 W
*C. calycosum**	970606-4	20						Texas - 29.54.84 N / 99.15.14 W
*C. calycosum**	970606-5	20	40					Texas - 30.06.85 N / 99.04.33 W
*C. calycosum**	970606-8a	20	40					Texas - 30.03.89 N / 99.33.02 W
*C. calycosum**	970607-1b	20						Texas - 30.00.21 N / 100.02.94 W
*C. calycosum**	970608-1a	20	40					Texas - 29.15.76 N / 100.45.39 W
*C. calycosum**	970608-3	20						Texas - 29.27.37 N / 100.57.67 W
*C. calycosum**	970608-5b	20	40					Texas - 30.10.80 N / 102.24.11 W
C. capense	MG98706	18	36	AY047789	AY047874	AF402194	AF402250	Mexico - 29.44.89 N / 114.44.53 W
C. erythraea	MG97760	20	40	AY047780	AY047865	AF402191	AF402247	California - 40.52.90 N / 123.32.89 W
C. erythraea	MG98809			AY047779	AY047864			Spain - 42.53 N / 03.40 E
C. erythraea	MG990242			AY047781	AY047866			France - 43.55 N / 04.25 E
C. erythraea	MG990243			AY047782	AY047867			France - 43.34 N / 04.10 E
C. erythraea	MG990703			AY047783	AY047868			Spain - 39.30 N / 03.00 E
C. erythraea	MG991001					AF402192	AF402248	Ecuador - Quito
*C. erythraea**	960609-2	20	40					California - 38.09.68 N / 120.38.91 W
*C. erythraea**	960609-3	20	40					California - 38.09.68 N / 120.38.91 W
C. exaltatum	MG981113	20	40	AY047721	AY047806	AF402154	AF402210	California - 34.58.47 N / 120.34.39 W
C. exaltatum	MG981114	20	40	AY047722	AY047807			California - 34.58.47 N / 120.34.39 W
C. exaltatum	MG981115	20	40	AY047723	AY047808			California - 35.06.70 N / 120.05.78 W
*C. exaltatum**	960615-2	20	40					California - 35.02.78 N / 120.11.01 W
C. glanduliferum	MG96518	20	40	AY047731	AY047816	AF402159	AF402215	Texas - 30.11.65 N / 102.53.41 W
*C. glanduliferum**	970609-1b							Texas - 30.29.08 N / 102.31.96 W
*C. glanduliferum**	970609-2							Texas - 30.23.15 N /102.26.60 W

Table 1. continued

*C. glanduliferum**	970609-3							Texas - 30.11.64 N / 102.53.38 W
C. madrense	MG990230	21	42	AY047742	AY047827	AF402166	AF402222	Mexico - 23.27.15 N / 105.49.91 W
C. mairei	MG991004	27#	54#	AY047790	AY047875	AF402195	AF402251	Niger - 16.54 N / 08.45 E
C. martinii	MG990213	21				AF402170	AF402226	Mexico - 19.42.30 N / 101.35.96 W
C. martinii	MG990215	21	42	AY047749	AY047834	AF402168	AF402224	Mexico - 19.52.41 N / 102.25.65 W
C. martinii	MG990219		42	AY047750	AY047835	AF402169	AF402225	Mexico - 19.53.75 N / 103.04.25 W
C. maryannum	MG96516	21	42	AY047729	AY047814	AF402158	AF402214	N Mexico - 32.48.24 N / 106.15.67 W
C. maryannum	MG96517			AY047730	AY047815			N Mexico - 33.30.60 N / 104.27.92 W
*C. maryannum**	970610-2	21	42					Texas - 31.58.76 N / 104.31.21 W
*C. maryannum**	970610-3	21	42					N Mexico
C. muehlenbergii	MG96513		40	AY047776	AY047861	AF402190	AF402246	California - 36.07.14 N / 118.49.55 W
C. muehlenbergii	MG97703	20	40	AY047785	AY047870			Texas - 29.56.49 N / 95.20.80 W
C. muehlenbergii	MG97761	20		AY047777	AY047862	AF402189	AF402245	California - 40.55.94 N / 123.37.51 W
C. muehlenbergii	MG97773	20	40	AY047778	AY047863			California - 37.16.36 N / 122.12.86 W
*C. muehlenbergii**	960607-1		40					California - 35.56.02 N / 122.41.72 W
*C. muehlenbergii**	960607-2		40					California - 35.56.02 N / 122.41.72 W
*C. muehlenbergii**	960607-3		40					California - 38.31.77 N / 122.39.08 W
*C. muehlenbergii**	960609-1		40					California - 38.26.49 N / 120.51.32 W
*C. muehlenbergii**	970603-1		40					Texas - 29.56.49 N / 95.20.80 W
*C. muehlenbergii**	970603-5		40					Texas - 29.49.53 N / 99.34.51 W
*C. muehlenbergii**	970625-1		40					California - 99.13.52 N / 121.25.81 W
*C. muehlenbergii**	970626-1		40					California - 39.37.26 N / 121.25.37 W
*C. muehlenbergii**	970626-2		40					California - 39.42.76 N / 121.39.78 W
*C. muehlenbergii**	970627-3		40					California - 40.56.98 N / 123.37.12 W
*C. muehlenbergii**	970628-1		40					California - 40.29.39 N / 123.58.47 W
*C. muehlenbergii**	970628-3		40					California - 40.15.97 N / 123.52.27 W
*C. muehlenbergii**	970628-4		40					California - 40.13.27 N / 123.47.68 W
*C. muehlenbergii**	970628-5		40					California - 40.06.99 N / 123.47.98 W
*C. muehlenbergii**	970629-4							California - 38.27.13 N / 122.21.68 W
*C. muehlenbergii**	970629-5		40					California - 38.02.86 N / 122.37.34 W
*C. muehlenbergii**	970629-6		40					California - 37.53.37 N / 122.28.50 W

Table 1. continued

C. multicaule	MG96515	21	42	AY047726	AY047811	AF402156	AF402212	Arizona - 33.37.15 N / 110.55.44 W
C. multicaule	MG97745	21	42	AY047727	AY047812	AF402157	AF402213	Texas - 30.03.54 N / 99.54.66 W
C. multicaule	MG98708	21		AY047728	AY047813			Mexico - 27.35.61 N / 105.20.70 W
C. namophilum	MG96512	17	34	AY047711	AY047796	AF402144	AF402200	Nevada - 36.22.18 N / 116.18.12 W
*C. namophilum**	960612-1	17	34					Nevada - 36.22.18 N / 116.18.12 W
C. nevadense	MG97753	17		AY047718	AY047803	AF402149	AF402205	Nevada - 40.35.36 N / 115.17.62 W
C. nevadense	MG97754	17		AY047719	AY047804	AF402150	AF402206	Nevada - 40.37.99 N / 115.14.71 W
C. nudicaule	MG990220	21		AY047745	AY047830			Mexico - 19.49.44 N / 103.39.40 W
C. nudicaule	MG990229	21	42			AF402173	AF402229	Mexico - 20.59.67 N / 104.28.22 W
C. pulchellum	MG97701			AY047786	AY047871			Texas - 29.56.49 N / 95.20.80 W
C. pulchellum	MG98505			AY047787	AY047872	AF402193	AF402249	France - 44.34 N / 06.04 E
C. pusillum	MG98709			AY047756	AY047841			Mexico - 22.05.55 N / 100.29.79 W
C. pusillum	MG990209			AY047755	AY047840	AF402174	AF402230	Mexico - 19.40.17 N / 100.40.13 W
C. quitense	MG97501					AF402175	AF402231	Belize - 16.44.02 N / 88.59.44 W
C. quitense	MG990201	21	42	AY047766	AY047851	AF402176	AF402232	Mexico - 20.47.43 N / 99.17.57 W
C. quitense	MG990202	22	44	AY047764	AY047849			Mexico - 20.35.60 N / 98.37.40 W
C. quitense	MG990204	21	42	AY047744	AY047829			Mexico - 16.48.87 N / 92.12.02 W
C. quitense	MG990206	22		AY047761	AY047846			Mexico - 17.01.25 N / 92.50.32 W
C. quitense	MG990207??	22	44					Mexico - 17.07.21 N / 92.52.33 W
C. quitense	MG990208??		44					Mexico - 17.02.76 N / 92.51.30 W
C. quitense	MG990210	21		AY047767	AY047852	AF402177	AF402233	Mexico - 19.40.39 N / 100.43.37 W
C. quitense	MG990214	22	44					Mexico - 19.42.30 N / 101.35.96 W
C. quitense	MG990218	22	44	AY047762	AY047847			Mexico - 19.53.75 N / 103.04.25 W
C. quitense	MG990222	22		AY047763	AY047848			Mexico - 19.46.36 N / 103.42.70 W
C. quitense	MG990224	21		AY047768	AY047853	AF402178	AF402234	Mexico - 20.11.87 N / 104.22.98 W
C. quitense	MG990226	21	42	AY047769	AY047854			Mexico - 20.14.78 N / 104.27.87 W
C. quitense	MG990236		42	AY047765	AY047850			Mexico - 23.26.90 N / 105.49.40 W
C. setaceum	MG990225	21		AY047748	AY047833	AF402181	AF402237	Mexico - 20.14.78 N / 104.27.87 W
C. setaceum	MG990232		42	AY047746	AY047831	AF402182	AF402238	Mexico - 23.27.15 N / 105.49.91 W
C. setaceum	MG990233	21		AY047747	AY047832	AF402183	AF402239	Mexico - 23.27.49 N / 105.49.63 W
C. sp.nov.2	MG990231	22	44	AY047743	AY047828	AF402167	AF402223	Mexico - 23.27.15 N / 105.49.91 W

Table 1. continued

C. sp.nov1	MG96507	37						California - 37.25.22 N / 118.24.59 W
C. sp.nov1	MG97750	37	74					Utah - 38.34 N / 109.32 W
C. sp.nov1	MG97751	37		AY047716	AY047801	AF402151	AF402207	Utah - 38.34 N / 109.32 W
C. sp.nov1	MG97755	37		AY047717	AY047802	AF402152	AF402208	Mexico - 30.46.27 N / 115.55.58 W
C. sp.nov1	MG98702	37		AY047720	AY047805	AF402153	AF402209	Mexico - 30.46.27 N / 115.57.58 W
C. sp.nov1 *	960611-1c	37						Nevada - 37.25.20 N / 118.24.59 W
C. sp.nov1 *	960611-2	37						California - 36.40.39 N / 118.05.59 W
C. sp.nov1 *	970620-1b	37						Utah - 38.34 N / 109.32 W
C. sp.nov1 *	970622-1	37	74					Nevada - 40.49.40 N / 115.45.82 W
C. sp.nov1 *	970624-1	37						Nevada - 39.31.76 N / 118.46.10 W
C. spicatum	MG98508	11#	22#	AY047792	AY047877	AF402197	AF402253	Morocco - 30.41.030 N / 06.16.150 E
C. spicatum	MG981005			AY047791	AY047876	AF402196	AF402252	France - 43.34 N / 04.10 E
C. strictum	MG990203		44					Mexico - 17.02.95 N / 98.08.36 W
C. strictum	MG990211		44	AY047757	AY047842			Mexico - 19.40.02 N / 100.49.39 W
C. strictum	MG990212		44	AY047751	AY047836	AF402179	AF402235	Mexico - 19.41.07 N / 100.50.17 W
C. strictum	MG990217		44	AY047758	AY047843	AF402180	AF402236	Mexico - 19.55.28 N / 103.02.16 W
C. strictum	MG990221	22	44	AY047752	AY047837			Mexico - 19.49.44 N / 103.39.40 W
C. strictum	MG990223		44	AY047759	AY047844			Mexico - 19.46.36 N / 103.42.70 W
C. strictum	MG990237							Mexico - 19.44.81 N / 100.12.40 W
C. strictum	MG990238		44					Mexico - 19.32.46 N / 100.22.40 W
C. tenuiflorum	MG98802	20#	40#	AY047773	AY047858	AF402187	AF402243	Morocco - 31.32.930 N / 07.35.810 E
C. tenuiflorum	MG98803	10#	20#	AY047774	AY047859	AF402188	AF402244	Morocco - 31.32.930 N / 07.35.810 E
C. tenuiflorum	MG990708			AY047775	AY047860			Spain - 39.30 N / 03.00 E
C. tenuifolium	MG990228		72	AY047772	AY047857	AF402186	AF402242	Mexico - 20.15.29 N / 104.28.87 W
C. texense	MG97708	20		AY047735	AY047820			Texas - 30.11.15 N / 97.43.27 W
C. texense	MG97714	20		AY047734	AY047819	AF402161	AF402217	Texas - 29.40.38 N / 98.50.08 W
*C. texense**	970604-2	20	40					Texas - 30.19.40 N / 97.46.40 W
*C. texense**	970605-2	20	40					Texas - 29.36.84 N / 98.43.33 W
*C. texense**	970605-3a	20	40					Texas - 29.37.18 N / 98.43.77 W
*C. texense**	970605-4	20	40					Texas - 29.39.65 N / 98.49.36 W
*C. texense**	970606-1b	20	40					Texas - 29.38.07 N / 99.11.06 W

Table 1. continued

C. trichanthum	MG96504	17	34	AY047710	AY047795	AF402143	AF402199	California - 38.41.58 N / 122.35.87 W
C. trichanthum	MG97768	17		AY047709	AY047794			California - 38.43.18 N / 122.30.33 W
*C. trichanthum**	970629-2	17	34					California - 38.42.64 N / 122.28.49 W
*C. trichanthum**	970629-3	17	34					California - 38.41.57 N / 122.26.74 W
*C. trichantum**	960608-2	17	34					California - 38.41.58 N / 122.35.87 W
C. venustum	MG96514	17	34	AY047713	AY047798	AF402146	AF402202	California - 33.44.34 N / 116.50.42 W
C. venustum	MG98704	17		AY047715	AY047800	AF402148	AF402204	Mexico - 32.01.28 N / 116.37.67 W
C. venustum	MG98707	17		AY047714	AY047799	AF402147	AF402203	California - 33.44.34 N / 116.50.42 W
C. wigginsii	MG990234	22	44	AY047753	AY047838	AF402171	AF402227	Mexico - 23.26.90 N / 105.49.40 W
C. wigginsii	MG990235		44	AY047754	AY047839	AF402172	AF402228	Mexico - 23.26.90 N / 105.49.40 W

Table 2. Summary of chromosome counts performed in the current study and comparison with the literature data. One asterisk shows discordance between present study and literature whereas species with two asterisks present a new chromosome number. N is the number of populations investigated for each species.

Taxon	Literature data		Present results		N
	n	2n	n	2n	
*C. arizonicum***			20	40	9
*C. beyrichii**	41[c]		20	40	7
*C. brachycalyx***	ca36[c]		36	72	1
*C. breviflorum***			20	40	1
*C. calycosum**	42[c]		20	40	20
*C. capense***			18	36	1
C. erythraea	20[a]	40[a]	20	40	3
*C. exaltatum**	40[c]		20	40	4
*C. glanduliferum***			20	40	4
C. madrense	21[c]		21	42	1
C. martinii	21[c]		21	42	3
C. maryannum	21[c]		21	42	3
C. muehlenbergii s.l.	20[c]		20	40	21
*C. multicaule***			21	42	3
C. namophilum var. *namophilum*	17[b]		17	34	2
C. namophilum var. *nevadense*	17[d]	34[d]	17	34	2
*C. sp.nov1***			37	74	10
C. nudicaule	21[c]		21	42	2
C. quitense	21[c]		21	42	6
C. quitense	22[c]		22	44	6
C. setaceum	21[c]		21	42	3
*C. strictum**	21[c]		22	44	7
*C. tenuifolium***			36	72	1
*C. texense**	21[c]		20	40	7
C. trichanthum	17[c]		17	34	5
*C. sp.nov2***			22	44	1
C. venustum	17[c]		17	34	3
C. wigginsii	22[c]		22	44	2

[a] Zeltner, 1970; [b] Reveal et al., 1973; [c] Broome, 1978; [d] Broome, 1981; [e] Turner, 1993

Table 3. Alignement and sequence characteristics of the different nrDNA and cpDNA regions investigated in *Centaurium*.

	ITS1 + ITS2	ITS1	ITS2	*trn*L+F	*trn*Lintron *trn*F-F spacer	
Length range (ingroup) (bp)	453-463	223-235	222-236	767-795	397-417	367-393
Length mean (ingroup) (bp)	460.2	229.8	230.5	772.3	401.7	370.6
G+C content (range %)	59.6-63.3	58.1-62.9	61.1-64.8	34.2-35.4	35.6-36.9	32.4-34.2
G+C content (mean %)	61.7	60.6	62.6	34.9	36.3	33.4
Aligned length	479	239	240	852	478	374
Number of excluded sites	17	17	0	51	19	32
Number of coded indels	0	0	0	8	5	3
Number of sites after exclusion[a]	462	222	240	793	419	374
Length of indels (range) (bp) 0	0	0	1-12	1-12	1-8	
Length of indels (total)	0	0	0	37	26	11
Parsimony informative	170	83	87	42	13	29
constant	226	108	118	702	391	318
uninformative	66	31	35	42	15	27
indels (number)	65	28	37	58	39	19
g1-stat	-0.57	-0.58	-0.56	-0.73	-0.75	-0.79

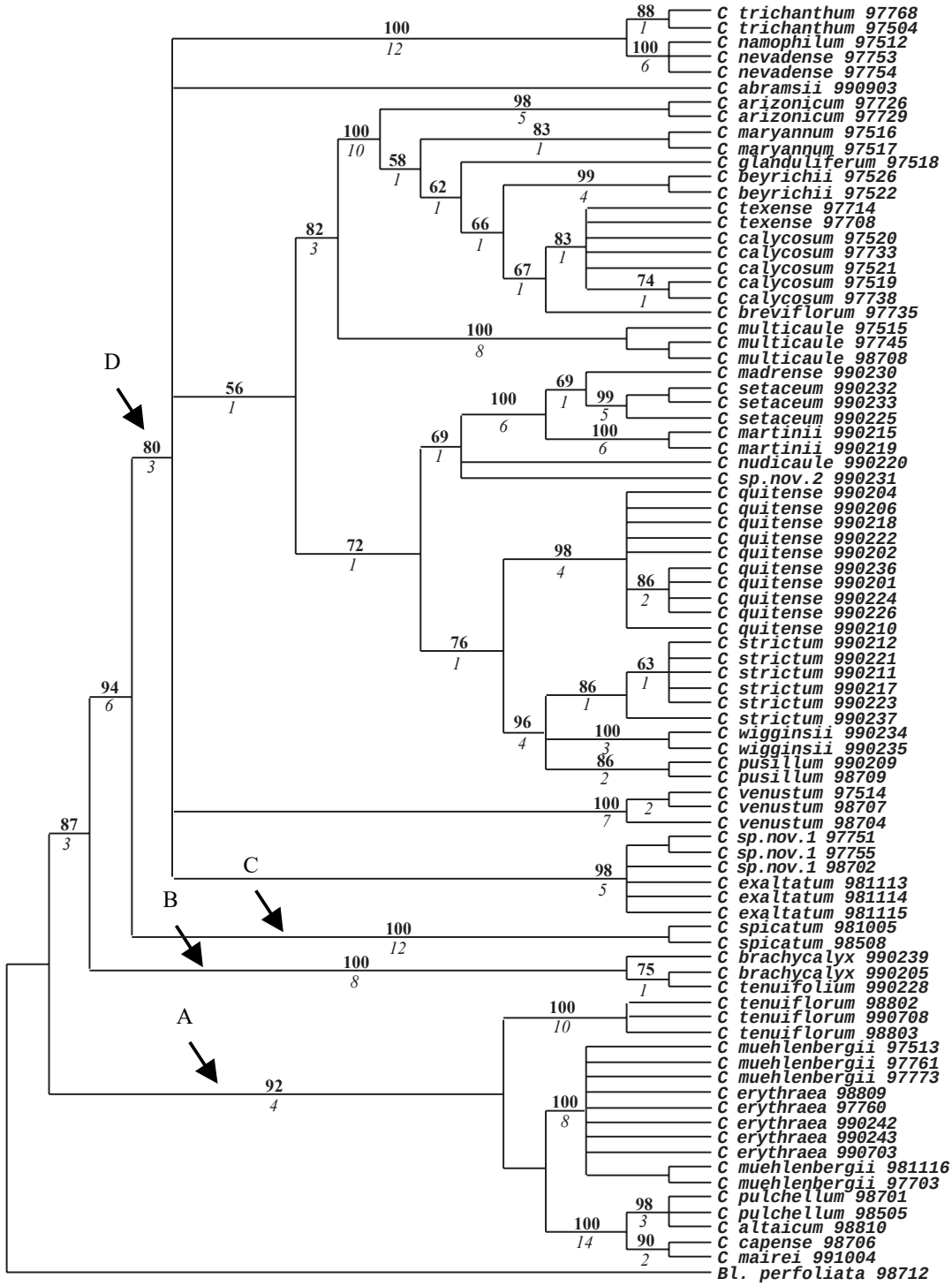


Figure 1. Strict consensus of 9000 MP trees (448 steps) obtained from parsimony analyses of ITS1 and ITS2 sequences combined. Numbers above and below branches indicate bootstrap values of 100 replicates (in %) and decay index, respectively. Letters A to D represent clades described in the text.

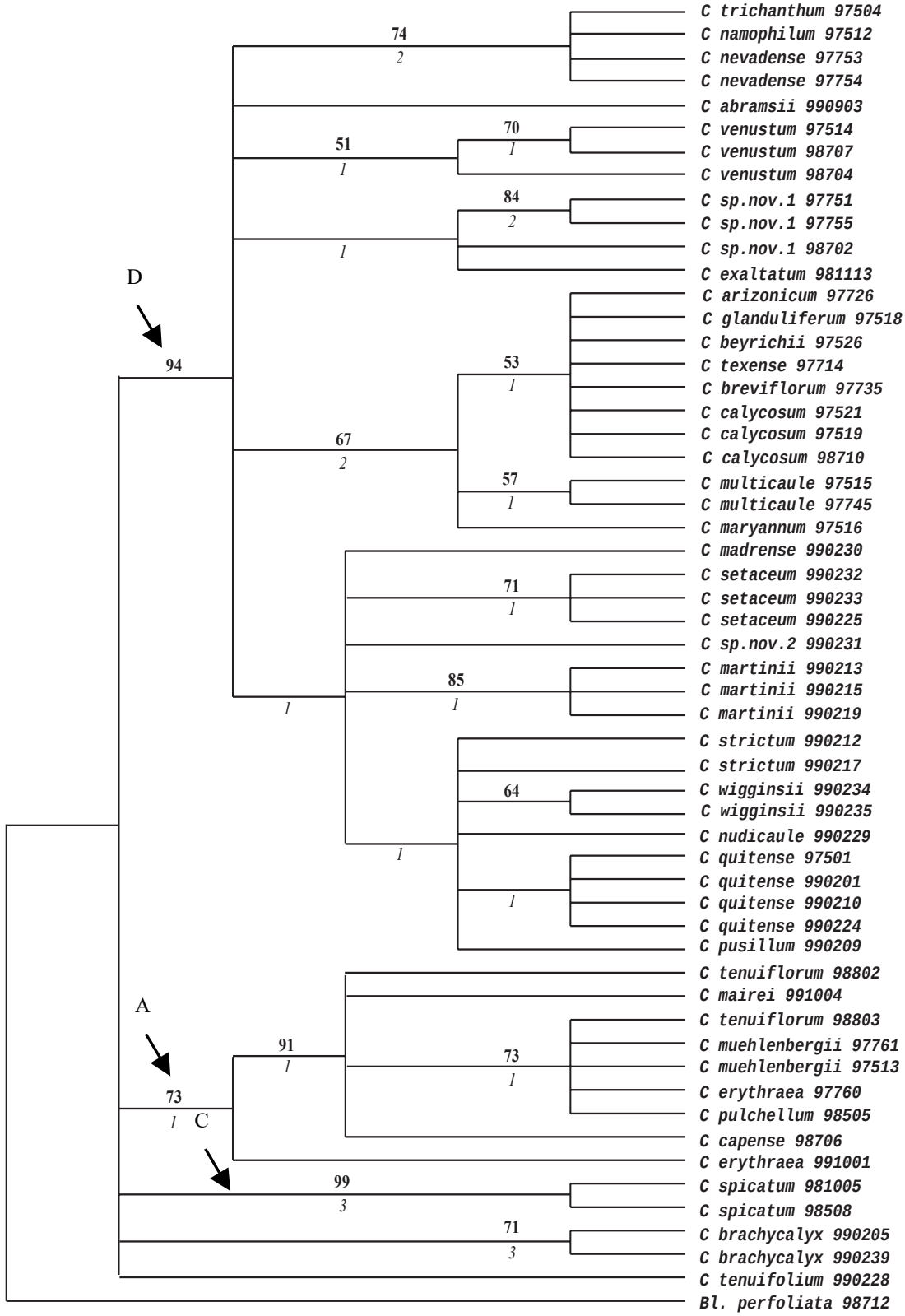


Figure 2. Strict consensus of 5000 MP trees (113 steps) obtained from parsimony coded analyses of *trn*L and *trn*F sequences combined. Numbers above and below branches indicate bootstrap values of 100 replicates (in %) and decay index, respectively. Letters A to D represent clades described in the text.

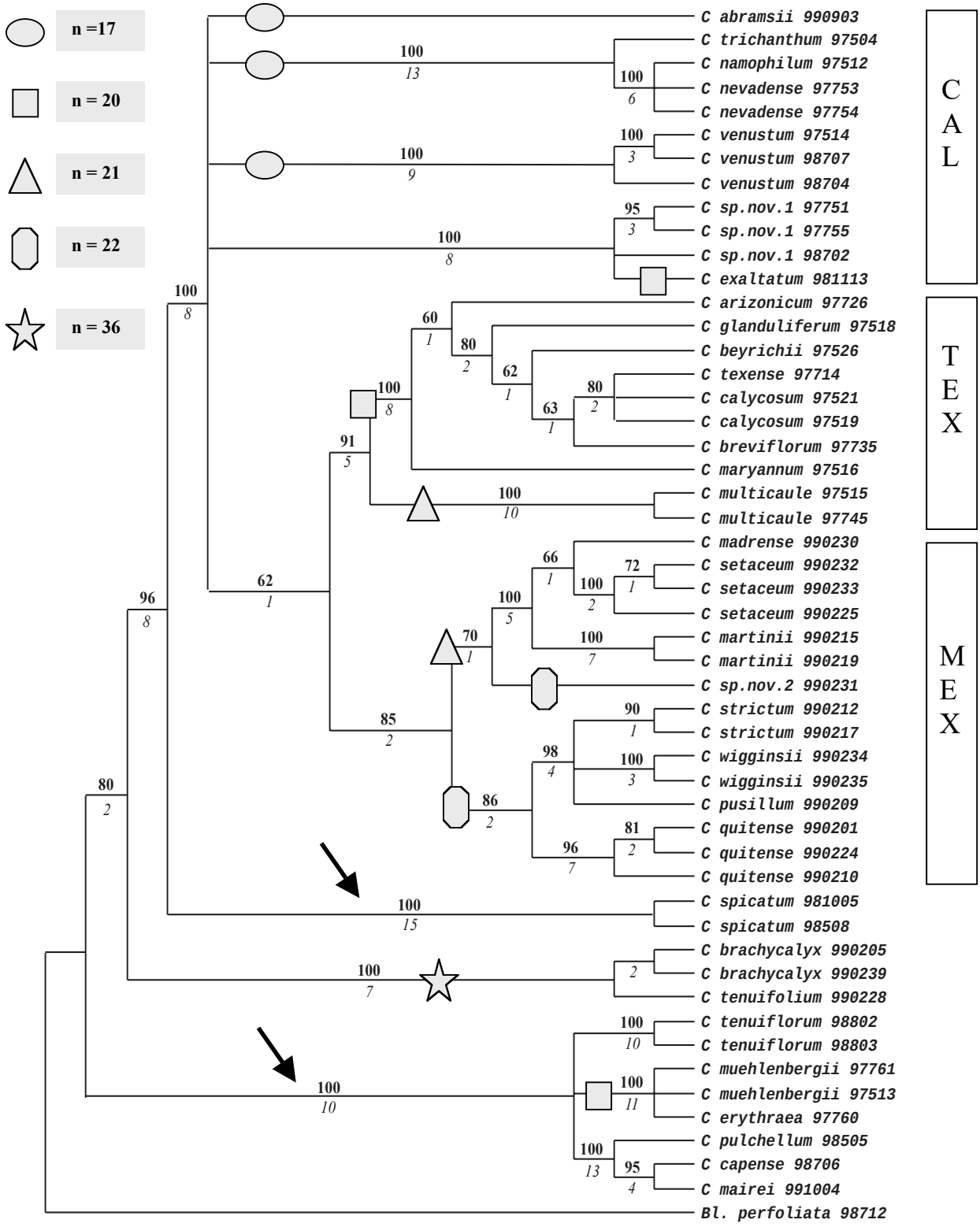


Figure 3. Strict consensus of 972 minimal trees (529 steps) obtained from parsimony coded analyses of ITS1, ITS2, *trn*L and *trn*F sequences combined. Numbers above and below branches indicate bootstrap values of 100 replicates (in %) and decay index, respectively. Geographical repartition is indicated on the right (CAL: California / TEX: Texas / MEX: Mexico). Symbols represent haploid chromosome numbers. Arrows signal Eurasian clades.

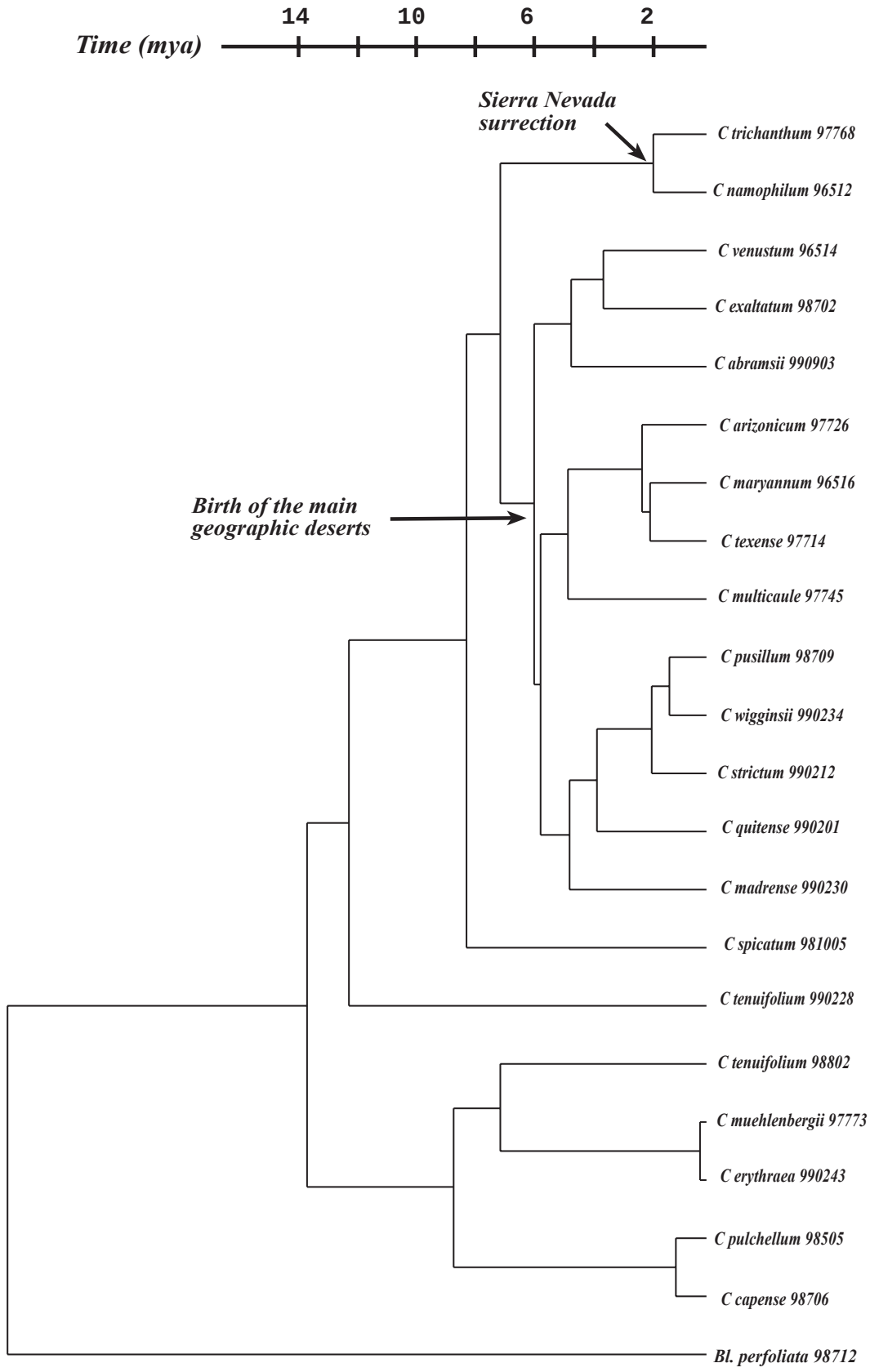


Figure 4 . Phylogram of the maximum likelihood ITS tree (HKY model) for *Centaurium*, under the assumption of a molecular clock

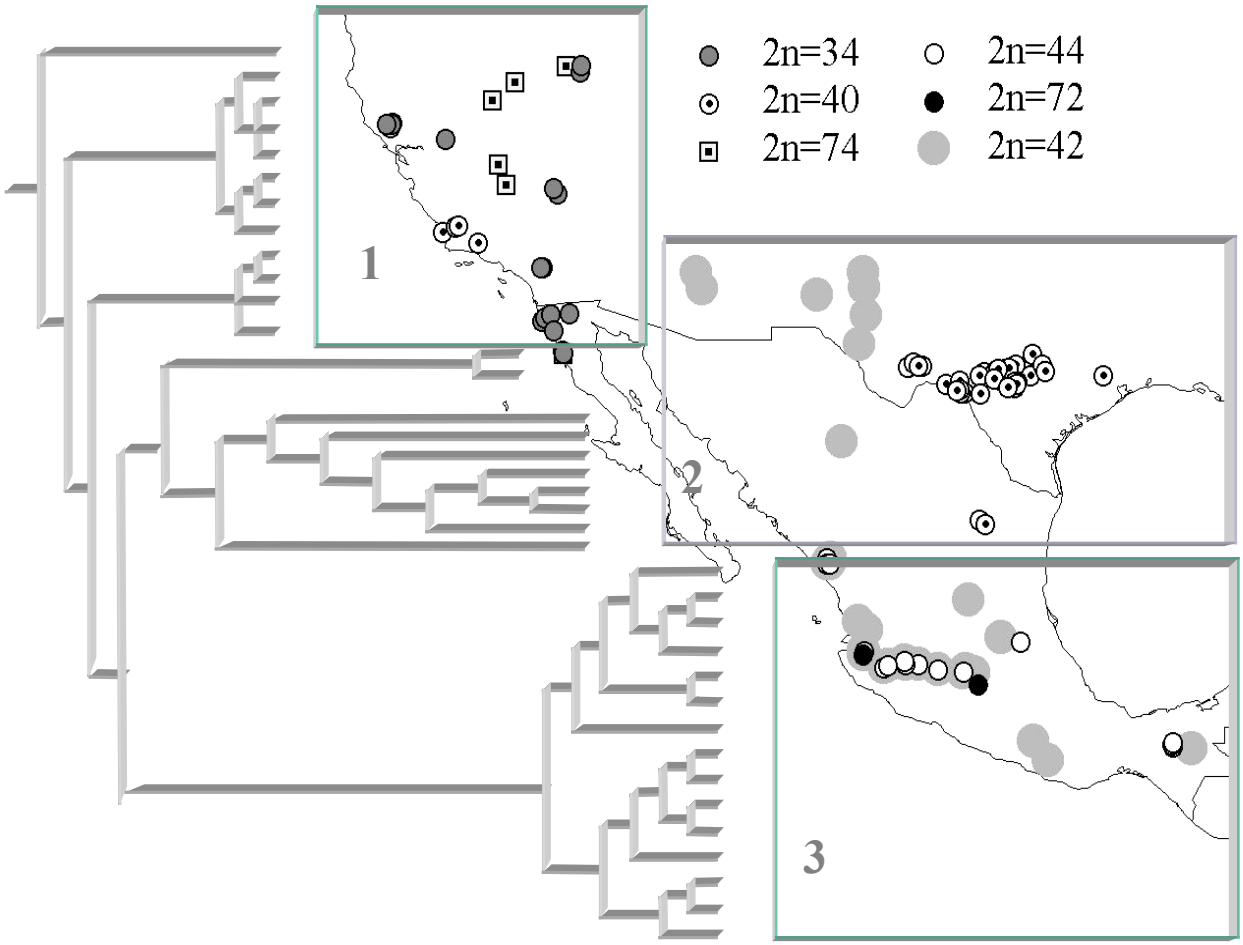


Figure 5. Repartition of the different populations sampled. Numbers represent the main geographic groups described (see text).

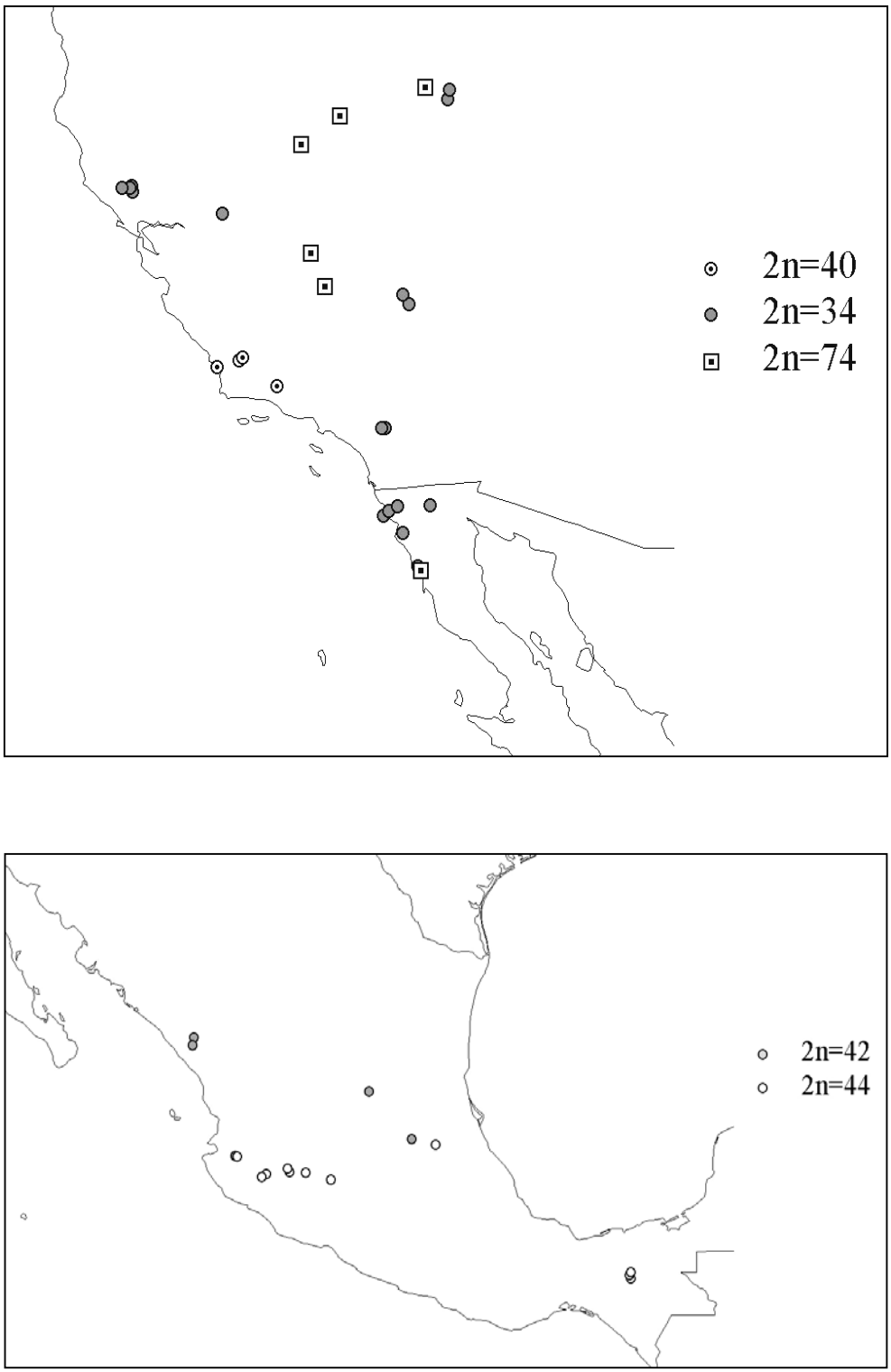


Figure 6. Map A: Geographic repartition of *Centaurium sp.nov.1,* a putative allopolyploid species (2n = 74). Map B: Geographic repartition of two cytotypes of *Centaurium quitense (*see text for details).

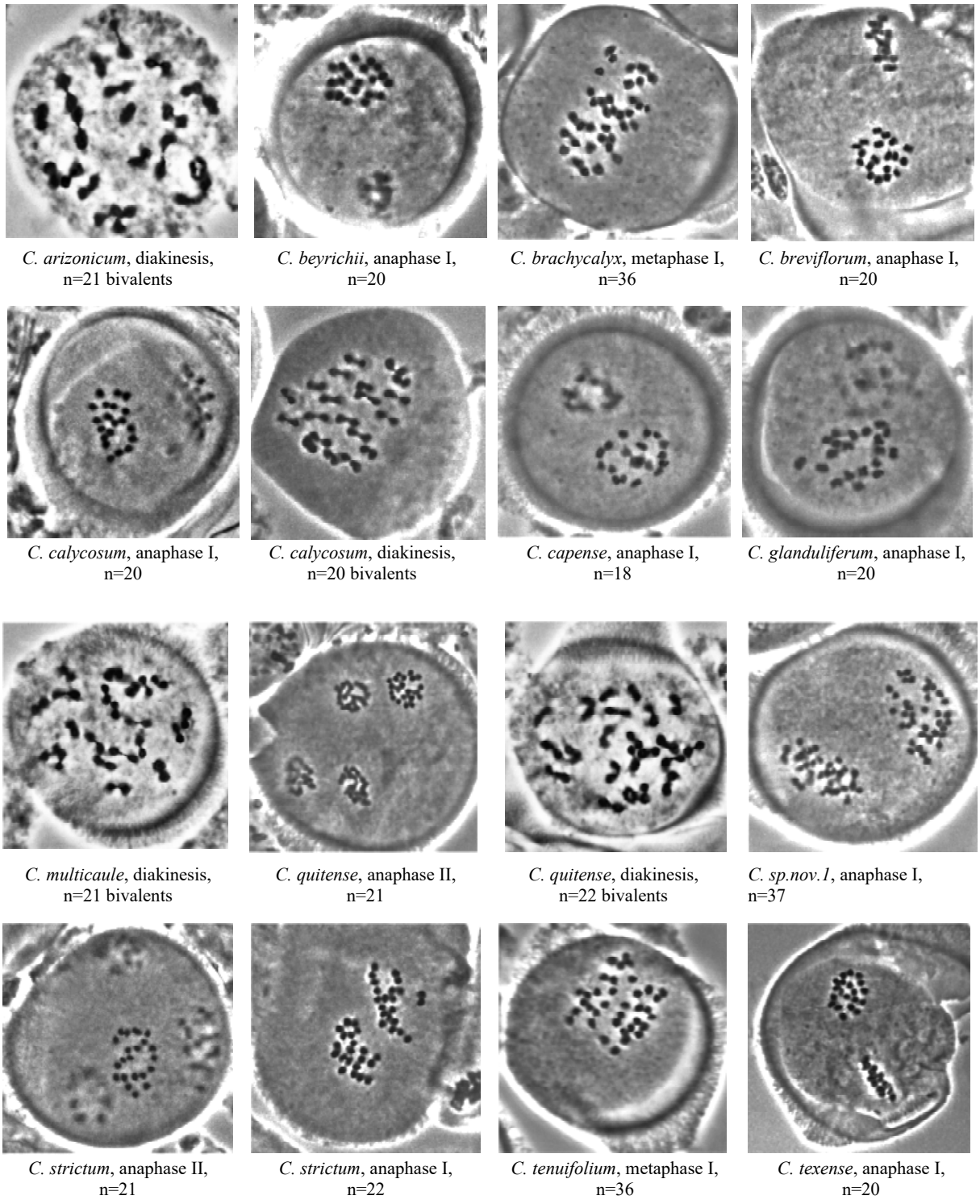


Figure 7. Microphotographs of North American *Centaurium* chromosomes. Each plate represents either a new or a different number compared to those previously reported.

Chapter 2

Biogeography and molecular evolution of the Old World species of *Centaurium* (Gentianaceae – Chironiinae)

Biogeography and molecular evolution of the Old World species of *Centaurium* (Gentianaceae – Chironieae)

ABSTRACT

The Eurasian group of *Centaurium* comprises around 30 species and several infra-specific taxa mainly distributed around the Mediterranean basin. In this area, polyploidy appears to be an important trend of evolution, for diploid, tetraploid and hexaploid species occur. On the other hand, hybridization phenomena may play a significant contribution in the speciation process, and moreover, frequently obscures taxa delimitation. The present study intends to infer phylogenetic relationships of Eurasian *Centaurium*, in order to understand the underlying evolutionary trends within the genus, to measure the importance of polyploidy process, and finally to contribute to the systematics treatment of the group. Moreover, a large sampling of taxa, collected for 30 years and consisting in more than 1400 well-delimited populations, with known chromosome valences, was used to reconstruct the biogeographic patterns of the Eurasian *Centaurium*. For molecular systematic purposes, sixty-six and seventy sequences of the nrDNA ITS region and cpDNA *trn*LF region were used, respectively. Phylogenetic trees inferred from the ITS data set depict five main clades. The most basal group, which corresponds to the taxonomic section Spicaria, is closer to the genus *Exaculum* than to the remaining species of *Centaurium*. Within that subset a main assemblage comprising *C. erythraea* and relatives taxa, is completely unresolved. The weak ITS sequence divergence within this clade, along with the occurrence of morphologically well-differentiated species, may indicate a recent evolution of the section and strong environmental pressures. The biogeographic history of the group, inferred under a molecular clock assumption, reveals that two main geological events, namely the Messinian-salinity-crisis and the Quaternary-ice-ages, have played a major role in the current repartition of the species. Nowadays, diploids occur in areas corresponding to ancestral glacial refugia, whereas repartition of polyploids is more peripheral and may depict post-glacial re-colonization patterns.

INTRODUCTION

Polyploidy (i.e. the presence of more than two genomes per cell) is a speciation process widely active in angiosperm evolution. Estimations of plant species that have experienced at least one polyploidization event in their lineage vary from 50% to 70% (Averett, 1980; Grant, 1981). Hence, polyploidy is now considered as one of the

major evolutionary trends in plant speciation (Thompson and Lumaret, 1992; Soltis and Soltis, 1993; Bretagnolle and Thompson, 1995; Ramsey and Schemske, 1998; Petit *et al.*, 1999). Different processes result in the emergence of polyploids depending on the degree of genetic differentiation between parental species (Bretagnolle *et al.*, 1998; Ramsey and Schemske, 1998). First, polyploids may result within a single diploid species, without hybridization event, and are called autopolyploids (Stebbins, 1947; Lewis, 1980). It has been stressed that the latter are of significant evolutionary importance (Soltis and Soltis, 1993). Second, polyploids may results from hybridization processes between genetically distinct diploid taxa. This process characterizes allopolyploidy and involves basically a hybridization process in addition to a polyploidization event.

The genus *Centaurium* (Gentianaceae) appears to be a useful model to appraise the process and pattern of polyploid evolution in Europe and surrounding areas. Presently, some polyploid complexes, comprising diploid, tetraploid, and sometimes hexaploid taxa, have been described and their distribution has been partially established (Zeltner, 1970). Some diploid species are restricted to areas that seem to correspond to some of the glacial refugia described (Vogel *et al.*, 1999). On the other hand, tetraploids generally show a more northern and peripheral repartition, suggesting pattern of post-glacial re-colonization (Hewitt, 1999).

Quaternary climatic changes have had a strong influence on the geographical distribution of plant species and possibly on *Centaurium*. During the last 2.3 million years, several continental glaciations occurred and are thought to be in phase with the Milankovitch cycles (Hays *et al.*, 1976). These cycles result from fluctuation in the Earth's orbit around the sun, producing variations in the amount of solar radiation received by the Earth (Berger, 1978). During a phase of low solar radiation, temperatures dropped and led to an advance of the ice sheets. In the same time, the sea level decreased, as water became locked in the ice, resulting in the emergence of new coastal areas and land bridges. With a lowering of temperatures, ice covered the main part of the land leading to a full glaciation. Following this, temperatures rose, ice began to retreat and sea levels increased, characterizing an interglacial period. Such alternation of glacial / interglacial cycles occurred several times during the Pleistocene (Ferris *et al.*, 1999). These climatic changes have greatly altered the environmental conditions and many species would either have to adapt to the changes, migrate or die. Many temperate plants have survived periods of glaciations in several refugia, i.e. areas where species persist, at any population density, during a cold stage (Bennett *et al.*, 1991). For European trees, 17 refugia have been identified, mainly in the South (Huntley and Birks, 1983), and have been later confirmed for the fern genus *Asplenium* (Vogel *et al.*, 1999). It may be possible that these southern refugia have been isolated from each others during each glaciation, allowing the populations to diverge (Ferris *et al.*, 1999). Then, with the rise of temperature during the interglacial

period, large vacant areas due to glaciers retreat became suitable for colonization, and the temperate flora rapidly migrated northward. This leading edge expansion, probably by long distance dispersal, and repeated many times would have led to loss of alleles and homozygoty (Hewitt, 1996).

These different factors, such as the repeated episodes of both genetic diversification in refugia and loss of alleles in re-colonizing populations, along with the occurrence of geographical barriers to gene flow, may have played a major role in the current range and genetic diversity of many Mediterranean species and particularly *Centaurium*. This hypothesis implies recent origin and establishment of polyploid taxa in the present area of distribution.

The goal of this study is to reconstruct the molecular phylogeny of Eurasian *Centaurium* in order to check their monophyly and the evolution of the putative polyploid complexes previously described. The different gene trees obtained will serve as a framework to evaluate the robustness of the existing systematic treatment (Melderis, 1972). Moreover, the later is presently incomplete, focusing only on European species, and lies partially on karyological data (Zeltner, 1970).

The present paper also intends to reconstruct the biogeographic scenario of this group. Several points must be discussed, such as the estimated age of Eurasian *Centaurium*, the relative importance of glacial refugia and the occurrence of re-colonization patterns, the congruence between phylogenetic patterns and the paleogeology of the Mediterranean region.

Sequences of the ITS and *trn*L-F regions of nrDNA and cpDNA, respectively, have proven to be useful to reconstruct phylogenic trees and corresponding area-cladograms in the New World *Centaurium* (Chapter 1). Thus they will be addressed here to infer phylogenetic relationships within the study group. On the other hand, a large set comprising ca.1400 well-located populations (GPS coordinates) of Eurasian *Centaurium*, with several information such as chromosome numbers or ecological data (Zeltner, unpublished data), will serve as a framework to test different biogeographic hypotheses.

MATERIAL AND METHODS

DNA extraction, amplification and sequencing---Sixty-six accessions of 27 *Centaurium* species were sequenced for the combined ITS1 and ITS2 regions and 70 for the cpDNA region. Plant material comes from natural populations (silica gel method) or from the herbarium of the university of Neuchâtel (NEU).

Total DNA extraction was performed from leaf tissues using either the CTAB method (Doyle and Doyle, 1987) with an additional cleaning step using ammonium acetate 7.5 M, or applying the DNAeasy-kit (Qiagen, Basel). Double stranded DNA were amplified by symmetric PCR using a Perkin-Elmer or a Biometra thermo-cycler with the following conditions: 1 cycle of 3 minutes at 94°C linked to 30 cycles of 10 sec at 94°C, 20 sec at 50-55°C, 1min30 at 72°C and then 4 min at 72°C to complete primer extension.

DNA region were amplified with 3 pairs of universal primers: ITS1 (GGA AGT AGA AGT CGT AAC AAG G) / ITS2 (TCC TCC GCT TAT TGA TAT GC) (White *et al.*, 1990) for the nrDNA Internal Transcribed Spacer region; *trn*Lc (CGA AAT CGG TAG ACG CTA CG) / *trn*Ld (GGG GAT AGA GGG ACT TGA AC) for the *trn*L (UAA) intron, and *trn*Le (GGT TCA AGT CCC TCT ATC CC) / *trn*Lf (ATT TGA ACT GGT GAC ACG AG) for the *trn*L-*trn*F intergenic spacer (Taberlet *et al.*, 1991). PCR products were first checked on a 1% agarose gel stained with ethidium bromide (in order to evaluate the quality and quantity of the amplified templates) and then purified using the QIAquick PCR purification kit (Qiagen). Variable ITS bands were separated directly on the gel, by cutting and purifying the desired region with the QIAEX II Gel Extraction Kit (Qiagen). When this separation was not possible, cloning methods were used. PCR products were ligated with plasmids (pGEM(r)T vector) and transformed into *Escherichia coli* competent cells, previously selected on LB plates containing ampicillin and X-gal (pGem(r)T and pGem(r)T easy Vector Systems, Promega). Sixteen white *E. coli* clones (in which ligation was successful) were picked and cultured on LB containing ampicillin broth, for isolating plasmids. Purified plasmids DNA or direct pure PCR product were sequenced automatically on an Applied Biosystems 310 DNA sequencer, using the dideoxy chain termination technique (BigDye Terminator Cycle Sequencing Ready Reaction Kit of Perkin Elmer). A five µl reaction was performed using 2ul of premix (containing the 4 labeled terminators, the deoxynucleoside triphosphate, the AmpliTaq DNA, $MgCl_2$ and Tris-HCl buffer), 0.2 µl of primer (10 ng/µl) and 2.8 µl of amplified DNA (about 10 ng/µl). The succeeding program was performed on a thermo-cycler (Biometra or Perkin Elmer) : 30 cycles of 10 sec at 96°C followed by 4 min at 60°C.

Sequences were carefully checked and edited using Sequence Navigator (Perkin Elmer Corporation*). Multiple alignment was produced with Clustal X (Higgins *et al.*, 1996), with subsequent manual arrangements. Sixty-six and seventy sequences of the ITS and *trn*LF regions were selected to produce data matrix, respectively.

Phylogenetic analyses

Maximum parsimony methods--All sequence data sets were analyzed using the maximum parsimony criterion as implanted in the PAUP 4.0b6 version (Swofford, 1999). The following computing options were used: simple sequence addition, TBR branch swapping, steepest algorithm ON, Mulpars ON and gaps treated as missing data. Branch support evaluation was performed using both bootstrapping (Felsenstein, 1985) and Bremer index calculation (Bremer, 1988). Bootstrap values (decay indices) were obtained using 100 replicates and resampling all the characters in the respective data matrix, with heuristic each using simple sequence addition (maxtree limit=100). Decay values were obtained using TreeRot.v2 (Sorenson, 1999).

Outgroup choice--Respective analyses were rooted with the genus *Blackstonia* (Chapter 1). Another taxon, *Exaculum*, was included in the analysis. This particular Eurasian genus is thought to be close to *Cicendia*, which is sister clade to *Blackstonia* (Thiv *et al.*, 1999; Struwe *et al.*, 200*). On the other hand, some morphological characters clearly discriminate between *Cicendia* and *Exaculum,* and the latter may also have affinities with *Centaurium*. Therefore, the systematic position of *Exaculum* needed to be investigated in the present study.

Phylogenetic signal and congruence--Phylogenetic signal within each data set was measured with the skewness test, by calculating the g1 value (Hillis and Huelsenbeck, 1992). The null hypothesis of this test assumes that the data has no signal and so the shortest tree is just the shortest tree for a symmetric distribution of tree length. Therefore, the distribution of the tree length will not be skewed if there is not any signal.

In order to check congruence between respective data sets before further combination, we performed the Partition Homogeneity tests as implanted in PAUP 4.0 (Farris *et al.*, 1995). The combined matrix was divided in two partition (nrDNA and cpDNA) and 1000 replicates were used for this test.

On the other hand, comparison between classical systematics treatment of Eurasian *Centaurium* and phylogenetic clades obtained in the respective analyses, were conducted using the Templeton's significantly less parsimonious test or SLP test (Templeton, 1983). The different sections and subsections were used as constraints to produce alternative topologies of the inferred molecular trees.

Sequence divergence and molecular clock--The ITS data matrix was analyzed under a maximum likelihood criterion, using the HKY 85 model of sequences substitution. The pairwise distances between selected groups, under the model chosen, were computed in PAUP. In order to check the null hypothesis of a molecular clock within the ITS sequences mutation rates, likelihood ratio tests were performed (Hasegawa *et al.*, 1985; Goldman, 1993; Yang *et al.*, 1995; Huelsenbeck and Crandall, 1997). Different subsets of the original ITS data matrix, comprising representative species for each clade, were first created. Phylogenetic trees were then evaluated with or without enforcing a molecular clock, and their respective scores were computed (Table 4). The calibration of

internal nodes was performed by using fast and slow mutation rates estimated from ITS sequences in some Gentianaceae groups (Table 3). The divergence time of the different groups can be calculated with the formula: T = D / 2r, D being the mean distance divergence (HKY85 model) and r, the mutation rate (Schnabel and Wendel, 1998). Lastly, estimated ages on the different nodes can be compared with geological events.

Data mapping---The ArcView program was used to transform the GPS coordinates of each populations investigated, in data point available for further mapping. We retained a total of ca. 1400 populations, sampled over 30 years (Zeltner, unpublished data), where several parameters such as chromosome number counts, phenology and ecological conditions were available.

RESULTS

Main characteristics of the DNA sequences (ITS / *trn*L-F)---The main characteristic of the different DNA regions sequenced are shown in Table 3. Total length of the ITS sequences, excluding the 5.8S cistron, ranged from 453 to 464 base pairs (bp) with a mean GC content of 62.23 %. Both the ITS1 and ITS2 spacers are quite similar in length (229 bp vs. 231 bp) and mean GC content (60.48 % and 63.96 %, respectively). The introduction of 45 gaps was necessary in the multiple alignment of 66 sequences, producing a matrix of 474 characters from which 115 were parsimony informative (24.2 %). Analyses were performed after deletion of 2 ambiguous characters (labile gaps) and 9 informative indels that were coded. The resulting data matrix comprised 469 characters (45 gaps and 119 informative characters). Sequences of the *trn*L intron and *trn*L-F spacer were relatively similar in length range (382 - 417 bp and 366 - 376 bp, respectively) and mean GC content (36.61 % and 33.64 %), respectively. For phylogenetic analyses, 9 ambiguous characters were excluded from the matrix and informative indels were coded. Aligned sequences were 787 bp long, with 16* gaps (2%) and 39 parsimony informative characters (5%).

Phylogenetic signal and congruence between respective data sets---In the present study respective data sets have skewness values ranging from g1= - 0.40 to g1 = -0.50. This significant negative skewness rejects the null hypothesis of no signal (Hillis and Huelsenbeck, 1992). Therefore ITS and *trn*LF data sets should be suitable for phylogenetic analyses.

Heuristic searches performed on ITS data provided an unknown number of trees (reaching computer memory limit) with 314 steps (C = 0.748, R = 0.25). Five main clades are supported by high bootstrap values and Bremer indices. Members of the *Spicaria* section (clade A), including the genus *Exaculum,* form the most basal assemblage (BS = 100 - DI = 7). The set *C. maritimum*-*C. pulchellum*-*C. malzacianum* (clade B: BS = 97 - DI = 5) is neighbor with the *C. tenuiflorum* assemblage (clade C: BS = 99 - DI= 6). Lastly, member of the *C. littorale* group (clade D: BS = 70 - DI = 1) come close the large and unresolved *C. erythraea* section (clade E: BS = 92).

Analyses of the chloroplast data set resulted in more than 5000 trees (L = 129 / C = 0.729 and R = 0.918) and the strict consensus tree obtained did not satisfactorily resolve the ingroup relationships (Fig. 2). Only the basal position of *C. spicatum* group is well supported (BS = 94 - DI = 3). *Exaculum pusillum* comes close the latter forming a weak supported assemblage (BS = 54), concordant with clade A of the ITS tree. Within the remaining large set, subclades may be drawn, such as *C. maritimum*-*C bianoris* (BS = 100 - DI = 7) and a large data set comprising mainly accessions of *C. tenuiflorum*, *C. pulchellum*, *C. muehlenbergii* and non-European populations of *C. erythraea* (BS = 81).

The partition homogeneity test revealed significant heterogeneity ($P < 0.001$) between the nuclear and chloroplast sequences of a combined data set comprising 43 taxa and 1259 aligned nucleotides. Nevertheless, in order to detect incongruence, this matrix was submitted to maximum parsimony analyses (informative indels coded, n=1232 characters), and inferred trees (L=467; C = 0.719 / R = 0.863) were highly similar to the ITS ones, except for the placement of *C. bianoris* in clade C (instead of clade B in ITS analysis). On the other hand, the five clades, highly congruent with clade A to E of the ITS analysis, are supported by bootstrap values ranging from 61 to 100 (analysis not shown).

The Templeton's significantly less parsimonious test (SLP test) was first performed on the trees obtained for each separate partition in the combined matrix. The ITS data set was used to test whether the most parsimonious *trn*LF tree is significantly less parsimonious than the ITS shortest tree. Probability values were highly significant ($P < 0.0001$) indicating that the ITS data statistically not support the *trn*L topology (Johnson and Soltis, 1998). On the other hand, results of the SLP test with different constraints resulted in generally significant differences with the most parsimonious tree, suggesting conflict between the traditional classification and the phylogenetic reconstruction (not shown).

Sequences divergence, molecular clock and biogeographic events timing---DNA divergence values between the ITS sequence were calculated by using the HKY85 parameters distances. The average sequence divergence

between the ingroup members was 6.48 +/- 0.01 % (4.91 +/- 0.01 %, excluding clade A). The highest divergence was 14.86 % (between diploid *C. spicatum* and *C. pulchellum*) within the ingroup and 25.46 % in the whole data set (between the respective outgroups *Blackstonia* and *Exaculum*). The lowest value was 0 between several members of the *C. erythraea* group. In clade E. the sequence divergence ranges from 0 to 1.77%.

The likelihood scores of the different subsets, obtained by pruning selected taxa, and their respective ratio tests are shown in Table 4. The hypothesis of a molecular clock could not be rejected in all the cases, e. g. after keeping or pruning the different genera *Blackstonia* and *Exaculum*, the *Spicaria* section, and the putative hybrid taxa.

Biogeographic repartition of the Eurasian *Centaurium*---Five main polyploid complexes occur within the Eurasian *Centaurium* clade. The *C. spicatum* group shows a large repartition with diploid taxa confined in the Mediterranean basin and tetraploids in the Southern Hemisphere such as Australia or Easter Island (Map 1). The *C. pulchellum* complex (Map 2) comprises some diploid populations restricted in the region of Israel. On the other hand, tetraploid populations (*C. pulchellum* var. *pulchellum* and var. *altaicum*) range from Finland coast to South Morocco, and from Atlantic coast to eastern China, whereas hexaploid (*C. mairei*) mainly occur in the southern part of the area. Finally, *C. centaurioides* subsp. *malzacianum* and subsp. *centaurioides* occupy peripheral and possibly vicariant areas in the Arabian Peninsula and India, respectively (Map 2). Within the *C. tenuiflorum* clade (group C), subspecies *acutiflorum* (2n = 2x = 20) may be divided in two diploid varieties occupying roughly western (var. *hermanii*) and eastern (var. *acutiflorum*) Mediterranean areas, respectively. Tetraploid populations (subsp. *tenuiflorum*) are circum-Mediterranean and radiate northerly (Map 3). In this case, tetraploid area seems smaller, more confined, than the diploid one. In the *C. littorale* complex, diploid (*C. scilloides*) may be encountered in the Atlantic coasts along with tetraploid *C. chloodes. Centaurium. littorale* subsp. *littorale* has a more northern repartition, and *C. littorale* subsp. *uliginosum*, an intermediate one, with a radiation to the east (Map 4). Lastly, the *C. erythraea complex* is a very large group comprising diploid to hexaploid taxa. Diploid repartition (Map 6) indicates that most of the taxa have either restrictive areas located in Iberian Peninsula (e. g. *C. favargeri* group, *C. majus* subsp. *majus*) and Morocco (*C. bernardii*, *C. suffruticosum*), or more expanded ones (*C. rumelicum*). In the *C. majus* assemblage, diploids have eastern repartition, with a clear fragmentation between the subspecies *majus* and *C. suffruticosum* in North Morocco, while tetraploids occur at the west of a line Sardinia / Sicily. *Centaurium erythraea* subsp. *erythraea* is relatively frequent in the Northern part of the Mediterranean basin, and some variants can be encountered on the margin of

its distribution (subsp. *capitatum* / *C. turcicum* etc). This taxon is often introduced in many countries, mainly on the coasts (Australia, North and South America, South Africa).

Putative hybrid species and parents have been mapped (Maps 2 to 5) and their origin is discussed below (Table 5).

DISCUSSION

Phylogenetic reconstruction of Old World *Centaurium* species and systematics implications---Inferred phylogenetic trees rooted with the genus *Blackstonia* reveal two major data sets within Eurasian *Centaurium*, namely the *C. spicatum* group with x=11 chromosomes, and all the remaining species with x = 10 or x = 9, momentary called *Centaurium* sensu stricto.

Systematic position of the *C. spicatum* group--Members of this well-supported clade comprise section *Spicatum* characterized by spiciform-cymed inflorescences and basal chromosome number of n = 11 (Zeltner, 1970). These species are mainly centered in the Mediterranean basin and have sometimes been introduced in foreign areas (Australia or North America). *Centaurium spicatum* was previously considered to be a diploid taxon in its whole distribution range. The present study reveals, for the first time, the occurrence, in the Southern hemisphere, of tetraploid populations in Australia and Chile. This biogeographic distribution suggests a North - South evolution within this particular polyploid complex and reinforces its marginal position within the Eurasian *Centaurium* set. Moreover, the affinity with the genus *Exaculum* is striking. This monospecific genus comprises annual, slender, species with tetramerous flowers forming a cyme and having 2n = 20 chromosomes (Favarger, 1960). The closeness of these taxa may be due to the restrictive sampling of related genera in the present study. *Exaculum pusillum* does not fit in the *Blackstonia* clade (high sequence divergence value between the two genera) and members of the clade A seem distant from both *Blackstonia* and *Centaurium* s. s. Moreover, previous study also demonstrate that *C. spicatum* can not be included in the New world *Centaurium* group (Mansion and Zeltner, unpublished), suggesting either the inclusion of *Exaculum* in *Centaurium*, close to the intermediate section Spicaria, or the exclusion of section *Spicaria* and generic rank elevation, with subsequent splitting of *Centaurium*. A global study is presently being performed to check the monophyly and intergeneric relationships of *Centaurium* (Chapter 3).

Systematic position of *Centaurium* sensu stricto--A second large clade comprises all Eurasian *Centaurium* with cymed-inflorescences and a basal chromosome number of x = 10. The four main groups depicted do not always

correspond to the taxonomic sections previously proposed (Zeltner, 1970; Melderis, 1972). A basal clade (B) comprises *C. maritimum* (2n = 2x = 10), *C. pulchellum* complex (n = 9), *C. centaurioides* var. *centaurioides* and var. *malzacianum* (2n = 6x = 28). The present study does not support the varietal ranks of the two last taxa because they belong to separate clades and differ by morphological and biogeographical characters. Their respective origin will be discussed below. *Centaurium malzacianum* Maire comprises vigorous plants, to 100 cm high, with linear leaves, large corolla (15 mm), and mainly occurs in the Arabian Peninsula and Sinai. *Centaurium centaurioides* (Roxb.) Rolla Rao & Hemadri does not exceed 40 cm high, possesses spatulate leaves and smaller corolla (10 mm), and may be encountered in India. This species shows close relationships with the hexaploid *C. mairei* (n = 27), which extend from the Canary Islands to India (Zeltner, 1990). A second subdivision in *Centaurium* s. s. (clade B) comprises *C. tenuiflorum*, *C. bianori*s and *C. serpentinicola*. The clear separation between *C. tenuiflorum* (x = 10) and *C. pulchellum* (x = 9) on the ITS tree is weakly supported by morphological grounds; the differentiation of these two species is not easy in the field. On the other hand chloroplast approaches may suggest a common ancestor for the two taxa. Nevertheless, the subsection *Parviflora* grouping *C. tenuiflorum* and *C. pulchellum*, on the base of rosette absence and flower size, is obviously polyphyletic. So is subsection *Vulgaria* comprising species with linear leaves and dense-tufted rosettes. Effectively, *Centaurium littorale* - *C. chloodes* - *C. scilloides* (clade D) and *Centaurium favargeri* - *C. gypsicola* - *C. barrelieri* form distinct assemblages, and their inclusion in a same subsection is not supported here. Lastly, the remaining unresolved clade (E) fits the subsection *Centaurium*, and contains *C. erythraea* and relatives, along with *C. muehlenbergii* from North America (Chapter 1). The position of *C. littorale* subsp. *uliginosum* in this clade may be striking for (1) this species is very close, morphologically, with *C. littorale*, and (2) it seems to possess a geographic vicariant repartition. Moreover, chloroplast sequences of two accessions of *C. uliginosum* appear to be relatively different and the position of this taxon remains unclear. An alternative hypothesis, based on the molecular data set, would be a close relationships with *C. erythraea* var. *capitatum* that occurs in the same northern range and shows morphological resemblance.

Biogeographic reconstruction of the Eurasian *Centaurium*---The hypothesis of a molecular clock could not be rejected within the *Centaurium* s. s. group and so branch length may be approximately proportional to time. The tree obtained under the molecular clock constraint (Fig. 3) reveals an early and important separation between two basal lineages (groups A and B) and the remaining clades. The latter form a numerically important group that appears to be of recent evolution. Depending on the mutation rate choice, the estimated age of the

basal nodes (node 1 and 2) range between 9.4 x 10^6 and 5 x 10^6 years, respectively. Therefore early divergence in *Centaurium* s. s. may be estimated during the late Miocene. On the other hand, the derived group (node 3) emerged between 3.2 and 1.7 mya, roughly at the beginning of the Quaternary ice ages (from 2.4 mya to the present).

During the geological history of the Mediterranean basin, two major events have had main effects on the vegetation and flora evolution of this region. First, the drying out of the Mediterranean, known as the Messinian salinity crisis (Ryan, 1973; Bocquet *et al.*, 1978), occurred between 5.57 - 4.93 mya and 5.96 - 5.33 mya, following recent estimations (DeJong, 1998; Krigsman *et al.*, 1999). During this period, xeric and cooler climatically conditions along with the apparition of new emerged bridges, connecting actual islands, have greatly contributed to important plant migrations and changes in the floristic composition (Bocquet *et al.*, 1978). Present distribution of *C. maritimum* and *C. tenuiflorum*, taxa well implanted in all the Mediterranean islands, seems to reflect past kinetic reconstruction of the Mediterranean during the Messinian (Maps 2 and 3).

A second major period, known as the Quaternary ice ages, has played an important and particularly dramatic role in the Mediterranean biota. The severe climatic oscillations, along with glacial flux and reflux produced great changes in species distribution (Hewitt, 2000). The importance of glacial refugia and the genetic consequences of Ice Ages, producing an accumulation of genome differences and adaptations by repeated contractions and post-glacial re-colonizations, have been well documented (Bennett *et al.*, 1991; Hewitt, 1996, Hewitt, 1999; Vogel *et al.*, 1999). Under the clock assumption, the ITS phylogram denotes that most of the polyploid species of *Centaurium* have arisen during the Ice Ages. Moreover, diploid taxa (*C. favargeri* group) were likely present before this period, and presently occupy possible glacial refugia (Map 6).

Applying the mutation rate found in the North American *Centaurium* ITS sequences (5.5 x 10^{-9} per site per year), one can propose the following biogeographic scenario for the Eurasian clade. Diploid species such as *C. maritimum* and *C. tenuiflorum* subsp *acutiflorum*, first arise by the late Miocene (node A = 7.21 +/- 0.07 mya and node B = 6.67 +/- 0.01 mya). These salt tolerant taxa, presently occurring in xeric stations, often near the sea (Zeltner, 1970), may have expand their range during the Messinian, as reflected by their present distribution (Maps 2 and 3). After this period, new diploid species appeared (node C = 1.98 +/- 0.47 mya), with probably more important distribution than today. Then, 2.4 million years ago, climate cooling and first glaciations occurred, and diploid may have survived in several refugia. Climatic oscillations have likely produced new taxa by re-colonization followed by important extinction, and retreats in new refugia. Recent estimations open up the possibility that these refugia remained isolated, not only during the period of glaciations but also during the

warm interglacial periods (Ferris *et al.*, 1999). It is therefore possible for populations of a species to have been isolated in a refugia for several glaciations, or even for the duration of the Pleistocene. Finally, by the end of the Würm period, remaining diploid species, occurred in restricted refugial habitats whereas polyploid extension is presently observed (hexaploid mainly occur in South or eastern parts of the Mediterranean basin).

Evolutionary trends in Eurasian *Centaurium*

Polyploid complexes repartition and *Centaurium* evolution--The present study confirms that most of the diploid species and morphologically close polyploid taxa form natural groups, suggesting that polyploidy may be a major evolutionary trend in *Centaurium*. All the well-defined clades comprises both diploid and polyploid taxa and may be considered as polyploid complexes .

In general, diploid species occur in strict a Mediterranean area, as shown with *C. spicatum*, *C. tenuiflorum* subsp *acutiflorum*, *C. majus* or *C. favargeri* groups. On the other hand, diploid populations of *C. pulchellum* have been located once in the region of Israel, whereas the *C. littorale* group finds its origin on the Atlantic coasts. In this area, a perennial and hypothesized relict taxon, *C. scilloides*, can be found (Zeltner, 1970). These respective areas fit glacial refugia demonstrated for some groups of plants or animals, in Europe (Huntley and Birks, 1983; Vogel *et al.*, 1999; Hewitt, 2000). The highest number of diploid taxa in Eurasian *Centaurium* may be found in the southern Iberian Peninsula (region E of Huntley & Birks - Map 6)**.** *Centaurium favargeri* occurs in the later, but also in the Durance region (regions J ranging from Avignon to the east of the Italian border), indicating an important fragmentation in its ancestral and probably large distribution area. Effectively, this presently very rare species, was collected in the Caucasian region, a century ago (Zeltner, 1987). The refugia area comprising Sardinia and Corsica correspond to some sympatric areas between diploid taxa such as, C. *tenuiflorum* subsp. *acutiflorum* and subsp. *hermanii* (Map 3) or diploid - tetraploid ones, namely *C. majus* subsp. *majus* and subsp. *rhodense* (Map 5). Lastly *C. scilloides* presently occupy a less important Atlantic refugia (Map 4).

Nowadays, polyploid taxa generally occupy a more peripheral range, but in many cases recover the ancestral diploid area and therefore, evolutionary patterns within the polyploid complexes are not easy to draw. Models have been proposed to infer tetraploid cytotype establishment in diploid populations (Bretagnolle *et al.*, 1998). In such simulation, the main factors of polyploid establishment is the degree of niche overlap.

On the other hand, the weak ITS sequence divergence and the lack of clade resolution encountered in the section *Centaurium* may be interpreted as a recent origin of this groups, probably during the Pleistocene ice ages, as suggested below by the molecular clock hypothesis. Moreover, the notable morphological variation encountered

in several taxa such as *C. barrelieri*, *C. gypsicola*, *C majus* or *C. bernardii* may be the reflect of a strong pressure due to environmental selection. Recent and abrupt climatic changes, which have occurred during the Quaternary glaciations, appear to be one of the factors favoring polyploid formation in this group.

In the Eurasian *Centaurium*, postglacial re-colonization patterns may be suggested, from ancestral refugia to new ecological areas, leading to a relatively rapid morphological evolution in this group. In such processes, one can considerate that only a small fraction of the overall genetic diversity in refugia may be dispersed into the areas vacated by the retreat of glaciers (Vogel *et al.*, 1999). Therefore, a relative low genetic diversity is expected in recolonizant and polyploid populations compared to those remain in refugia. Such correlation between polyploid complexes, genetic diversity, and glacial refugia locations have been demonstrated for ferns (Vogel *et al.*, 1999).

Polyploidy and putative hybrid origins--Many species in genus *Centaurium* seem to have an allopolyploid origin through hybridization. DNA markers are suitable tools to prove such kind of patterns and, in some cases, to suggest parent identity. (Rieseberg and Soltis, 1991; Rieseberg and Wendel, 1993; Arnold, 1997)

The more striking example is *C. bianoris*, an insular endemic species occurring in the Balearic Islands (Majorca). According to Zeltner (1978), this tetraploid species with either salmon, yellow or pink flowers, is supposed to be an hybrid between diploid *C. maritimum* (yellow flowers - n = 10) and *C. tenuiflorum* subsp *acutiflorum* (pink flowers - n = 10). These putative hybrids populations are encountered only in Majorca Island, where the two possible parents species also occur. Most of the *C. bianoris* plants generally resemble *C. maritimum* in their vegetative outlook and seem to have a similar phenology. Notwithstanding, the salmon color of the corolla, clearly intermediate between yellow and pink, and the tetraploid chromosome valence argue for an hybrid origin. Moreover, the presence of 3 patterns of color in the hybrid flower (respectively, yellow, salmon or pink), have been considered as the result of the Mendelian inheritance of this character (Zeltner, 1978). The present work confirms *C. tenuiflorum* to be the pollen parent and *C. maritimum* to be the recipient one. Nevertheless, tetraploid forms with yellow or pink flowers may also be autotetraploid. Moreover, multiple recurrent origins of these allopolyploids may be tested.

Another undescribed case of speciation by hybridization can be noted with *C. malzacianum*. This hexaploid taxon (n=28) falls near *C. maritimum* (n = 10) in the cpDNA analyses and close to the *C. pulchellum* group (n = 18) in the trees inferred from nrDNA sequences. Presently, the mother parent, *C. maritimum*, in contrast to the other taxa implicated, does not occur in the Arabian Peninsula (Oman, Yemen). Much evidence such as basal position in phylogenetic analyses, restricted ecology or diploid chromosome valence, support ancient origin for

C. maritimum. Indeed, the present geographic distribution reveals a fragmented area ranging from the Atlantic coast to Turkey in its eastern limit, but some populations of *C. maritimum* in Egypt were signaled 50 years ago (Tackholm, 1956). Therefore, one can expect an original widespread repartition for this taxon ranging as far as Arabian Peninsula.

Along with *C. malzacianum*, *C. centaurioides* shows the same putative allopolyploid chromosome number (n = 28) and occurs in Indian region. Several hypotheses concerning its origin can be drawn: (1) a scenario similar to the *C. malzacianum* one, but with *C. pulchellum* as recipient parent; (2) a vicariant origin with *C. malzacianum*, regarding the respective repartition; and (3) a dysploid origin from *C. mairei* (n = 27), for the two taxa are morphologically very close and often occur in sympatric populations (Map 2).

Some alternative explanations may be proposed to test the tetraploid origin of the widespread *C. pulchellum*. The only diploid population of this taxon (n = 9), encountered in Israel (Newé Ativ: Mt Hermon), is known to possess an asymmetric karyotype with one pair of large chromosomes and eight small remaining pairs (Zeltner, 1985). Mitosis analysis of the tetraploid populations (n = 18) reveal again one pair of big chromosomes. This does not support an autotetraploid origin for *C. pulchellum*, for 2 pairs would have been expected. An allopolyploid origin should be possible, with another taxon having a gamete number of n = 9 chromosomes, but no asymmetric karyotype. Such populations have not been encountered in the localities investigated (Zeltner, 1985). Lastly, one can invoke the possibility of hybridization followed by polyploidy with taxa having n = 10, with subsequent chromosome rearrangements (e. g., Robertionan fusion) giving rise to a n = 18 taxon from an ancestral n = 19 one. These hypotheses need to be fully investigated by chromosome banding techniques.

Another putative hybrid species has been described from Turkey. *Centaurium serpentinicola* (2n = 4x = 40) is characterized by a robust and erect outlook, with large flowers, and may shows affinities with *C . erythraea* and *C. tenuiflorum* (Carlström, 1986). Molecular analyses support *C. tenuiflorum* subsp. *acutiflorum* to be a parental species in both ITS and *trn*L-F trees, and therefore, the maternal origin of this taxon is confirmed. *Centaurium rumelicum*, a diploid taxon more or less frequent in the area, may be the pollen donor (Zeltner, personal communication). On the other hand, *C. serpentinicola* comes close *C. bianoris* in the ITS analyses and morphological resemblance between both the taxa is striking. This may suggest that the pink forms of *C. bianoris* and *C. serpentinicola* to be either hybrid combinations - *C. tenuiflorum* (diploid / mother parent) x *C. maritimum* (diploid / pollen parent) - or autotetraploid derivatives of *C. tenuiflorum*.

Lastly, *C. somedanum*, a rare and endangered species of the *C. littorale* group, shows close morphological affinity with *C. scilloides*. Both taxa are perennial, with decumbent stems and showy flowers. One can suggest

autopolyploidy process in *C. scilloides* (2n = 2x = 20) giving *C. somedanum* (2n = 4x = 40). On the other hand, basal position of *C. somedanum* in the ITS clade and close proximity with *C. gypsicola* in cpDNA analyses better fits an allopolyploid origin (*C. scilloides* x *C. gypsicola*) : respective repartitions of these taxa (Map 4) support this alternative scenario.

CONCLUSION

In the present chapter, the simultaneous use of independent DNA sources (ITS and the *trn*LF regions, respectively) permits us to refine some previously unresolved questions concerning the Eurasian species of *Centaurium*. The phylogenetic reconstruction of this group reveals five well-supported clades in trees rooted by the genus *Blackstonia*. The more basal group comprises the section *Spicaria*, whose relationships with the remaining Eurasian species is subject to caution in the present work. A large set of data, based either on molecular analyses or on classical characters, such as gross morphology or karyology, tend to support an independent origin for this section. The remaining clade shared a common ancestors and are generally well resolved, expect for a derived large set comprising *C. erythraea* and relatives.

Based on the phylogenetic trees topology and other evidences, the existing systematics treatments of this groups have been questioned, revealing that some sections appeared to be artificial. On the other hand, molecular analyses confirm that most of the polyploid complexes proposed so far are of natural origin, indicating the importance of this evolutionary trend in the Eurasian *Centaurium*. In this group, an important part of the taxa originated via hybridization and allopolyploidy processes, and different hypotheses have been proposed in the light of the present molecular results.

The clock-like behavior of the present data set permits us to roughly calibrate the internal nodes of the phylogenetic trees, and to infer the evolutionary history of this group. By using the same ITS mutation rate as previously proposed for the American clade of *Centaurium* (Chapter 1), we found that two geological events, namely the Messinian salinity crisis and the Pleistocene ice ages, are likely to have played a major role in the current distribution and in the diversity of the present group. Diploid species are encountered in possible glacial refugia, where they may have diverge independently for most of the Pleistocene ice ages, whereas polyploids have more peripheral distribution, suggesting post-glacial recolonization scenarios.

REFERENCES

ARNOLD, M. L. 1997. *Natural hybridization and evolution*. Oxford University Press. New York.

AVERETT, J. H. 1980. Polyploidy in plant taxa: summary. in *Polyploidy, biological relevance* : 262-273. Lewis Editor. Plenum Press. New York.

BENNETT, K. D., P. C. TZEDAKIS, and K. J. WILLIS. 1991. Quaternary refugia of North European trees. *Journal of Biogeography* 18 : 103-115.

BERGER, A. 1978. Long term variation of calorific insolation resulting from the Earth's orbital elements. *Quaternary Research* 9 : 139-167.

BOCQUET, G., B. WIDLER, and H. KIEFER. 1978. The Messinian model. A new outlook for the floristic and systematics of the Mediterranean area. *Candollea* 33 : 269-287.

BREMER, K. 1988. The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. *Evolution.* 42 : 795-803.

BRETAGNOLLE, F., F. FELBER, F. G. CALAME, and P. KÜPFER. 1998. La polyploidie chez les plantes. *Botanica Helvetica* 108 : 5-37.

BRETAGNOLLE, F. and J. D. THOMPSON. 1995. Gametes with somatic chromosome number: their mechanisms of formation and role in the evolution of autopolyploid plants. *The New Phytologist* 129 : 1-22.

CARLSTRÖM, A. 1986. New taxa and notes from the SE Aegean area and SW turkey. *Willdenowia* 16 : 73-78.

DeJONG, H. 1998. In search of historical biogeographic patterns in the Western Mediterranean terrestrial fauna. *Biological Journal of the Linnean Society* 65 : 99-164.

DOYLE, J. J. and J. L. DOYLE. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical bulletin* 19 : 11-15.

FARRIS, J. S., M. KALLERSJÖ, A. G. KLUGE, and C. BULT. 1995. Testing significance of incongruence. *Cladistics* 10 : 315-319.

FAVARGER, C. 1960. Etude cytologique du *Cicendia filiformis* et du *Microcala pusilla* (Gentianacées). *Bulletin de la Societé Botanique de France* 107 : 94-98.

FELSENSTEIN, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 17 : 223-263.

FERRIS, C., R. A. KING, and G. M. HEWITT. 1999. Isolation within species and the history of glacial refugia. in *Molecular systematics and plant evolution*. 20-34. P. M. Hollingsworth, R. M. Bateman and R. J. Gornal Editors. Taylor & Francis. London.

GOLDMAN, N. 1993. Statistical tests of models of DNA substitution. *Journal of Molecular Evolution* 36 : 182-198.

GRANT, V. 1981. *Plant speciation*. Columbia University Press. New York.

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

HAYS, J. D., J. Imbrie, and N. H. Schakleton. 1976. Variations in the Earth's orbit: pacemaker of the ice ages. *Science* 194 : 1121-1132.

HEWITT, G. 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* 58 : 247-276.

HIGGINS, D. G., J. D. THOMPSON, and T. J. GIBSON. 1996. Using CLUSTAL for multiple sequence alignments. *Methods in Enzymology* 266 : 383-402.

HEWITT, G. 2000. The genetic legacy of the Quaternary ice ages. *Nature* 405 : 907-913.

HEWITT, G. M. 1999. Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society* 68 : 87-112.

HILLIS, D. M. and J. P. HUELSENBECK. 1992. Signal, noise and reliability in molecular phylogenetic analyses. *Journal of Heredity* 83 : 189-195.

HUELSENBECK, J. P. and K. A. CRANDALL. 1997. Phylogeny estimation and hypothesis testing using Maximum likelihood. *Annual Review of Ecology and Systematics* 28 : 437-466.

HUNGERER, K. B. and J. W. KADEREIT. 1998. The phylogeny and biogeography of G*entiana* L. sect. *Ciminalis* (Adans.) Dumort: a historical interpretation of distribution ranges in the European high mountains. *Perspectives in Plant Ecology, Evolution and Systematic* 1 (1) : 121-135.

HUNTLEY, B. and H. J. B. BIRKS. 1983. *An atlas for past and present pollen maps for Europe : 0 - 13 000 years ago*. Cambridge University Press. Cambridge.

JOHNSON, L. A. and D. E. SOLTIS. 1998. Assessing incongruence: Empirical examples from molecular data. in *Molecular systematics of plants 2: DNA sequencing*. D. E. Soltis, P. S. Soltis and J. J. Doyle Editors. Kluwer. Boston.

KRIJGSMAN, W., F. J. HILGEN, I. RAFFI, F. J. SIERRO, and D. S. WILSON. 1999. Chronology, causes and progression of the Messinian salinity crisis. *Nature* 400 (6745) : 652.

LEWIS, W. H. 1980. Polyploidy in angiosperms: dicotyledons. in *Polyploidy: biological relevance* : 241-268. W. Lewis Editor. Plenum Press. New York.

MELDERIS, A. 1972. Taxonomic studies on the European species of the genus *Centaurium* Hill. *Botanical Journal of the Linnean Society.* 65 (2) : 224-250.

PETIT, C., F. BRETAGNOLLE, and F. FELBER. 1999. Evolutionary consequences of diploid-polyploid zones in wild species. *Trends in Ecology and Evolution.* 14 (8) : 306-311.

RAMSEY, J. and D. W. SCHEMSKE. 1998. Pathways, mechanisms, and rate of polyploid formation in flowering plants. *Annual Review of Ecology and Systematic.* 29 : 467-501.

RIESEBERG, L. and J. F. WENDEL. 1993. Introgression and its consequences in plants. in *Hybrid zone and the evolutionary process*. R. G. Harrisson Editors. Oxford University Press. Oxford.

RIESEBERG, L. H. and D. E. SOLTIS. 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary trends in plants.* 5 (1) : 65-83.

RYAN, W. B. F. 1973. Geodynamic implications of the Messinian crisis of salinity. in *Messinian events in the Mediterranean*. 7: 26-38. C. W. Drooger Editor. Amsterdam

SCHNABEL, A. and J. F. WENDEL. 1998. biogeography of *Gleditsia* (Leguminosae) based on *ndh*F and *rpl*16 chloroplast gene sequences. *American Journal of Botany.* 85 (2) : 1753-1765.

SOLTIS, D. E. and P. S. SOLTIS. 1993. Molecular data and the dynamic nature of polyploidy. *Critical Reviews in Plant Science.* 12 : 243-273.

SORENSON, M. D. 1999. *TreeRot version 2*. Boston University. Boston.

STEBBINS, G. 1947. Types of polyploids: their classification and significance. *Advances in Genetics.* 1 : 403-429.

STRUWE, L., J. W. KADEREIT, J. KLACKENBERG, S. NILSSON*, et al.* 200*. Systematics, character evolution, and biogeography of Gentianaceae, including a new tribal and subtribal classification. in *Gentianaceae - Systematics and natural history*. L. Struwe and V. A. Editors. Cambridge University Press. Cambridge.

SWOFFORD, D. L. 1999. *PAUP* Phylogenetic Analysis Using Parsimony (*and other methods) Version 4.* Sinauer Associates. Sunderland.

TABERLET, P., L. GIELLY, G. PAUTOU, and J. BOUVET. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology.* 17 : 1105-1109.

TACKHOLM, V. 1956. *Student's flora of Egypt*. Anglo-Egyptian Bookshop. Cairo.

TEMPLETON, A. R. 1983. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of human and apes. *Evolution.* 37 : 221-244.

THIV, M., L. STRUWE, and J. W. KADEREIT. 1999. The phylogenetic relationships and evolution on the Canarian laurel forest endemic *Ixanthus viscosus* (Aiton) Griseb. (Gentianaceae) : Evidence from *mat*K and its sequences, and floral morphology and anatomy. *Plant Systematics and Evolution.* 218 : 299-317.

THOMPSON, J. D. and R. LUMARET. 1992. The evolutionary dynamics of polyploid plants: origin, establishment and persistence. *Trends in Ecology and Evolution.* 7 (9) : 302-307.

VOGEL, J. C., F. J. RUMSEY, J. J. SCHNELLER, J. A. BARRETT, and M. GIBBY. 1999. Where are the glacial refugia in Europe? Evidence from Pteridophyte. *Biological Journal of the Linnean Society.* 66 : 23-37.

von HAGEN, K. B. and J. W. KADEREIT. 2001. The phylogeny of *Gentianella* (Gentianaceae) and it colonization of the Southern hemisphere as revealed by nuclear and chloroplast DNA sequence variation. *Organisms, Diversity and Evolution.* 1 : 61-79.

WHITE, T. J., T. BRUNS, S. LEE, and J. TAYLOR. 1990. Amplification and direct sequencing of fungal ribosomal DNA for phylogenetics. in *PCR protocols: a guide to methods and applications*. : 315-322. M. A. Innis, D. H. Gelfund, J. J. Sninsky and T. J. White Editors. Academic Press. San Diego.

YANG, Z., N. GOLDMAN, and A. E. FRIDAY. 1995. Maximum Likelihood trees from DNA sequences: a peculiar statistical estimation problem. *Systematic Biology.* 44 : 384-399.

ZELTNER, L. 1970. Recherches de biosystématique sur les genres *Blackstonia* Huds. et *Centaurium* Hill (Gentianaceae). *Bulletin de la Société Neuchâteloise des Sciences Naturelles.* 93 : 1-164.

ZELTNER, L. 1978. Recherches sur le *Centaurium bianoris* (Sennen) Sennen. *Revue de Biologie et Ecologie Méditerranéenne.* 5 (1) : 51-58.

ZELTNER, L. 1985. Etude cytotaxonomique et cytogéographique du *Centaurium pulchellum* (Swartz) Druce, sensu lato. *Botanica Helvetica.* 95 : 47-57.

ZELTNER, L. 1987. Contribution a l' étude cytotaxonomique et cytogéographique des genres *Blackstonia* Huds. et *Centaurium* hill en URSS. *Bulletin de la Société Neuchâteloise des Sciences Naturelles*.110 : 17-35.

ZELTNER, L. 1990. Sur quelques *Centaurium* Hill (Gentianaceae) d'Egypte, de la Peninsule Arabique et des Indes. *Bulletin de la Société Neuchâteloise des Sciences Naturelles.* 113 : 255-270.

Table 1. Accession data for taxa sampled for karyological study and phylogenetic analyses of the internal transcribed spacer (ITS1 and ITS2) of nrDNA and *trn*L-F region (*trn*L intron and *trn*L-F IGS spacer) of cpDNA.

				Sequence accessions				
Taxon	Voucher number	n	2n	DNA extract	ITS	*trn*L	*trn*F	Location
Blackstonia perfoliata	Fri9807	*	*	MG98712	+	+	+	Frigolet, France - 43.47 N / 04.40 W
C. barrelieri	LZ2058	10	*	MG981106	+	+	+	Spain - 41.07.100 N / 00.36.890 W
C. barrelieri	980813	*	*	MG010101	+	+	-	Spain
C. bernardii	LZ268	*	20	MG991006	-	+	+	Morocco – 34.02.910 N / 06.34.280 W
C. bianoris	B15	20*	40*	MG001012	+	+	+	Majorca, Spain
C. bianoris	20-4-2S	20*	40*	MG001013	-	+	+	Majorca, Spain
C. bianoris	20-4-2J	20*	40*	MG001014	-	+	+	Majorca, Spain
C. bianoris	B18	20*	40*	MG001015	-	+	+	Majorca, Spain
C. capense	980606-1	18	36	MG98706	+	+	+	Mexico - 29.44.89 N / 114.44.53 W
C. centaurioides	LZ1678	28	*	MG990241	+	+	+	Rewa, India - 24.17 N / 81.21 E
C. centaurioides	LZ1657	*	56	MG991003	+	-	-	Bihar, India - 24.10 N / 85.17 E
C. chloodes	LZ2592	20	*	MG981112	+	+	+	Spain - 42.49.660 N / 09.06.050
C. clementii	LZ2588	22	44	MG01201	+	+	+	Australia – 24.42.218 S / 115.18.518 E
C. erythraea	970627-1	20	40	MG97760	+	+	+	California - 40.52.90 N / 123.32.89 W
C. erythraea	Espa0798	*	*	MG98809	+	-	-	Spain - 42.53 N / 03.40 E
C. erythraea	981002-1	*	*	MG990242	+	-	-	Poulx, France - 43.55 N / 04.25 E
C. erythraea	980926-4	*	*	MG990243	+	-	-	Espiguette, France - 43.34 N / 04.10 E
C. erythraea	990619-2B	*	*	MG990703	+	-	-	Majorc, Spain - 39.30 N / 03.00 E
C. erythraea	LA0899	*	*	MG991001	-	+	+	Near Quito, Ecuador - 00.13 S / 78.30 W
C. erythraea	LZ2587	20	*	MG001211	+	-	-	Australia
C. erythraea	001120-5	20	*	MG001217	-	+	+	Australia
C. erythraea	001119-1a	20	*	MG001221	-	+	+	Australia
C. ery. subcapitatum	lZ2869	20	*	MG010201	+	+	+	Ireland – 53.14.870 N / 09.37.750 W
C. ery. subcapitatum	lZ2876	20	*	MG010202	+	+	+	Ireland
C. favargeri	LZ2045	10	*	MG98815	+	+	+	Spain - 41.21.070 N / 03.52.470 W
C. favargeri	LZ2044	10	*	MG981105	+	+	+	Spain - 41.51.210 N / 00.56.430 W
C. gypsicola	LZ2081	10	*	MG990244	+	+	+	Spain - 40.08.800 N / 03.29.640 W
C. littorale	LZ1816	20	*	MG98904	+	+	+	England - 55.41 N / 01.48 E
C. mairei	LZ1740	27	*	MG991004	+	+	+	Aïr, Niger - 16.54 N / 08.45 E
C. majus	LZ2066	10	*	MG97411	+	+	+	Spain, 38.12.460 N / 02.40.980 W
C. majus	990622-1a	*	*	MG990715	+	-	-	Majorc, Spain
C. majus	990622-1b	*	*	MG990716	+	-	-	Majorc, Spain

Table 1. continued

				Sequence accessions				
Taxon	Voucher number	n	2n	DNA extract	ITS	*trn*L	*trn*F	Location
C. majus	LZ1714	*	20	MG010211	-	+	+	Spain – 42.07.21 N / 02.58.34 W
C. majus rhodense	NS0005			MG001022	+	+	+	Sicily, Italia
C. malzacianum	LZ1586	28	56	MG98901	+	+	-	North Yemen - 15.30 N / 45.20 E
C. malzacianum	LZ1584	*	56	MG990905	+	-	-	North Yemen - 15.40 N / 43.52 E
C. malzacianum	LZ1585	28	*	MG991002	-	+	+	North Yemen - 17.12 N / 43.31 E
C. malzacianum	000404-2a			MG000402	+	-	-	Oman - 23.27.45 N / 57.04.96 E
C. malzacianum	000405-1			MG000408	+	+	+	Oman - 24.44.80 N / 56.14.25 E
C. malzacianum	000405-3			MG000410	+	+	+	Oman - 24.13.49 N / 56.20.75 E
C. malzacianum	000406-1b			MG000412	+	+	+	Oman - 23.43.52 N / 56.59.23 E
C. maritimum	Ban0698	*	*	MG98703	+	+	+	Banyuls, France - 42.28 N / 03.07 E
C. maritimum	PK2000			MG001011	+	+	+	Italy
C. muehlenbergii	960613-1		40	MG96513	+	+	+	California - 36.07.14 N / 118.49.55 W
C. muehlenbergii	970627-2	20		MG97761	+	+	+	California - 40.55.94 N / 123.37.51 W
C. muehlenbergii	970630-1	20	40	MG97773	+	-	+	California - 37.16.36 N / 122.12.86 W
C. muehlenbergii	970603-3	20	40	MG97703	+	+	+	Texas - 29.56.49 N / 95.20.80 W
C. pulchellum	gap0798	*	*	MG98505	+	+	+	Near Gap, France - 44.34 N / 06.04 E
C. pulchellum	970603-1			MG97701	+	+	+	Texas - 29.56.49 N / 95.20.80 W
C. pulchellum	950817-1	*	*	MG98810	+	+	-	Sinkiang, China - 42.22.49 N / 86.19.56 W
C. rumelicum	LZ1899	10	*	MG991007	+	+	+	Spain - 42.19.810 N / 07.13.090 W
C. rumelicum	LZ1769	*	20	MG010206	+	+	+	Spain - 42.19.810 N / 07.13.090 W
C. scilloides	LZ1853	*	20	MG98502	+	+	-	Azores - 38.29 N / 27.21 W
C. scilloides	LZ2593	10	*	MG981111	+	+	+	Spain - 43.18.120 N / 08.39.450 E
C. serpentinicola	LZ2589	20	*	MG001017	+	+	+	Turkey – 36.59.513 N / 28.39.053 E
C. somedanum	LZ1875	20	*	MG98501	+	+	+	Leon, Spain - 43.07.39 N / 06.15.23 W
C. spicatum	LZ1756	11	22	MG98508	+	+	+	Morocco - 30.41.030 N / 06.16.150 E
C. spicatum	980926-5	*	*	MG981005	+	+	+	Espiguette, France - 43.34 N / 04.10 E
C. spicatum	LZ2581	*	44	MG001016	+	+	+	Chile – 27.07.500 S / 109.22.700 W
C. spicatum	LZ2582	*	44	MG001203	+	+	+	Australia – 33.17.822 S / 115.42.378 E
C. spicatum	LZ2584	22	44	MG001206	+	+	+	Australia – 30.11.848 S / 115.01.477 E
C. spicatum	LZ2585	22	44	MG001209	+	+	+	Australia – 26.24.102 S / 114.09.920 E
C. spicatum	LZ2583	22	44	MG001210	+	-	-	Australia – 35.01.172 S / 116.56.717 E
C. suffruticosum	LZ1940	20	*	MG010208	-	+	+	Spain - 41.52.410 N / 04.47 W
C. tenuiflorum	LZ1767	20	40	MG98802	+	+	+	Morocco - 31.32.930 N / 07.35.810 E

Table 1. continued

				Sequence accessions				
Taxon	Voucher number	n	2n	DNA extract	ITS	*trn*L	*trn*F	Location
C. tenuiflorum	LZ1766	10	20	MG98803	+	+	+	Morocco - 31.32.930 N / 07.35.810 E
C. tenuiflorum	990620-3	*	*	MG990708	+	-	-	Majorc, Spain - 39.30 N / 03.00 E
*C. tenuiflorum***	LZ2586	20	40	MG001212	+	-	-	Australia – 30.03.643 S / 115.12.024 E
C. turcicum	LZ1687	*	60	MG991008	+	+	+	Georgia - 42.03 N / 43.49 E
C. turcicum	LZ2590	30	*	MG001018	+	+	+	Turkey – 37.55.136 N / 29.07.380 E
C. turcicum	LZ2591	30	*	MG001019	+	+	+	Turkey – 37.49.043 N / 31.38.991 E
C. uliginosum	LZ1699	20	*	MG991005	+	+	+	Sweden - 57.25 N / 18.54 E
C. uliginosum?	LZ2873		40	MG010205	-	+	+	North Ireland – 55.11.070 N / 06.43.170 W

Table 2. Sequence characteristics of the different nrDNA and cpDNA regions investigated in Eurasian species of *Centaurium*.

	ITS1 + ITS2	ITS1	ITS2	*trn*L+F	*trn*L intron	*trn*L-F spacer
Length range (ingroup) (bp)	453 - 464	226 - 236	227 - 230	750 - 792	382 - 417	366 - 376
Length mean (ingroup) (bp)	460	231	229	770	398	372
G+C content (mean %)	62.23	60.48	63.96	35.17	36.61	33.64
G+C content (range %)	59.6 - 63.7	56.6 - 62.6	61.8 - 65.5	34.5 - 36.4	35.5 - 38.3	33.1 - 35.6
Aligned length	474	243	231	796	420	396
Number of excluded sites	2	2	0	9	9	0
Number of coded indels	9	5	4	4	2	2
Number of sites after exclusion[a]	472	241	231	787	411	396
Length of indels (range) (bp)	1 - 5	1 - 5	1	1 - 11	11	1 - 6
Length of indels (total)	12	8	4	29	22	7
Parsimony informative	119	*	*	39	*	*
constant	289	*	*	676	*	*
uninformative	68	*	*	47	*	*
indels (number)	45	29	16	54	43	11
g1-stat	-0.59	-0.46	-0.58	-0.45	-0.36	-0.40

Table 3. Mutation rates estimated from geological or from fossil data and presently available for some genera of the Gentianaceae

Mutation rate (per site per year)	Genus	Calibration	Reference
4.48×10^{-9}	*Gentianella*	Geological evidence	von Hagen & Kadereit, 2001
5.50×10^{-9}	*Centaurium*	Geological evidence	Chapter 1
8.41×10^{-9}	*Gentianella*	Fossil evidence	von Hagen & Kadereit, 2001
1.6×10^{-8} - 2.2×10^{-8}	*Gentiana (Ciminalis)*	Geological evidence	Hungerer & Kadereit, 1999

Table 4. Likelihood ratio test (LRT) performed on several subsets of Eurasian *Centaurium*. The letters represent the clade described in the text. df is the degree of freedom (number of sequences minus 2). P values > 0.05 indicate no significant differences between the null hypothesis of a clock like behavior and the alternative hypothesis.

groups	n (taxa)	df	LRT	P
ABCDE - 2x/4x	20	18	18	> 0.1
BCDE - 2x/4x	18	16	16.82	> 0.1
BCDE - 2x	14	12	16.7	> 0.1
BCDE - 2x/4x/h	22	20	27.52	> 0.1

Table 5. Possible origins of some polyploids taxa in the Eurasian species of *Centaurium*. ([1]) parental origin: (p) = pollen ; (s) = seed ; (?) = unknown

Taxon	Possible parent 1 ([1])	Possible parent 2 ([1])	Allo polyploidy	Morpho-logical evidences	Biogeo-graphical evidences	Other scenario possible
C. bianoris	*C. maritimum (s)*	*C. tenuiflorum (p)*	yes	yes	yes	*
C. malzacianum	*C. maritimum (s)*	*C. pulchellum (p)*	yes	possible	possible	*
C. centaurioides	*C. pulchellum (s)*	*C. maritimum (?)*	possible	possible	possible	Dysploidy from *C. mairei*
C. serpentinicola	*C. tenuiflorum 2x (s)*	*C. maritimum (?)*	possible	yes	yes	Autopolyploidy *C. tenuiflorum*
C. somedanum	*C. scilloides (?)*	*C. gypsicola (?)*	possible	yes	yes	Autopolyploidy *C. scilloides*
C. mairei 6x	*C. pulchellum 2x (?)*	*C. pulchellum (?)*	possible	yes	yes	*
C. pulchellum 4x	*C. pulchellum 2x (?)*	*C. maritimum (p?)*	+ dysploidy	yes	possible	*

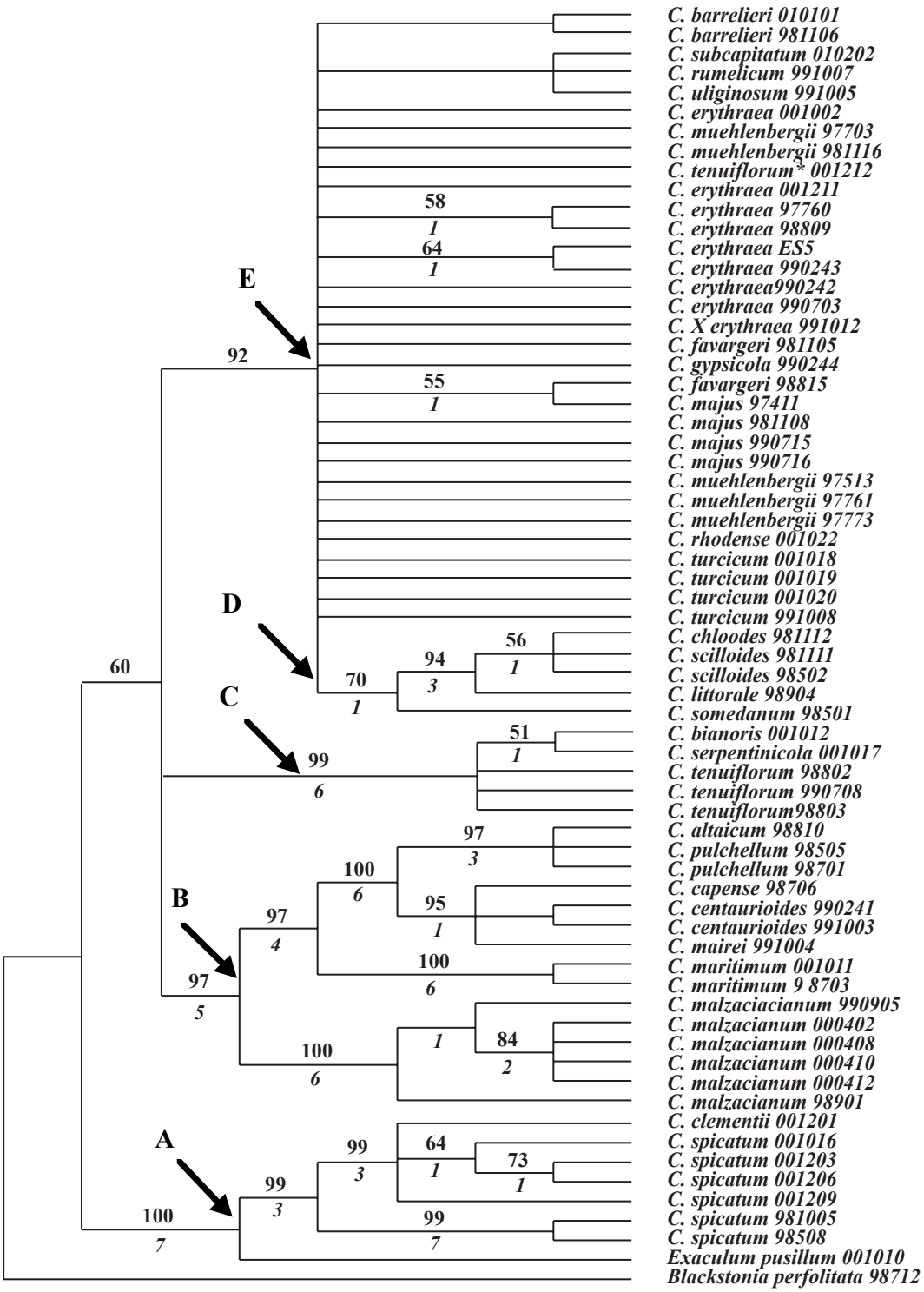


Figure 1. Strict consensus tree (L = 314) obtained from parsimony analyses of ITS1 and ITS2 sequences combined. Numbers above and below branches indicate bootstrap values of 100 replicates (in %) and decay index, respectively. Letters indicate clades mentioned in the discussion (see text).

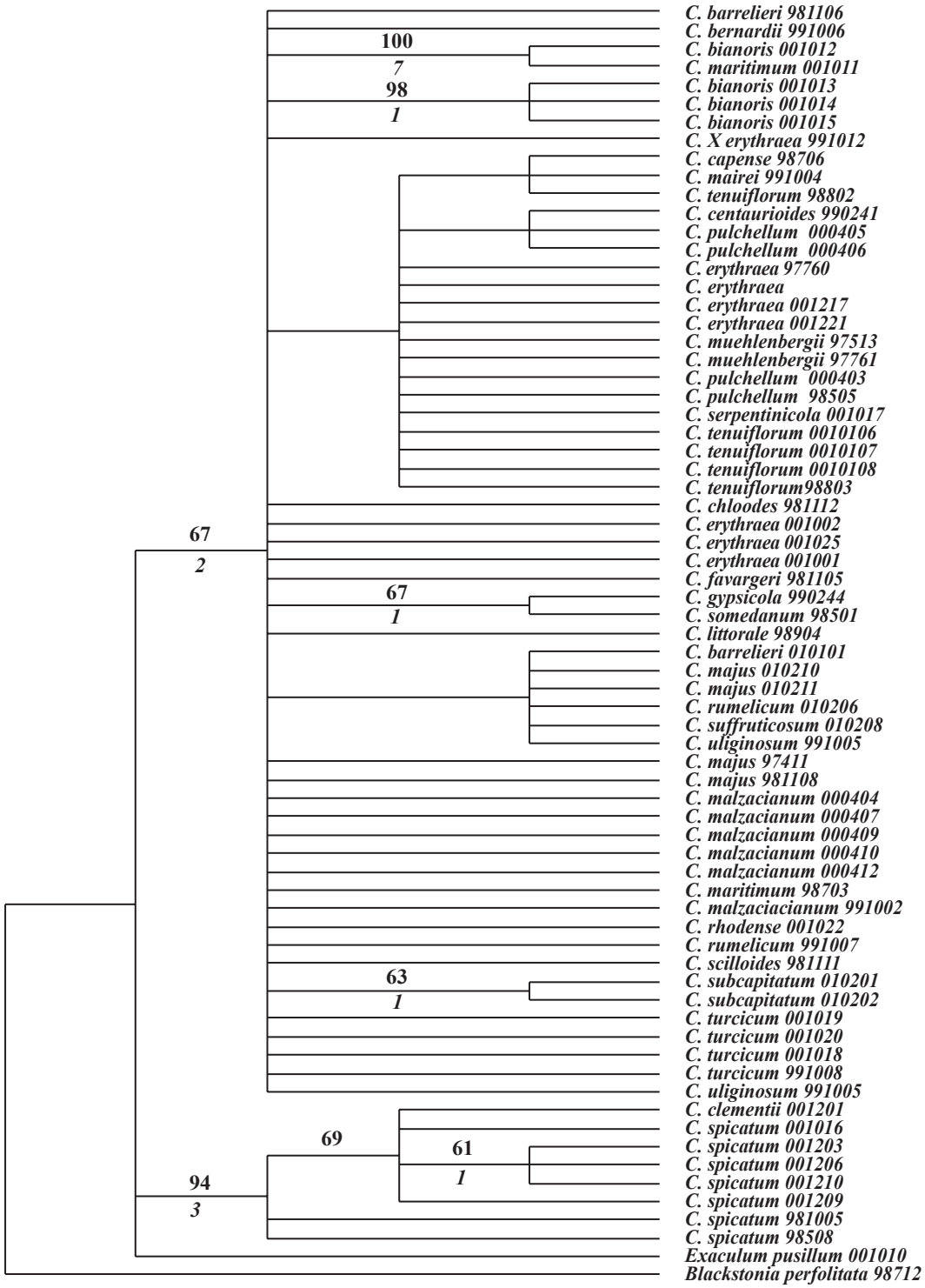


Figure 2. Strict consensus tree (129 steps) obtained from parsimony analyses of *trn*L intron and *trn*L-F spacer sequences combined. Numbers above and below branches indicate bootstrap values of 100 replicates (in %) and decay index, respectively.

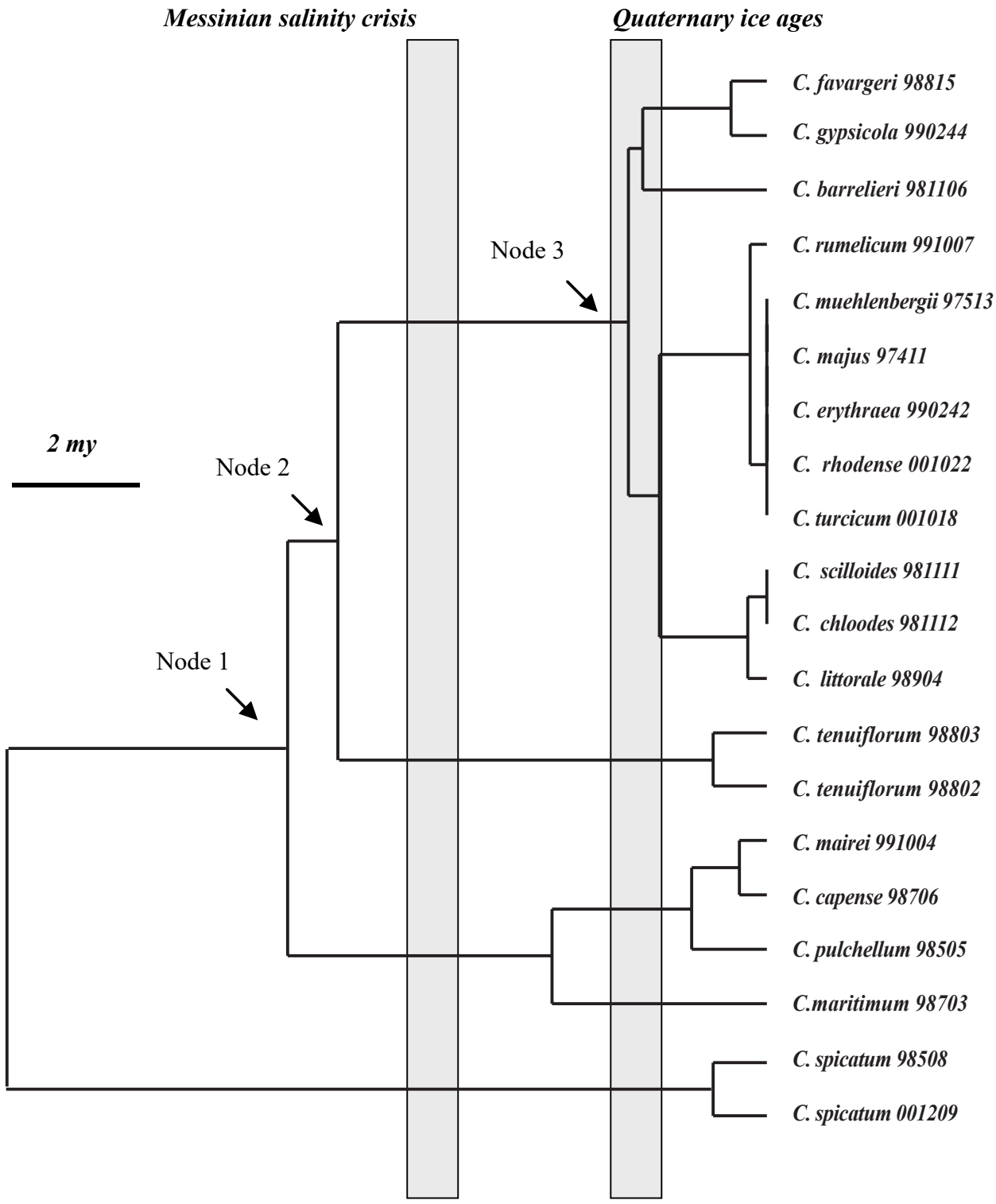
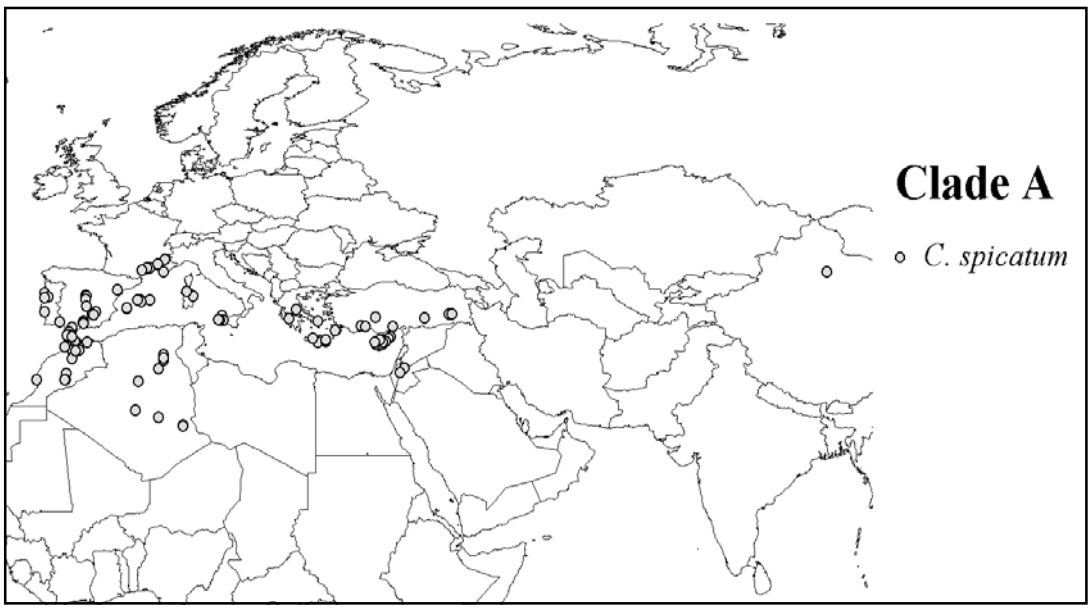
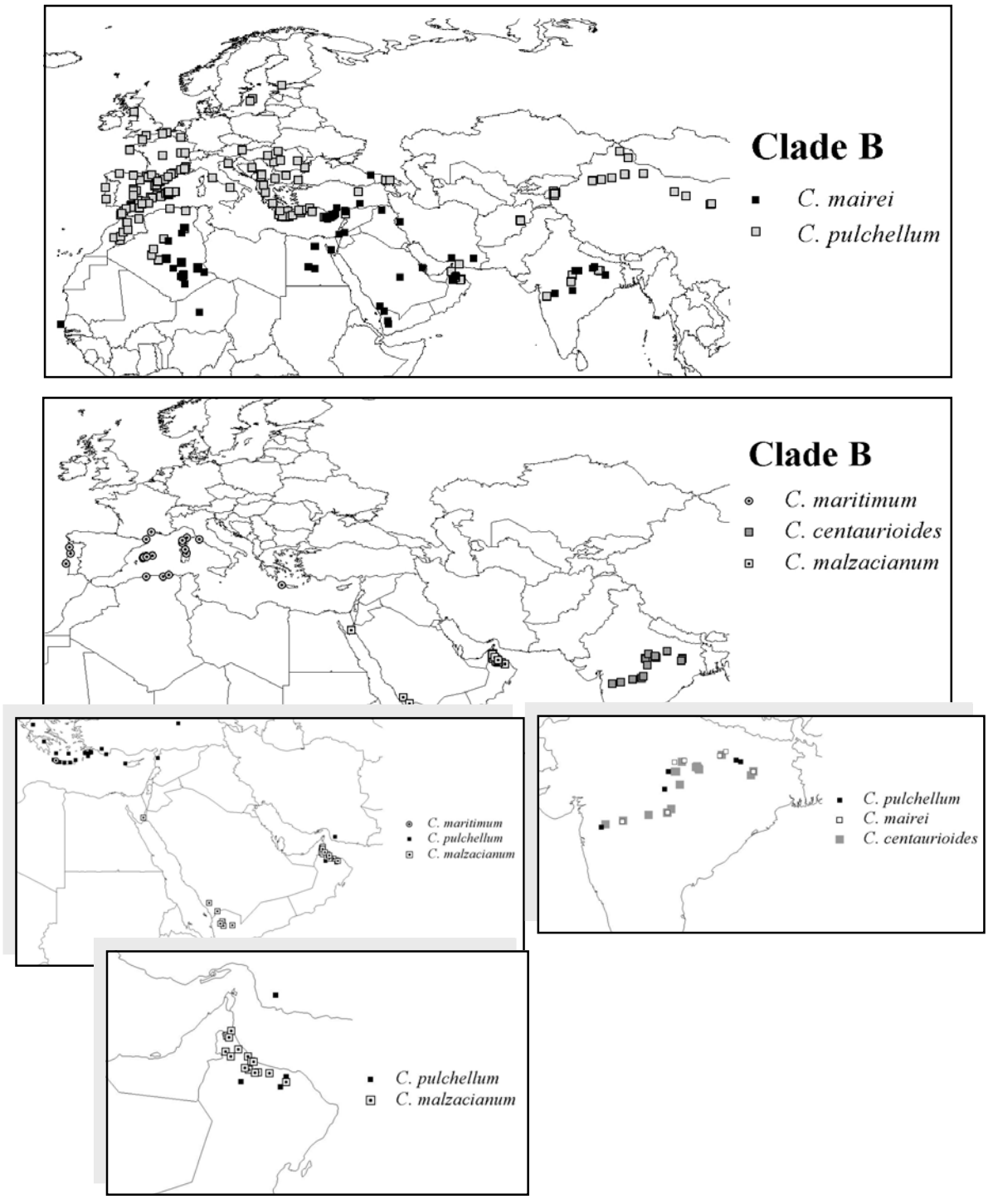


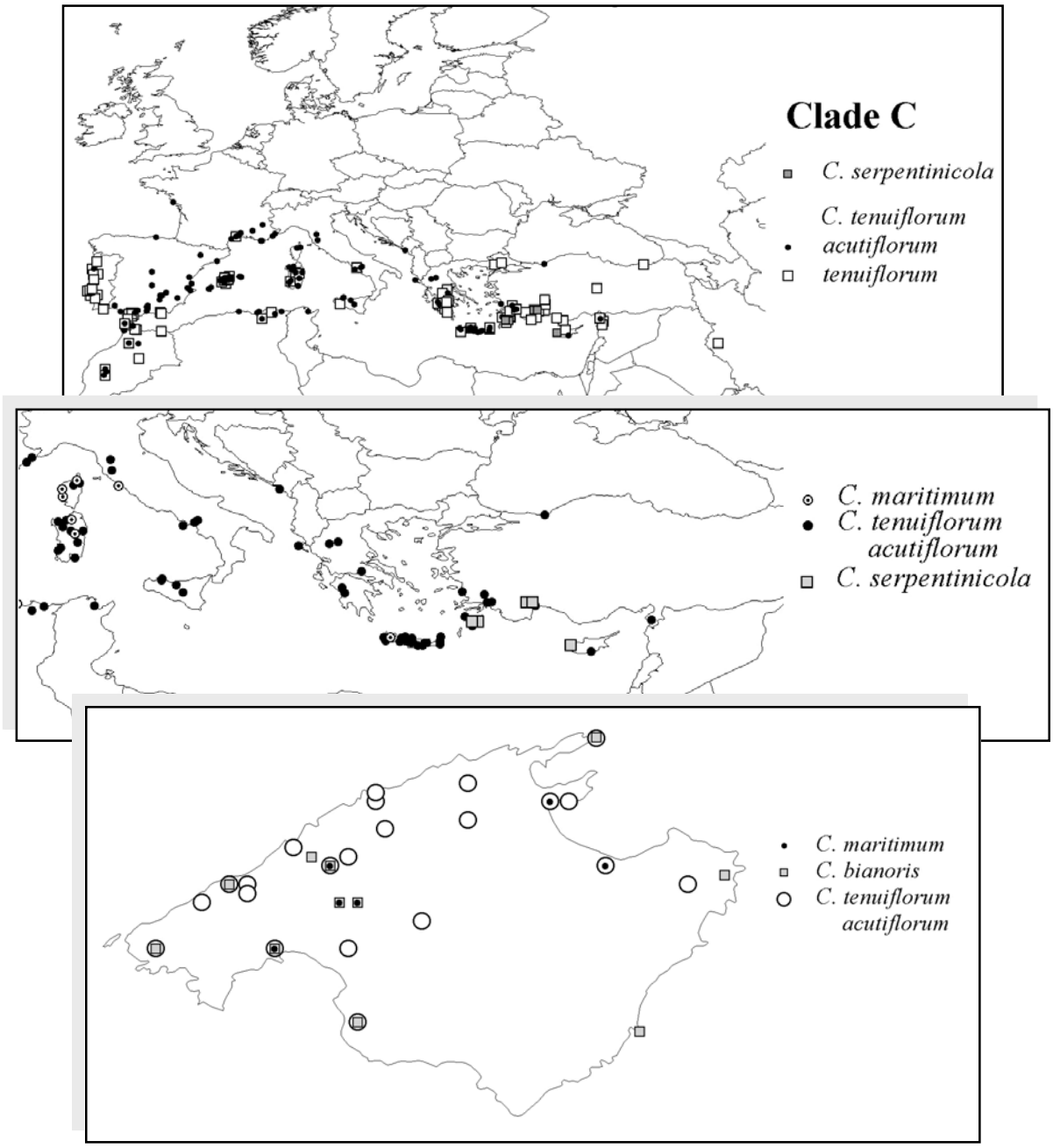
Figure 3. Phylogram of the maximum likelihood ITS tree (HKY85 model) for Eurasian species of *Centaurium*, under the assumption of a molecular clock. Node ages have been estimated with geological events (see text).



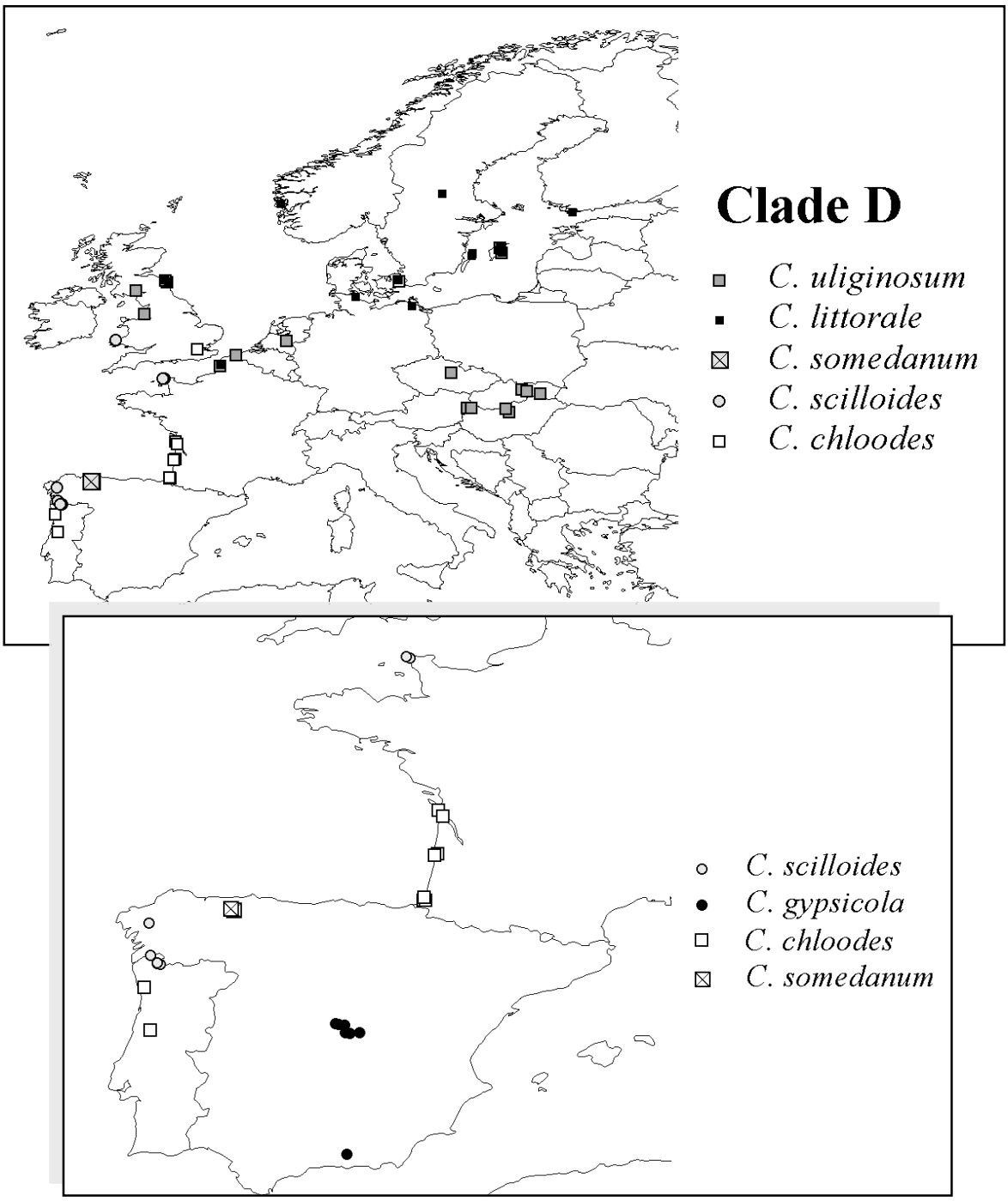
Map 1. Geographical repartition of the *Centaurium* populations sampled for karyological and molecular analyses. The composition of the clade A of the ITS cladogram is given in the text.



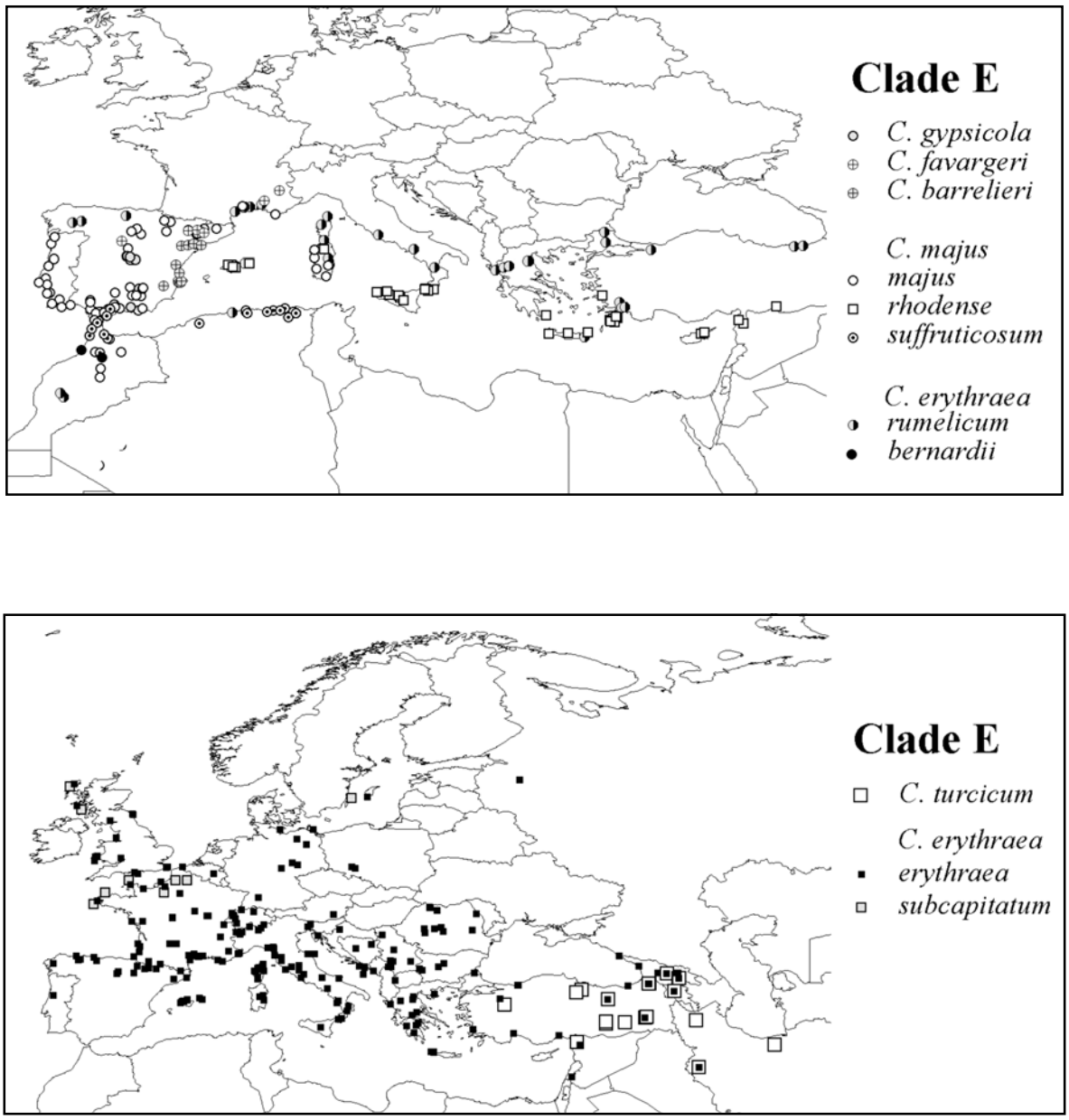
Map 2. Geographical repartition of the *Centaurium* populations sampled for karyological and molecular analyses. The composition of the clade B of the ITS cladogram is given in the text. Zooms represent present distribution of possible hybrids species (*C. malzacianum* or *C. centaurioides*) and their putative parents



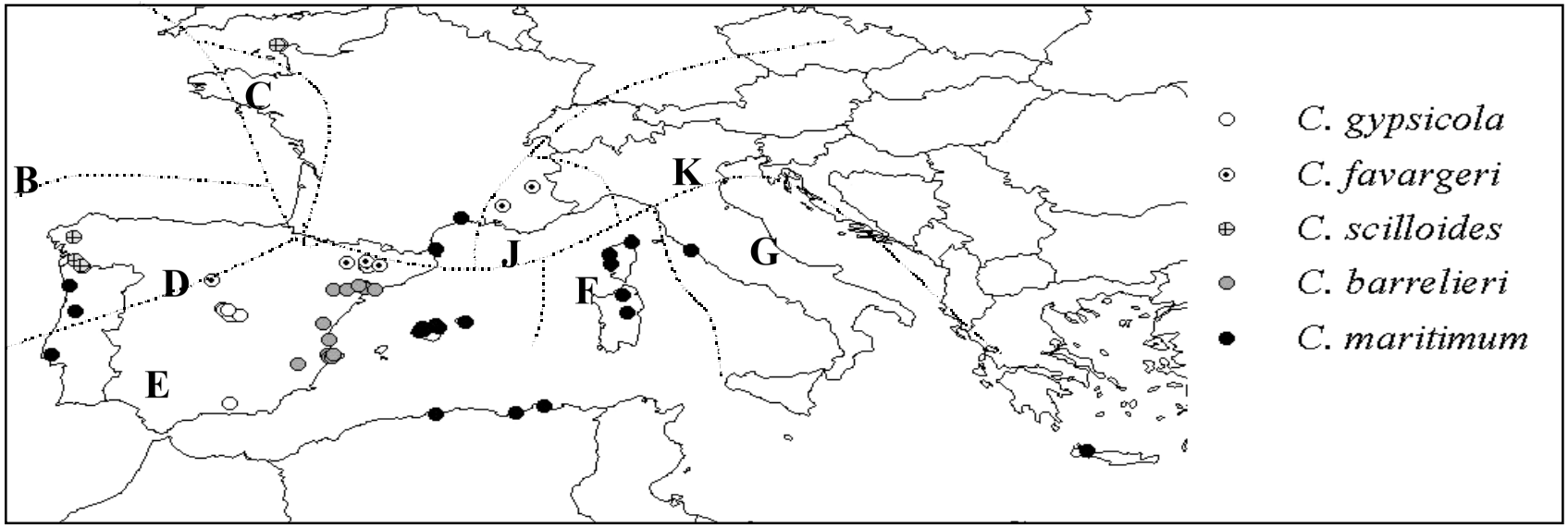
Map 3. Geographical repartition of the *Centaurium* populations sampled for karyological and molecular analyses. The composition of the clade C of the ITS cladogram is given in the text. Zooms represent the present distribution of possible hybrid species (*C. bianoris* and *C. serpentinicola*) and their putative parents.



Map 4. Geographical repartition of the *Centaurium* populations sampled for karyological and molecular analyses. The composition of the clade D of the ITS cladogram is given in the text. Zooms represent the present distribution of possible hybrid species (*C. somedanum*) and its putative parents



Map 5. Geographical repartition of the *Centaurium* populations sampled for karyological and molecular analyses. The composition of the clade E of the ITS cladogram is given in the text. The respective maps represent the diploid (above) and polyploid (below) species occurring in this group



Map 6. Repartition of some diploid species in the Eurasian group *of Centaurium.* Letters represent glacial refuge areas according to Huntley and Birks (1983)

Chapter 3

Generic delimitation and phylogenetic relationships within the subtribe Chironiinae (Gentianaceae), with special reference to *Centaurium*. Evidences from nrDNA and cpDNA sequences

Generic delimitation and phylogenetic relationships within the subtribe *Chironiinae* (*Chironieae*: Gentianaceae), with special reference to *Centaurium*. Evidences from nrDNA and cpDNA sequences

ABSTRACT

The temperate herbaceous genus *Centaurium* has been traditionally regarded as a well-delimited assemblage on the basis of the coiling of the anthers at the anthesis. To better understand the evolutionary history of the genus as a whole and the relationships with related genera of the subtribe Chironiinae, molecular analyses were performed using 146 sequences of the Internal Transcribed Spacer, and 120 accessions of the *trn*L-F chloroplast region. The results reveal two main clades of *Centaurium*, with an Eurasian group sharing affinities with the *Chironia* - *Orphium* assemblage, and an American assemblage nested near *Sabatia* and *Eustoma*. Moreover, the independent position of two other sections of *Centaurium*, sect. *Spicaria* and *Gyrandra*, respectively, add strong support for the polyphyly of *Centaurium*. This surprising result is supported by both data sets, which otherwise appear to be partially incongruent, on the base of several statistical tests.

The conflict, between molecular and morphological data not only affects *Centaurium* but other taxa as well, such as *Cicendia* and *Exaculum*, previously regarded as sister genera. Nevertheless, these accessions appear on long branches in the ITS phylogram, and this possible increase of mutation rate may have perturbed both Maximum parsimony reconstruction and molecular clock tests performed on the group. The occurrence of genetically different assemblages within *Centaurium* suggests taxonomic problems. If sections *Spicaria* and *Gyrandra* appear to be well defined and may support either a new genus or the inclusion in existing ones, the main Eurasian and American groups are particularly difficult to segregate. Few synapomorphies have been detected with classical markers (morphology, karyology or palynology), suggesting that these clades may have evolved independently during a main part of their history with parallel evolutionary constraints. The polyphyly of *Centaurium*, not only complicates the systematics within the Chironiinae but also have important implications for the biogeography of the clade. Several episodes of intercontinental dispersal, from a Mediterranean center of origin have been proposed, always involving a clade of the artificial group *Centaurium*.

INTRODUCTION

The intergeneric classification of the Gentianaceae has been early established by Endlicher (1838), Grisebach (1839), Bentham and Hooker (1876), Knoblauch (1894) and Gilg (1895). The latter taxonomic treatment, based mainly on pollen characters, was used as a reference for at least a century. Recently, with the soar of molecular tools, a new classification combining phylogenetic approaches and traditional data has been proposed (Struwe *et al.*, 200*).

In the light of these investigations, the genus *Centaurium*, previously placed in the subtribe Erythraeinae sensu Gilg, is presently included in subtribe Chironiinae of the Gentianaceae-Chironieae. This tribe is partially congruent with subtribe Chironiinae and Erythraeinae of Gilg's treatment, but nevertheless comprises genera such as *Eustoma* or *Ixanthus*, excluded by Gilg. Of the 23 genera comprising the subtribe Chironiinae, *Centaurium* appears to be the largest, even though the exact number of species in *Centaurium* is not clearly established. On the base of field experience and on different herbaria collections examined, a good estimation appears to be 55 - 60 species (Personal observations).

Presently the phylogenetic position of *Centaurium* in the Chironieae is not clear. Trees inferred from combined sequences of *mat*K region and *trn*L intron tend to support close relationships with *Sabatia*, whereas ITS results showed *Centaurium* to be sister clade to *Chironia* - *Orphium* group (Thiv *et al.*, 1999). The systematic treatment recently proposed (Struwe *et al*., 200*) follows ITS results and places *Centaurium* in a sister clade with *Chironia* and *Orphium* genera, and close to *Sabatia* and *Eustoma*.

On the other hand, only a few representative species (mainly Eurasian ones) have been investigated, and previous results clearly do not reflect the whole diversity of the subtribe. Another problem recently underlined concerns the monophyly of *Centaurium* (Chapter 1 and 2). Previous investigations focusing on particular subgroups of this genus have shown some important infrageneric separation. Taxa belonging to Eurasian and American clades, or to particular sections, seem to have independent histories.

The main goal of the present paper is to infer the position of the different subgroups of *Centaurium* in the subtribe and consequently to investigate if *Centaurium* is monophyletic. For this purpose, sequences of the Internal Transcribed Spacer of nrDNA and of the trnL-F region of cpDNA will be used. In order to check intergeneric relationships of *Centaurium* within the Chironiinae, large data sets will be analyzed with parsimony analyses.

The phylogenetic reconstruction of the Chironiinae, covering the whole of *Centaurium* diversity, will serve as a framework to infer biogeographic scenarios and character evolution in this group.

MATERIAL AND METHODS

DNA extraction, amplification and sequencing---Different accessions of *Centaurium* species and related genera (*Blackstonia*, *Chironia*, *Eustoma*, *Exaculum*, *Orphium* and *Sabatia*), were sequenced for the ITS (144 accessions) and *trn*L-F regions (120 accessions), respectively. The species have been determined in the field, on the base of Zeltner or Melderis' systematic treatment (Zeltner, 1970; Melderis, 1972). All vouchers specimen are deposited in the herbarium of the University of Neuchâtel (NEU). Some ITS sequences were retrieved from GenBank (*Cicendia filiformis*, ref. AJ011463 / AJ011473; *Ixanthus viscosus*, ref. AJ011471 / AJ011481; and *Sabatia angularis*, ref. AJ011467 / AJ011477).

DNA was extracted from fresh leaves, dried with silica gel, of from herbarium specimens, using CTAB methods or special kits (DNeasy, Qiagen). Amplification and sequencing methods are those previously used (Chapter 1 and 2).

Sequence analysis---DNA sequences were initially aligned using the program Clustal W (Higgins *et al.*, 1996). The multiple alignment obtained was then manually adjusted by sequential pairwise comparisons. The boundaries of ITS regions were identified based on a comparison with some reference sequences, e.g. *Gentiana frigida* (Yuan *et al.*, 1996) or *Centaurium tenuiflorum* clone (Mansion, unpublished data). Different parameters such as sequence length, transition / transversion ratio, GC content, number of informative characters or pairwise distance divergence values, were computed using Paup 4.0b6 version (Swofford, 1999).

Phylogenetic analysis---For outgroup choice, midpoint rooting was first performed on the respective data sets. In this preliminary analysis of ITS sequences, *Cicendia filiformis* occupies a basal position near the *Blackstonia* - *Ixanthus* group. This result is congruent with previous independent analysis of Thiv *et al.* (1999). Therefore, the genus *Cicendia* was used to root all the ITS analyses. Unfortunately, the full *trn*L-F sequence of this genus was lacking, and therefore, the cpDNA data set was rooted with *Blackstonia*.

Since the respective matrices were of large size (144 and 120 accessions, respectively), heuristic searches were performed under the Fitch Parsimony criterion. In each matrix, ambiguous characters were excluded and

informative indels were either deleted and mapped on the strict consensus tree or coded (presence / absence), appending binary character at the end of the data matrix.

Branch swapping with Tree Bisection-Reconnections (TBR), saving all minimal trees (Mulpars), with accelerated transformations of characters (ACCTRAN), and SIMPLE stepwise sequence addition were performed for each search, using PAUP 4.0b6. In order to speed up the search, the following strategy was used (Struwe, pers. comm.). First, when accessions had identical sequences, all but one was removed using the FILTER TAXA option in MacClade 3.07 (Maddison and Maddison, 1997). Under Paup*, the COLLAPSE BRANCHES option was then turned ON for branches with a minimum length of zero. This step reduced tree numbers because lot of terminal polytomies were unresolved on one most parsimonious tree (MP tree), instead of being poorly and ambiguously resolved on several MP trees (only branches with unambiguous support are of interest in maximum parsimony reconstruction). Under these conditions, 200 replicates with RANDOM ADDITION sequence, TBR branch swapping, MULPARS on, were performed saving no more than 5 MP trees per replicate (step 1). A new search was then started, using the same parameters than in the step 1, but using the trees obtained previously as Starting Trees, and saving as many MP trees as possible (step 2).

Finally, the strict consensus trees obtained from the respective steps were compared. The lack of difference means that all the topological variation was obtained in the first random search.

Branch support was performed using Jackknifing method (Farris *et al*, 1996) and decay analyses (Bremer, 1988). Bremer's indices were calculated using TreeRot V2 (Sorenson, 1999).

Characters mapping (Biogeographical area / xanthone patterns / morphological characters etc.) was performed using TRACE CHARACTER function as implemented in MacClade* (Maddison and Maddison, 1997).

The molecular clock hypothesis was tested on the ITS data set, using the likelihood ratio test, under the HKY 85 model of sequences substitution (Hasegawa *et al.*, 1985). The model choice was a compromise between the best model obtained using the likelihood ratio method and the time used for computation. The original data set was reduced in a matrix of 26 sequences, keeping only representative taxa from each well-supported clade. Several tests were performed with or without pruning taxa in different clades, in order to check possible conflicting species.

In order to check the clock-like behavior of selected taxa pairs within the group, relative rate tests were performed (Li and Tanimura, 1987; Bousquet *et al.*, 1992). In this method, the rate of nucleotide substitution between two taxa and their respective common ancestor is calculated, in reference to a third outgroup species. For this purpose, the RRT-v1.1 program was used (Robinson *et al.*, 1998).

Analysis of congruence--Several methods, all implanted in Paup*, were performed to test phylogenetic congruence between respective data sets. First, strict consensus trees were independently examined and Jackknife values compared for all the branches, allowing initial identification of potentially localized sources of conflict. The partition homogeneity test (PHT), or incongruence length difference test (ILD), was then used to assess whether or not the respective data sets were only partitions of a same large assemblage (Farris *et al.*, 1995). It has been suggested that PHT p-values greater than p = 0.01 reflect congruent data sets that, if combined will either improve or will not negatively affect phylogenetic accuracy (Cunningham *et al.*, 1998).

Finally the Wilcoxon signed-rank test (WSR) or Templeton's significantly less parsimonious test (SLP test) implanted in Paup*, was applied to estimate whether the respective data sets reject the constraint tree over the MP tree (Templeton, 1983; Johnson and Soltis, 1998). The constraints employed were (1) the rival MP tree, (2) the combined tree, and (3) several trees with forced clades, respectively.

RESULTS

DNA sequence variation--A summary of sequence characteristics (length and GC content) for the ITS and *trn*L-F regions of *Centaurium* and allied genera is presented in Table 2.

Sequence length within the ITS region ranges from 398 to 463 base pairs (bp) in *Centaurium* and from 398 to 469 bp in the outgroup genera. The mean GC content is 61.80 % in the whole data set and 61.93 % in *Centaurium* sensu lato. The alignment of 144 ITS sequences produced a matrix of 477 characters from which 248 were parsimony informative (52 %). Analyses were performed after deletion of 24 ambiguous characters (labile gaps or autapomorphies), and after removing the identical sequences in MacClade 3.07. The resulting matrix was composed of 129 taxa and 453 characters from which 230 parsimony informative (50.77 %).

In the *trn*L-F region, sequence length ranges from 757 to 790 bp, with a mean GC of 34.75 in *Centaurium*. A matrix of 92 sequences and 780 aligned characters, with 73 informative ones (9.36 %) was obtained after pruning 28 taxa with identical sequences and coding informative indels as binary characters.

Phylogenetic analyses---g1 negative values were 0.21 and 0.31 for ITS and *trn*L-F data sets, respectively. For the ITS matrix, preliminary heuristic searches with simple sequences addition reached Maxtree limit and 5000 MP trees were obtained (L = 839 / C = 0.546 / R = 0.897). Then, the two-steps strategy employed to avoid local

optima, failed to discover shorter trees. The first random addition step (200 replicates, collapse on, TBR branch swapping), keeping only 5 trees per replicates, provided 93 trees of 839 steps. These trees were then used are starting trees for a new heuristic search (same parameters as in step 1) and 4403 trees were obtained (L = 839 / C = 0.546 / R = 0.897). The strict consensus trees obtained in the respective steps were identical in topology and parsimony scores, indicating that the first random search allowed a good examination of the trees islands diversity (Maddison, 1991).

An examination of the ITS strict consensus tree reveals an important separation between the *Cicendia* - *Blackstonia* - *Ixanthus* group and the remaining genera (Fig. 1). The branch support of the ingroup clade (i.e., all the genera but *Cicendia*, *Blackstonia* and *Ixanthus*) was strong, with Jackknives values (J) of 92 and decay index (DI) of 7. The latter clade may be then subdivided in several groups. A first clade (group A, J = 93, DI = 4) comprises the genera *Chironia* and *Orphium*, whereas *Eustoma*, *Sabatia*, and section *Gyrandra* of *Centaurium* (set C2), form a weakly supported group B. All the remaining species of *Centaurium* are distributed in 3 independent lineage, namely group C1 (Eurasian species, J = 75, DI = 2), group C3 (section *Spicaria*; J = 99, DI = 6) and group C4 (American species; J = 78, DI = 3). Lastly, *Exaculum* is nested between clades C3 and C4 (Fig. 1).

Inferred trees from *trn*L-F sequences analyses are poorly resolved and generally received weak branches supports (Fig. 2). Nevertheless, the polyphyly of *Centaurium* is partially supported and some groups have good Jackknife values, e.g., C3 (J = 91), C4 (J = 91), and C2, excluding *C. tenuifolium* (J = 87).

Analysis of congruence between respective data sets---Visual inspection of the respective ITS and *trn*LF strict-consensus trees for topologies and branches supports can lead to the following observations: (1) most of the "non-*Centaurium*" genera (e. g. *Chironia*, *Orphium*, or *Sabatia*), well-defined on the ITS cladogram, are not discriminated and received weak or no support for branches in the cpDNA tree and (2), within *Centaurium* s.l., only section *Spicaria* (C3) and the North American clade (C4) receive strong branch support in both the nrDNA and cpDNA trees. Describe conflict in the well-supported clades.

The partition homogeneity test performed on the combined data set, with 1000 replicates, a Maxtree limit of 100, and using the NNI branch swapping, indicated that the nuclear and chloroplast DNA matrix were not congruent (p = 0.001). It has been suggested that PHT p-values greater than p = 0.01 reflect congruent data sets that, if combined will either improve or will not negatively affect phylogenetic accuracy (Cunningham *et al.*, 1998).

The Templeton's significantly less parsimonious test (SLP test) first performed on the most parsimonious trees of each respective data sets and on their respective strict consensus trees (Table 1), reveals incongruence between all the combinations, with very significant values between the cpDNA data sets an the combined one ($P < 0.0001$). Nevertheless, when topological constraints were applied to the analyses, most of the resulting trees do not significantly differ from the most parsimonious ones (Table 1). Only the monophyly of *Centaurium (*constraint C1 + C2 + C3 + C4) was not supported, and generated trees significantly less parsimonious cladograms.

Sequence divergence - Molecular clock---Mean DNA sequence divergences within the main clades obtained with the ITS data set, were calculated by using the HKY85 model (Table 2). Divergence values range from zero, between several species of the section *Centaurium* (Chapter 2) to a maximum of 29.79 % between *Blackstonia perfoliata* and *Exaculum pusillum* (Table 3). Sequence divergences computed for several combinations between well-supported clades are shown in Table 3.

In order to test the assumption of a molecular clock in the ITS mutation rate, within the Chironiinae, a reduced matrix comprising 26 representative taxa was analyzed.

Under the HKY85 model of sequence substitutions, likelihood score were calculated for respective data sets obtained after pruning one or several taxa, with or without enforcing a molecular clock (Table 4). The different ratio tests performed reveal that a the clock assumption must be rejected in some cases, i.e. when taxa such as *Cicendia*, *Blackstonia* or *Exaculum*, were present in the analyses (Table 4). Nevertheless, in the majority of the analyses, a clock-like behavior of the data set is supported.

Moreover, the relative rate tests performed show that most of the combinations involving *Cicendia*, *Blackstonia* or *Exaculum*, have significant values (Table 5), and thus confirm the LRT results.

Therefore we can assume that the whole ingroup clade (i.e., excluding the basal *Cicendia* - *Blackstonia* - *Ixanthus* set), minus *Exaculum*, follows a clock-like evolution.

Biogeographic reconstruction---Despite many criticisms of the dispersalist approach (see Morrone and Crisci, 1990, for review), some authors considered that understanding ancestral areas for an individual group is a valid part of the study of the natural history of this group (Bremer, 1992). Therefore, we have tested here the Bremer's procedure to identify the ancestral area of the Chironiinae, based on the information given by the reduced taxon area of this group (Fig. 3). Considering each area as a binary character, and using the Camin-Sokal parsimony

criterion, we can compare the number of gain and losses for each character and evaluate the most likely ancestral area. The results clearly demonstrate a Mediterranean origin for the Chroniinae with a subsequent radiation to the Americas (Table 6).

The character "distribution" optimized on the ITS strict consensus tree supports a Mediterranean origin of the Chroniinae group (Fig. 3).

DISCUSSION

Utility of the respective data sets in the Chroniinae---The utility of the ITS sequences to resolve relationships within the Chroniinae is clearly demonstrated in the present study. The strict consensus trees generally show well-resolved groups, and branch support is relatively high for most of the clades. Moreover, this phylogenetic reconstruction is helpful to check the different systematics treatments proposed for the Chroniinae, and will allow us to further test some evolutionary hypotheses. Given the poor resolution of the *trn*L-F data set (9.36 % of informative sites), little can be inferred about taxa relationships within the Chroniinae. There is a good or moderate support for some groups and the polyphyletism of *Centaurium* is strongly supported, based on the chloroplast trees. Moreover, the simultaneous utilization of both nrDNA and cpDNA in molecular analyses has become a necessity for some biological processes such as introgression, gene duplication or lineage sorting, are known to disturb phylogenetic reconstruction (Doyle, 1992; Wendel and Doyle, 1998). In *Centaurium*, topological incongruence between phylogenetic cladograms inferred from ITS and *trn*L-F data sets, have proven good tools to estimate reticulate evolution and detection of allopolyploidy (Chapter 1 and 2).

Finally, regarding at the phylogenetic relationships between the groups inferred from the nrDNA and cpDNA data sets, one can draw the following conclusions. First, the different genera investigated in the present study, form a clade congruent with the Chroniinae subtribe defined by Struwe *et al.*, 200*. Second, the genus *Centaurium* is clearly paraphyletic and four well-separated clades are present in the ITS strict consensus tree. These groups are supported by the cpDNA tree as well (Fig. 2). Third, the Old World species of *Centaurium* s.l do not form a clade for, quite all the Eurasian species are found in clade C1, whereas members of the section *Spicaria* belong to group C3. Lastly, American centauries may be considered as a good natural group (J = 78 % / DI = 3) if we exclude the Mexican section *Gyrandra*, represented here by *C. tenuifolium* and *C. brachycalyx*, and nested in the *Sabatia* - *Eustoma* clade. The latter received poor Jackknife support (< 50) whereas the *Sabatia* clade alone shows high Jackknife and decay values (J = 99 / DI = 8). Finally, the neighbor relationships between *Cicendia* and *Exaculum* hypothesized earlier (Struwe *et al.*, 200*), is not supported by ITS and *trn*LF analyses.

These respective genera are separated by many internal nodes on the respective cladograms, and appear on long branches in the Neighbor Joining tree (not shown).

Congruence between data sets---With the development of molecular systematics, numerous DNA regions of nuclear and organelle genomes are presently available for phylogenetic reconstruction. Moreover, others characters, obtained by classical methods (e. g. morphological or chemical data), can be added to the pool of DNA sequences available, and thus multiple data sets may be incorporated into phylogenetic studies. Three alternative strategies have been proposed for handling multiple data sets, and gave rise to considerable debate, regarding the advantages or limitations of the combined, the conditional or consensus approaches, respectively (de Queiroz *et al.*, 1995; Huelsenbeck *et al.*, 1996). Whatever the good strategy, the real issue hinges on determining the source of conflict between data sets, and more important, the evolutionary explanation of the existing, if any, differences. Several statistical approaches have been recently proposed to assess whether data were a mere partition of a global data set, thus sharing the same natural history and evolutionary processes, or not. (Johnson and Soltis, 1998).

Presently, the partition homogeneity and SLP tests performed over the respective MP trees (ITS, *trn*LF and combination of both), show significant values, suggesting that the cpDNA and nrDNA data sets should not be combined and thus appear to be incongruent. On the other hand, topological constraints performed on the ITS tree, reject the monophyly of *Centaurium*, as suggested by the *trn*LF and ITS data sets, indicating some congruence between the two approaches.

Phylogenetic relationships within the Chironiinae---The present study supports the basal position of the *Blackstonia* / *Ixanthus* clade, close to the outgroup *Cicendia*. A previous analysis of this group, using ITS sequences and rooted with *Symphyllophyton* supports this topology (Thiv *et al.*, 1999). Some morphological features such as the presence of perfoliate bracts, calycine colleters, or a yellow corolla, are shared by *Blackstonia* and *Ixanthus*. The pale yellow to cream color of the flower is found in *Cicendia* as well. Nevertheless, these genera of Mediterranean origin possess very different chromosome numbers, n = 10 or 20 for *Blackstonia* (Zeltner, 1970), n = 13 for *Cicendia* (Favarger, 1960), and n = 31 for *Ixanthus* (Thiv *et al.*, 1999). Moreover, the close relationships previously hypothesized between *Cicendia* and *Exaculum* is not supported by either the ITS or the *trn*LF analyses. Despite a similar vegetative outlook (slender herbs, more or less ramose, with minute flowers) and a particular affinity for the same ecological places (Favarger, 1960), these unrelated

genera are separated by a deep phylogenetic drift on the ITS strict consensus tree. Some floral characters clearly differentiate the taxa. The flowers of *Exaculum* are pink and possess a very short calyx tube with the teeth exceeding the corolla tube (as in *Centaurium spicatum*), the corolla tube being longer than the corolla lobes (equal length in *Cicendia*). The stigmata and capsule shapes greatly differ between the two taxa (*Cicendia*: subcapitate stigmata and oblong capsule; *Exaculum*: bilobate stigmata and elliptic capsule), along with the chromosome valence (n = 10 for *Exaculum pusillum*; Favarger, 1960). Anyway, one must pay attention to the fact that both *Cicendia* and *Exaculum* appear on long branches in the ITS phylogram (not shown). This may indicate that these short-lived annual species have higher mutation rates that the other members of the Chironiinae subtribe (or maybe shorter generation time), and thus perturb the clock like behavior of the whole group. Moreover, the long-branch attraction a phenomenon is known to perturb phylogenetic reconstruction based on Maximum parsimony criterion (Felsenstein, 1978), and may explain the important divergence between these morphologically and ecologically close taxa.

Most of the species belonging to the respective *Chironia / Orphium* and Eurasian *Centaurium* groups are found in Mediterranean climates. One can encounter *Chironia* in grassy places of southern Africa, and the South African endemic genus *Orphium* in Fynbos vegetation. The latter is presently thought to be monospecific, but our ITS results reveal close relationships with *Chironia baccifera*, suggesting either a broader circumscription of *Orphium* or a paraphyletic *Chironia*. Presently *Chironia* (ca. 30 species) has not been investigated with molecular tools and its monophyly is not clearly established. These African taxa along with clade C1 of *Centaurium* appears to be sister clade, sharing pink to purple (rarely white or yellow) flowers or coiling anthers after dehiscence (Marais and Verdon, 1963).

Another clade comprising the American genus *Sabatia* and *Centaurium* section *Gyrandra* (J = 69) comes close to *Eustoma*, but the resulting cluster received no or weak branch support (J < 50). *Sabatia* is presently a numerically important genus within the Chironiinae (ca 20 taxa), comprising very distinctive taxa classified in various sections on the base of morphological, karyological or artificial crosses grounds (Wilbur, 1955; Perry, 1971). The section *Gyrandra* of *Centaurium* has been first described as a distinct genus (Grisebach, 1839), and presently contains five species (Broome, 1973). *Eustoma* is a small genus, with 3 species occurring in the Southern United States, Mexico and Greater Antilles (Shinners, 1957). All these taxa share similar chromosome number (n = 36 is found in the 3 genera), a generally large corolla (except *Sabatia arenicola* and *C. brachycalyx*), and their respective geographic repartition (Mexico and Eastern North America) may support a common ancestral origin.

Generic delimitation of *Centaurium* and systematics implications---In the earliest classification of the Gentianaceae, *Centaurium* was placed in the subtribe Erythraeinae, based on pollen characters such as smooth or finely granular, 3-colpate pollen (Gilg, 1895). In this treatment, four sections were depicted and two of them (*Euerythraea* and *Trichostylus*) appear to be very artificial. Presently, the paraphyletism encountered in *Centaurium*, not only perturbs the infrageneric classification, but the whole intergeneric relationships, as well. Alternative possibilities may be proposed to adapt the *Centaurium* variability in a systematic classification.

One way is to consider a broad genus *Centaurium*, in order to keep the monophyly of the group. In this case, some other members of the subtribe (*Sabatia*, *Eustoma* and *Exaculum*) must be included in it, and probably *Chironia* and *Orphium*, as well. Such recognition of expanded *Centaurium* poorly reflects the subtribe diversity and is not very practical in a systematic point of view. On the other hand, a large genus such as *Gentiana* comprises many sections, and quite as many clades, all of them very distinctive in morphological, karyological, and biogeographical or ecological characters. The gentian group presently contains more than 360 species, much more than a large genus *Centaurium* (ca 130 taxa). The generic limits largely rest on subjective interpretation. If we assume that a goal of the classification is to recognize monophyletic genera, one must define the most reduced clade corresponding to such grades. In this perspective, the fusion of very different clade in one expanded *Centaurium* is not satisfactory.

Alternatively, a reductionist approach would be to attribute a generic status to each well-marked *Centaurium* clade. In this case, the clade C1 only may be considered to represent *Centaurium* s. s., for the genus was typified by *C. littorale*, a north European species (Gillett, 1963). In this reductionist point of view, one must find morphological features (or other) to support the genetic diversity. The clade C2 corresponding to the section *Spicaria* in previous treatments (Grisebach, 1839; Gilg, 1895), is particularly well defined by its gross morphology, e. g. spiciform inflorescence, large and elliptic leaves, crowded at the base, chromosome number (n=11) or global repartition (see Chapter 2). This is also the case for the section *Gyrandra*, comprising *C. tenuifolium* and *C. brachycalyx*, and possibly four other taxa not present in the present molecular analysis (Broome, 1973). Members of this group seems to possess a high chromosome number (n = 36), particular morphological features (see Chapter 1), and are centered in the central part of Mexico, with a southern radiation for *C. brachycalyx*.

On the other hand, it seems more problematic to find synapomorphic characters in the American clade s. s. (clade C4) and to propose good morphological features to distinguish it from the Eurasian species. Until now, no

global systematic treatment has been proposed for *Centaurium* s. l. and therefore no keys based on morphological characters are available to discriminate between the different groups.

Finally, a more parsimonious solution would be to include some of the separate clade in their respective sister groups, which may imply some major generic rearrangements within the Chironiinae.

For example, species of the section *Gyrandra* (or Grisebach's genus *Gyrandra*) could fit within a monophyletic genus *Sabatia*, an assumption that needs to be confirmed. Therefore, a phylogenetic study of *Sabatia* is presently greatly needed before further inclusion of *Gyrandra* within it.

Characters evolution in the Chironiinae and implications in the systematics of *Centaurium*---One of the more striking result in this study is the apparently incongruence between morphological and molecular characters. Whether some phylogenetic relationships have been previously suggested between, e.g. *Centaurium chironioides* and *Chironia linoides* (Broome, 1973) or *Centaurium* s.l and *Sabatia* (Weaver and Rüdenberg, 1975; Wilbur, 1955), the paraphyly of *Centaurium* has never been proposed, as far we know. On the contrary, *Centaurium* appeared to be one of the best defined genus, based on the particular coiling of the anther after pollen release, character which is also present but at a lesser degree, in some *Chironia, Orphium* and *Sabatia* species. As mentioned above, only few systematics treatments of the whole genus have been proposed, and no previous segregation between American and Eurasian species derived form these works.

In one early study (Gilg, 1895), four sections have been proposed, from which one (*Spicaria*) appear to be a well-marked clade. Gilg's section *Trichostylus* was based on the capitate form of the stigmata, but currently appears to be artificial, because of the grouping of Californian (*C. trichanthum*), Indian (*C. centaurioides*) or Mexican species (*C. nudicaule* / *C. tenuifolium*), belonging to different clades (Chapters 1 and 2). Lastly, section *Euerythraea* comprised all the remaining taxa with no further subdivision, indicating the difficulty of finding evident morphological or anatomical discriminating character.

The genus *Gyrandra*, previously proposed by Grisebach (1839), and lowered to sectional rank by Gray (1878), does not appear in Gilg's classification. This taxon, basically grouping *C. chironioides* and *C. tenuifolium*, is characterized by large, nearly rotate corollas and oval-elliptic capsules (vs. cylindrical and oblong in *Centaurium*). Based on these morphological grounds and other evidences (karyology, DNA), some other species must be included in it, such as *C. pauciflorum* or *C. pterocaule* not included in the present molecular analysis. Nevertheless, *C. brachycalyx* is relatively distinguishable in this group because of the smallness of its flowers and a general outlook reminiscent of *Zygostigma australe*, a rare Gentianaceae included in the Chironiinae.

As mentioned above, the generic status of section *Gyrandra* needs to be confirmed. Moreover, the phylogenetic position of *Gyrandra* is also supported by the unique stamen insertion, occurring between the sinus of the rotate corolla and not in the tube (personal observations), character also present in *Sabatia* (Struwe *et al.*, 200*).

Presently, the particular phylogenetic position of *Gyrandra* and *Spicaria* taxa (see Chapter 1 and 2), confirmed by some additional characters, supports their exclusion from *Centaurium*. On the other hand, the American and Eurasian clades (C1 and C4) are very difficult to separate morphologically, and their systematic segregation appears to be problematic.

Character mapping of some morphological features on the ITS strict consensus tree, reveals that the strict anther coiling is reversal in few taxa. On the other hand, different degrees of anther coiling may be encountered in most of the genera of the Chironiinae except *Eustoma* and *Exaculum*.

It has been proposed that the different stigma shapes encountered in *Centaurium* taxa could be a potentially good taxonomic character (Dunn, 1967, Broome, 1973). Stigmata with fan forms generally appear in the American *Centaurium*, whereas Eurasian species mainly have broad, elliptic or reniform ones. Nevertheless, this character does not appear to be synapomorphic for these groups (personal observations). In return, the degree division in the upper part of the style clearly separates the New World and Old World groups. Moreover, the ratio style / ovary is generally lower than one in the Eurasian *Centaurium.*

Recent and ongoing palynological study performed on all the taxa representative of the Chironinae subtribe (Nilsson and Mansion, unpublished), reveals that pollen ultrastructure does not greatly differ between the separate clades of *Centaurium* and *Sabatia* as well.

The pollen grains have previously been reported as 3-colporate, and finely reticulate in 14 species of *Centaurium* (Köhler, 1905). A *C. pulchellum* pollen type (3-colporate, striate to reticulo-striate, with muri abruptly changing direction, and striae perforate) has been described for some taxa, including *C. erythraea*, *C. littorale*, *C. maritimum*, *C. pulchellum*, *C. scilloides*, *C. spicatum* and *C. tenuiflorum* (Punt and Nienhuis, 1976). This does not favor a separation between clade C1 and C2 of the Eurasian *Centaurium*. On the other hand, American species have been poorly studied, and the description of *C. quitense* and *C. tenuiflorum* pollen grains (Elias and Robyns, 1975) do not really differ with the *C. pulchellum* group. Significant variation has been observed regarding the aperture and exine features of six species (Rao and Chinnappa, 1983), and such kind of study is presently enlarged to all the representative taxa of the Chironiinae (Nilsson and Mansion, unpublished).

Other characters such as chromosome number are known to possess a high level of homoplasy and, if they are helpful to segregate between morphologically close species, they do not appear to be informative for generic or

sectional delimitation in *Centaurium* (Fig. 3). For example, the number 2n = 40 appears frequently in the Eurasian section *Centaurium*, in the *C. tenuiflorum* group (Chapter 2) and in the Texan species related to *C. calycosum* as well (Chapter 1).

Phytochemical characters have been extensively used for taxonomic purposes in the Gentianaceae, and were sometimes helpful to discriminate between genera or sections (Hostettmann-Kaldas *et al.*, 1981; Gottlieb, 1982). Particular molecules are known to occur in aerial parts and roots of many gentians, namely xanthones (producing yellow dye) and secoiridoids (bitter compounds). The xanthones possess diverse oxidation patterns and are not distributed equally between the different taxa. A comprehension of the polarity of this character (i. e. are poly-oxygenated xanthones primitive or derived compounds?) would be helpful for taxonomic purposes. Presently, xanthones have been reported in some Eurasian species of *Centaurium* forming the clades C1 (e. g. *C. pulchellum* or *C. tenuiflorum*) and C3 (*C. spicata*). Furthermore, these compounds occur in the Texan species of the clade C4, but have not been identified yet (Mansion and Rodriguez, unpublished data). The presence / absence of this character have been mapped on the ITS tree, but these data are fragmental, most of them provide from the literature and not from the population investigated, and the polarity of the character (evolution of the oxidation pattern) seems to be more important than the simple occurrence. To conclude, these data should be helpful to support different *Centaurium* groups and relative ones, but are not currently available.

The ITS tree mapping of selected taxonomic characters such as, floral morphology, karyology, pollen features or chemical content, poorly help to discriminate between the Eurasian and the American assemblages of *Centaurium*, and thus indicates that the synapomorphies are rare if not absent in the group.

It is noteworthy that the synapomorphies are rare in the Chironiinae and absent, or hitherto undetected, in the Chironieae. A cladistic analysis of the group based on morphological characters would be helpful.

However, most of the clades depicted here are strongly supported by different molecular data sets (Struwe *et al.*, 200*).

Similar cases, where molecular analyses do not support an apparent morphological homogeneity, have been reported in the Polemoniaceae within the genus *Linanthus* (Bell and Patterson, 2000). In this group, which has mainly diverge in response to the spreading drought that occurred in California during the Pliocene, similarly to the North American *Centaurium* clade (Chapter 1), no clear morphological characters was available to segregate between the different molecular clade. Several other genera may be cited, and particularly *Lotus* (Fabaceae), a group showing a striking similar geographical distribution with *Centaurium*, and where two very distinct North American and Eurasian clades have been depicted by ITS analyses (Allan and Porter, 2000).

Biogeography and evolutionary history of the Chironiinae---The current distribution of the genera investigated in this study shows a predominantly northern-temperate distribution for the Chironiinae, and only a few such as *Orphium*, *Chironia* or *Zygostigma* inhabit the Southern Hemisphere. As shown with the ancestral area evaluation and the character "distribution" mapping on the ITS strict consensus tree (Fig. 3)**,** the Mediterranean region appears to be the center of origin and diversification of the Chironiinae. Then, many independent radiations, with possible back migrations, may be hypothesized, (1) to North and Central America, via the Bering land bridge or the boreotropics - Greenland bridge; (2) throughout Africa to the Cape region of South Africa; (3) to Australia and Hawaiian Islands.

(1) Most of the genera encountered in the Mediterranean / North American scenario are restricted either to the Mediterranean region and surrounded areas (*Blackstonia* / *Ixanthus* / *Exaculum* / *Centaurium* clade C1) or to the North America (*Centaurium* clade C4 / *Sabatia* / *Gyrandra* / *Eustoma*). Only *Cicendia* is present in both the areas, along with some naturalized species of *Centaurium* (*C. erythraea* / *C. pulchellum* and *C. capense* / *C. spicatum*). This fact may indicate that a first separation between Old World and New World common ancestors may have occurred early in the subtribe evolution, depending on the divergence between the two species of basal genus *Cicendia*. Alternatively, this split may be dated later if we suppose an introduction by human of *C. quadrangularis*, in the New World.

In all cases, the ITS strict consensus tree indicates that alternation of radiation to the Americas and back migrations were frequent in the Chironiinae history. Two major routes are known to have provided connections between the Old and New worlds namely, the Bering land bridge and the North Atlantic land bridge. Whereas the former was available throughout the Tertiary, the latter involved at least four geographical links (two between North America and Greenland, one between Fennoscandinavia and Greenland, and one between Greenland and Europe) and thus was more or less suitable for respective exchanges (McKenna, 1983). Nevertheless, the Greenland bridge appeared to be available in the early Eocene (ca 50 mya), which seems too early for the present group. The divergence between two basal genera such as *Ixanthus* and *Orphium* has been estimated to 30 myrs on the base of a *mat*K - *trn*L data set (Thiv, unpublished data).

On the other hand, during the Miocene, the Bering land bridge appears to be the main connection for most of the exchanges between temperate deciduous plants. The temperatures at these latitudes were to cool for many of the herbaceous groups (Tiffney, 1985). Nevertheless, movements via the Bering land bridge remained possible by colonization of disturbed sites (or alternatively by a southern Aleutian bridge), or spread by long-distance

dispersal (Tiffney, 1985). Considering that many of the Chironiinae taxa presently often occur in disturbed sites, where competition is weak, one can propose the first solution to be more likely. Moreover, the prevailing wind direction in the Bering region, and the seed type encountered in the Chironiinae, do not favor a long-distance dispersion.

The ancestral area cladogram may indicate that a Mediterranean ancestor first colonized Eastern America (*Sabatia* / *Eustoma*) and central Mexico (*Gyrandra*). Then, a more complicated scenario (which may imply extinct taxa) evokes both a back-radiation to the Mediterranean area (*Exaculum* / *Spicaria*) and perhaps Australia (*Spicaria*), along with an important diversification in the Western North America (C*entaurium* clade C4). In this North American assemblage, three geographic groups have been isolated by the early Pliocene and are presently separated by insuperable barriers such as deserts or mountains (Chapter 1).

(2) The relationships of South-African and Eurasian floras have been documented, and ca. 4 % of the genera encountered in South Africa are thought to have affinities with Eurasian taxa (Goldblatt, 1978). One of the significant features of the Tertiary history of Africa is the union of the African-Arabian block with Eurasia, closing the Tethys sea, by the mid-Miocene (Axelrod and Raven, 1978). Direct migration of the Eurasian flora into South Africa was then possible. Moreover, a new uplift in East Africa, by the Miocene, implied the formation of cooler and drier belts, and thus significant routes for migration and exchanges between Mediterranean flora and South African one. Then, by the Pliocene and Pleistocene, exchanges with South Africa seem to be impossible due to a possible tropical - equatorial barrier (Quezel, 1978). Therefore, one can hypothesize a probable Miocene origin (25 - 5 mya) for a *Chironia* - *Orphium* divergence from a Mediterranean ancestor. This range is congruent with an estimation time of ca. 10 myrs for the phylogenetic node separating (*Chironia – Orphium*) clade and the Eurasian *Centaurium*, using the mutation rate previously obtained in Chapter 1 (results non shown).

(3) Most of the Chironiinae presently encountered in Australia (*Blackstonia*, *Centaurium* and *Cicendia*) are of European origin, and have been more or less naturalized in the western part of the Island. Only *Centaurium spicatum* is widespread in damp and sandy places and occurs in all the Australian states (Adams, 1996). The present distribution the *Spicaria* section shows diploid taxa widespread around the Mediterranean basin and radiating throughout Asia and Japan, and tetraploid ones in western Australia, Eastern Islands and Hawaii (Carr, 1978). If the presence of European *Centaurium* in Australia can be interpreted as anthropozoogenic introduction, an alternative scenario may be proposed for *C. spicatum*. A long-distance dispersal of ancestral diploid to the Pacific islands remains possible. Moreover, tetraploid taxa occurring in

Australia are genetically different from the European diploid ones, but we are not able to differentiate between the respective phenotypes, mainly because of the current reduced sampling. A global study covering the whole morphological and molecular diversity of this group is greatly needed. The presence of diploid taxa in Australia needs to be checked to corroborate biogeographical hypotheses. *Centaurium spicatum* is supposed to be an ancestral taxon that originated in the middle Miocene (Chapter 1 and 2). The migration of diploid populations, accompanying the invasion of Asian plants in northeastern Australia, could have been possible during this period by the New Guinea bridge (Hill, 1994). Presently the occurrence of *C. spicatum* var. *japonicum*, in Taiwan and Japan support this view.

CONCLUSION

The current phylogenetic reconstruction, based on large sets of independent molecular data, indicate that within the Chironiinae, the core genus *Centaurium* is clearly paraphyletic. The ITS strict consensus tree reveals that Eurasian and American species of *Centaurium* form two distinct clades, whereas two sections, namely *Gyrandra* and *Spicaria*, are independent lineage. The cladograms inferred from chloroplast DNA sequences confirm the polyphyly of *Centaurium*, but otherwise appear to be incongruent with the nuclear DNA results. The weak phylogenetic signal encountered in the *trn*LF region may explain topological differences, particularly concerning the non-resolved position of genera such as *Chironia* or *Sabatia*. On the other hand, the phylogenetic position of the genus *Exaculum*, nested with section *Spicaria*, in both the approaches, rises up some problems. First, in the ITS cladogram, this genus appear to be very distant with *Cicendia*, despite a gross morphological similarity. Particular floral characters may help to discriminate both the genera, but do not explain the deep phylogenetic drift between them. Moreover, most of the incongruence detected within the ITS and *trn*LF data sets, often involves *Exaculum* or *Cicendia*. The long branch encountered for these taxa in the ITS phylogram may have perturbed parsimonious analyses or the molecular clock behavior of the group.

From a systematic point of view, the polyphyly of *Centaurium* failed to be supported by strong and evident morphological characters. If the segregation from *Centaurium* of the sections *Spicaria* and *Gyrandra* may be supported by several features, attempts to find synapomorphies in both Eurasian and American clades often fails.

The polyphyly of *Centaurium* also has important biogeographical consequences. Several patterns of dispersal, with possible back migrations, may be hypothesized from a probable Mediterranean center of origin to the Americas, and alternatively to South Africa and Australia.

REFERENCES

ADAMS, L. G. 1996. *Gentianaceae*. in *Flora of Australia*, vol. 28: 73-103. A. Wilson, editor. CSIRO. Melbourne.

ALLAN, G. J. and J. M. PORTER. 2000. Tribal delimitation and phylogenetic relationships of Loteae and Coronilleae (Faboideae: Fabaceae) with special reference to *Lotus*: evidence from nuclear ribosomal ITS sequences. *American Journal of Botany* 87 (12) : 1871-1881.

AXELROD, D. I. and P. H. RAVEN. 1978. Late Cretaceous and Tertiary vegetation history of Africa. in *Biogeography and ecology of Southern Africa*. M. J. A. Werger Editors. W. Junk. The Hague.

BELL, C. D. and R. W. PATTERSON. 2000. Molecular phylogeny and biogeography of *Linanthus* (Polemoniaceae). *American Journal of Botany* 87 (12) : 1857-1870.

BENTHAM, G. and J. HOOKER. 1876. Gentianeae. in *Genera Plantarum*. 2**:** 799-820. Williams and Norgate Editors. Lovell Reeve. London.

BOUSQUET, J., S. H. STRAUSS, and L. PENG. 1992. Complete congruence between morphological and *rbc*L-based molecular phylogenies in birches and related species (Betulaceae). *Molecular Biology and Evolution* 9 (6) : 1076-1088.

BREMER, K. 1988. The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. *Evolution* 42 : 795-803.

BREMER, K. (1992). Ancestral areas: a cladistic reinterpretation of the center of origin concept. *Systematic Biology* 41 (4) : 436-445.

BROOME, C. R. (1973). Systematics of *Centaurium* (Gentianaceae) of Mexico and Central America. Thesis. Duke University. Duke.

CARR, G. D. 1978. Chromosome number of Hawaiian flowering plants and the significance of cytology in selective taxa. *American Journal of Botany* 65 : 236-242.

CUNNINGHAM, C. W., K. E. OMLAND, and T. H. OAKLEY. 1998. Reconstructing ancestral character states: a critical reappraisal. *Trends in Ecology and Evolution* 13 (9) : 361-366.

de QUEIROZ, A., M. J. DONOGHUE, and J. KIM. 1995. Separate versus combined analysis of phylogenetic evidence. *Annual Review of Ecology and Systematic* 26 : 657-681.

DOYLE, J. J. 1992. Gene trees and species trees: molecular systematics as one character taxonomy. *Systematic Botany* 17 : 144-163.

DUNN, R. A. (1967). A revision of the genus *Centaurium* of continental United States. Thesis. Catholic University of America. Washington D. C.

ELIAS, T. S. and A. P.-I. ROBYNS. 1975. Family 160. Gentianaceae. in *Flora of Panama*. 8 : 61-101. R. E. Woodson and R. W. Schery Editors. Annals of the Missouri Botanical Garden. St Louis

ENDLICHER, S. L. 1838. *Genera plantarum secundum ordines naturales disposita*. Vienna.

FARRIS, J. S., M. KALLERSJÖ, A. G. KLUGE, and C. BULT. 1995. Testing significance of incongruence. *Cladistics* 10 : 315-319.

FARRIS, J. S., V. A. ALBERT, M. KÄLLERSJÖ, D. LIPSCOMB, A. G. KLUGE. 1996 Parsimony jackknifing outperforms neighbor-joining *Cladistics* 12 (2) : 99-124.

FAVARGER, C. 1960. Etude cytologique du *Cicendia filiformis* et du *Microcala pusilla* (Gentianacées). *Bulletin de la Societé Botanique de France* 107 : 94-98.

FELSENSTEIN, J. 1978. Cases in which parsimony and other compatibility methods will be positively misleading. *Systematic Zoology* 27 : 401-410.

GILG, E. 1895. Gentianaceae. in *Die Naturlichen Pflanzenfamilien*. 4**:** 50-108. A. Engler and K. Prantl Editors

GILLETT, J. M. 1963. *The gentians of Canada and Greenland*. Research Branch, Canada Department of Agriculture. Ottawa.

GOLDBLATT, P. 1978. An analysis of the Flora of Southern Africa: its characteristics, relationships and origins. *Annals of the Missouri Botanical Garden* 65 : 369-386.

GOTTLIEB, O. R. 1982. Evolution of xanthones in Gentianaceae and Guttiferae. in *Micromolecular evolution, Systematics and ecology* : 89-95. Springer Verlag. Berlin-Heidelberg-New York

GRAY, A. 1878. Gamopetalae after Compositae. in *Synoptical flora of North America*. 2. Ivison, Blakeman and Taylor. New York.

GRISEBACH, A. H. R. 1839. *Genera et Species Gentianearum adjectis observationibus quibusdam phytogeographycis* : 116-125. J. G. Cotta. Stuttgart et Tübingen.

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

HIGGINS, D. G., J. D. THOMPSON, and T. J. GIBSON. 1996. Using CLUSTAL for multiple sequence alignments. *Methods in Enzymology* 266 : 383-402.

HILL. R. S. 1994. *History of the Australian vegetation: Cretaceous to recent*. R. S. Hill Editor. Cambridge University Press. Cambridge.

HOSTETTMANN-KALDAS, M., K. HOSTETTMANN, and O. STICHER. 1981. Xanthones, flavones and secoiridoids of American *Gentiana* species. *Phytochemistry* 20 : 443-446.

HUELSENBECK, J. P., J. J. BULL, and C. CUNNINGHAM. 1996. Combining data in phylogenetic analysis. *Trends in Ecology and Evolution* 11 (4) : 152-157.

JOHNSON, L. A. and D. E. SOLTIS. 1998. Assessing incongruence: Empirical examples from molecular data. in *Molecular systematics of plants 2: DNA sequencing*. D. E. Soltis, P. S. Soltis and J. J. Doyle Editors. Kluwer. Boston.

KNOBLAUCH, E. 1894. Beiträge zur Kenntniss der Gentianaceae. *Bot Centralbl* 60 : 321-401.

KÖHLER, A. 1905. Der Systematische Wert der Pollenbeschaffenkeit bei den Gentianaceen. *Mitteilungen aus der Botanischen Museum der Universität Zürich*: 25. (Dissertation)

LANYON, S. 1985. Detecting internal inconsistence in distance data. *Systematic Zoology* 34 : 397-403.

LI, W.-H. and M. TANIMURA. 1987. The molecular clock runs more slowly in man than in apes and monkeys. *Nature* 326 (5) : 93-96.

MADDISON, D. R. 1991. The discovery and importance of multiple islands of most parsimonious trees. *Systematic Zoology* 40 (3) : 315-328.

MADDISON, W. P. R. and D. R. MADDISON. 1997. *MacClade 3.07*. Sinauer Associates. Sunderland.

MARAIS, W. and C. VERDON. 1963. Gentianaceae. in *Flora of southern Africa* 26 : 171-243. R. A. Dyer, L. E. Codd, and H. B. Rycroft Editors. Department of Agricultural Technical Services, Republic of South Africa, Pretoria

McKENNA, M. C. 1983. Cenozoic paleogeography of North Atlantic land bridges. in *Structure and development of the Greenland - Scotland ridge*. M. H. P. Bott, S. Saxov, M. Talwani, and J. Thiede Editors. Plenum Press. New York.

MELDERIS, A. 1972. Taxonomic studies on the European species of the genus *Centaurium* Hill. *Botanical Journal of the Linnean Society* 65(2) : 224-250.

MORRONE, J. J. and J. V. CRISCI. 1990. Historical biogeography: introduction to methods. *Annual Review of Ecology and Systematic* 26 : 373-401.

PERRY, J. D. 1971. Biosystematic studies in the North American genus *Sabatia* (Gentianaceae). *Rhodora* 73 (795) : 309-369.

PUNT, W. and W. NIENHUIS. 1976. The northwest European pollen flora. 1 (6) : Gentianaceae. in *Review of Palaeobotany and Palynology.* 21 : 89-123. Punt. W. T. Editor. Elsevier. Amsterdam.

QUEZEL, P. 1978. Analysis of the flora of Mediterranean and Saharan Africa. *Annals of the Missouri Botanical Garden* 65 : 479-534.

RAO, S. K. and C. C. CHINNAPPA. 1983. Pericolporate pollen in Gentianaceae. *Canadian Journal of Botany* 61 : 174-178.

ROBINSON, M., M. GOUY, C. GAUTIER, and D. MOUCHIROUD. 1998. Sensitivity of the relative-rate test to taxonomic sampling. *Molecular Biology an Evolution* 15 (9) : 1091-1098.

SHINNERS, L. H. 1957. Synopsis of the genus *Eustoma* (Gentianaceae). *The Southwestern Naturalist* 2 : 38-43.

SORENSON, M. D. 1999. *TreeRot version 2*. Boston University. Boston.

STRUWE, L., J. W. KADEREIT, J. KLACKENBERG, S. NILSSON*, et al.* 200*. Systematics, character evolution, and biogeography of Gentianaceae, including a new tribal and subtribal classification. in *Gentianaceae - Systematics and natural history*. L. Struwe and V. Albert Editors. Cambridge University Press. Cambridge.

SWOFFORD, D. L. 1999. *PAUP* Phylogenetic Analysis Using Parsimony (*and other methods) Version 4.* Sinauer Associates. Sunderland.

TEMPLETON, A. R. 1983. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of human and apes. *Evolution* 37 : 221-244.

THIV, M., L. STRUWE, and J. W. KADEREIT. 1999. The phylogenetic relationships and evolution on the Canarian laurel forest endemic *Ixanthus viscosus* (Aiton) Griseb. (Gentianaceae) : evidence from *mat*K and ITS sequences, and floral morphology and anatomy. *Plant Systematics and Evolution* 218 : 299-317.

TIFFNEY, B. H. 1985. The Eocene North Atlantic land bridge: its importance in Tertiary and modern phytogeography of the Northern Hemisphere. *Journal of the Arnold Arboretum* 66 : 243-273.

TIFFNEY, B. H. 1985. Perspectives on the origin of the floristic similarity between eastern Asia and eastern North America. *Journal of the Arnold Arboretum* 66 : 73-94.

WEAVER, R. E. and L. RÜDENBERG. 1975. Cytotaxonomic notes on some Gentianaceae. *Journal of the Arnold Arboretum* 56 : 211-222.

WENDEL, J. F. and J. J. DOYLE. 1998. Phylogenetic incongruence: window into genome history and molecular evolution. in *Molecular systematics of plants 2. DNA sequencing.* D. E. Soltis, P. M. Soltis, and J. J. Doyle Editors. Kluwer Academic. Boston.

WILBUR, R. L. 1955. A revision of the North American genus *Sabatia* (Gentianaceae). *Rhodora* 57 (673 / 674) : 1-33 and 43-104.

YUAN, Y.-M., P. KÜPFER, and J. J. DOYLE. 1996. Infrageneric phylogeny of the genus *Gentiana* (Gentianaceae) inferred from nucleotide sequences of the internal transcribed spacer (ITS) of nuclear ribosomal DNA. *American Journal of Botany.* 83 (5) : 641-652.

ZELTNER, L. 1970. Recherches de biosystématique sur les genres *Blackstonia* Huds. et *Centaurium* Hill (Gentianaceae). *Bulletin de la Société Neuchâteloise des Sciences Naturelles* 93: 1-164.

Table 1. Results of the Templeton's significantly less parsimonious test (SLPt) with different constraints Probability values less than 0.05 indicate that the alternative topology is significantly less parsimonious than the shorter tree of the data set being tested. L_0 is the length of the MP trees in the original data set. L_1 is the length of the MP trees under the specific constraint. Letters correspond to clades described in the text.

Data set (L_0)	Constraint tree (L_0)	Constraint clade	L_1	Probability (2-tailed)
ITSred (949)	MP *trn*LFred trees (1223)	*	*	P < 0.0001
ITSred (949)	MP combined trees (913)	*	*	P < 0.05
ITS (839)	*	(*Exaculum* / *Cicendia*)	850	NS
ITS (839)	*	(C1 / C2 / C3 / C4)	859	P < 0.001
		(C3 / C4 / *Exaculum*)	841	NS
ITS (839)	*	(C3 / C4)	839	NS
ITS (839)	*	(B / C3 / C4 / *Exaculum*)	839	NS

Table 2. Sequence characteristics of the different nrDNA (ITS) and cpDNA (*trn*LF) regions investigated in species of the Chironiinae subtribe.

	ITS 1 and ITS 2 region combined			***trn*L intron – *trn*L -F spacer**		
	L (range)	GC (range)	GC (mean%)	L (range)	GC (range)	GC (mean%)
Total (144/120)	398-463	59.4-63.8	61.81	757-790	33.9-35.8	34.77
Centaurium (130/111)	452-463	59.8-63.8	61.8	757-790	33.9-35.8	34.75
Clade C1 (58/57)	453-461	60.2-63.8	62.55	757-790	34.3-35.8	34.70
Clade C2 (3/3)	455-459	63.2-63.6	63.44	775-780	33.9-34.5	34.31
Clade C3 (7/8)	457-461	59.8-61.52	60.35	772-775	34.7-35.1	34.80
Clade C4 (62/43)	452-463	59.8-62.8	61.36	771-790	34.3-35.3	34.85
Blackstonia (3/1)	459-461	59.4-59.7	59.59	763	-	35.12
Ixanthus (1/0)	461	60.52	-	-	-	-
Cicendia (1/0)	454	63.44	-	-	-	-
Chironia (2/2)	456-469	60.5-61.5	61.03	773-774	34.9-35.2	35.04
Orphium (1/0)	462	61.09	-	-	-	-
Sabatia (4/4)	398-460	61.3-62.4	61.79	777-787	34.7-34.8	34.72
Eustoma (1/1)	462	62.12	-	764	-	35.21
Exaculum (1/1)	460	-		763	-	35.12

Table 3. Sequence divergence for ITS sequences (HKY 85 model) between the respective clade of the Chironiinae subtribe

	Cicendia	*Blackstonia*	*Ixanthus*	*Chironia*	*Orphium*	Clade C1	Clade C2
Cicendia	*						
Blackstonia	27.09 (0.001)	*					
Ixanthus	23.71 (0.001)	11.84 (0.001)	*				
Chironia	20.27 (0.02)	13.33 (0.001)	19.52 (0.05)	*			
Orphium	20.84 (-)	18.00 (0.002)	19.99 (-)	5.01 (0.01)	*		
Clade C1	20.41 (0.01)	20.23 (0.01)	18.99 (0.01)	11.05 (0.01)	9.49 (0.01)	*	
Clade C2	18.95 (0.005)	22.30 (0.004)	20.84 (0.004)	11.78 (-)	11.06 (-)	12.76 (0.01)	*
Sabatia	25.50 (0.004)	25.68 (0.02)	25.87 (0.024)	16.64 (0.022)	16.48 (0.012)	16.42 (0.02)	13.39 (0.02)
Eustoma	21.98 (-)	21.72 (0.001)	20.64 (-)	13.08 (0.001)	12.71 (-)	13.14 (0.005)	12.09 (0.01)
Clade C3	20.53 (0.001)	21.97 (0.005)	21.45 (0.007)	12.05 (0.001)	12.56 (0.002)	13.57 (0.01)	12.19(0.001)
Exaculum	24.20 (-)	29.79 (0.003)	28.77 (-)	17.86 (0.044)	19.09 (-)	19.09 (0.01)	17.96(-)
Clade C4	21.91 (0.02)	23.23 (0.01)	22.77 (0.01)	15.61 (0.02)	14.38 (0.01)	13.66 (0.01)	13.25 (0.01)

Table 3. **(continued)**

	Sabatia	*Eustoma*	Clade C3	*Exaculum*	Clade C4
Sabatia	*				
Eustoma	15.32 (0.04)	*			
Clade C3	17.21(0.001)	14.21(0.006)	*		
Exaculum	21.11(0.012)	18.54(-)	12.22 (0.003)	*	
Clade C4	17.12 (0.01)	14.79 (0.02)	9.50 (0.02)	12.95 (0.01)	*

Table 4. Likelihood ratio test (LRT) performed under the HKY 85 model of sequence substitution for different combinations of a reduced data set of 26 representative taxa. (df) degree of freedom

Groups	N taxa	LRT value (df)	Statistical difference
Whole data set - (*Cicendia)*	25	66 (23)	P < 0.01
Ingroup	22	40 (20)	P <0.01
Ingroup - (*Exaculum*)	21	30 (19)	NS
Ingroup - (*Exaculum* / C3)	19	26 (17)	NS
C1 + A	10	12 (8)	NS
C1	8	8 (6)	NS
B + C3 + C4 + *Exaculum*	14	22 (12)	P < 0.05
C4 + C3 + *Exaculum*	7	3	NS
C4	5*	1*	NS
B + C3 + *Exaculum*	9	20 (7)	P < 0.01

Table 5. Results of relative rate tests performed between different pairs of taxa with selected outgroups (*Cicendia*, *Sabatia* or *Eustoma*). K1 (K2) represents the fraction of nucleotide substituted among taxa 1 (2) and the reference taxon. dk is the difference between K1 and K2 and sd_dk, the standard deviation of this difference. Shaded areas represent significant differences between the respective lineages.

lineage1	lineage2	reference	K1	K2	dK	sd_dK	ratio	P
Blackstonia	*Chironia*	*Cicendia*	24.56%	16.54%	8.02%	2.03%	395.31%	0.01%
Blackstonia	*Sabatia*	*Cicendia*	24.56%	17.92%	6.63%	2.14%	310.48%	0.19%
Blackstonia	clade C4	*Cicendia*	24.56%	15.98%	8.57%	2.18%	393.54%	0.01%
Blackstonia	clade C1	*Cicendia*	24.56%	18.44%	6.12%	2.01%	304.33%	0.23%
Blackstonia	clade C3	*Cicendia*	24.56%	15.01%	9.54%	2.23%	428.37%	0.00%
Blackstonia	*Exaculum*	*Cicendia*	24.56%	10.96%	13.60%	2.69%	506.14%	0.00%
Chironia	*Sabatia*	*Cicendia*	16.54%	17.92%	-1.38%	1.35%	-102.60%	30.49%
Chironia	clade C4	*Cicendia*	16.54%	15.98%	0.56%	1.56%	35.86%	71.99%
Chironia	clade C1	*Cicendia*	16.54%	18.44%	-1.90%	1.09%	-173.95%	8.20%
Chironia	clade C3	*Cicendia*	16.54%	15.01%	1.53%	1.55%	98.45%	32.49%
Chironia	*Exaculum*	*Cicendia*	16.54%	10.96%	5.58%	2.06%	270.54%	0.68%
Sabatia	clade C4	*Cicendia*	17.92%	15.98%	1.94%	1.46%	132.91%	18.38%
Sabatia	clade C1	*Cicendia*	17.92%	18.44%	-0.51%	1.41%	-36.42%	71.57%
Sabatia	clade C3	*Cicendia*	17.92%	15.01%	2.91%	1.55%	188.23%	5.98%
Sabatia	*Exaculum*	*Cicendia*	17.92%	10.96%	6.97%	1.98%	352.43%	0.04%
clade C4	clade C1	*Cicendia*	15.98%	18.44%	-2.46%	1.62%	-151.37%	13.01%
clade C4	clade C3	*Cicendia*	15.98%	15.01%	0.97%	1.23%	78.81%	43.07%
clade C4	*Exaculum*	*Cicendia*	15.98%	10.96%	5.02%	1.71%	293.08%	0.34%

Table 5. (continued)

lineage1	lineage2	reference	K1	K2	dK	sd_dK	ratio	P
clade C1	*clade C3*	*Cicendia*	*18.44%*	*15.01%*	*3.43%*	*1.68%*	*203.80%*	*4.16%*
clade C1	*Exaculum*	*Cicendia*	*18.44%*	*10.96%*	*7.48%*	*2.19%*	*341.84%*	*0.06%*
clade C3	*Exaculum*	*Cicendia*	*15.01%*	*10.96%*	*4.05%*	*1.82%*	*222.45%*	*2.61%*
Sabatia	*clade C4*	*Chironia*	*12.30%*	*13.13%*	*-0.83%*	*1.45%*	*-57.50%*	*56.53%*
Sabatia	*clade C1*	*Chironia*	*12.30%*	*7.69%*	*4.60%*	*1.43%*	*321.94%*	*0.13%*
Sabatia	*Exaculum*	*Chironia*	*12.30%*	*13.48%*	*-1.18%*	*1.45%*	*-81.28%*	*41.63%*
clade C4	*clade C1*	*Chironia*	*13.13%*	*7.69%*	*5.44%*	*1.60%*	*340.53%*	*0.07%*
clade C4	*Exaculum*	*Chironia*	*13.13%*	*13.48%*	*-0.35%*	*1.08%*	*-32.33%*	*74.64%*
clade C1	*Exaculum*	*Chironia*	*7.69%*	*13.48%*	*-5.78%*	*1.64%*	*-353.56%*	*0.04%*
clade C4	*clade C3*	*Sabatia*	*14.43%*	*13.04%*	*1.40%*	*1.24%*	*112.25%*	*26.16%*
clade C4	*Exaculum*	*Sabatia*	*14.43%*	*17.97%*	*-3.54%*	*1.77%*	*-200.21%*	*4.53%*
clade C3	*Exaculum*	*Sabatia*	*13.04%*	*17.97%*	*-4.93%*	*1.80%*	*-274.35%*	*0.61%*
Sabatia	*clade C4*	*Eustoma*	*11.80%*	*16.23%*	*-4.43%*	*1.64%*	*-269.63%*	*0.70%*
Sabatia	*clade C3*	*Eustoma*	*11.80%*	*14.10%*	*-2.30%*	*1.66%*	*-138.88%*	*16.49%*
Sabatia	*Exaculum*	*Eustoma*	*11.80%*	*18.56%*	*-6.76%*	*2.12%*	*-318.84%*	*0.14%*
clade C4	*clade C3*	*Eustoma*	*16.23%*	*14.10%*	*2.13%*	*1.32%*	*161.48%*	*10.64%*
clade C4	*Exaculum*	*Eustoma*	*16.23%*	*18.56%*	*-2.33%*	*1.86%*	*-125.22%*	*21.05%*
clade C3	*Exaculum*	*Eustoma*	*14.10%*	*18.56%*	*-4.46%*	*1.89%*	*-236.41%*	*1.81%*

Table 6. Results of the Ancestral Area Approach for the Chironiinae (Bremer, 1992) performed on a reduced data set of 22 representative taxa present in the whole geographical range of the Chironiinae

Area	Gains	Losses	Gains / Losses	Ancestral Area Score
Mediterranea	5	3	1.7	1
Cape region	1	4	0.25	0.15
Australia	1	5	0.20	0.12
North West America	4	6	0.66	0.39
North East America	2	5	0.40	0.24
Central / South America	3	9	0.33	0.19

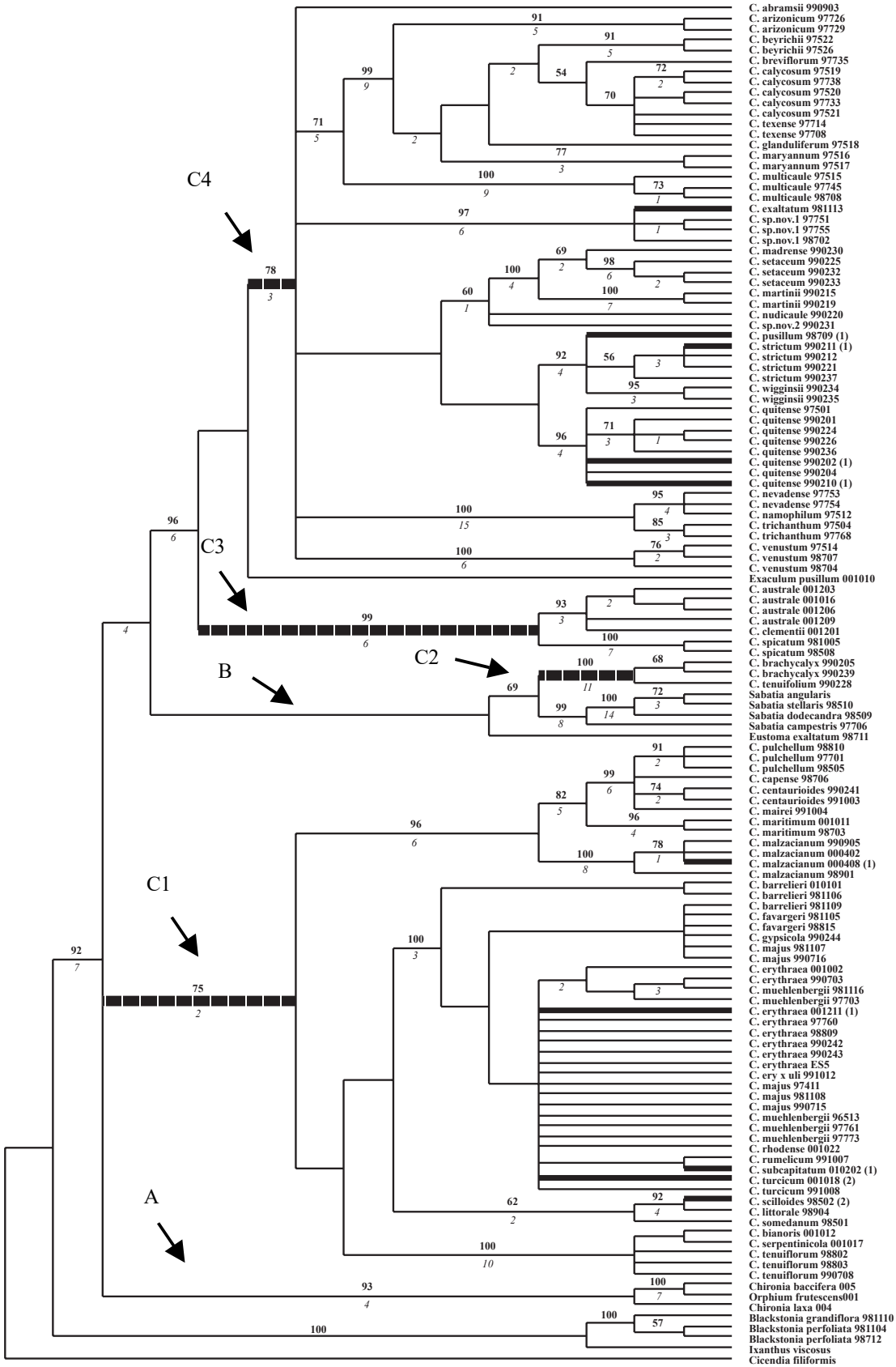


Figure 1. Strict consensus of 4403 minimal trees (839 steps) obtained from parsimony analyses of ITS1 and ITS2 sequences combined. Numbers above and below branches indicate jackknife values of 100 replicates (in %) and decay index, respectively. Thick branches indicate more than one sequence (number in brackets). Dashed lines are *Centaurium* groups

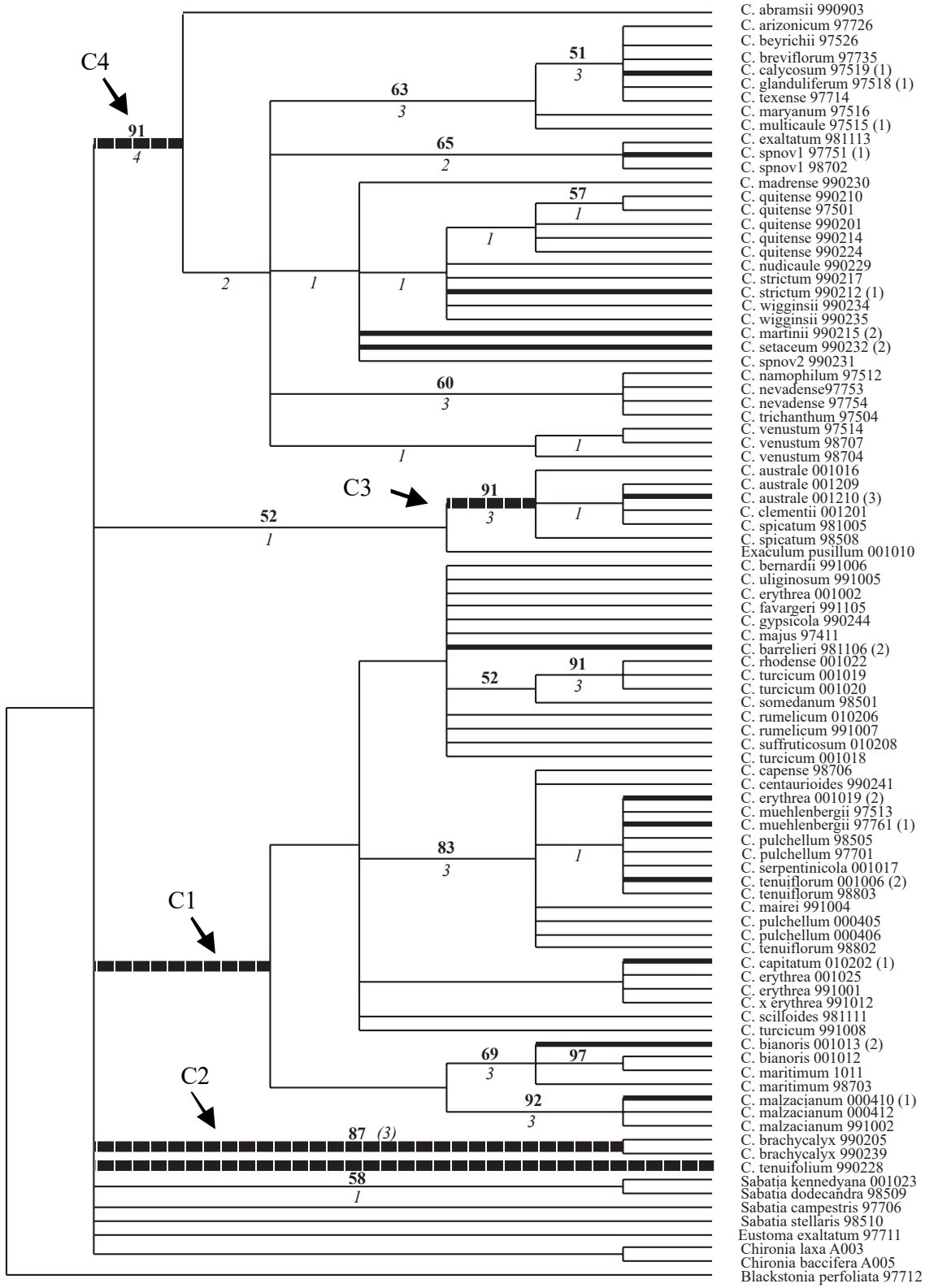


Figure 2. Strict consensus of 66 minimal trees (220 steps) obtained from parsimony analyses of *trn*L intron and *trn*L-F spacer sequences combined. Numbers above and below branches indicate jackknife values of 100 replicates (in %) and decay index, respectively. Thick branches indicate more than one sequence (number in brackets). Dashed lines are *Centaurium* groups

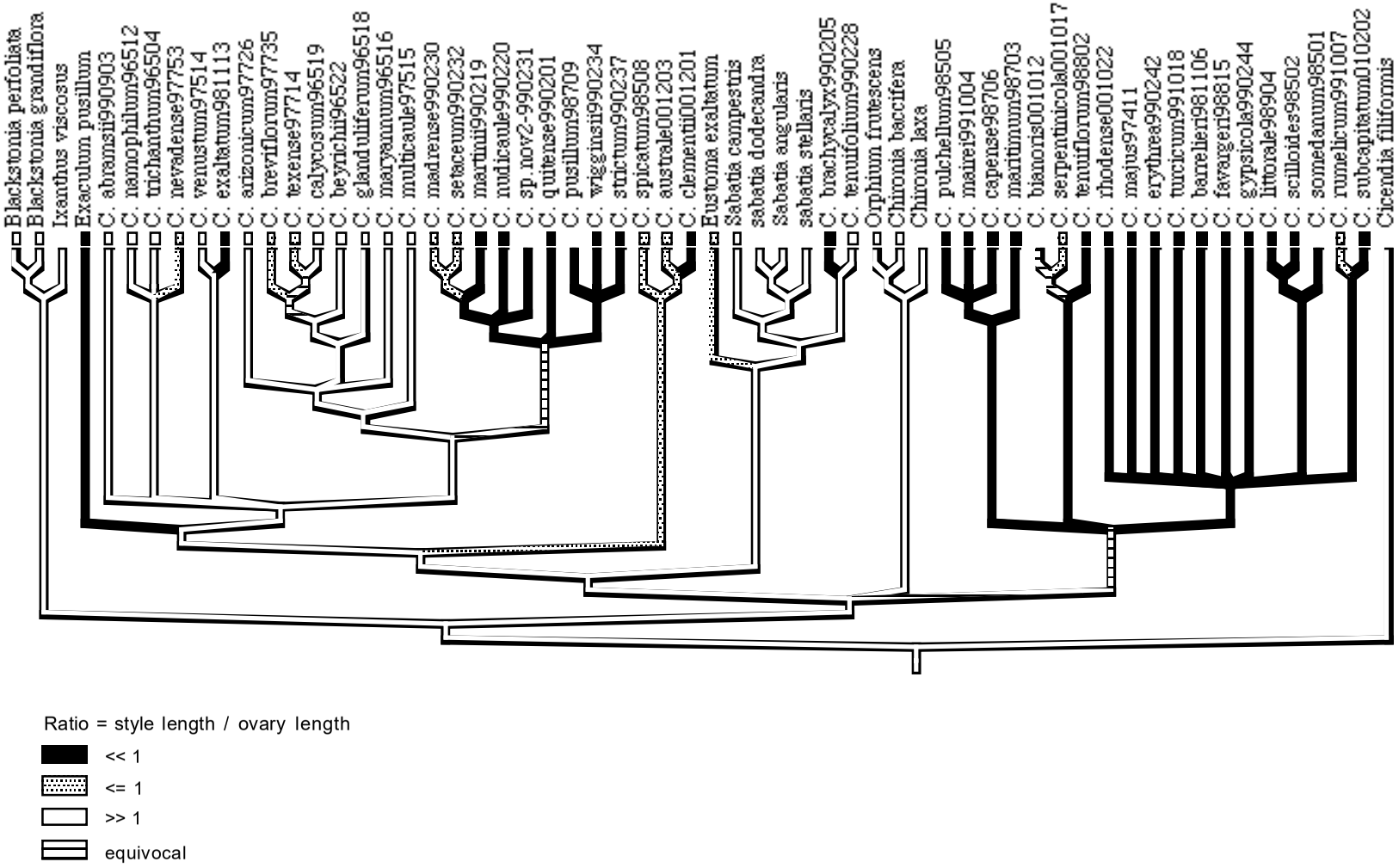


Figure 3. Mapping of morphological characters on ITS strict consensus tree of 60 representative taxa of the Chironiinae

Figure 3 (continued). Mapping of morphological characters on ITS strict consensus tree of 60 representative taxa of the Chironiinae

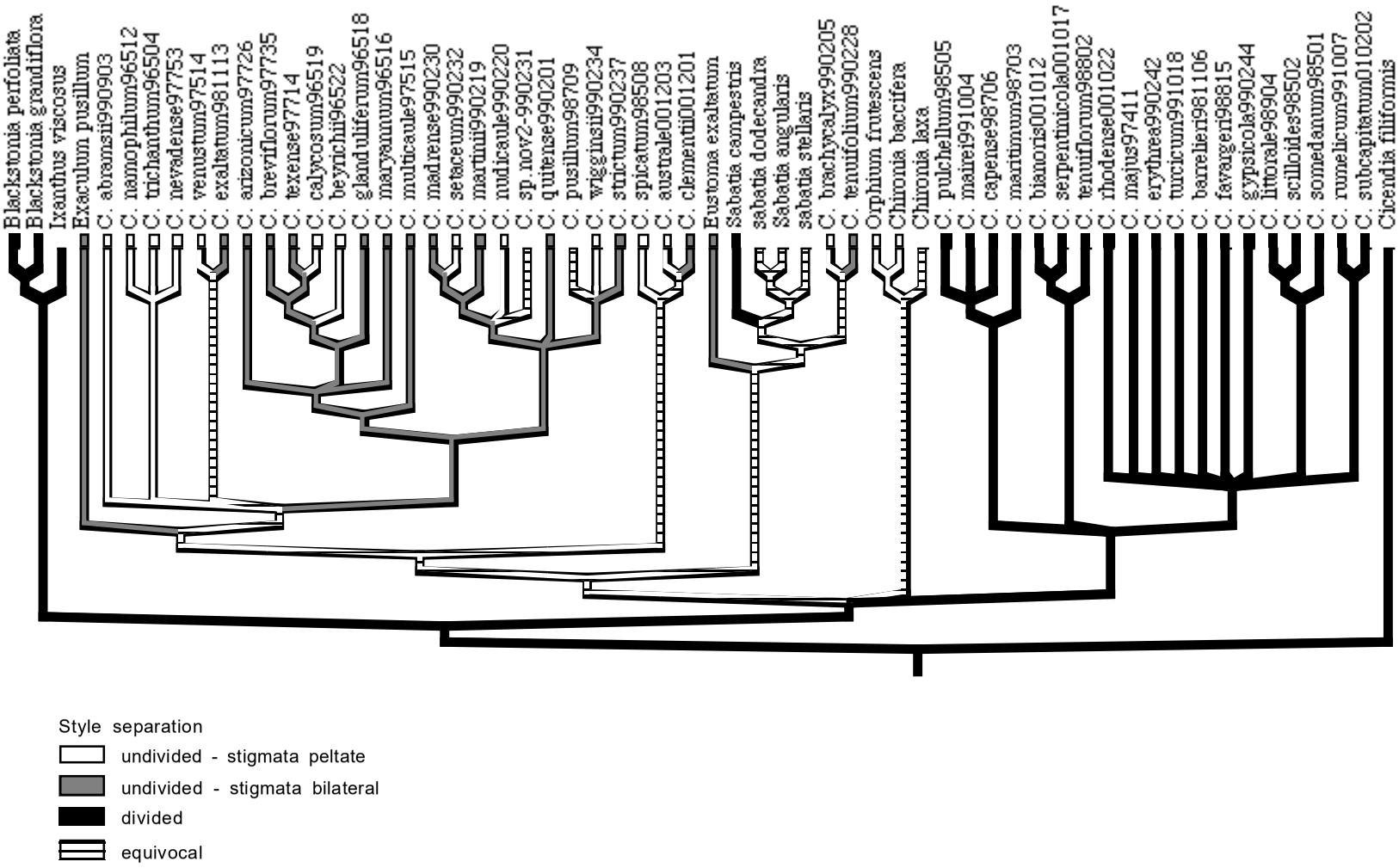


Figure 3 (continued). Mapping of morphological characters on ITS strict consensus tree of 60 representative taxa of the Chironiinae

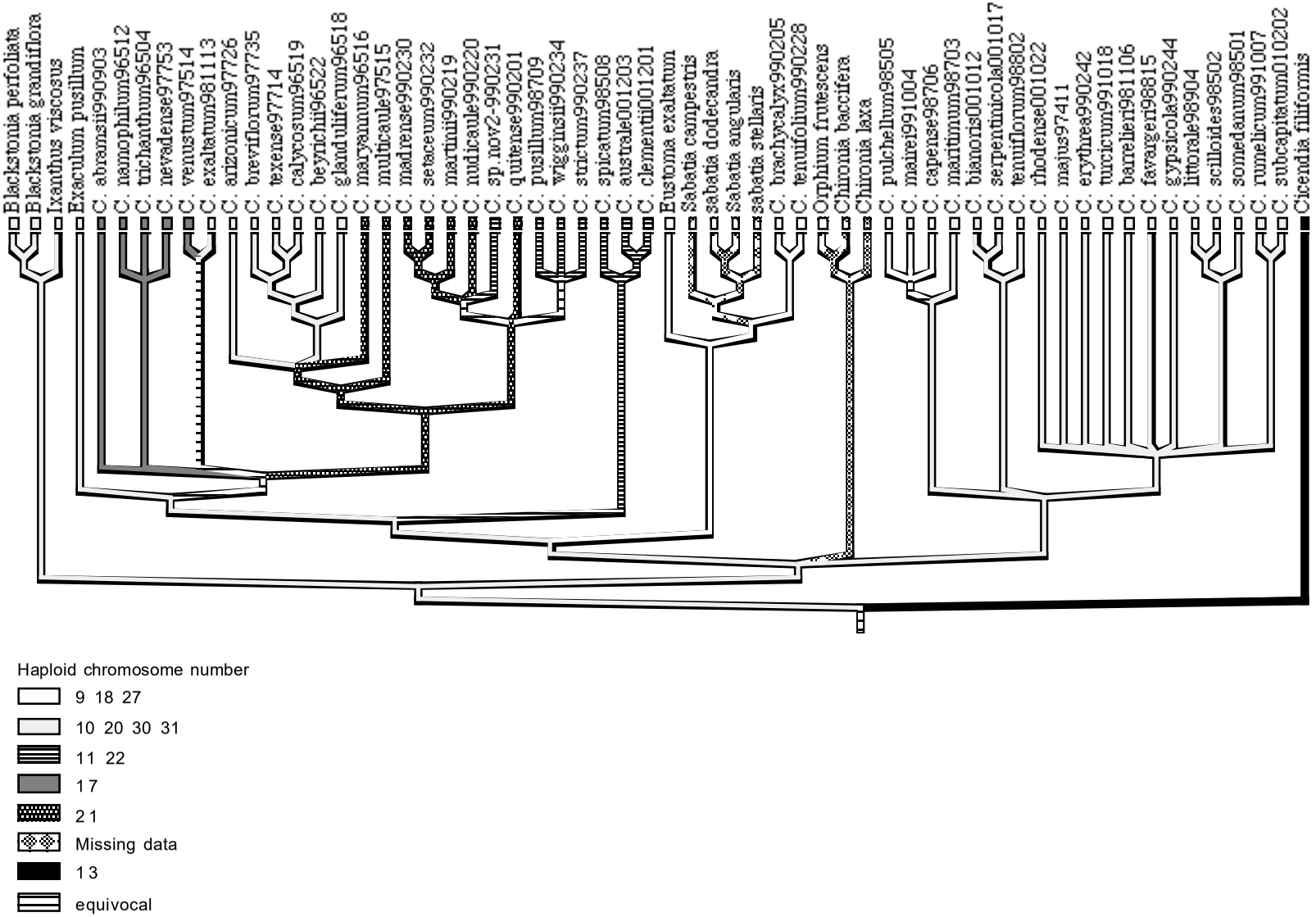


Figure 4. Mapping of karyological characters on ITS strict consensus tree of 60 representative taxa of the Chironiinae

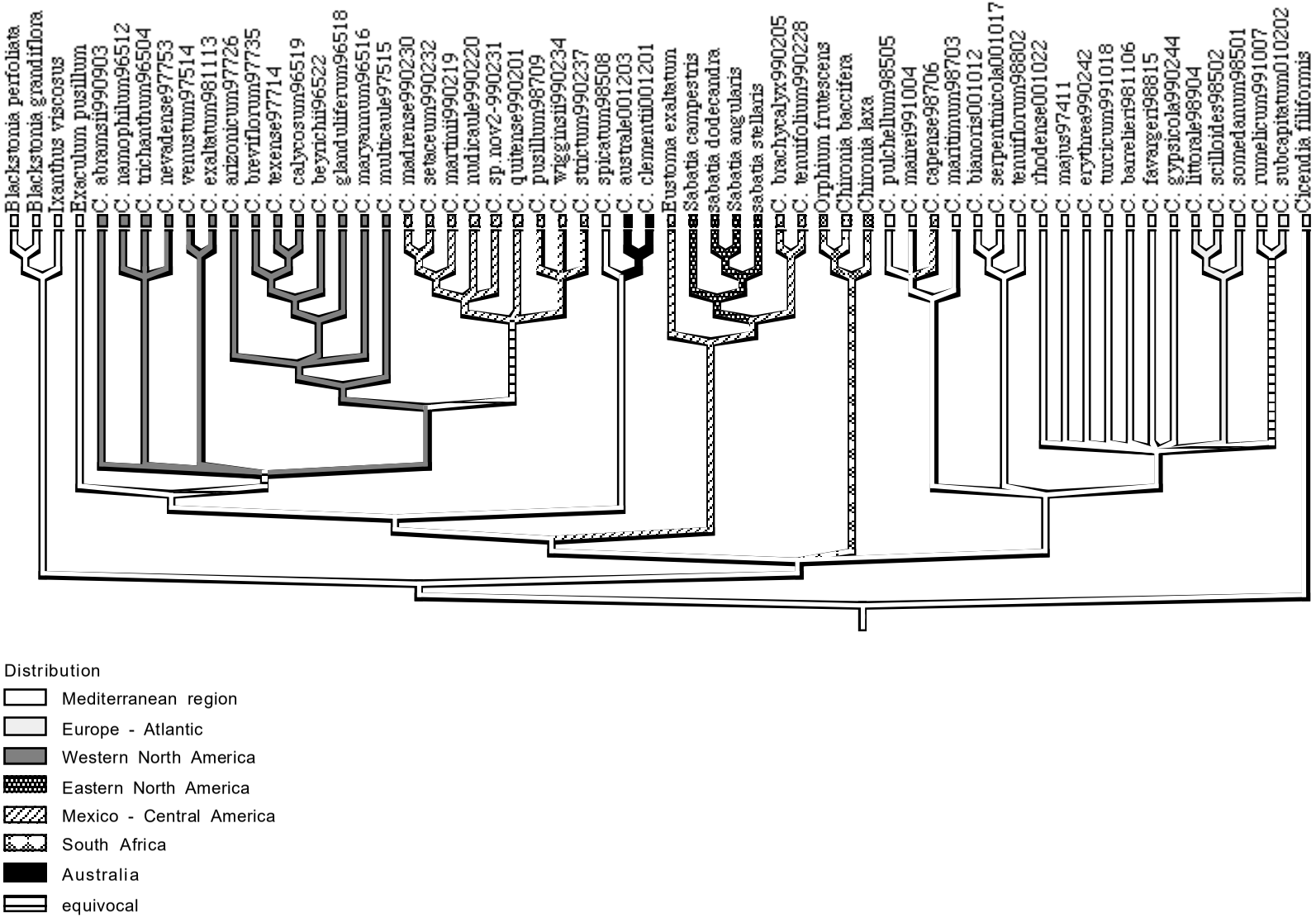
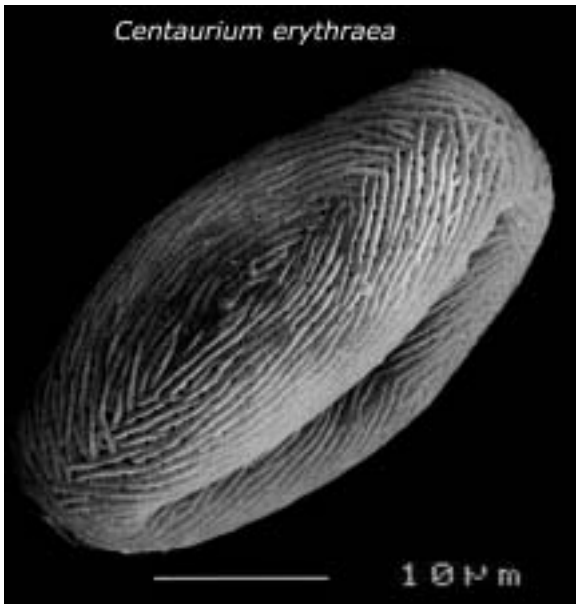
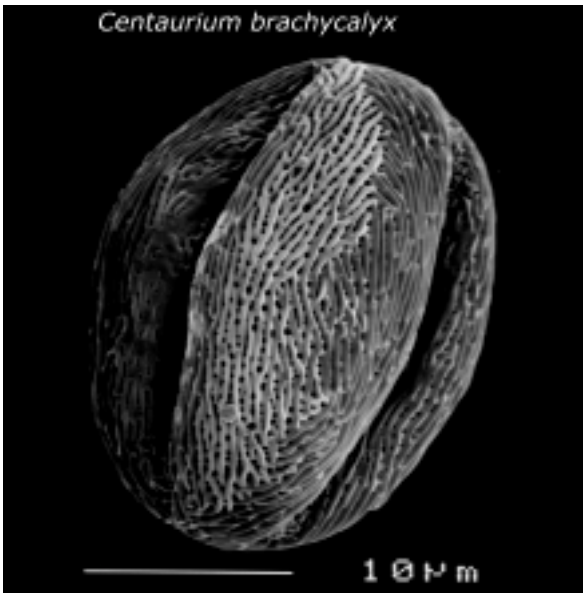


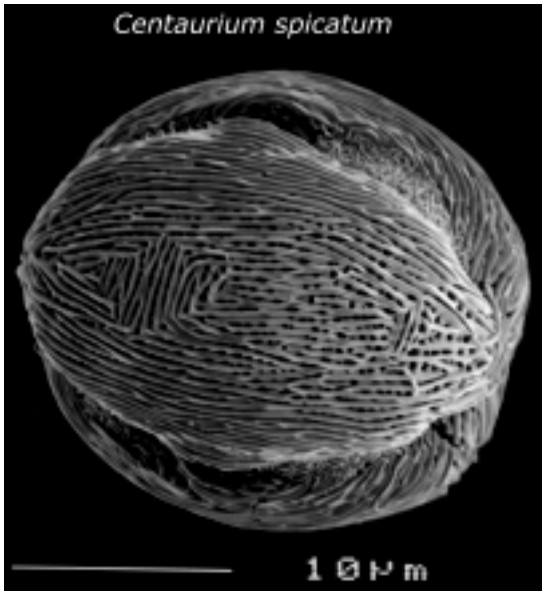
Figure 4 (continued). Mapping of biogeographical characters on ITS strict consensus tree of 60 representative taxa of the Chironiinae



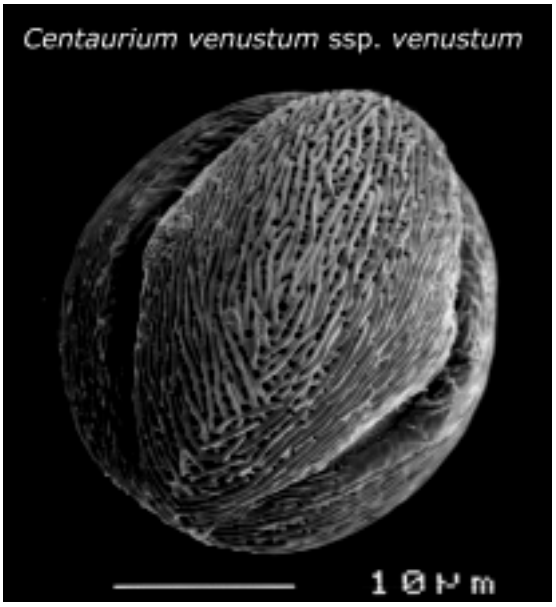
Centaurium erythraea (Clade C1)



Centaurium brachycalyx (Clade C2)



Centaurium spicatum (clade C3)



Centaurium venustum (Clade C4)

Figure 5. SEM microphotographs of *Centaurium* s.l. pollen grains. Clades C1 to C4 are described in the text.

Chapter 4

Phylogenetic evaluation of the 5S-NTS nrDNA sequences in the subtribe Chironiinae (Gentianaceae), with a special focus on the genus *Centaurium*

Phylogenetic evaluation of the 5S-NTS nrDNA sequences in the subtribe Chironiinae (Gentianaceae), with a special focus on the genus *Centaurium*

ABSTRACT

The conception of a monophyletic genus *Centaurium* that has prevailed since the creation of this group has been recently questioned with DNA markers. Presently, *Centaurium* appears to be an artificial assemblage comprising four groups, well supported by independent molecular data sets. On the other hand, morphological characters poorly sustain the fragmentation of the genus and subsequently, the delimitation of the different clades depicted. The aim of the present study is to infer phylogenetic relationships between *Centaurium* and related genera forming the subtribe Chironiinae, by using sequences of the 5S-NTS region. On the other hand, further comparisons with previous molecular and morphological results will provide an opportunity to check the phylogenetic utility of this poorly investigated region. The results obtained do not argue for a good suitability of the 5S spacer for the current scale of taxa divergence. Indeed, the sequences of this region appear to be extremely variables and thus difficult to align without ambiguity. Nevertheless, inferred cladograms, obtained with several analyses of the data matrix, were generally similar in topology and branches support, but otherwise showed important incongruence with ITS results. The main causes of divergence between the two spacers rely on technical causes or on biological processes occurring at the genomic level, such as the presence of paralogous sequences in the 5S locus or the weak efficiency of concerted evolution among 5S repeat units and arrays. On the other hand, congruent relationships have been detected between the data sets, and both clearly support the paraphyly of *Centaurium*, thus favoring previous molecular conclusions over global morphological evidences.

INTRODUCTION

The genus *Centaurium* in its current conception is regarded as a natural assemblage of annuals and biennials herbs, occurring in moist and damp places or disturbed and open sites. This group shows a clear disjoint repartition with a main part of the diversity occurring in the Mediterranean region and the other one in the western North America, but no global study has been hitherto performed to evaluate relationships between these separate assemblages. Moreover, the systematic of each of these geographic groups of *Centaurium* has been partially resolved (Grisebach, 1839; Gilg, 1895; Melderis, 1931, 1972; Zeltner, 1970).

Recently, molecular studies using several DNA data sets have been performed at both infrageneric level, within the North American and Eurasian *Centaurium* (Chapter 1 and 2), and intergeneric one (Chapter 3). These studies revealed for the first time the polyphyly of the genus, with a clear genetic gap between the American and the Eurasian clade, and the exclusion of two well-marked sections as well. Nevertheless, despite the strong support of these respective clades, only a few, if not, morphological characters were available to draw clear limits between them. Moreover, some incongruence have been detected between the different data sets investigated so far, and mainly relied on the phylogenetic ability of the DNA regions investigated to resolve relationships at the infrageneric level, or from biological sources of conflict such as introgression or reticulate evolution. Therefore, within the subtribe Chironiinae, some groups have not been well resolved yet, and other poorly supported assemblages needed further confirmations (e. g. the position of *Exaculum*, *Eustoma*).

Within the DNA regions investigated, the Internal Transcribed Spacer has proven its phylogenetic utility at several systematic levels, with few groups remained unresolved, notably the Eurasian section *Centaurium*. On the other hand, cpDNA was useful to discriminate between genera or well-marked group, or to detect some hybridization events (Chapter 1 and 2), but generally poorly resolved interspecific relationships in the Gentianaceae (Gielly and Taberlet, 1996, Thiv *et al.*, 1999).

Presently, only a few DNA regions are known to evolve faster than the ITS ones or to have similar utility in the phylogenetic reconstruction at the interspecific level.

The 5S region of ribosomal DNA consists of multiple copies occurring in tandem arrays. Each copy is composed of a highly conserved gene, making this region relatively easy to amplify, and is separated from the following one by a non-transcribed spacer (NTS). Sequences of the 5S spacer are variable in size and generally evolve rapidly, thus giving this region a good potential to resolve phylogenetic relationships as proven by several studies performed either at the intergeneric level, e.g. the *Callistachys* group of Fabaceae (Crisp *et al.*, 1999) or

the Beaufortia suballiance of Myrtaceae (Ladiges *et al.*, 1999), or at an interspecific one, e. g. *Eucalyptus* (Myrtaceae) Udovic *et al.*, 1995), *Alibertia* (Rubiaceae) (Persson, 2000), or *Calamus* (Palmae) (Baker *et al.*, 2000). Recently, sequences of the 5S-NTS region have proven better resolution in the phylogeny of the genus *Macrocarpea* (Gentianaceae) than did the ITS ones (Grant & Struwe, unpublished).

Nevertheless, some potential problems are known to appear when using the 5S-NTS region for phylogenetic reconstruction. Within a locus of a multigene family, concerted evolution is thought to homogenize the repeated units. This homogenization is rapid and generally complete for ITS, but appears to occur weakly and episodically in the 5S-DNA region, and not necessary between loci (Sanderson and Doyle, 1992), mainly explaining the relatively poor utilization of this region in molecular systematics.

Another known problem with multiple gene copies is the paralogy. In this case, one copy may be sequenced from one taxon and a different copy, i. e. not truly homologue, from another taxon.

The present study intend to use the phylogenetic potential of the 5S nrDNA region and particularly the intergenic spacer, to better resolve intergeneric relationships within the Chironiinae and to check the results previously obtained with ITS and *trn*LF sequences, particularly the polyphyly of *Centaurium*.

MATERIAL AND METHODS

DNA extraction and amplification---A total of 57 sequences of *Centaurium* species and relatives genera (*Blackstonia*, *Chironia*, *Eustoma*, *Exaculum*, *Orphium* and *Sabatia*) was obtained for the 5S-NTS region. For most of the species, DNA was extracted from fresh leaves collected in the field and immediately dried with silica gel, or from herbarium specimens. The CTAB methods (Doyle and Doyle, 1987) or special kits (DNeasy, Qiagen) were used, depending on the quantity and quality of available material. DNA amplification was performed by performing three successive PCR and using the following 25 µl mix: 2.5 µl BUFFER 10X (PROMEGA), 1.5 µl $MgCl_2$ 25 mM, 0.25 µl dNTP 10 nM, 0.5 µl of each 10 mM primer, 0.20 µl Taq-DNA-polymerase 5 U/µl (PROMEGA) and 2 µl DNA (30 ng/µl). The following primers, designed by Cox, Benett and Dyer (1992), were employed: 5S Forward [5' – TGG GAA GTC CTY GTG TTG CA – 3'] and 5S Reverse [5' – KTM GYG CTG GTA TGA TCG CA – 3'].

The first reaction generally gave several bands. A fragment of ca. 400 base pairs (bp), common to all the accessions, was directly cut and purified from the gel using the QIAEX II Gel Extraction Kit (Qiagen), then diluted 40 times and re-amplified using the same initial protocol. The operation was repeated until the

amplification of only one fragment of ca. 400 bp. Further purification and sequencing were finally performed as previously described (Chapter 1 and 2).

Sequences alignment---DNA sequences were initially aligned using the program Clustal W (Higgins *et al.*, 1996). The alignment obtained was then manually adjusted by sequential pairwise comparisons. For some regions appeared to be very variable and difficult to align, several strategies were employed. First, different aligned matrices were analyzed using the same parameters and inferred tree scores were compared. Second, the ambiguous regions were excluded from the analyses, following the recommendations of Swofford *et al.* (1996). Because of the difficulty to delimitate these regions unambiguously, the program Soap* was used to produce culled matrices (Loytynoja and Milinkovitch, 2001). This package allows the obtainment of different alignments under several parameters (penalties for gaps, gap extension), then compare all the alignments obtained with a reference one, arbitrary chosen, and finally delete the regions were 50 % (or more) sites appear to be ambiguously aligned. For the Soap* program allows only a few sequences to be used, we reduced the initial data set to 26 5S-NTS accessions were ITS data were also available, for further comparisons (see below). Finally, the elision method (Wheeler, 1995) was performed, as an alternative of the culling strategy. Indeed, the deletion of all but alignment-insensitive regions, generally produce data sets with few characters, thus quickly becoming uninformative for phylogenetic reconstruction (Gatesy *et al.*, 1993). In elided matrix, a character that is alignment invariable appears as an identical column in each of the separate data matrices, and the total number of data set weights each transformation in this column combined. Consequently, characters of alignment-variable regions receive less weight. In the current study, 13 initial matrices, produced under different parameters, were combined in a large matrix of 6790 characters.

Sequences parameters---Different parameters such as sequence length, GC content, number of informative characters or pairwise distance divergence values, were computed using PAUP 4.0b6 version (Swofford, 1999).

Phylogenetic analyses---Heuristic searches were performed under the Fitch Parsimony criterion, implanted in PAUP*, using TBR branch swapping, accelerated transformations of characters (ACCTRAN), SIMPLE stepwise sequence addition and saving all minimal trees (Mulpars). A two-steps analytic strategy, with collapse branch option ON, was employed, as described in Chapter 3. Branch support was evaluated using the Jackknifing

method (Farris et al., 1996), with 100 replicates and SIMPLE sequence addition. As previously suggested (Chapter 3), all the analyses were rooted with the genus *Blackstonia*.

Congruence with other data set---In order to evaluate the phylogenetic potential of the 5S-NTS region, a data set of 26 sequences was compared with an ITS matrix comprising the same accessions. For this purpose, congruence tests implanted in PAUP* were conducted. The partition homogeneity test (PHT), was first used to assess whether or not the respective data sets were only partitions of a same large assemblage (Farris *et al.*, 1995). The Wilcoxon signed-rank test (WSR) or SLP test in PAUP* (Templeton, 1983), was then applied between the respective most parsimonious trees obtained.

RESULTS

Sequences characteristics---The sequences characteristics of the 5S-NTS matrix are shown in Table 1. For comparison, the main features of the ITS sequences used in Chapter 3 are also presented. The sequence length within the 5S-nrDNA spacer ranges from 354 base pairs (bp) in *Blackstonia perfoliata* to 409 bp in *Centaurium maritimum*. The GC content mean is 58.75 % in the whole data, ranging from 49.47 % in *Centaurium tenuifolium* (section *Gyrandra*) to 62.75 % in *C. spicatum* (section *Spicaria*). The mean sequence divergence between the outgroup *Blackstonia* and the remaining taxa is 58.97 % +/- 0.62, ranging otherwise from a minimal value of 0 between *C. maritimum* and *C. bianoris* and from a maximum of 88.29 % between *Blackstonia* and *Sabatia*.

Alignments and phylogenetic reconstruction---Several alignments of 57 sequences of the 5S-NTS region were produced. Different analyses performed on the respective matrices, and keeping all the characters do not greatly change tree topologies (results not shown). A matrix of 338 aligned bp, with 242 informative characters (71.38%), 52 uninformative and 44 constant ones was then chosen. In this data set, 15 informative indels, coded with binary presence / absence characters, were added at the end of the matrix and 37 ambiguous characters were deleted. The heuristic searches then performed on this matrix produced 20 MP trees of 1064 steps (C= 0.516 / R = 0.717). The examination of the strict consensus tree (Fig. 1) reveals a basal position of the genus *Eustoma*, along with a major and partially unresolved assemblage of 6 well-supported groups (J values ranging from 71 in

C to 99 in B). In this clade, only the group F, comprising the North American species of *Centaurium* is resolved. On the other hand, the Eurasian *Centaurium* are divided 2 distinct lineage (B and C) not corresponding to sections or clade described elsewhere (Chapter 2), whereas section *Spicaria* form an independent set E. Lastly, *Chironia* and *Orphium* cluster together (group A), whereas *Sabatia* is nested with *Centaurium* section *Gyrandra* (group D).

It is noteworthy, however, that the tree topology varies depending on the analysis parameters. Strict consensus tree obtained with either a non coded matrix with ambiguous characters excluded (391 bp), or keeping all the characters (428 bp), are fully revolved (Fig. 1 - Tree 2). In this case, groups (B - D) and (C - E) form 2 sister clades close to the basal group A, whereas the clade F is still well apart. On the other hand, the 50 % deletion of ambiguously aligned characters in Soap does not greatly alter the tree topology, even though the numbers of characters in the resulting matrices become weak (Fig. 2).

Finally, the elision approach gave quite the same topology than the one encountered without coding the informative indels and branch supports were generally high (Fig. 4)

Congruence between different data sets---The partition homogeneity test gave significant values ($P < 0.001$) indicating that the respective 5S-NTS and ITS matrices could not be considered as a part of the same original data set. The Templeton's Significantly Less Parsimonious test (SLPt), performed between the most parsimonious trees of the respective 5S-NTS (2 trees) and ITS matrices (1 tree) resulting from the same alignment parameters, also reveals that both data sets are fully incongruent each other ($P < 0.0001$).

When comparing topologies and the branch supports of the respective 5S-NTS and ITS strict consensus trees, the following divergences can be observed on the 5S cladograms: the genus *Eustoma* is basal and close to the outgroup; the Eurasian *Centaurium* form a paraphyletic assemblage, sometimes comprising *Exaculum*, *Sabatia*, section *Spicaria* and *Gyrandra*; lastly, within the North American *Centaurium* (clade F), topological differences with ITS appear.

DISCUSSION

The utility of the 5S-NTS region in phylogenetic reconstruction of the Chironiinae---A first and obvious ascertaining, in the present study, was the difficulty of sequence alignments in the 5S data set. Only a few regions were constant between the different sequences (e. g., 13 % in the 338-bp matrix), thus non-favoring

unequivocal alignments. On the other hand, others regions of the data matrix appeared to be hyper-variable and impossible to align. Lastly, insertion / deletion events often occurred, making the data set definitively heterogeneous, and thus delicate to analyze.

Consequently, sequences divergence level was very high, with a maximum of 88 % between *Blackstonia* and *Sabatia*, indicating that this region appear to be not suitable for intergeneric relationships in this group. Nevertheless, within each group, the resolution was generally better than with ITS, suggesting a good potential of 5S-NTS sequences at a specific level or below. Moreover, the number of informative sites is significantly higher in the 5S data set (ca 73 %), but alternatively, the rate of evolution is faster and certainly too high relative to the scale of taxa divergence. In such case, the phylogenetic signal may be obscured by homoplasy (Wendel and Doyle, 1998).

Sequences of the 5S nrDNA region also reveal a high GC content (mean 58 %), similar to the one encountered in ITS (62 %). Consequently, these regions are relatively difficult to amplify, compared to the *trn*LF region of cpDNA with low GC content (Chapter 1 and 2). Moreover, the high variability located at the beginning of the 5S spacer may cause problems for the primer to anneal.

Lastly, the technical difficulties encountered to produce 5S-NTS sequences, with either multiple purification or sometimes cloning methods, do not argue for the good suitability of this region for phylogenetic inferences.

Concluding, in regard to the multiple problems encountered with the 5S-NTS sequences, it appears that this region is not suitable to analyze the present scale of taxa divergence in the Chironiinae. Nevertheless, its potential for interspecific or populational levels may be of interest.

Phylogenetic inferences obtained with 5S-NTS---The trees resulting from diverse analyses of the 5S-NTS data matrices do not significantly change in topology and branch support. Even the deletion of 50 % of the ambiguously aligned characters in Soap* produces similar results, indicating the phylogenetic robustness of the clades obtained. The latter, however, received high Jackknife values. On the other hand, the coding of informative indels produced less resolved clade (Fig. 1), but otherwise, the strict consensus tree obtained is similar in topology with the Jackknife tree (not shown). This may indicate that some branches, present in other strict consensus trees, where variable regions were not deleted or coded, are the consequences of a high level of homoplasy. This is clearly the case for the branch joining the *Sabatia* clade (group D), along with *Exaculum pusillum*, to a part of the Eurasian *Centaurium* (group B); similarly, the inclusion of section *Spicaria* (group E) in the other Eurasian assemblage of *Centaurium* (group C), is weakly supported (Fig. 2)

Whereas 5S-NTS and ITS data sets appear to be highly incongruent (SLP test), some groups are generally present in the respective cladograms (Fig. 3). The *Chironia* - *Orphium* clade is well supported (J = 85) and appear to be basal in the reduced 5S and ITS trees (Fig. 4). This is also the case with the *Sabatia* / *Gyrandra* group (J = 72) and above all, for the American clade of *Centaurium* (J = 97).

The 5S data set, whatever the level of incongruence with ITS, confirm the polyphyly of *Centaurium*, otherwise suggested by the *trn*LF region of cpDNA (Chapter 3). The independent origin of the North American clade, along with the inclusion of section *Gyrandra* in the *Sabatia* clade, are supported by the three molecular data sets hitherto investigated, namely ITS, 5S-NTS and *trn*LF (see Chapter 1 and 3). On the other hand, a possible position of section *Spicaria* in the Eurasian group, along with the split of the latter is not congruent with the previous phylogenetic analyses (chapter 2 and 3).

Results of the ITS and *trn*LF analyses led us to envisage a possible generic rank for the section *Spicaria*, rank supported by other parameters such as the gross morphology, the particular karyological number or the geographic distribution (Chapter 1, 2, and 3). Presently, trees inferred from the elision analysis of the 5S data do not favor the segregation of *C. spicatum* and include it the Eurasian group C (Fig. 3). This result, however is not supported by the coded analysis which confirm section *Spicaria* as an independent lineage (J = 96).

The elision method has been proposed as an alternative of the culling strategy performed in Soap*, by keeping all the alignment produced under different parameters, and joining them in a same large data matrix (Wheeler, 1995). One advantage of this method is the higher weight received by non-ambiguous sites compare to ambiguous ones whose position will vary in the different matrices. Nevertheless, a major problem concern the homology of the characters which may consequently affect tree topology (Wheeler, 1995).

The genus *Exaculum* remains problematic in the 5S data set. The relationship with section Spicaria, as suggested by the previous molecular analyses (Chapter 2 and 3), is not confirmed here. Moreover, this taxon is often placed in the Eurasian clade, but otherwise, this branch received no or weak Jackknife support (Fig. 2). Morphologically, *Exaculum* shares some external resemblance with *Cicendia* (Chapter 3), but unfortunately this taxon could not be included in the present analysis. On the other hand, floral characters reject this association and better fit close affinities with *Centaurium spicatum* (e. g., long calyx teeth, and subcapitate shape of the stigmata).

The phylogenetic position of *Eustoma* also greatly diverge with the conclusions obtained with ITS and *trn*LF data sets. In these approaches, *Eustoma* was either included in the *Sabatia* clade, but with weak Jackknife support, whereas no resolution was obtained with the cpDNA data set (Chapter 3). The close relationship

between *Eustoma* and *Blackstonia* obtained in the 5S analysis is surprising. Both the taxa share similar vegetative outlook, with connate and glaucous leaves and the resemblance of non-flowering individuals of these genera, in the field, is striking. Furthermore, the pollen of *Eustoma* is particular and clearly differs from the type encountered in *Sabatia* (Nilsson and Mansion, unpublished). Nevertheless, floral characters such as the stamen insertion at the sinus of the lobes, and karyological or biogeographical ones better support the phylogenetic neighborhood with *Sabatia* and *Gyrandra*, as suggested by ITS (Chapter 3).

Possible causes of incongruence between 5S and ITS data sets---With the proliferation of molecular tools within the last decade, a new widespread practice consists to apply several data sets to a common assemblage of taxa and to compare the gene trees obtained. The more congruent these cladograms appear, the better will be the knowledge of the natural history of the group, inferred from molecular evidences.

Nevertheless, a general conclusion arises from this appealing approach: the phylogenetic incongruence is not rare and, moreover, it is almost the rule rather than the exception. This has led to many controversies between partisans of the combined analyses and defenders of the separate ones (e. g., Miyamoto, 1985; Hillis, 1987; Doyle, 1992; de Queiroz *et al.*, 1995). On the other hand, many efforts have been made to measure and evaluate the potential of incongruence of competing data set, resulting in the elaboration of many statistical tests (e. g., Templeton, 1983; Mason-Gamer and Kellogs, 1996). Lastly, the phylogenetic incongruence may be seen as an open window into genome history and evolution (Wendel and Doyle, 1998). The possible reasons of phylogenetic incongruence have been classified into three main phenomena, namely (1) the technical causes, (2) the organism-level processes and (3) the gene-level processes (Wendel and Doyle, 1998). Generally, conflict between data sets, as observed with the 5S-NTS and the ITS matrices, results from more or less important interactions between the several causes described.

Presently, a first concern appears to be the gene choice, which must be optimized to the scale of divergence level between the taxa under investigation. If the ITS region gave satisfactory results at intergeneric and interspecific levels, it failed to discriminate between closely related species or recently evolved ones, e. g. between diploid and polyploids taxa in section *Centaurium* (Chapter 2). On the other hand, 5S-NTS appears to evolve too fast to resolve relationships within the Chironiinae, and the problem of excessive homoplasies appearing as synapomorphies may be encountered, thus perturbing phylogenetic reconstruction (Felsenstein, 1978). Another technical problem with the 5S spacer is the sequence length, with a present mean of 390 bp, which is clearly shorter than the ITS one (mean 430 bp). Moreover, the deletion of ambiguous characters, gave even shorter

aligned matrices, which may decrease the potential phylogenetic signal (de Queiroz *et al.*, 1995). However, it has been argued that sequence length always need to be considered with the evolutionary rates (Hillis, 1996), which is presently fast, thus compensating the weak length of the 5S sequences.

At organism level, one of the most significant sources of molecular conflict appears to be hybridization and introgression processes. Incongruent trees resulting from chloroplast and nuclear genes have confirmed or unveiled putative hybrid species in *Centaurium* (Chapter 1 and 2). The pollen parents of *C. malzacianum* (*C. maritimum*) or *C. serpentinicola* (*C. tenuiflorum*), suggested by the ITS data set, are currently confirmed with the 5S sequences. Nevertheless, conflict exists concerning *C. bianoris*. On the base of morphological evidences and topological incongruence between cpDNA and nrDNA data sets, this hybrid was thought to have arisen from hybridization between *C. maritimum* as seed parent and *C. tenuiflorum* as pollen genitor (Zeltner, 1978). Presently, *C. maritimum* and *C. bianoris* share the same 5S-NTS sequence. One explanation could be that the hybrid species may possess paralogous sequences of both parents for the 5S locus along with the ITS sequences of *C. tenuiflorum*. Then, DNA amplification performed on the hybrid *C. bianoris* could have revealed *trn*LF and 5S-NTS sequences of *C. maritimum*, the recipient species, along with the ITS sequences of the pollen parent, *C. tenuiflorum*.

Other potential sources of conflicts have been described at the genome level, such as paralogy and inter-locus interactions (Doyle, 1992; Sanderson and Doyle, 1992).

Paralogy is defined by the occurrence of two loci, arising from gene duplication, in the same individual genome. Therefore, the sampling of orthologous and paralogous sequences in the same data set will lead to phylogenetic incongruence. Presently, several putative paralogous alleles may be expected in the 5S-NTS region, for PCR amplifications generally gave 3 or more bands, generally well separated. The purification performed by isolating the fragment of interest by direct cutting, do not exclude the occurrence of alleles of similar size, and thus different paralogous sampling in the data set. Further cloning methods need to be applied in order to investigate the range of alleles variation within the 5S locus, in the Chironiinae.

Furthermore, additional complexity occur in the multigene families where interlocus homogenization, also known as concerted evolution (Arnheim, 1983), may be either complete, as generally encountered in the ITS region, or partial. Weak homogenization has been documented in 5S region, not only between loci (Sanderson and Doyle, 1992), but also within a locus (Kellogs *et al.*, 1996; Cronn *et al.*, 1996).

This case has been sometimes encountered in 5S-NTS data set, where a single purified fragment gave clearly polymorphic sequences.

CONCLUSION

The present study was undertaken to apprise the phylogenetic relationships within the Chironiinae subtribe, and furthermore, to check the paraphyly of the genus *Centaurium* previously revealed by two molecular data sets of independent origin (Chapter 3). Moreover, another aim of this work was to test the efficiency of a fast evolving region of ribosomal DNA, namely the 5S Non Transcribed Spacer (5S-NTS), in phylogeny reconstruction. The sequences obtained for 57 accessions of *Centaurium* and relative genera, appear to be very variable and thus difficult to align without ambiguity. Several analyses performed, with different aligned matrices, using culling or elision methods, gave generally similar results. The polyphyly of the genus *Centaurium* was confirmed in all the inferred cladograms, which otherwise reveal major incongruence with ITS phylogenies. A first explanation appeared to be the fast mutation rate of the 5S spacer which did not fit the level of taxonomic divergence of the Chironiinae group, resulting in difficulty to analyze data sets, with likely a high level of homoplasy. Other causes of incongruence with the 5S-NTS region were the possible presence of several paralogues alleles within a locus, along with a weak inter-locus concerted evolution. Lastly, the frequent occurrence of hybridization within the genus *Centaurium* definitively complicates the phylogenetic reconstruction using 5S-NTS sequences.

REFERENCES

ARNHEIM, N. D. 1983. Concerted evolution of multigene family. in *Evolution of genes and proteins*. M. Nei and R. K. Koehn Editors. Sinauer Associates. Sunderland.

BAKER, W. J., T. A. HEDDERSON, and J. DRANSFIELD. 2000. Molecular phylogenetics of *Calamus* (Palmae) and related rattan genera based on the 5S nrDNA spacer sequence data. *Molecular Phylogenetics and Evolution* 14 (2) : 218-231.

COX, A. T., M. D. BENNETT, and M. A. DYER. 1992. Use of polymerase chain reaction to detect spacer size heterogeneity in plant 5S-rRNA gene clusters and to locate such clusters in wheat (*Triticum aestivum* L.). *Theoretical and Applied Genetics* 83 : 684-690.

CRISP, M. D., R. APPELS, S. M. SMITH, and W. M. S. KEYS. 1999. Phylogenetic evaluation of 5S ribosomal RNA gene and spacer in the *Callistachys* group (Fabaceae: Mirbalieae). *Plant Systematic and Evolution* 218 : 33-42.

CRONN, R. C., X. ZHAO, H. PATERSON, and J. F. WENDEL. 1996. Polymorphism and concerted evolution in a multigene family: 5S ribosomal DNA in polyploids and allopolyploids cottons. *Journal of Molecular Evolution* 42 : 685-705.

De QUEIROZ, A., M. J. DONOGHUE, and J. KIM. 1995. Separate versus combined analysis of phylogenetic evidence. *Annual Review of Ecology and Systematics* 26 : 657-681.

DOYLE, J. J. 1992. Gene trees and species trees: molecular systematics as one character taxonomy. *Systematic Botany* 17 : 144-163.

DOYLE, J. J. and J. L. DOYLE. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical bulletin* 19 : 11-15.

FARRIS, J. S., M. KALLERSJÖ, A. G. KLUGE, and C. BULT. 1995. Testing significance of incongruence. *Cladistics* 10 : 315-319.

FARRIS, J. S., V. A. ALBERT, M. KÄLLERSJÖ, D. LIPSCOMB, and A. G. KLUGE. 1996 Parsimony jackknifing outperforms neighbor-joining *Cladistics* 12 (2) : 99-124.

FELSENSTEIN, J. 1978. Cases in which parsimony and other compatibility methods will be positively misleading. *Systematic Zoology* 27 : 401-410.

GATESY, J., R. DeSALLE, and W. WHEELER. 1993. Alignment-ambiguous nucleotide sites and the exclusion of systematic data. *Molecular Phylogenetics and Evolution* 2 : 152-157.

GIELLY, L. and P. TABERLET. 1996. A phylogeny of the European gentians inferred from chloroplast *trn*L (UAA) intron sequences. *Botanical Journal of the Linnean Society* 120 (1) : 57-75.

GILG, E. 1895. Gentianaceae. in *Die Naturlichen Pflanzenfamilien*. 4 (2) : 50-108. A. Engler and K. Prantl Editors. Verlag von Wilhelm Engelmann. Leipzig.

GRISEBACH, A. H. R. 1839. *Genera et Species Gentianearum adjectis observationibus quibusdam phytogeographycis* : 116-125. J. G. Cotta. Stuttgart et Tübingen.

HIGGINS, D. G., J. D. Thompson, and T. J. Gibson. 1996. Using CLUSTAL for multiple sequence alignments. *Methods in Enzymology* 266 : 383-402.

HILLIS, D. M. 1987. Molecular versus morphological approaches to systematics. *Annual Review of Ecology and Systematic* 18 : 23-42.

HILLIS, D. M. 1996. Inferring complex phylogenies. *Nature* 383 : 23-42.

KELLOGS, E. A., R. APPELS, and R. MASON-GAMER. 1996. When genes tell different stories: incongruent gene trees for diploid genera of Triticeae (Gramineae). *Systematic botany* 21 : 231-247.

LADIGES, P. Y., G. I. McFADDEN, N. MIDDLETON, D. A. ORLOVICH, N. TRELOAR*, et al.* 1999. Phylogeny of *Melaleuca*, *Callistemon* and related genera of *Beaufortia* suballiance (Myrtaceae) based on 5S and ITS1 spacer regions of the nrDNA. *Cladistics* 15 : 151-172.

LOYTYNOJA, A. and M. C. MILINKOVITCH. 2001. SOAP, a program for cleaning multiple alignments from unstable blocks. *Bioinformatics* (in press)

MASON-GAMER, R. J. and E. A. KELLOGS. 1996. Testing for phylogenic conflict among molecular data set in the tribe Triticea (Gramineae). *Systematics Biology* 45 (4) : 524-545.

MELDERIS, A. 1931. Genetical and taxonomical studies in the genus *Erythraea* Rich. *Acta Horti. Bot. Univ. Latv.* 6 : 123-256.

MELDERIS, A. 1972. Taxonomic studies on the European species of the genus *Centaurium* Hill. *Botanical Journal of the Linnean Society* 65 (2) : 224-250.

MIYAMOTO, M. M. 1985. Consensus cladograms and general classifications. *Cladistics* 1 : 186-189.

PERSSON, C. 2000. Phylogeny of the tropical *Alibertia* group (Rubiaceae), with emphasis on the genus *Alibertia*, inferred from ITS and 5S ribosomal DNA. *American Journal of Botany.* 87 (7) : 1018-1028.

SANDERSON, M. J. and J. J. DOYLE. 1992. Reconstruction of organismal and gene phylogenies from data on multigene families: concerted evolution, homoplasy, and confidence. *Systematic Biology.* 41 : 4-17.

SWOFFORD, D. L. 1999. *PAUP* Phylogenetic Analysis Using Parsimony (*and other methods) Version 4.* Sinauer Associates. Sunderland.

SWOFFORD, D. L., G. K. OLSEN, P. J. WADDDELL, and D. M. HILLIS. 1996. Phylogeny reconstruction. in *Molecular Systematics, second edition*. D. M. Hillis, C. Moritz and B. K. Mable Editors. Sinauer Associates. Sunderland.

TEMPLETON, A. R. 1983. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of human and apes. *Evolution.* 37 : 221-244.

THIV, M., L. STRUWE, and J. W. KADEREIT. 1999. The phylogenetic relationships and evolution on the Canarian laurel forest endemic *Ixanthus viscosus* (Aiton) Griseb. (Gentianaceae) : evidence from *mat*K and ITS sequences, and floral morphology and anatomy. *Plant Systematics and Evolution.* 218 : 299-317.

UDOVIC, F., G. MCFADDEN, and P. Y. LADIGES. 1995. Phylogeny of *Eucalyptus* and *Angophora* based on 5S rDNA spacer sequence data. *Molecular Phylogenetics and Evolution.* 4 (3) : 247-256.

WENDEL, J. F. and J. J. DOYLE. 1998. Phylogenetic incongruence: window into genome history and molecular evolution. in *Molecular systematics of plants 2. DNA sequencing*. D. E. Soltis, P. M. Soltis and J. J. Doyle Editors. Kluwer Academic. Boston.

WHEELER, W. C. 1995. Sequence alignment, parameters sensitivity, and the phylogenetic analysis of molecular data. *Systematic Biology.* 44 : 321-331.

ZELTNER, L. 1970. Recherches de biosystématique sur les genres *Blackstonia* Huds. et *Centaurium* Hill (Gentianaceae). *Bulletin de la Société Neuchâteloise des Sciences Naturelles.* 93 : 1-164.

ZELTNER, L. 1978. Recherches sur le *Centaurium bianoris* (Sennen) Sennen. *Revue de Biologie et Ecologie Méditerranéenne.* 5 (1) : 51-58

Table 1. Sequences characteristics of the 5S-NTS and ITS regions investigated in *Centaurium* and relative genera of the Chironiinae subtribe. (*) ITS sequences data provide from chapter 3 and are shown here for comparison. Parsimony characteristics are given for the whole data matrix (428 bp for 5S-NTS and 453 bp for ITS, respectively).

Sequence characteristics	**5S-NTS**	**ITS***
Length range (bp)	354 (*Blackstonia*) – 409 (*C. maritimum*)	398 (*Sabatia angularis*) – 463 (*Centaurium*)
Length mean (bp)	390	430
G+C content (mean %)	58.75	61.81
G+C content (range %)	49.47 (*C. tenuifolium*) – 62.75 (*C. spicatum*)	59.4 (*Blackstonia*) –63.8 (*Centaurium majus*)
Aligned length	428	477
Number of excluded sites	37	24
Number of coded indels	15	-
Number of sites after exclusion	393	453
Length of indels (range) (bp)	1-22	-
Length of indels (total - bp)	68	-
Parsimony informative	316 (73.8 %)	232 (48.6 %)
constant	63 (14.7 %)	74 (15.5 %)
uninformative	90 (21 %)	156 (32.7)
indels (total number)	223	155
g1-stat	-0.44	-0.22

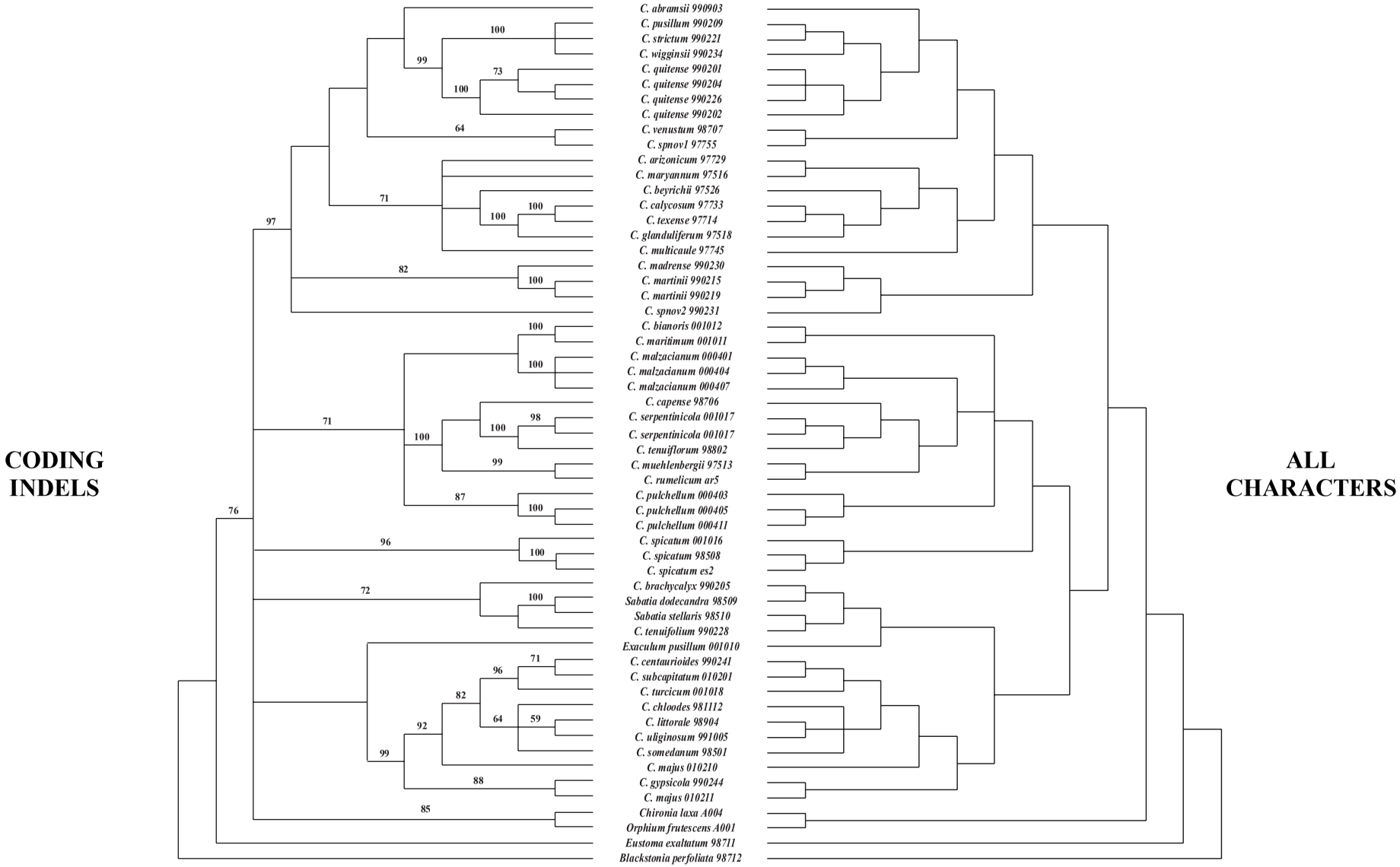


Figure 1. Strict consensus of 20 (1064 steps) and 12 (1454 steps) minimal trees, respectively, obtained from parsimony analyses of 57 sequences from the 5S-NTS region of rDNA. Tree 1 results from coding strategy of the informative indels, whereas all characters are kept in Tree 2 (see text). Numbers above branches indicate jackknife values of 100 replicates (%).

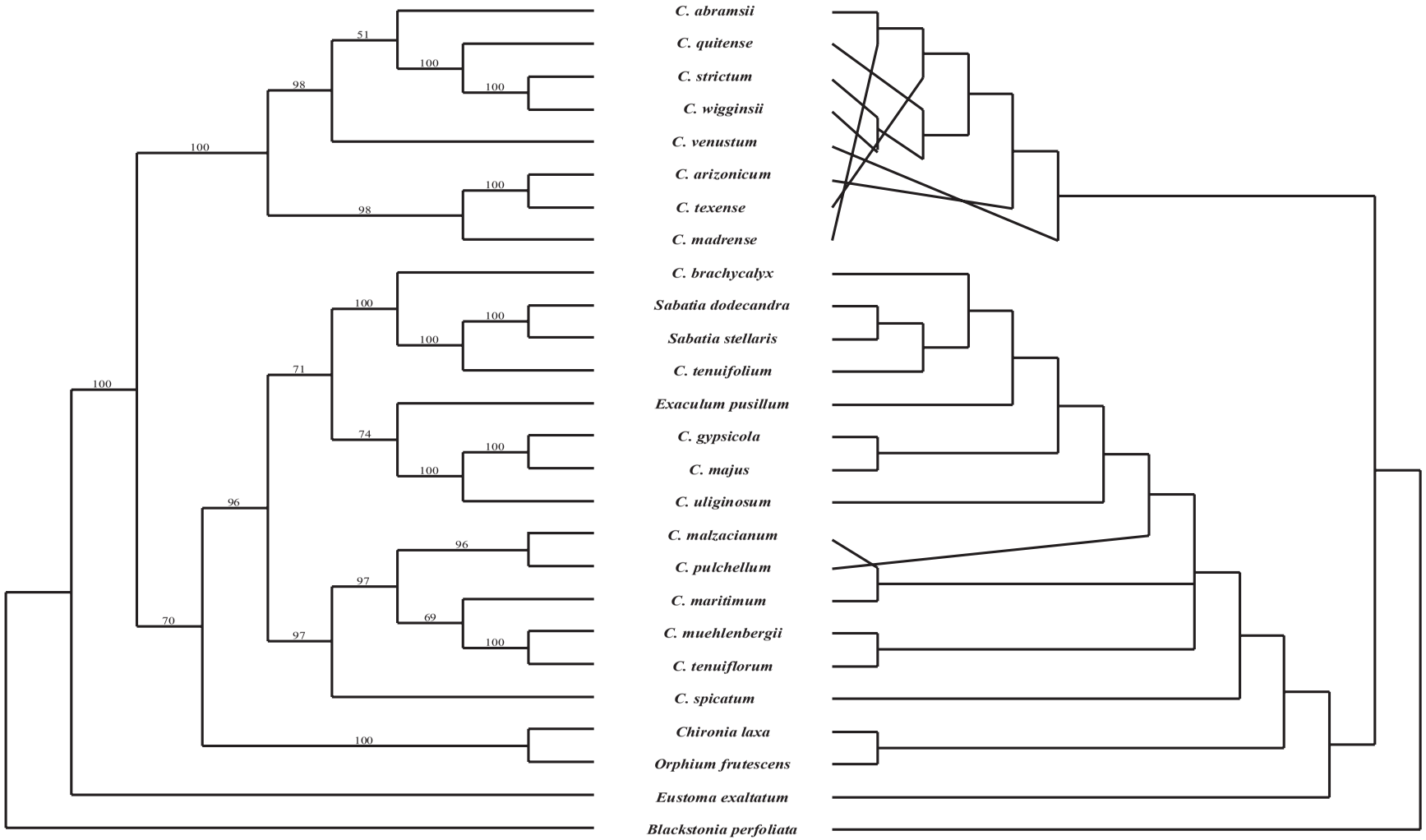


Figure 2. Topology comparison between the Jackknife tree obtained from the elision method (right) and one most parsimonious tree resulting from the culling of 50 % ambiguously aligned sites (see text). Deviations of branches indicate the main conflicts between the trees.

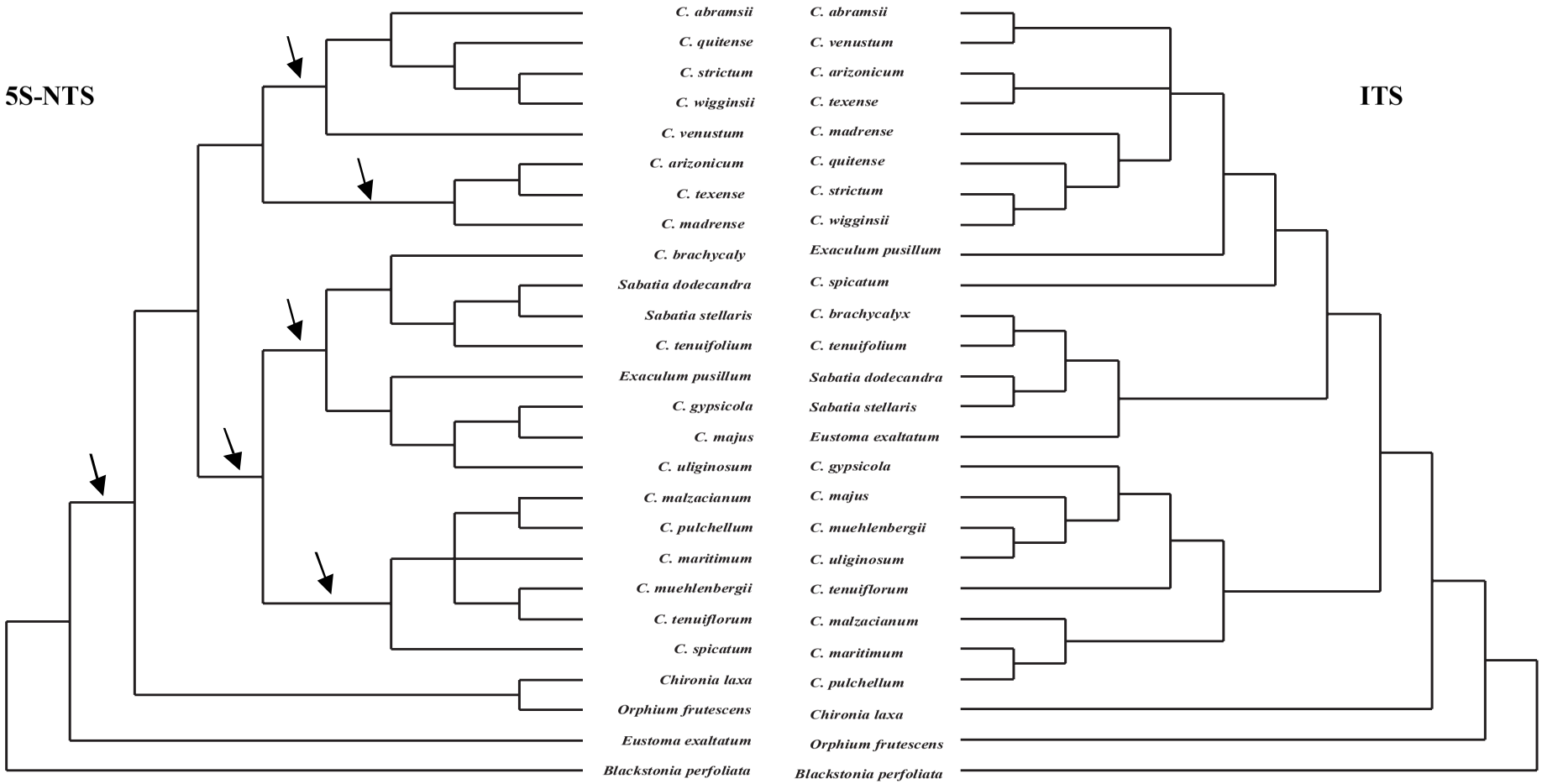


Figure 3. Topology comparison between most parsimonious trees obtained from a reduced data set of 26 5S NTS sequences (left), and ITS sequences (right), respectively. Arrows indicate incongruence between the internal branches.

Chapter 5

RAPD assessments in *Centaurium* with a particular focus on the *C. muehlenbergii* complex

RAPD assessments in *Centaurium* with a particular focus on the *C. muehlenbergii* complex

ABSTRACT

The *Centaurium muehlenbergii* complex occurs mainly in the western part of the United States and is currently seen as a particularly difficult taxonomic group. Many forms have been described, ranging from small plants with relatively big corollas (1.5 - 2 cm) to erect plants of important size (> 50 cm), with all the intermediates possible. Moreover, several synonyms have been proposed to fit this diversity, but generally complicate the systematics of the group. Lastly, the occurrence of European introduced species morphologically closed to some forms encountered in *C. muehlenbergii*, along with an ambiguous typification of this taxon, definitively make unclear its delimitation. The use of RAPD markers to clarify the systematics of this group has been performed on several accession of *C. muehlenbergii* sampled in California and Texas, along with close related taxa with similar outlook, namely *C. erythraea* and *C. tenuiflorum*, collected in the Mediterranean region, the United States and Australia. The first results indicate that populations of *C. erythraea* and *C. tenuiflorum* native from Europe differ in RAPD patterns, with those introduced in North America and Australia. Otherwise, the naturalized forms of *C. tenuiflorum* appear to be indistinguishable with *C. muehlenbergii*, on the base of morphological or molecular characters. Moreover, Australian and American populations of *C. erythraea* show clinal variation in morphology between *C. erythraea* forms natives of Europe and *C. muehlenbergii*. These preliminary observations suggest a possible hybrid origin for *C. muehlenbergii*, along with introgression patterns with *C. erythraea*. Furthermore, RAPD patterns of introduced populations of *C. tenuiflorum* are similar with *C. muehlenbergii*, and clearly differ with *C. tenuiflorum* of the Mediterranean region. The role of the latter as a putative parent of *C. muehlenbergii* remains unclear, mainly due to the reduced size of the present sample. Further investigation involving a larger data set of the taxa currently investigated, along with the inclusion of native species of California, alternatively described under the binomial *C. muehlenbergii*, are necessary for a better resolution of this group.

INTRODUCTION

Within *Centaurium* s. l., many natural groups are thought to be of recent evolution and some of them are particularly difficult to discriminate with morphological or karyological evidences (Melderis, 1931; Zeltner, 1970). Under various ecological constraints, different taxa may show convergent similarities (e.g. *C. tenuiflorum* and *C. pulchellum* assemblages), whereas close-relative ones exhibit very different outlooks (e. g. *Spicaria* section). Moreover, recent molecular studies performed on the main groups of this polyphyletic genus, and using different DNA region sequences, reveal unresolved clade (Chapters 1 and 2). This may suggests close genetic affinities between otherwise morphologically well-marked species. The important variability observed within *Centaurium*, at phenotypic or genotype levels, is increased by the frequent occurrence of introgression or reticulate evolution (Chapters 1 and 2). One consequence of this apparently lack of delimitation in some *Centaurium* taxa, is the rise of an abundant and redundant taxonomy, comprising many synonyms or erroneous determinations.

The *Centaurium muehlenbergii* complex (including *C. curvistamineum*, *C. floribundum*, and *C. tenuiflorum*) may be viewed as one of such difficult group. Several names have been proposed to describe a hypothesized Californian native taxon, often confounded with native relatives (e. g. *C. davyi*, *C. exaltatum*, or *C. trichanthum*) or with introduced species (e. g. *C. erythraea* or *C. pulchellum*). The taxonomic history of *C. muehlenbergii* is a difficult one and much of the confusion has arisen both with the typification of the taxon and with the recent report of *C. tenuiflorum*, a Mediterranean native, in North America (Pringle, unpublished). In his original description of *Erythraea muehlenbergii*, Grisebach (1839) cited several specimens collected by Muhlenberg throughout the United States, including the European *C. pulchellum*. Latter, some authors reduced *C. muehlenbergii* to the Western populations originally collected (Hooker and Arnott, 1830-1841; Gray, 1878), whereas others treated these populations as new taxa, namely *C. floribundum* (Benth.) B. L. Robins (Dunn, 1967). This combination has further been applied to designate several misidentified Californian taxa. The schizotypification of *C. muehlenbergii* was finally completed by Piper (1906). On the other hand, annotations on herbarium sheets mention *C. tenuiflorum* as a synonym of *C. muehlenbergii* or *C. floribundum* (Broome, unpublished data). *Centaurium tenuiflorum* is of Mediterranean origin, but has also been reported as an abundant naturalized species in Australia (Adams, 1996), and very recently in North America (not before 1990). Some authors may have possibly confounded *C. muehlenbergii* with *C. tenuiflorum*, treating the alternate form (i.e. the typified *C. muehlenbergii*) under *C. curvistamineum* (Abrams, 1951; Munz and Keck, 1959). Lastly, some confusion with *C. erythraea*, another European species, has definitively made unclear the *C. muehlenbergii* delimitation.

Recent molecular studies performed on the American and Eurasian species of *Centaurium* (Chapter 1 and 2) suggested the inclusion of *C. muehlenbergii* in the Old World group, close to *C. erythraea*, but well apart from the *C. tenuiflorum* clade. The latter comprises both diploids and tetraploids collected around the Mediterranean basin (Chapter 2). On the other hand, sequence divergence was weak between *C. erythraea* and *C. muehlenbergii*, for the ITS and *trn*LF regions investigated, resulting in unresolved clade. Hence, a hybrid origin involving *C. erythraea* and *C. tenuiflorum*, or the occurrence of introgression patterns, has been suggested to explain the origin of *C. muehlenbergii* in California (Chapter 1).

Nevertheless, taxa sampling in these study may not have covered the whole diversity of *C. muehlenbergii*. Moreover, other molecular tools need to be performed, in other to resolve the phylogenetic relationships within this group.

Over the past decade, Random Amplified Length Polymorphic DNAs (RAPDs) have been widely used to assess genetic polymorphism, providing hundred of studies at intergeneric, inter or intra-specific or populational levels (Wolfe and Liston, 1998). These PCR-based techniques, using single, arbitrary, decamer primers, are technically simple and have attracted many researchers for their speed to provide data sets. On the other hand, some disadvantages are known (Harris, 1999), including technical problems (reproducibility, primer structure, product competition) or data interpretation (homology, dominance, allelic variation, genome sampling or non-independence of loci). Nevertheless, these problems can be overcome in many cases, and a high number of studies have generated biologically meaningful data, thus attesting to the applicability of the technique (Wolff and Morgan-Richards, 1999).

In the present study, we intend to use RAPD markers (1) to provide a genetic delimitation of *Centaurium muehlenbergii* (2) to check for patterns of putative reticulate evolution or introgression, and (3) to infer phylogenetic relationships with putative parent species, *C. erythraea* and *C. tenuiflorum*. Moreover, Australian populations of *C. tenuiflorum* have been added in order to cover a main part of the present geographical distribution of this taxon.

MATERIAL AND METHODS

Taxa sampling---Because of the difficulty to discriminate between different accessions sampled in the fields, the name *C. muehlenbergii* was attributed to all the population of this taxon, collected in California and Texas. Indeed, it was often encountered a great variation within each populations. Moreover, the morphs previously described, such as those with erect branching, compact inflorescence, sessile and small flowers, known as the C. "tenuiflorum" forms, and others more ramified, with lax inflorescence and more or less pedicellate flowers, show many intermediates.

Sixteen populations of "*C. muehlenbergii*" and two accessions of "*C. erythraea*" were sampled in California and Texas (June 1996 / June 1997), whereas two populations of "*C. tenuiflorum*" and 8 of "*C. erythraea*" were of Australian origin (November 2000). European native populations of *C. tenuiflorum* along with *C. erythraea* and some members of the section *Centaurium* (e.g., *C erythraea* subsp. *rumelicum* and *C. majus* subsp. *rhodense*) were added to the analyses. The respective origin and collection number of these different accessions are shown in Table 1.

DNA extraction and RAPD amplification---For most taxa, fresh leaves collected in natural populations were dried with silica gel; in some cases, herbarium samples were used. When possible, only one individual was selected per population, but in some cases, leaves of several small individuals were used. Preliminary analyses do not reveal an important intra-populational variability (data not shown).

Total DNA extraction was performed from leaf tissues using either the CTAB method (Doyle and Doyle, 1987) with an additional cleaning step using ammonium acetate 7,5 M, or using a DNeasy® kit (Qiagen, Basel). Template DNA concentration was checked by visual observation of staining intensity relative to a molecular weight marker and found to be ~0.5 ng/µl.

We used PCR mix of 25 µl containing 1X PCR buffer (Qiagen AG, Basel), 1.5 mM of $MgCl_2$, 2 µl of Q-solution (0.4 X), 0.2 mM of each dNTP, 0.2 µM of a decamer primer (Operon Technologies), 0.75 U of Taq DNA polymerase (Qiagen AG, Basel) and 1 ng / µl template DNA. Amplifications were performed in a Biometra I thermo-cycler, with an initial denaturation at 94°C for 5 min, followed by 38 cycles of 1 min at 93°C, 1 min at 45°C and 1 min at 72°C, ending with a final step extension of 5 min at 72°C. PCR products were separated on a 1.6 % (w/v) agarose gel stained with ethidium bromide and visualized under UV light. Sixty primers of the series OPB 1-20, OPP 1-20 and OPT 1-20 (Operon Technologies, California 94501) were tested on 5 populations and 14 primers amplifying specific markers for each species, were selected (OPP8/P9/P17 - OPT1/T5/T7/T8/T13/T15/T16/T20 - OPB5/B6/B7). Samples were run twice for each selected primer, in order to check the reproducibility of the method.

Data analyses---Only the well-amplified products on agarose gel were coded in a binary matrix of presence / absence characters. The RAPD distance measures of all pairs of individuals were then calculated using UPGMA and Neighbor Joining (NJ) methods, implanted in PAUP* 4.0b6 (Swofford, 1999) and CLUSTER* (http://www.biology.ualberta.ca/jbrzusto) programs. Moreover, branch support was estimated by bootstrapping

method (Felsenstein, 1985), using 100 replicates, simple sequence addition, TBR branch swapping, and 75 resampling of the original matrix. Under the distance option implanted in PAUP*, the Euclidean metric was performed, $E = n[1 - (n_{xy} / n)]$, where n_{xy} is the number of bands shared by 2 individuals, n the number of polymorphic bands (Excoffier *et al.*, 1992). CLUSTER* was used to compute the Jaccard's coefficient, $J = (n_{xy} / n - m_{xy})$, which explicitly excludes m_{xy}, the number of shared absences (Jaccard, 1908), and to repeat analysis of the Euclidean distances. Indeed, it has been shown that different computer packages may produce different trees with the same algorithm (Backeljau *et al.*, 1995). Therefore, similar results obtained from PAUP* and CLUSTER* programs would provide additional support for the inferred phenograms.

In the light of the respective advantages and limitations of the RAPD technique, our results will be based on the following assumptions: (1) each RAPD marker comprises a single dominant locus with alleles being amplified (present) or not (absent); (2) the majority of the RAPD markers are randomly distributed throughout the genome (despite the frequent GC rich content of the primers); (3) amplified products of the same size represent homologous fragments.

RESULTS

In the 50 populations investigated, the 17 primers used generated a total of 74 polymorphic and one monomorphic RAPD fragments. The 75 RAPD markers were analyzed by the UPGMA clustering method, using both the Euclidean distances and the Jaccard's coefficient. UPGMA phenograms generated by PAUP* and CLUSTER* packages, for the Euclidean distance matrix, were identical in topology (Fig. 1). On the other hand, the UPGMA clustering method performed by CLUSTER*, and using the Jaccard's coefficient, generated a dendrogram whose global topology differ slightly from the Euclidean one (Fig. 2).

The UPGMA tree based on Euclidean distances between all pairwise comparisons of the 50 populations and calculated from the 75 markers, depicted 3 clusters (Fig. 1). A basal group (set A) contained all the Mediterranean accessions of *C. tenuiflorum*. A second assemblage comprised European populations of *C. erythraea*, along with *C. rumelicum* and *C. rhodense* (set B). Finally, all the populations of *C. muehlenbergii* formed a third group (set C) with naturalized populations of *C. erythraea* (California and Australia) and those of *C. tenuiflorum* (Australia).

The UPGMA dendrogram obtained with Jaccard's coefficient supported the "*C. muehlenbergii*" third clade. Nevertheless, *C. erythraea* and *C. tenuiflorum* form polyphyletic groups in this reconstruction (Fig. 2).

In both the approaches, naturalized populations of either *C. erythraea* or *C. tenuiflorum* do not belonged to the same cluster than the European native populations, except for one Australian population of *C. erythraea* (eryau14).

In the UPGMA bootstrap analysis performed with PAUP*, only some clade received phylogenetic support, indicating that the inferences based on a strictly bifurcating may be problematic for evaluating phylogenetic relationships within the *C. muehlenbergii* complex. On the other hand, the groups depicted on the UPGMA phenograms, and based on Euclidean distances, received better supports than those obtained with the Jaccard coefficient.

Finally, when the respective trees were compared with Parsimony or ML scores, we found that the use of Jaccard's coefficient of similarity provide a less parsimonious tree than the Euclidean approach (L = 615 vs. L = 606).

DISCUSSION

Comparison of the respective topologies under different distance metrics---RAPD markers have proven to be useful tools for interspecific comparisons (review in Wolfe and Liston, 1998). Presently, these markers satisfactory discriminate between populations of closely related species, whose relationships could not have been inferred with ITS sequences (chapter 1 and 2). Nevertheless, some divergences occur in the current analysis, depending on the type of distance employed. Clustering methods gave slightly different trees based on either the Euclidean metrics or the Jaccard's coefficient of similarity. The former calculates the overall similarity of two taxa, including both shared presence and absence. The Jaccard's coefficient excludes the shared absence, an assumption that seems to be closer to the biological reality.

Nevertheless, groups delimited by the Euclidean distances appear to be more compatible with the results obtained with parsimonious analyses of DNA sequences (Chapter 2). In this study, the *C. tenuiflorum* clade appeared to be basal compared to the *C. erythraea* one, comprising *C. erythraea* subsp. *rumelicum* and *C. majus* subsp. *rhodense*. Phenograms based on Jaccard's distance tend to include the latter taxa in the Mediterranean group of *C. tenuiflorum* (Fig. 2). However, the *C. muehlenbergii* cluster is present in the respective phenograms but receives weak branch support (Bootstrap Value: BV < 50).

Clustering methods reconstruct trees without consideration of the underlying phylogenetic relationships, and thus bootstrapping method have been performed to provide a measure of the reliability of the different sets. Presently, the

clusters obtained with the Euclidean distances are better supported (BV > 50) than those deduced with the Jaccard's coefficient.

On the other hand, a quick lose of the utility of RAPD have been reported when the taxonomic level comparison increases (Vander Kloet and Paterson, 2000). In the present analysis, C. tenuiflorum accessions have relatively distinct RAPD profiles, supporting an increase of the genetic divergence with the remaining taxa. With such a distance augmentation between the respective populations, it becomes difficult to differentiate co-migrating non-homologous bands (Rieseberg, 1996). Notwithstanding, some authors have shown that errors due to non-homology of RAPD bands are random, and thus have little effects on the relative similarities (Adams and Rieseberg, 1998).

***Centaurium muehlenbergii* delimitation---**In the present analysis, all the populations of *C. muehlenbergii* collected in California and Texas cluster together in the respective UPGMA dendrograms. The current sampling, with 16 populations collected in California and Texas, clearly dos not cover the whole range of *C. muehlenbergii* distribution, but may be considered as representative of the taxon diversity.

On a morphological point of view, many forms have been collected, ranging from small plants with relatively showy flowers, to tall and erect individuals showing terminal crowded inflorescence, and intermediate ones with a more or less lax branching. Nevertheless, these respective phenotypes do not seem to correspond to particular clusters obtained with RAPD markers. The populations constituting the group M1 (Fig. 1 and 2), correspond to the "*C. tenuiflorum*" form, previously described. This set contains Texan populations collected near Houston, and first determinate by us as *C. tenuiflorum*, along with similar forms (i.e., erect plants with crowded-cymes) collected in California and in Australia. RAPDs data confirm the differences previously observed between the naturalized forms of *C. tenuiflorum* and the Mediterranean ones. Previous phylogenetic reconstruction, based on ITS sequence data, greatly supported an independent *C. tenuiflorum* clade, comprising diploid and tetraploids populations and occurring in the Mediterranean basin. This natural assemblage was well separated from a large "*C. erythraea* clade" (section *Centaurium*), containing original and introduced populations of *C. erythraea*, *C. muehlenbergii*, and Australian populations of *C. tenuiflorum* (Chapter 2). On the other hand, a second "*C. muehlenbergii*" group called M2 (Fig. 1 and 2), is present in both UPGMA cladograms obtained with Euclidean and Jaccard's coefficient. This set includes populations with larger inflorescence or smaller individuals with, proportionally larger flowers. Nevertheless, the erect "*C. tenuiflorum*" forms are sometimes presents.

Patterns of introgression and reticulate evolution in *Centaurium muehlenbergii*--A first statement, resulting from field experience, is that *C. tenuiflorum* forms present in Australia, and *C. muehlenbergii* of California and Texas, often occur on coastal ranges where *C. erythraea* has also been introduced. The current RAPD analysis confirms close genetic relationships between these taxa, for they always cluster together in the respective dendrograms. Moreover, different forms of *C. muehlenbergii*, "*C. tenuiflorum*" and *C. erythraea* (erect or lax plants) may be encountered in several regions of the Americas or Australia, and no particular phenotype seems to occur in a given area. This apparently lack of morphological and biogeographical partitioning of *C. muehlenbergii* populations with biparental inherited RAPD markers may be explained by several factors.

First, one can invoke the small sample size and subsequent sampling error. Because the whole range of the hitherto known populations of *C. muehlenbergii* has not been covered, notably the northern and central states of the US (Holmes, 1996), we may fail to detect some existing biogeographic or morphological patterns. Moreover, other populations of the "*C. tenuiflorum*" forms such as those present in eastern Australia (Adams, 1996) or probably in South America (Hunziker 3644 - K), need to be investigated.

A second explanation may be that *C. muehlenbergii* is of recent formation and has not yet stabilized the morphological variability of one or other putative parent. On all evidences, *C. muehlenbergii* derive from introduced populations of *C. erythraea* and the respective taxa are often not easily recognizable in shared geographic areas. In California, populations with intermediate characters, such as an important rosette of broad and elliptic leaves, large and erect stems (*C. erythraea*) along with more or less compact inflorescence of small flowers ("*C. tenuiflorum*"), frequently occur (e.g., population eryca760). The same case has recently been described in Australia (Chapter 2). "*Centaurium tenuiflorum*" forms, often sympatric with introduced populations of *C. erythraea*, can not be separated from "erect" *C. muehlenbergii*, using either classical markers (morphology, karyology) or molecular ones (DNA sequences, RAPDs markers). On the other hand, naturalized populations of *C. erythraea* can be distinguished from indigenous ones, on the base of nuclear DNA markers, suggesting reciprocal introgression with either *C. muehlenbergii* or naturalized *C. tenuiflorum*. The occurrence in Australia of possibly non-introgressed populations of *C. erythraea* (e. g. population eryaus14), supports this view.

A preliminary scenario advocated to explain the origin of *C. muehlenbergii* in the Americas (Chapter 1), might be partially confirmed or refined here. The hybrid origin of the so-called *C. muehlenbergii* - *C. tenuiflorum* forms, is confirmed, not only in the Americas but in Australia as well, and *C. erythraea* appears to be one of the genitors.

Further studies including South American or Insular populations of these respective taxa are fully needed in order to check the importance of such phenomena.

Nonetheless, European populations of *C. tenuiflorum*, the other putative parent, do not appear to introgress, in the light of nuclear markers (Chapter 1) or RAPDs. Moreover, most of morphological characters delimiting *C. tenuiflorum* (e.g., acute lobes, white eye in the center of the corolla, bigger stamens) do not appear in the naturalized hybrid populations investigated. The maternal origin of *C. tenuiflorum* for *C. muehlenbergii*, is not excluded by chloroplast DNA sequences, but the lack of resolution in the inferred phylogenetic analyses (chapter 2), does not permit further conclusions.

The apparent weak homology between the Mediterranean populations the *C. tenuiflorum* and the naturalized ones, generated by the same primer, could be interpreted as an increased genetic distances between these taxa (Rieseberg, 1996). Nevertheless, a more important sampling should be necessary to test this interpretation. Moreover, ancient episodes of hybridization invoking *C. erythraea* and "true" punctually introduced *C. tenuiflorum* populations, followed by reciprocal introgression and elimination of one parental genotype, here *C. tenuiflorum*, can not be excluded. The phylogeography of the whole group based on cpDNA haplotypes would be helpful to conclude.

The amplification pattern of parental nuclear markers in *C. muehlenbergii* could be compatible with criteria expected in an homoploid derived species, such as the absence of unique marker relative to the parents and the additivity of molecular fragments (Allan *et al.*, 1997).

Systematics implications for the C. muehlenbergii complex---If we admit that the present sampling covers the main diversity of the *C. muehlenbergii* group, one can draw the following conclusions.

First, this taxon is likely of hybrid origin, *C. erythraea* being one of the more likely parents. Reciprocal introgression occurring between parental and hybrids populations indicate a recent scenario, compatible with the great morphological variability, and the similarity in RAPDs patterns, encountered both in *C. muehlenbergii* and parental introduced *C. erythraea*.

Second, the general outlook of most of the *C. muehlenbergii* form indicates close relationships with the Eurasian *C. tenuiflorum*. Nevertheless, no populations of the latter, either diploid or tetraploid ones, have been detected in our North American investigations. Indeed, *C. muehlenbergii* seems to differ from the pure Mediterranean populations of *C. tenuiflorum*, by always having a *C. erythraea* signature such as broader leaves, non-acute corolla lobes or short stamens. Moreover, natural hybrids between *C. erythraea* and *C. tenuiflorum* are rare in the geographic areas where

both the taxa occur. *Centaurium litardieri* (*C. erythraea* x *C. tenuiflorum*) have been described in few localities of Corsica and Sardinia. This hybrid is hypothesized to have risen from crosses between tetraploid populations of *C. erythraea* and *C. tenuiflorum*, or alternatively, diploid *C. tenuiflorum* with unreduced gametes (Zeltner, 1970).

Third, from a taxonomic point of view, the synonymy between *C. muehlenbergii* and *C. tenuiflorum* is rejected here, both taxa having morphological and molecular differences. Moreover, the present investigation does not support the segregation between lax forms, often called *C. floribundum*, and the erect, crowed-cyme ones (herbarium samples: Mason. 8358-102 (G); Benson. 47 (K)). The RAPDs variations detected between *C. muehlenbergii* populations apparently refer to a more or less important introgression with *C. erythraea*. One might better considerate *C. muehlenbergii* as a recent species of origin hybrid, clearly not yet stabilized with respect of sorting and segregation of the parental markers (e.g., morphological or molecular ones).

Lastly, an important fact concerns the real definition of the *C. muehlenbergii* taxa according to the different authors. In other worlds, do we really have sampled C. *muehlenbergii* populations? Photograph examinations of the respective type of *Erythraea muehlenbergii* (Gila River, Le Porter. 00297590 - NY) and *E. floribundum* (Sacramento, Harweg. 00297569 - NY), led us to conclude that these two taxa differ morphologically. On the other hand, other exemplars with ramified and large stems, having few but relatively big flowers, have been often named *C. muehlenbergii*. Nevertheless, the mentioned taxa clearly vary with the populations sampled by us in the field.

As pointed by Pringle, "those who have seen the real *C. muehlenbergii* required another name for it" (Pringle, personal communication). There are clearly different conceptions of *C. muehlenbergii*. On one hand, *C. muehlenbergii* seems to refer to the erect and branched forms we have collected. On the other hand, a more compact form, with showy and pedicellate flowers, may correspond to one of the type forms mentioned above. Unfortunately, we never have found the latter taxa in the fields, and they appear to be infrequent in the Herbarium collections. So are *Centaurium davyi*, a striking similar taxon found in the same damp places, near the sea (Abrams, 1951), and to a lesser extant, coastal populations of *C. exaltatum* (see Chapter 1). The long pedicel of the flower characterizes all these taxa, and some populations, with more or less large corollas, are often confounded with *C. venustum*. Further molecular analyses of these different taxa would be helpful to discriminate between these related taxa, and particularly between *C. davyi* and the compact form of "*C. muehlenbergii*".

CONCLUSION

In the light of the present and preliminary study, the taxa of *C. muehlenbergii* should be refined. For the last century, the name has been widely cited to describe several different taxa and most of the confusion relied on the lack of clarity with the type of the species. The herbarium sheets investigated (but not the type) often reveal similar forms with different names *(C. muehlenbergii*, *C. tenuiflorum, C. floribundum*, *C. curvistamineum*, and to a lesser extant*, C. trichanthum*, *C. venustum* or *C. davyi*), and vice-versa.

RAPD analyses performed on the most frequent form known as *C. muehlenbergii*, and recently renamed *C. tenuiflorum* (Pringle, personal communication), reveal that all the American populations sampled, along with Australian ones, may have diverse degree of introgression patterns with introduced populations of *C. erythraea*. However, *C. muehlenbergii* appears to be morphologically different with *C. tenuiflorum* from the Mediterranean, and molecular phylogenies clearly separate these taxa. Therefore, the name *C. tenuiflorum* is not valid for the North American and Australian accessions, and must be changed.

Conversely, the binomial *C. muehlenbergii* may fit a likely native taxon of California. This alternative form of *C. muehlenbergii* represents a less erect taxon with pedicellate and larger flowers. The specific rank of this taxon and the particular differentiation with close relatives species, such as *C. davyi* or *C. exaltatum* (n = 20), must be checked before further systematics conclusions.

REFERENCES

ABRAMS, L. R. 1951. An illustrated flora of the Pacific States, Washington, Oregon and California. in *An illustrated flora of the Pacific States, Washington, Oregon and California.* 3. Stanford University Press. Stanford

ADAMS, L. G. 1996. *Gentianaceae*. in *Flora of Australia*, vol. 28 : 73-103. A. Wilson Editor. CSIRO. Melbourne.

ADAMS, R. P. and L. H. RIESEBERG. 1998. The effect of non-homology in RAPD bands of similarity and multivariate statistical ordination in *Brassica* and *Helianthus*. *Theoretical and Applied Genetics* 97 : 323-326.

ALLAN, G. J., C. CLARK, and L. H. RIESEBERG. 1997. Distribution of parental DNA markers in *Encelia virginensis* (Asteraceae: Heliantheae), a diploid species of putative hybrid origin. *Plant Systematic and Evolution* 205 : 205-221.

BACKELJAU, T., L. DE BRUYN, H. DE WOLF, K. JORDAENS, S. VAN DONGEN, and R. VERHAGEN. 1995. Random Amplified Polymorphic DNA (RAPD) and parsimony methods. *Cladistics* 11 : 119-130.

DOYLE, J. J. and J. L. DOYLE. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical bulletin* 19 : 11-15.

DUNN, R. A. 1967. A revision of the genus *Centaurium* of continental United States. Thesis. Catholic University of America. Washington D. C.

EXCOFFIER, L., P. E. SMOUSE, and J. M. QUATTRO. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 86 : 991-1000.

FELSENSTEIN, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 17 : 223-263.

GRAY, A. 1878. Gamopetalae after Compositae. in *Synoptical flora of North America*. 2. Ivison, Blakeman and Taylor. New York

GRISEBACH, A. H. R. 1839. *Genera et Species Gentianearum adjectis observationibus quibusdam phytogeographycis* : 116-125. J. G. Cotta. Stuttgart et Tübingen.

HARRIS, S. A. 1999. RAPDs in systematics - a useful methodology? in *Molecular systematics and plant evolution* : 211-228. P. M. Hollingsworth, R. M. Bateman, and R. J. Gornall Editors. Taylor & Francis. London

HOOKER, W. J. and G. A. W. Arnott. 1830-1841. *Erythraea*. in *The botany of Captain Beechey's voyage*. 8 The Secretary of State for Colonial Affairs. London

JACCARD, P. 1908. Nouvelles recherches sur la distribution florale. *Bulletin de la Société Vaudense des Sciences Naturelles* 44 : 223-270.

MELDERIS, A. 1931. Genetical and Taxonomical studies in the genus *Erythraea* Rich. *Acta Horti. Bot. Univ. Latv.* 6 (123-256).

MUNZ, P. A. and D. D. KECK. 1959. *A California Flora*. University of California Press. Beckerley

PIPER, C. V. 1906. *Flora of the State of Washington*. Contributions from the United States National Herbarium 11

RIESEBERG, L. H. 1996. Homology among RAPD fragments in interspecific comparisons. *Molecular Ecology* 5 : 99-105.

SWOFFORD, D. L. 1999. *PAUP* Phylogenetic Analysis Using Parsimony (*and other methods) Version 4*. Sunderland, Sinauer Associates.

VANDER KLOET, S. P. and I. G. PATERSON. 2000. RAPD assessment of novelties resulting in a new species of *Vaccinium* L. (Ericaceae) from Vietnam. *Botanical Journal of the Linnean Society* 134 : 575-586.

WOLFE, A. D. and A. LISTON. 1998. Contribution of PCR based-methods to plant systematics and evolutionary biology. in *Molecular systematics of plants 2. DNA sequencing*. D. E. Soltis, P. S. Soltis and J. J. Doyle Editors. Kluwer Academic. Boston

WOLFF, K. and M. MORGAN-RICHARDS. 1999. The use of RAPD data in the analysis of population genetic structure: case of *Alkanna* (Boraginaceae) and *Plantago* (Plantaginaceae). in *Molecular systematics and plant evolution*: 51-73. P. M. Hollingsworth, R. Bateman, M. and R. J. Gornall Editors. Taylor & Francis. London

ZELTNER, L. 1970. Recherche de biosystematique sur les genres *Blackstonia* Huds. et C*entaurium* Hill (Gentianaceae). *Bulletin de la Société Neuchâteloise des Sciences Naturelles* 93 : 1-164.

Table 1. Different populations of *Centaurium* sampled for Random Amplified polymorphic DNA analyses. RAPD numbers refer to one or more individuals sampled in each population. Collection indicates either Herbarium accession or collects number. n and 2n are chromosome numbers reported in the respective populations.

Species	RAPD number	Collection	DNA number	n	2n	Location
C. ery. rumelicum	ru19	980713	MG19			Roumania
C. ery. rumelicum	ru43	LZ2385	MG43			Turkey
C. erythraea	ery42	981002-1	MG990242			Poulx, France - 43.55 N / 04.25 E
C. erythraea	ery43	980926-4	MG990243			Espiguette, France - 43.34 N / 04.10 E
C. erythraea	erypo1	PO1	MGpo1			Poulx, France - 43.55 N / 04.25 E
*C. erythraea**	eryaus11	LZ	MG001211			Australia -
*C. erythraea**	eryaus13	LZ	MG001213			Australia -
*C. erythraea**	eryaus14	LZ	MG001214			Australia -
*C. erythraea**	eryaus15	LZ	MG001215			Australia -
*C. erythraea**	eryaus16	LZ	MG001216			Australia -
*C. erythraea**	eryaus17	LZ	MG001217			Australia -
*C. erythraea**	eryaus20	LZ	MG001220			Australia -
*C. erythraea**	eryaus21	LZ	MG001221			Australia -
*C. erythraea**	eryca760	970627-1a	MG97760	20	40	California - 40.52.90 N / 123.32.89 W
*C. erythraea**	eryca760*	970627-1b	MG97760	20	40	California - 40.52.90 N / 123.32.89 W
C. majus rhodense	rh42	LZ2384	MG42			
C. majus rhodense	rh1024	NS0005	MG001024?			Sicily, Italy
C. majus rhodense	rh7	990621	MG010211		20	Spain – 42.07.21 N / 02.58.34 W
C. muehlenbergii	mueh513	960613-1	MG97513		40	California - 36.07.14 N / 118.49.55 W
C. muehlenbergii	mueh703	970603-3	MG97703	20	40	Texas - 29.56.49 N / 95.20.80 W
C. muehlenbergii	mueh761	970627-2	MG97761	20		California - 40.55.94 N / 123.37.51 W
C. muehlenbergii	mueh773	970630-1	MG97773	20	40	California - 37.16.36 N / 122.12.86 W
C. muehlenbergii	mueh757	970625-1	MG97757	20	40	California - 99.13.52 N / 121.25.81 W
C. muehlenbergii	mueh3	960607-2	MG3	20	40	California - 35.56.02 N / 122.41.72 W
C. muehlenbergii	mueh506	960609-1	MG97506	20	40	California - 38.26.49 N / 120.51.32 W
C. muehlenbergii	mueh705	970603-5	MG97705	20	40	Texas - 29.49.53 N / 99.34.51 W
C. muehlenbergii	mueh758	970626-1	MG97758	20	40	California - 39.37.26 N / 121.25.37 W
C. muehlenbergii	mueh759	970626-2	MG97759	20	40	California - 39.42.76 N / 121.39.78 W
C. muehlenbergii	mueh762	970627-3	MG97762	20	40	California - 40.56.98 N / 123.37.12 W
C. muehlenbergii	mueh763	970628-1	MG97763		40	California - 40.29.39 N / 123.58.47 W
C. muehlenbergii	mueh764	970628-2	MG97764			California
C. muehlenbergii	mueh766	970628-4	MG97766		40	California - 40.13.27 N / 123.47.68 W
C. muehlenbergii	mueh771	970629-4	MG97771	20		California - 38.27.13 N / 122.21.68 W
C. muehlenbergii	mueh772	970629-5	MG97772	20	40	California - 38.02.86 N / 122.37.34 W
C. tenuiflorum	tenma1	990620-1b1	MGma1			Majorc, Spain
C. tenuiflorum	tenma2	990620-1b2	MGma2			Majorc, Spain
C. tenuiflorum	tenma7	990619-2a	MGma7			Majorc, Spain
C. tenuiflorum	tensi1	501	MG000501			Sicily, Italy
C. tenuiflorum	tensi2	000502	MG000502			Sicily, Italy
C. tenuiflorum	tensi4	000504	MG000504			Sicily, Italy
C. tenuiflorum	tensi5	000505	MG000505			Sicily, Italy
C. tenuiflorum	tensi6	000506	MG000506	20	40	Sicily, Italy
C. tenuiflorum	tensi7	000507	MG000507			Sicily, Italy
C. tenuiflorum	tensi9	000509	MG000509			Sicily, Italy
C. tenuiflorum	tensi10	000510	MG000510			Sicily, Italy
C. tenuiflorum	tensi11	000511	MG000511			Sicily, Italy
C. tenuiflorum	tenar1	?	MGar1			France - Aresquiers
*C. tenuiflorum**	tenaus4*	001204	MG001204			Australia
*C. tenuiflorum**	tenaus12*	LZ2586	MG001212			Australia - 30.03.643 S / 115.12.024 E
C. muehlenbergii	mueh767	970628-5	MG97767			California - 40.06.99 N / 123.47.98 W

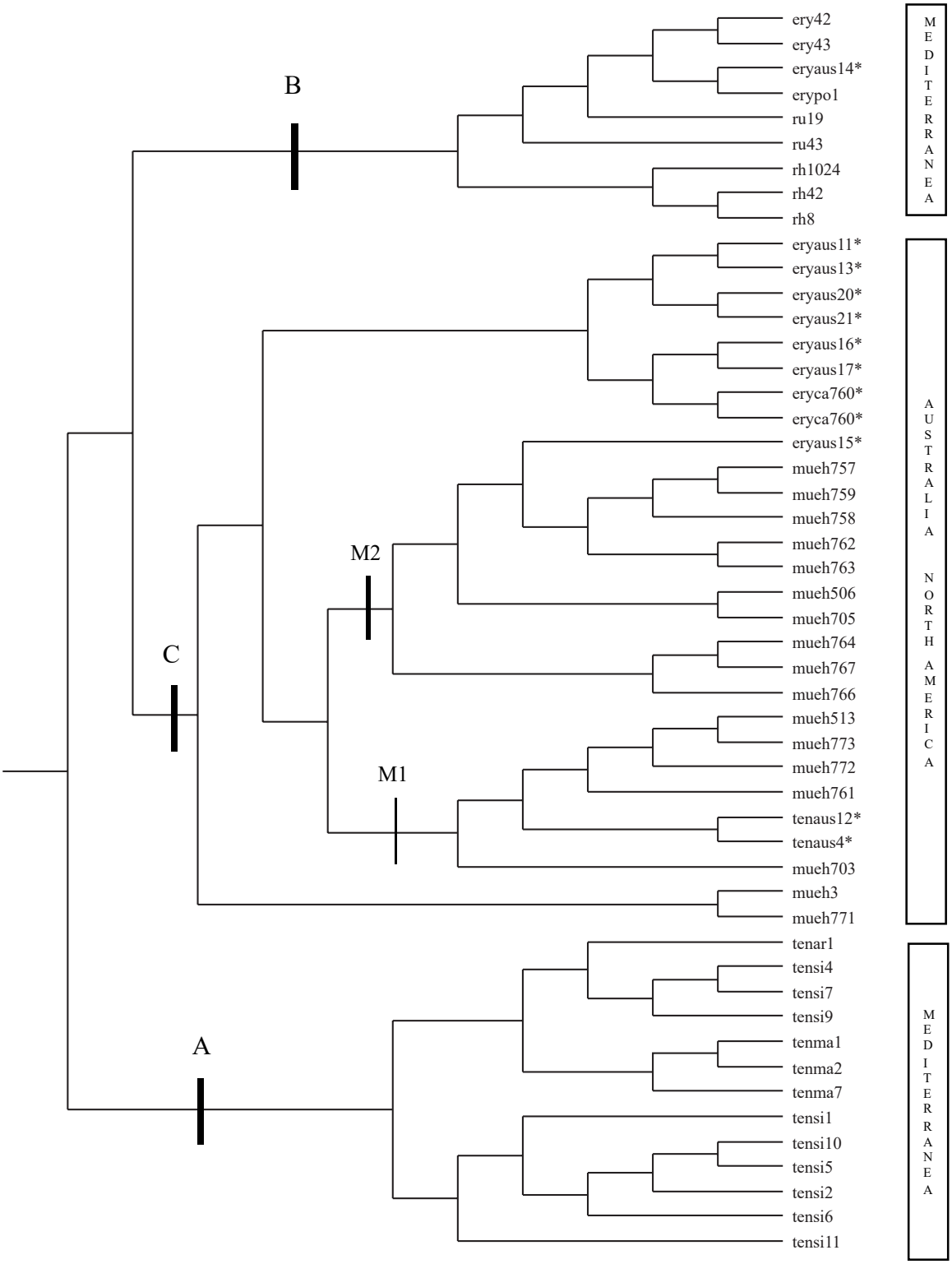


Figure 1. UPGMA phenogram for 50 populations of *Centaurium* based on Euclidean distances. Clusters A, B, C, M1 and M2 are described in the text. Asterisks signal naturalized populations.

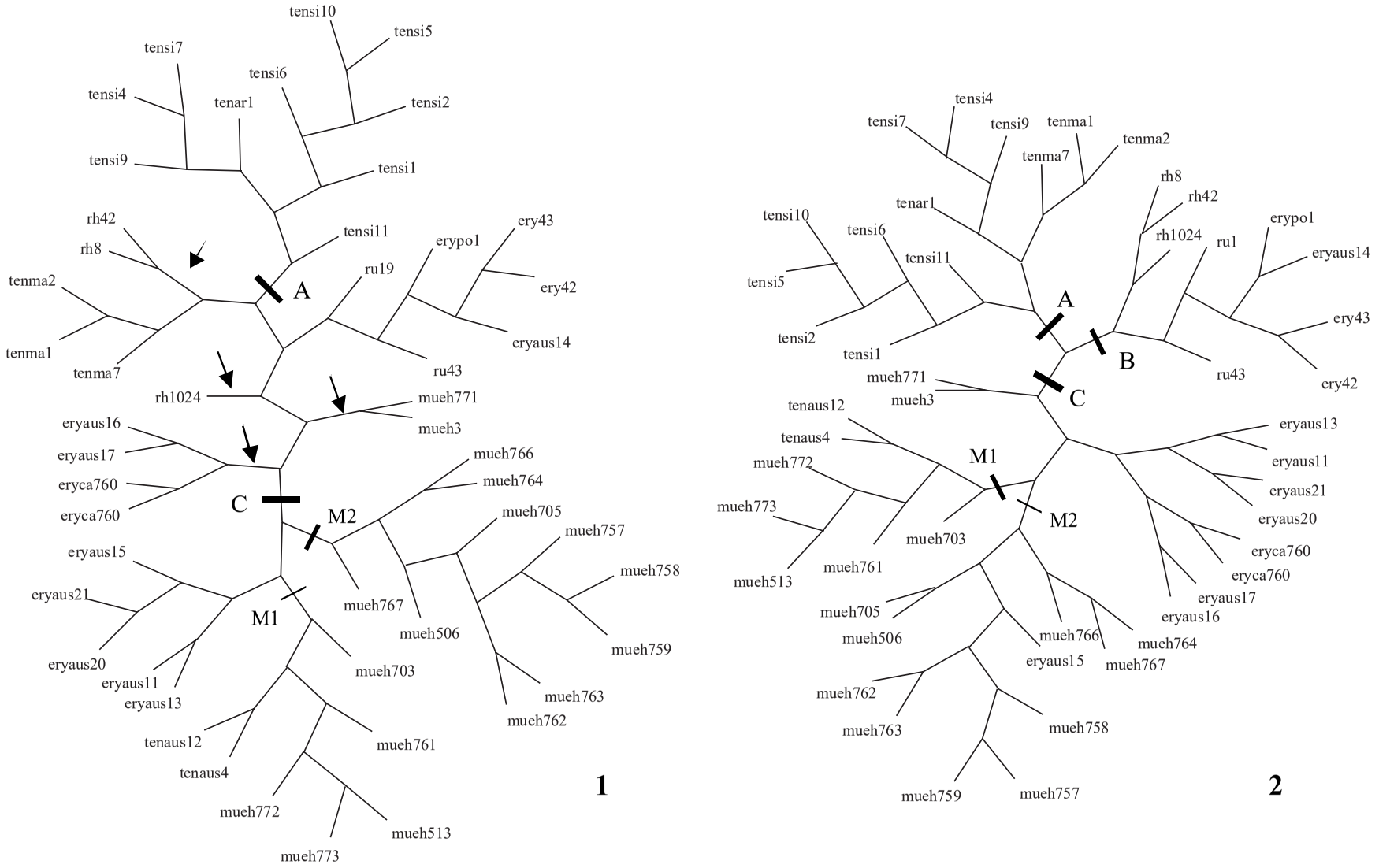


Figure 2. Unrooted UPGMA trees for 50 populations of *Centaurium* based on Jaccard (1) or Euclidean (2) distances. Arrows indicate topological divergence between the respective topologies. Letters are groups defined in the text.

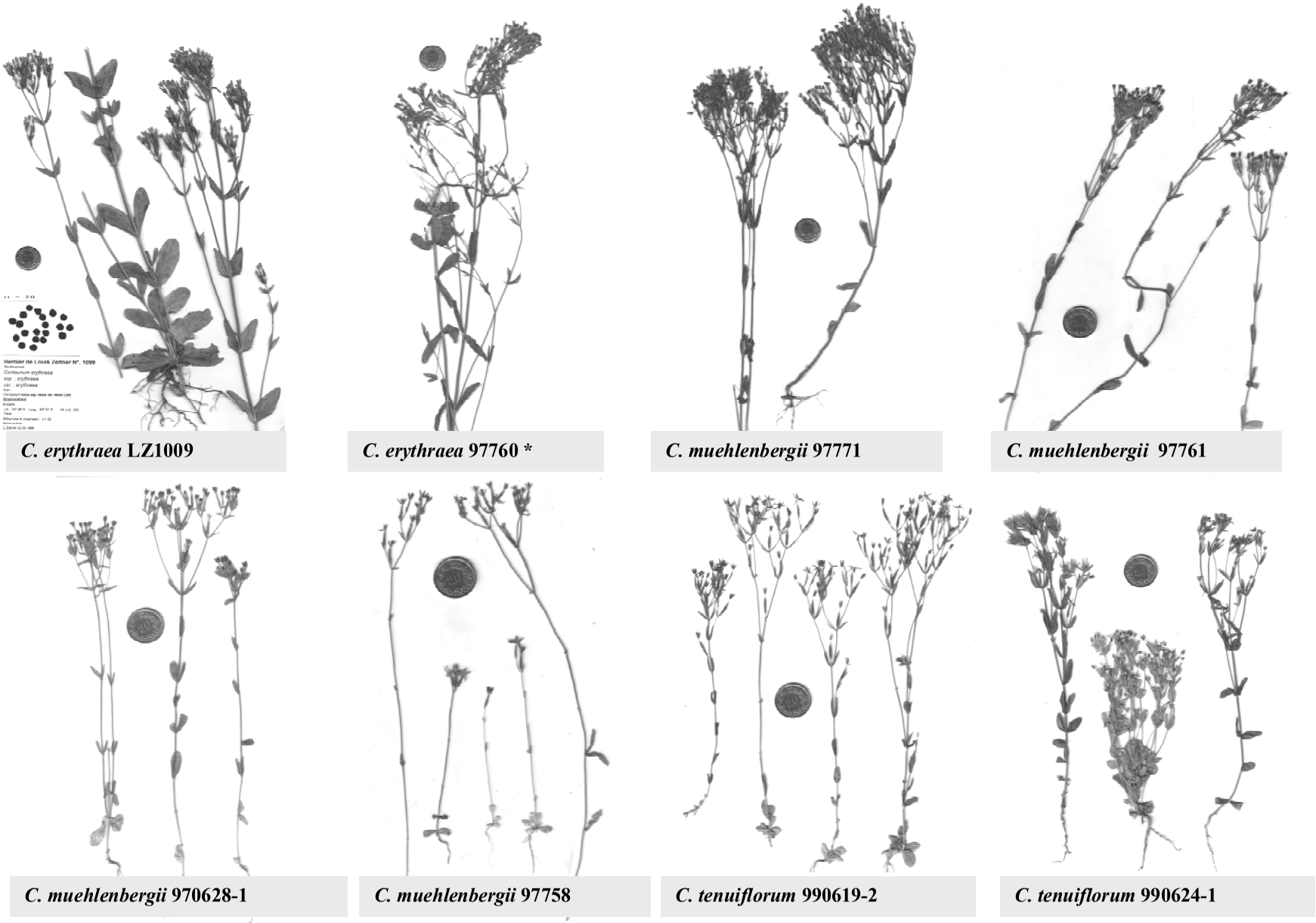


Figure 3. Different forms of *C. erythraea*, *C. muehlenbergii* and *C. tenuiflorum* investigated in the present study. Asterisk represents naturalized population of *C. erythraea.* The coin represents two centimeters.

Annexes

Annex 1: List of *Centaurium* species investigated in the present thesis and number of DNA sequences obtained for the different primers. (*) Taxonomic combination used in the present dissertation (**) synonymy found in the literature (A96: Adams, 1996; B73/77/81: Broome, 1973, 1977, 1981; M72: Melderis, 1972; R74: Reveal *et al.*, 1974; T93: Turner, 1993; Z70/78/85/91: Zeltner 1970, 1978, 1985, 1991; C1, C2: unusual or unpublished combinations from Chapter 1 and 2).

SPECIES (*)	**SYNONYM (**)**	**ITS1 - ITS2**	***trn*L intron**	***trn*LF spacer**	**5S-NTS**
Centaurium abramsii (Munz) Mansion (C1)	*Centaurium venustum* (Gray) Robinson var. *abramsii* Munz (B73)	1	1	1	1
Centaurium arizonicum (A. Gray) Heller (T93)	*Centaurium calycosum* (Buckl.) Fern. var. *arizonicum* (Gray) Tidestrom (B73)	2	2	1	1
Centaurium australe Druce (C2)		4	3	3	1
Centaurium barrelieri (Duf.) F.Q. et Rothm. (Z70)	*Centaurium linariifolium* (Larn.) G. Beck (M72)	2	2	1	
Centaurium bernardii (Maire et Sauvage) Mansion (C2)	*Centaurium minus* subsp. *bernardii* (Maire et Sauvage) Zeltner (Z70)	0	1	1	0
Centaurium beyrichii (Torrey & Gray) B. L. Robinson (B73)		2	2	1	1
Centaurium bianoris (Sennen) Sennen (Z78)	*Centaurium bianoris* (Sennen) Sennen var. *bianoris* (Z78)	1	4	4	1
Centaurium brachycalyx Standley & L. O. Williams (B73)		2	1	2	1
Centaurium breviflorum (Shinners) B. L. Turner (T93)	*Centaurium calycosum* (Buckl.) Fernald var. *breviflorum* Shinners (B73)	1	1	1	0
Centaurium calycosum (Buckl.) Fernald (T93)	*Centaurium calycosum* (Buckl.) Fernald var *calycosum* / var *nanum* (Gray) Robinson (B73)	5	2	4	1
Centaurium capense Broome (B77)		1	1	1	1
Centaurium centaurioides (Roxb.) Rolla Rao and Hemadri (C2)	*Centaurium centaurioides* (Roxb.) Rolla Rao and Hemadri subsp. *centaurioides* (Z91)	2	1	1	1
Centaurium chloodes (Brot.) Samp. (Z70)		1	1	1	1
Centaurium clementii (Domin) L.G. Adams (A96)		1	1	1	0
Centaurium erythraea Rafn (M72)	*Centaurium erythraea* Rafn subsp. *erythraea* (M72)	7	3	3	0
Centaurium exaltatum (Griseb.) W. F.Wight ex Piper (B73)		3	1	1	0
Centaurium favargeri Zeltner (L70)		2	2	2	0
Centaurium glanduliferum (Correll) B. L. Turner (T93)	*Centaurium beyrichii* (Torrey & Gray) B. L. Robinson var. *glanduliferum* Correll (B73)	1	1	1	1

Annex 1 (continued):

Centaurium gypsicola (Boiss. et Reut.) Ronniger (L70)	*Centaurium triphyllum* (W. L. E. Schmidt) Melderis (M72)	1	1	1	1
Centaurium littorale (D. Turner) Gilmour (M72)	*Centaurium vulgare* Rafn. (Z70)	1	1	1	1
Centaurium madrense (Hemsl.) Robinson (B73)		1	1	1	1
Centaurium mairei Zeltner (Z91)		1	1	1	0
Centaurium majus (Hoffmans. & Link) Ronn. subsp. *majus* (Z70)	*Centaurium erythraea* Rafn subsp. *majus* (Hoffmanns. & Link) Melderis (M72)	3	2	2	2
Centaurium majus (Hoffmans. & Link) Ronn. subsp. *rhodense* (Boiss. & Reut.) Zeltner (Z70)	*Centaurium erythraea* Rafn subsp. *rhodense* (Boiss. & Reut.) Melderis (M72)	1	1	1	0
Centaurium malzacianum Maire (C2)	*Centaurium centaurioides* (Roxb.) Rolla Rao and Hemadri subsp. *malzacianum* (Maire) Zeltner (Z91)	6	4	3	3
Centaurium maritimum (L) Fritch ap. Janchen (Z70)		2	2	2	1
Centaurium martinii Broome (B73)	*Centaurium phenax* R. F. Martin ex Broome (B73)	3	3	3	2
Centaurium maryannum B. L. Turner (T93)		2	1	1	1
Centaurium muehlenbergii (Griseb.) W. Wight ex Piper (B73)	*Centaurium floribundum* (Benth.) B. L. Robinson (B73)	4	3	4	1
Centaurium multicaule B. L. Robinson (B73)		3	2	2	1
Centaurium namophilum Reveal, Broome & Beatley var. *namophilum* (R74)		1	1	1	0
Centaurium namophilum Reveal, Broome & Beatley var. *nevadense* Broome (B81)		2	2	2	0
Centaurium nudicaule (Engelm.) B. L. Robinson (B73)		2	1	1	0
Centaurium pulchellum (Sw.) Druce (Z70)	*Centaurium pulchellum (Sw.) Druce* var. *pulchellum* (Z70)	3	3	2	3
Centaurium pulchellum (Sw.) Druce var. *altaicum* (Griseb.) Kitag. et Hara (Z85)		1	1	?	0
Centaurium pusillum Eastw. (B73)		2	1	1	1
Centaurium quitense (Kunth.) B. L. Robinson (B73)		14	8	7	4
Centaurium rumelicum (Velen.) Mansion (C2)	*Centaurium erythraea* Rafn subsp. *rumelicum* (Velen.) Melderis (M72)	2	2	2	1
Centaurium scilloides (L. fil.) Samp. (Z70)		2	2	1	0

Annex 1 (continued):

Centaurium serpentinicola A. Carlström (Z85)		1	1	1	1
Centaurium setaceum (Benth.) Robinson (B73)		3	3	3	1
Centaurium somedanum Lainz (C2)		1	1	1	1
Centaurium sp.nov1 (C1)		5	3	3	1
Centaurium sp.nov2 (C1)		1	1	1	1
Centaurium spicatum (L) Fritch ap. Janchen (Z70)		2	2	2	2
Centaurium strictum (Schiede) Druce (B73)		7	5	6	1
Centaurium subcapitatum (C2)	*Centaurium erythraea* Rafn subsp. *erythraea* var. *capitatum* (Willd.) Melderis (M72)	2	2	2	1
Centaurium suffruticosum (Griseb.) Ronniger (M72)	*Centaurium majus* subsp. *majus var. suffruticosum* (Salzmann) Zeltner (Z70)	0	1	1	0
Centaurium tenuiflorum (Hoffmanns. & Link) Fritch subsp. *acutiflorum* (Schott) Zeltner (Z70)		4	2	2	1
Centaurium tenuiflorum (Hoffmanns. & Link) Fritch subsp. *tenuiflorum* (Z70)					
Centaurium tenuifolium (Mart. & Gal.) Robinson (B73)		1	1	1	1
Centaurium texense (Griseb. ex Hook.) Fern. (B73)		2	1	1	1
Centaurium trichanthum (Griseb.) B. L. Robinson (B73)		2	1	2	0
Centaurium turcicum (Velen.) Ronniger (C2)	*Centaurium erythraea* Rafn subsp. *turcicum* (Velen.) Melderis (M72)	3	3	3	1
Centaurium uliginosum (Waldst. & Kit.) G. Beck ex Ronniger (C2)	*Centaurium littorale* (D. Turner) Gilmour subsp. *uliginosum* (Waldst. & Kit.) Melderis (M72)	1	2	2	1
Centaurium venustum (Gray) B. L. Robinson (B73)	*Centaurium venustum* (Gray) B. L. Robinson subsp. *venustum* (B73)	3	3	3	1
Centaurium wigginsii Broome (B77)		2	2	2	1

Annex 2. Preliminary systematic treatment of the genus *Centaurium* sensu lato based on molecular and karyological results. * Name adopted in the present dissertation / ** Genera name to be proposed / # new combination (provisional proposal)

*** "Eurasian *Centaurium*" ***
**** *Centaurium* Hill ****

Section *Parviflora* Ronniger

-*Centaurium tenuiflorum* (Hoffmanns. & Link) Fritch subsp. *acutiflorum* (Schott) Zeltner
-*Centaurium tenuiflorum* (Hoffmanns. & Link) Fritch subsp. *tenuiflorum*
-*Centaurium serpentinicola* A. Carlström

Section *Pulchella* Mansion #

-*Centaurium bianoris* (Sennen) Sennen
-*Centaurium capense* Broome
-*Centaurium centaurioides* (Roxb.) Rolla Rao and Hemadri
-*Centaurium mairei* Zeltner
-*Centaurium malzacianum* Maire
-*Centaurium maritimum* (L) Fritch ap. Janchen
-*Centaurium pulchellum* (Sw.) Druce (including var. *altaicum*)

Section *Littorale* Mansion #

Subsection *Littorale* Mansion #

-*Centaurium chloodes* (Brot.) Samp.
-*Centaurium littorale* (D. Turner) Gilmour
-*Centaurium scilloides* (L. fil.) Samp.
-*Centaurium somedanum* Lainz

Subsection *Erythraea* Mansion #

-*Centaurium bernardii* (Maire et Sauvage) Mansion #
-*Centaurium erythraea* Rafn
-*Centaurium erythraea* subsp. *subcapitatum* (Corb.) Zeltner
-*Centaurium majus* (Hoffmans. & Link) Ronniger subsp. *majus*
-*Centaurium majus* (Hoffmans. & Link) Ronniger subsp. *rhodense* (Boiss. & Reut.) Zeltner
-*Centaurium muehlenbergii* (Griseb.) W. Wight ex Piper (cf. Chapter 5)
-*Centaurium rumelicum* (Velen.) Mansion #
-*Centaurium turcicum* (Velen.) Ronniger

Subsection *Favergera* Mansion #

-*Centaurium barrelieri* (Duf.) F.Q. et Rothm.
-*Centaurium favargeri* Zeltner
-*Centaurium gypsicola* (Boiss. et Reut.) Ronniger

*** Section *Spicaria* Griseb. ***
**** Genus *Spicarium* Mansion # ****

-*Centaurium australe* Druce / *Spicarium australe* (Druce) Mansion #
-*Centaurium clementii* (Domin) L.G. Adams / *Spicarium clementii* (Domin.) Mansion #
-*Centaurium sebaeoides* (Griseb.) Druce / *Spicarium sebaeoides* (Griseb.) Mansion #
-*Centaurium spicatu*m (L) Fritch / *Spicarium mediterranense* Mansion #

*** Section *Gyrandra* Gray ***
**** Genus *Gyrandra* Griseb. ****

-*Centaurium brachycalyx* Standley & L. O. Williams / *Gyrandra brachycalyx* (Standley & L. O. Williams) Mansion #
-*Centaurium chironioides* (Griseb.) Druce / *Gyrandra chironioides* Griseb.
-*Centaurium pauciflorum* (Martens & Galcotti) B.L. Robinson / *Gyrandra pauciflora* (Martens & Galcotti) Mansion #
-*Centaurium tenuifolium* (Martens & Galcotti) B.L. Robinson / *Gyrandra speciosa* Bentham

N.B. This genus may be included in *Sabatia* (cf. Chapter3)

*** "North American *Centaurium*" ***
**** Genus *Zeltnera* Mansion # ****

Section *Californica* Mansion #

-*Centaurium abramsii* (Munz.) Mansion / *Zeltnera abramsii* Mansion #
-*Centaurium exaltatum* (Griseb.) W. F.Wight ex Piper / *Zeltnera exaltata* (Griseb.) Mansion #
-*Centaurium namophilum* Reveal, Broome & Beatley var. *namophilum* / # *Zeltnera namophila* (Reveal, Broome & Beatley) Mansion var. *namophila* #
-*Centaurium namophilum* Reveal, Broome & Beatley var. *nevadense* Broome / *Zeltnera namophila* (Reveal, Broome & Beatley) Mansion var. *nevadense* Broome #
-*Centaurium trichanthum* (Griseb.) B. L. Robinson / *Zeltnera trichantha* (Griseb.) Mansion #
-*Centaurium sp.nov1* / *Zeltnera* ******** Mansion # (close to *Z. exaltata*)

N.B. Some species such as *C. davyi* or *C. muehlenbergii* s. s. (cf. Chapter 5), not included in the present phylogenetic investigation, may fit this section.

Section *Texana* Mansion #

-*Centaurium arizonicum* (A. Gray) Heller / *Zeltnera arizonica* (A. Gray) Mansion #
-*Centaurium beyrichii* (Torrey & Gray) B. L. Robinson / *Zeltnera beyrichii* (Torrey & Gray) Mansion #
-*Centaurium breviflorum* (Shinners) B. L. Turner / *Zeltnera breviflora* (Shinners) Mansion #
-*Centaurium calycosum* (Buckl.) Fernald / *Zeltnera calycosa* (Buckl.) Mansion #
-*Centaurium glanduliferum* (Correll) B. L. Turner / *Zeltnera glandulifera* (Correl) Mansion #
-*Centaurium maryannum* B. L. Turner / *Zeltnera maryanna* (B. L. Turner) Mansion #
-*Centaurium multicaule* B. L. Robinson / *Zeltnera multicaule* (B. L. Robinson) Mansion #
-*Centaurium texense* (Griseb. ex Hook.) Fernald / *Zeltnera texense* (Griseb. ex Hook) Mansion #

Section *Mexicana* Mansion #

-*Centaurium madrense* (Hemsl.) B. L. Robinson / *Zeltnera madrense* (Hemsl.) Mansion #
-*Centaurium martinii* Broome / *Zeltnera martinii* (Broome) Mansion #
-*Centaurium nudicaule* (Engelm.) B. L. Robinson / *Zeltnera nudicaule* (Engelm.) Mansion #
-*Centaurium pusillum* Eastwood / *Zeltnera pusilla* (Eastwood) Mansion #
-*Centaurium quitense* (Kunth.) B. L. Robinson / *Zeltnera quitense* (Kunth.) Mansion #
-*Centaurium setaceum* (Benth.) Robinson / *Zeltnera setacea* (Benth.) Mansion #
-*Centaurium sp.nov2* / *Zeltnera* ***** Mansion #
-*Centaurium strictum* (Schiede) Druce / *Zeltnera stricta* (Schielde) Mansion #
-*Centaurium wigginsii* Broome / *Zeltnera wigginsii* (Broome) Mansion #

N.B. The new genus *Zeltnera* is dedicated to Louis and Nicole Zeltner, who have greatly contributed to the systematics of the genus *Centaurium*.

**** Unclassified species ****

-*Centaurium cachanlahuen* (Molina) B.L. Robinson
Most of the specimens seen were misidentified collections of *C. erythraea*, a taxon introduced in South America. No conclusions can be drawn without field collection of this taxon and examination of the type.

-*Centaurium davyi* (Jepson) Abrams.
This species may be close to *C. exaltatum*, *C. sp.nov.1* and *C. muehlenbergii* sensu Pringle (Chapter 5). Closer investigations within this group are needed

-*Centaurium gentryi* Broome
According to R.C. Broome, this species may fit the genus *Gyrandra*.

-*Centaurium lomae* (Gilg) Fabris
Maybe a form of *C. pulchellum* introduced in South America?

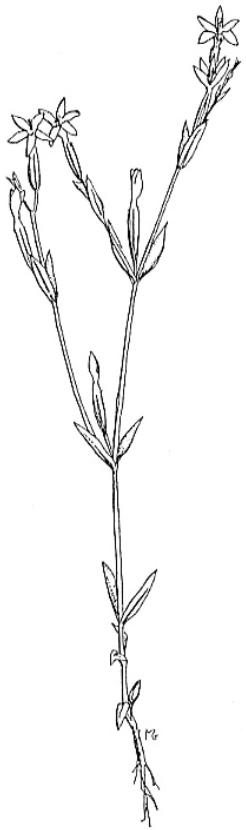
-*Centaurium pterocaule* Broome
This species is known from only one population. According to R. C. Broome, it may fit the genus *Gyrandra*.

ILLUSTRATIONS

"Eurasian *Centaurium*"

(Genus *Centaurium* Hill)

C. maritimum



C. bianoris

C. malzacianum

C. pulchellum

2 cm

"North American *Centaurium*"

(Genus *Zeltnera* Mansion)

C. nudicaule

C. sp nov2

C. madrense

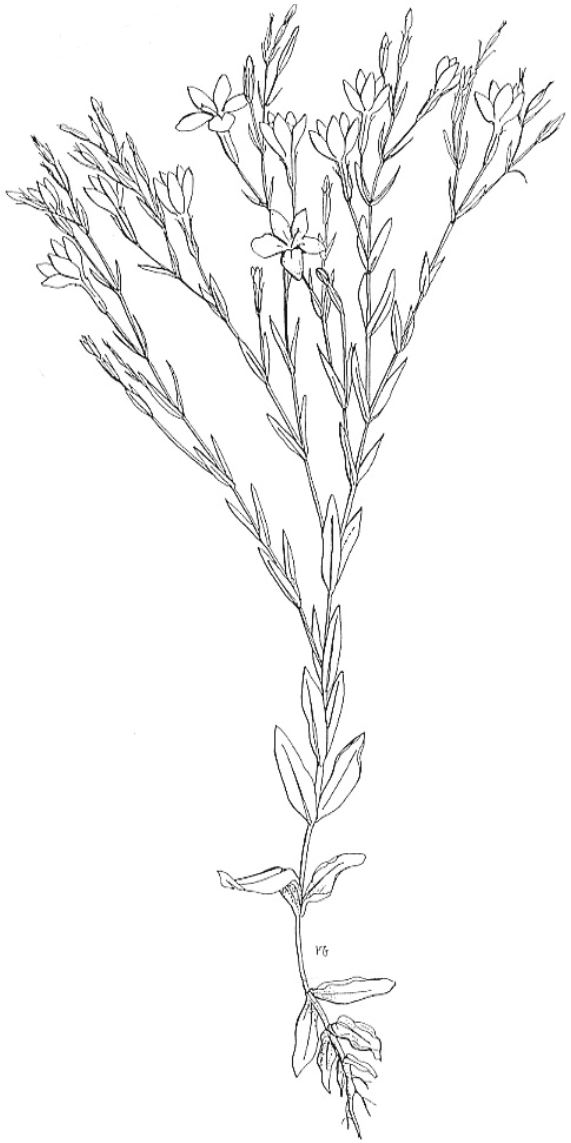
C. martinii

C. wigginsii

2 cm

"North American *Centaurium*"

(Genus *Zeltnera* Mansion)



C. arizonicum

C. texense

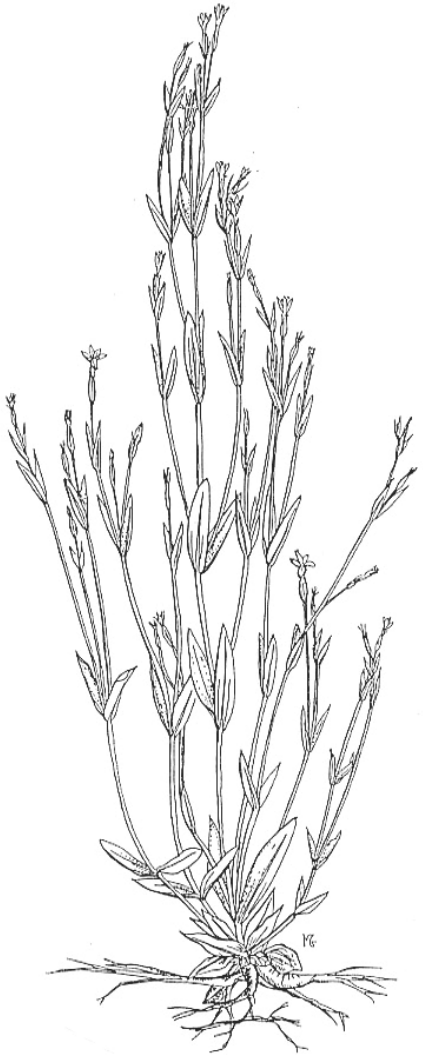
C. breviflorum

2 cm

"North American *Centaurium*"

(Genus *Zeltnera* Mansion)

C. venustum



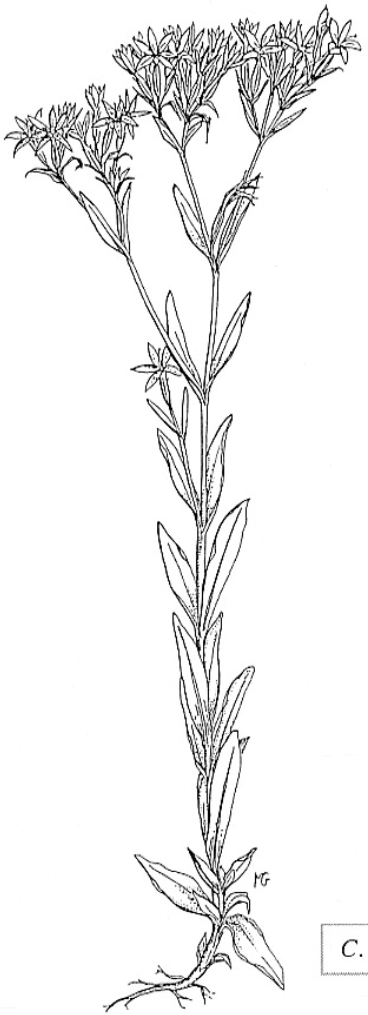
C. exaltatum

C. sp nov 1

2 cm

"North American *Centaurium*"

(Genus *Zeltnera* Mansion)



C. trichanthum

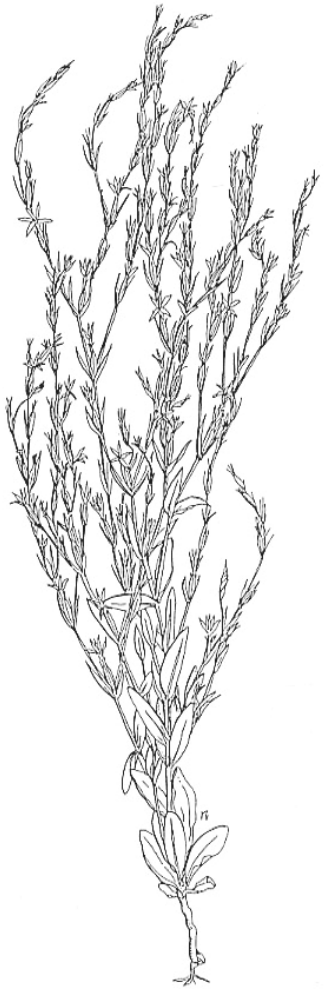
C. namophilum

C. abramsii

2 cm

"Section Spicaria Griseb."

(Genus *Spicarium* Mansion)



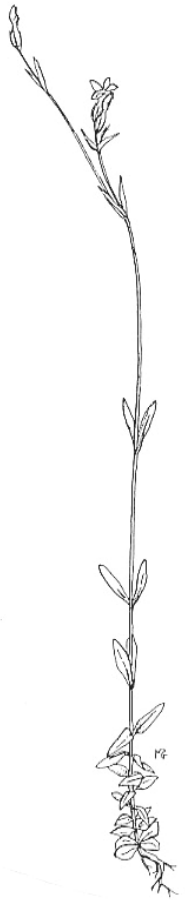
C. spicatum

C. clementii

"Section *Gyrandra* Gray"

(Genus *Gyrandra* Griseb.)

C. tenuifolium

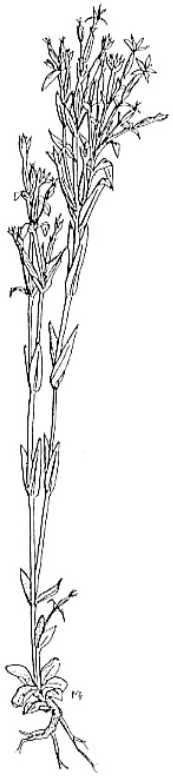


C. brachycalyx

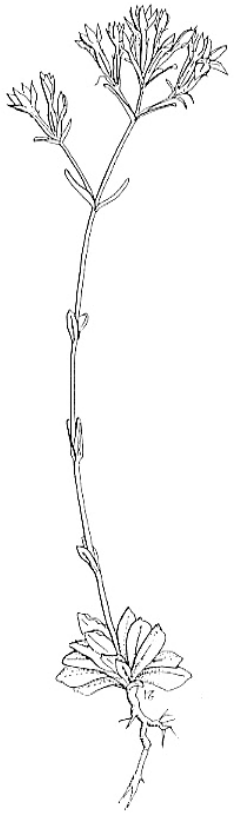
2 cm

"Eurasian *Centaurium*"

(Genus *Centaurium* Hill)



C. serpentinicola



C. majus subsp. rhodense

C. suffruticosum

C. rumelicum

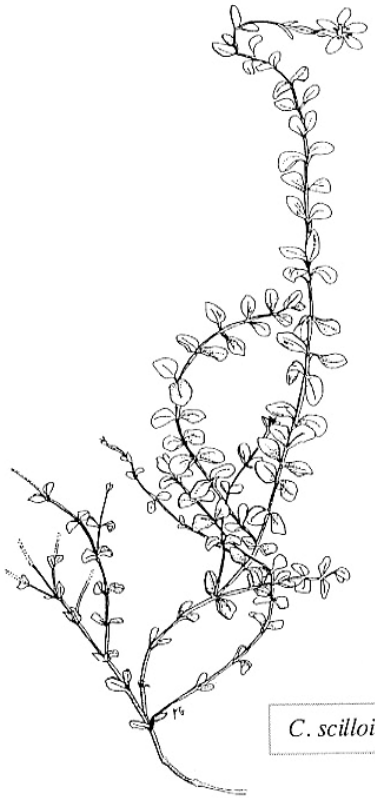
2 cm

"Eurasian *Centaurium*"

(Genus *Centaurium* Hill)

C. favargeri

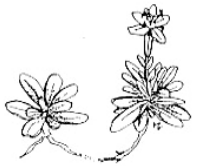
C. gypsicola



C. scilloides

C. somedanum

C. littorale



C. chloodes

2 cm