The use of low-copy nuclear genes in the radiation of the Macaronesian Crassulaceae Sempervivoideae – Phylogeny and evolutionary processes

Dissertation

Korinna Esfeld
Born in Lutherstadt Wittenberg

2009

Dissertation
submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

presented by
Diplom-Biologin Korinna Esfeld
born in: Lutherstadt Wittenberg
Oral-examination: .....

The use of low-copy nuclear genes in the radiation of the Macaronesian Crassulaceae Sempervivoideae – Phylogeny and evolutionary processes

Referees:

Prof. Dr. Marcus Koch

Prof. Dr. Claudia Erbar

... für meine Familie und meine Freunde,

die geduldig daran geglaubt haben!

Am Ende ist alles gut und wenn es nicht gut ist, dann ist das nicht das Ende....

Contents

Figures

Tables

0.1. Summary

Speciation and evolution of species are two of the most exciting topics in biology. Radiations, with their wide morphological and physiological variety, provide a promising tool to understand speciation and diversity of species. Numerous studies have revealed that the high morphological diversity of radiated species is not represented at the molecular level. Neutral markers like rDNA nrITS and chloroplast (cp) DNA evolve slowly compared to speciation in radiations and thus, may not provide enough information to resolve phylogenetic relationships. In contrast, low-copy nuclear genes evolve faster and may help to resolve relationships. This is supported by the hypothesis that accelerated changes in regulatory genes, as opposed to structural genes, can explain the evolution of species.

To contribute to this ongoing discussion, the radiation of the Macaronesian Crassulaceae Sempervivoideae (MCS) was studied. The polyploid species of the MCS are mainly distributed on the Canary Islands and comprise more than 70 species in three genera (*Aeonium*, *Aichryson*, and *Monanthes*) that display a huge morphological (e.g., flower color, number of floral organs, growth-form) and physiological (e.g., CAM activity) variety.

Two regulatory genes, homologs of the floral homoeotic genes *APETALA1* and *APETALA3*, and the structural gene encoding for phosphoenolpyruvate carboxylase (*PEPC*) were analyzed with respect to the following aims: 1) to evaluate the use of the low-copy nuclear genes to reconstruct phylogenies and to compare genealogies with the species phylogeny; 2) to estimate the impact of the studied genes in the speciation process and elucidate differences between the roles of regulatory and structural genes; 3) to determine if gene duplications occurred and to distinguish duplicates into orthologs and paralogs, and 4) to calculate the selection pressure (Ka/Ks-values) acting on the respective gene copies.

The three analyzed low-copy nuclear genes both support and contradict the phylogenetic relationships inferred by other markers. The selection acting on the studied low-copy genes is in contrast with the neutral evolution of nrITS and cpDNA markers and may explain observed differences. In particular *APETALA3* seems to be a promising marker for resolving species relationships.

In addition, the studied genes may have had an influence in speciation since individually they exhibit accelerated Ka/Ks-values compared to mean Ka/Ks-values estimated for regulatory and structural genes. Their Ka/Ks-values are also much higher than those obtained for other genes in studies with comparable experimental designs. Accelerated evolutionary rates were estimated for the regulatory genes as opposed to the structural gene *PEPC*. However, summarizing all observations, the impact of these genes may be limited. Further study is recommended to evaluate their true impact.

For all studied genes duplications were observed and emphasize the greatest challenge of working with low-copy nuclear genes – the differentiation of orthologs and paralogs. The observed duplication pattern suggests that the gene duplications are the result of polyploidization, a phenomenon to which the island colonization of the MCS species was connected previously.

In addition, all gene copies were under purifying selection pressure, even if the estimated Ka/Ks-values for the respective copies varied. Rate differences were estimated for *PEPC* and *APETALA3*; the latter also showed significant differences in the Ka/Ks-values comparing copy A and copy B. For *APETALA1* similar evolutionary rates and highest Ka/Ks-values were found.

Altogether, this thesis offers a promising approach to study speciation and evolution in the radiation of the MCS and is a valuable basis for further studies.

0.2. Zusammenfassung

Artbildung und Evolution gehören zu den spannendsten Themen der Biologie. Für deren Verständnis bieten Radiationen mit ihrer morphologischen und physiologischen Variation eine wichtige Grundlage. Zahlreiche Studien zeigten dabei, dass die morphologische Diversität radiierter Arten nicht mit ihrer molekularen Diversität korreliert. Neutrale molekulare Marker (rDNA nrITS, Chloroplasten-DNA) evolvieren im Vergleich zur Artbildung in Radiationen langsam und eignen sich demzufolge zumeist nicht, um die verwandtschaftlichen Beziehungen wiederzugeben. Kodierende Kerngene jedoch, die in einer geringen Kopiezahl vorliegenden, evolvieren schneller und könnten Verwandtschaftsbeziehungen radiierter Arten widerspiegeln. Zudem wird ein größerer Einfluss von regulatorischen Genen im Vergleich zu Strukturgenen auf die Artbildung diskutiert.

Um zu dieser Diskussion beizutragen, wurde die Radiation der Makaronesischen Semperviven (MCS) untersucht. Diese hauptsächlich auf den Kanaren verbreitete Artengruppe umfasst mehr als 70 polyploide Arten in den Gattungen *Aeonium*, *Aichryson* und *Monanthes*. Die Arten zeichnen sich durch hohe morphologische (z.B. Blütenfarbe, Anzahl der Blütenorgane, Wuchsform) und physiologische (z.B. Unterschiede in CAM-Aktivität) Variation aus. Zwei regulatorische Gene, Homologe der Blütenmorphologiegene *APETALA1* und *APETALA3* sowie das für die Phosphoenolpyruvat-Carboxylase kodierende Strukturgen (*PEPC*) wurden auf ihren Einfluss bezüglich der Artbildung hin untersucht. Die Ziele waren dabei 1) die Eignung der untersuchten Kerngene für phylogenetische Analysen zu bestimmen sowie Genbäume und Artbaum zu vergleichen, 2) den Einfluss der Gene auf den Artbildungsprozess zu bestimmen und Unterschiede zwischen Regulator- und Strukturgenen zu determinieren, 3) Genduplikationen, unterschieden in Orthologe und Paraloge, nachzuweisen und 4) den auf die jeweiligen Genkopien wirkenden Selektionsdruck (Ka/Ks-Werte) zu ermitteln.

Die analysierten Kerngene spiegeln nur teilweise die Artphylogenie wider und Unterschiede zwischen Genbäumen und dem Artbaum wurden beobachtet. Hierbei spielt die auf den Kerngenen wirkende Selektion eine wesentliche Rolle, da diese im Gegensatz zu der Evolution von nrITS und Chloroplasten-DNA Markern nicht neutral ist. Speziell *APETALA3* ist jedoch ein vielversprechender phylogenetischer Marker.

Ein Einfluss der analysierten Gene auf die Artbildung ist denkbar. Ihre durchschnittlichen Ka/Ks-Werte liegen über denen anderer Regulator- und Strukturgene und sind höher als Ka/Ks-Durchschnittswerte anderer Gene, die in ähnlichen Studien ermittelt wurden. Auch konnte eine höhere Evolutionsrate der Regulatorgene im Vergleich zu dem Strukturgen ermittelt werden. Insgesamt ist der Einfluss jedoch wohl limitiert und weitere Untersuchungen notwendig, um den tatsächlichen Einfluss zu ermitteln.

Genverdopplungen, die bei allen Genen beobachtet wurden, verdeutlichen, dass die Unterscheidung in Orthologe und Paraloge eine wesentliche Herausforderung bei der Arbeit mit Kerngenen darstellt. Das dabei beobachtete Muster legt nahe, dass die Verdopplungen auf Polyploidisierung, die mit der Inselbesiedlung gekoppelt war, zurückzuführen sind.

Alle Genkopien unterliegen unterschiedlich starker reinigender Selektion. Unterschiedliche Evolutionsraten wurden bei *PEPC* sowie bei *APETALA3* gefunden, wo zudem signifikant unterschiedliche Ka/Ks-Werte für die Kopien beobachtet wurden. *APETALA1* Genkopien haben die höchsten Ka/Ks-Werte, zeigen jedoch hierbei sowie bei der Evolutionsrate keine signifikanten Unterschiede.

Die vorliegende Arbeit bietet einen sehr vielversprechenden Ansatz für die Untersuchung von Artbildung und Evolution in der Radiation der MCS und bildet eine wichtige Basis für weiterführende Untersuchungen.

1. Introduction

The spectacular diversity of plant and animal species is an ongoing marvel for human beings. Speciation processes and evolution fascinate scientists from all fields and raise numerous questions and discussions. For example, traditionally adaptation and speciation was discussed due to fixation of many genes with small effects (see, e.g., Schemske and Bradshaw 1999). However, extensive studies were done recently to define major forces acting in speciation and examples are known where few traits and genes highly influence evolution. One of the key morphological differences between maize and teosinte is encoded by the *teosinte-branched-1* locus (for reference see Ford 2002). Innovation of nectar spurs lead to the extensive radiation of *Aquilegia* since this key character allowed specialization to different pollinators, a well-known speciation mechanism (Hodges and Arnold 1994a, Hodges 1997, Whittall and Hodges 2007). In the genus *Mimulus*, a single homozygous mutation in a gene of the flavonoid pathway causes a shift in flower color resulting in different pollinator preferences (Schemske and Bradshaw 1999, see also Durbin et al. 2003). Regarding these studies, it seems that a small number of genes may be responsible for the evolution of major differences in growth-form and flower morphology of plants. Central to speciation seems to be the accelerated evolution of regulatory genes, which has been proposed as a main explanation for the morphological development of species (King and Wilson 1975, Remington and Purugganan 2002, Durbin et al. 2003, Purugganan and Robichaux 2005).

Radiations, with their wide morphological and ecological variety, offer a promising tool to answer questions about speciation, species diversity, and evolution (Whittall et al. 2006). They are common on islands where extraordinary speciation is favored due to high diversity of habitats and lack of competition after colonization (MacArthur and Wilson 1967, Arnedo et al. 1996). Therefore, islands are seen as natural laboratories and model systems to study speciation (e.g., Francisco-Ortega et al. 1996, Baldwin et al. 1998). In the past, most studies concerning plant and animal evolution focused on the Pacific islands of Galapagos, Hawaii or Juan Fernandez that are several thousand kilometers away from their respective continent (Baldwin et al. 1998, see also Jorgensen and Olesen 2001). Recently, also Atlantic islands have become the focus of interest. Macaronesia, comprising the Azores, Canary Islands, Cape Verde, Madeira, and Salvages (fig. 1), serves as an ideal system to understand the origin

and evolution of island biota and the consequences of colonization and isolation. The seven main islands and several islets of the Canaries (fig. 1) are of particular interest due to their geographical proximity to the African continent. They are situated in the northeast Atlantic Ocean between 27°37' and 29°25'N and 13°20' and 18°10'W. Fuerteventura is closest to the continent, approximately 110 km away, and also La Palma is situated only 460 km away from the northwest African mainland (Carracedo 1994, Juan et al. 2000). Despite known sea mountains close to the sea level between the Canaries and the mainland of Portugal (Francisco-Ortega et al. 2000), a connection with the continent was never given (Andrus et al. 2004 and references therein, Sanmartín et al. 2008).

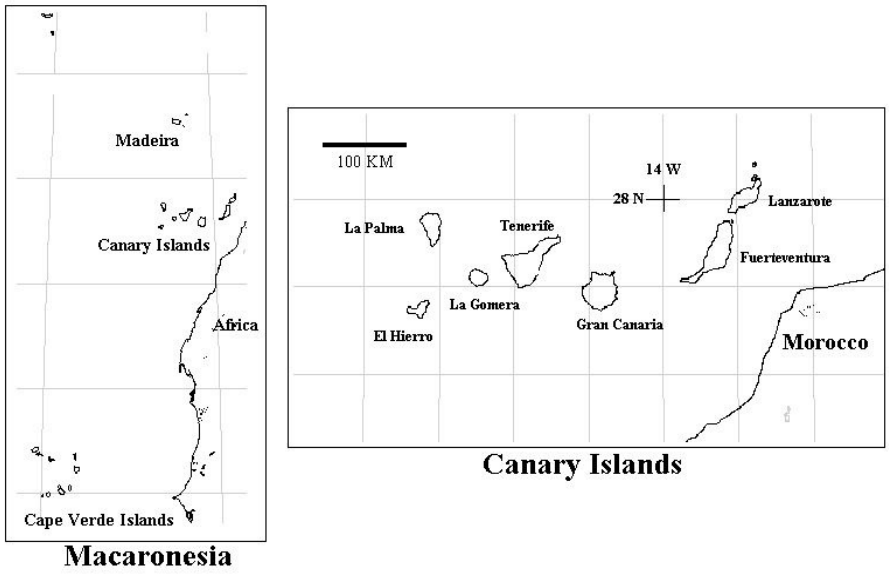


Fig. 1: Map of Macaronesia and the Canary Islands after Mort et al. (2002). Taken from the website http://www.eiu.edu/~bio_data/posters/2002/poster_016.htm.

The Canary Islands are of volcanic origin and provide a broad range of geological ages. Nearly ordered in a line, El Hierro is the youngest and westernmost island and Fuerteventura the oldest and easternmost. Ages of the seven main islands are 20.7 My for Fuerteventura, 15.5 My for Lanzarote, 13.9-16 My for Gran Canaria, 11.6 My for Tenerife, 10-12.5 My for La Gomera, 1.5-2 My for La Palma, and > 0.7 My for El Hierro (Carracedo 1994, Arnedo et al. 1996, Kim et al. 1996).

The combination of important factors such as geology, trade winds, elevation, and inclination has lead to distinct vegetation zones on the Canary Islands providing the fundament for evolutionary processes (Lems 1960, Baldwin et al. 1998). The humid and cool northeastern trade winds lead to two main climatic regions: one humid due

to the influence of these trade winds and the other not and therefore more arid. Ecological zones range from coastal desert and lowland scrub to humid laurel forest, pine forest, and alpine desert (Francisco-Ortega et al. 1996, Jorgensen and Frydenberg 1999).
Additionally volcanic activity has provided unstable conditions, which highly influences fast and divergent speciation by founder events, genetic bottlenecks, and genetic drift (Kull 1982, Nyffeler 1995, Jorgensen and Olesen 2001). Also inter-island dispersal and colonization of similar ecological zones on different islands has influenced speciation. Thus, the Canary Islands exhibit well-known examples of large radiated groups: animal species like geckos, several genera of Coleoptera, e.g., *Nesotes* or the spider genus *Dysdera*, and plant genera such as *Aeonium*, *Argyranthemum*, *Echium*, *Sideritis*, and *Sonchus* (Kull 1982, Lösch 1990, Arnedo et al. 1996, Böhle et al. 1996, Francisco-Ortega et al. 1996, Kim et al. 1996, Barber et al. 2000, Juan et al. 2000, Emerson 2002, Mort et al. 2002).

One of the most famous of the radiated plant groups are the Macaronesian Crassulaceae Sempervivoideae (MCS). This group comprises approximately 70 species within the three genera *Aeonium* Webb & Berthel. (including the former separated genus *Greenovia* Webb & Berthel.), *Monanthes* Haw., and *Aichryson* Webb & Berthel. and is a well supported group of the Crassulaceae (Berger 1930, Praeger 1932, Lems 1960, Liu 1989, Nyffeler 1992, Mes 1995, Mort et al. 2002).
The family of the **Crassulaceae** is monophyletic, part of the Saxifragales clade, and comprises six subfamilies, 35 genera, and 1500 species (Berger 1930, Liu 1989, van Ham and ‘t Hart 1998, Fishbein et al. 2001, Mort et al. 2001). The species in this group are morphologically diverse. Common features are succulent leaves and pentamerous, radially symmetrical flowers with one or two whorls of stamens (van Ham and ‘t Hart 1998, Mort et al. 2001). Crassulaceae are mainly herbaceous but for several genera like *Aeonium*, *Cotyledon*, *Crassula*, *Kalanchoe*, and *Sedum* woody species are known (Mes et al. 1996). Members of the family predominately inhabit semiarid to arid and mountainous habitats (Mes et al. 1996, van Ham and ‘t Hart 1998). They are adapted to low water supply by their succulent leaves and their special CO_2 fixation pathway, the Crassulacean Acid Metabolism (CAM; Lösch 1990). Crassulaceae are distributed worldwide but predominately found in subtropical and temperate zones (van Ham and ‘t Hart 1998). The origin of the family lies in

Southern Africa or in the Mediterranean area with centers of diversity in Mexico, South Africa, Himalaya, the Mediterranean region, and Macaronesia (van Ham and ‘t Hart 1998, Mort et al. 2001).

Within the Crassulaceae Berger (1930) defined six subfamilies and 33 genera based on the number and arrangement of floral parts, degree of sympetaly, and phyllotaxis, but five of his described subfamilies are polyphyletic (van Ham and ‘t Hart 1998, Mort et al. 2001). He defined two lineages, a “*Crassula* lineage” and a “*Sedum* lineage”, mainly separated by geography (van Ham and ‘t Hart 1998, Mort et al. 2001). This well-supported split at the base of the Crassulaceae nowadays defines the separation into two subfamilies: Crassuloideae and Sedoideae (van Ham and ‘t Hart 1998, Mort et al. 2001). The Crassuloideae contains only the *Crassula* clade whereas the Sedoideae comprises the remaining subfamilies (van Ham and ‘t Hart 1998). Four clades could be found within the Sedoideae: *Acre*, *Aeonium*, *Kalanchoe*, and *Leucosedum*. While evidence also exists for separate *Telephium* and *Sempervivum* clades, the genus *Sedum* is highly polyphyletic (Mort et al. 2001).

Berger (1930) also defined a **Sempervivoideae** clade, classifying species with polymerous flowers. It comprises the genera *Sempervivum*, *Aeonium*, *Aichryson*, *Greenovia*, and *Monanthes*. Another classification of the Sempervivoideae was done by Praeger (1932) who recognized these five sections, 32 species, one variety, and 25 hybrids (Liu 1989). Nevertheless, the Sempervivoideae are not monophyletic and a separation between the genus *Sempervivum* and the monophyletic MCS, comprising the remaining four genera, has been proposed (Mes et al. 1996, Mort et al. 2001).

As indicated by their name, the species of the **Macaronesian Crassulaceae Sempervivoideae** are nearly endemic in Macaronesia and mainly restricted to the Canary Islands (Mort et al. 2001). Nevertheless, *Aeonium leucoblepharum* and *A. stuessyi* are known in eastern Africa (Liu 1989). This disjunct distribution pattern brought up a controversial debate about the origin of the group (Mes 1995, Mes et al. 1996). Traditionally species of *Aeonium* were regarded as descendants of African progenitors of Tertiary resembling (Bramwell 1976, Kim et al. 1996), but the expansion of the Sahara desert resulted in extinction of these ancestral species. Only a few *Aeonium* species survived in northern Africa and western Morocco, dispersed

to the Canary Islands and gave rise to the present diversity through island speciation and adaptive radiation (Berger 1930, Lems 1960, Liu 1989, Lösch 1990).
Recently, molecular markers have revealed another picture. It now seems indisputable that the MCS arose from a single colonization event from northern Africa and diversified on the Canary Islands into the present monophyletic species group (Mes 1995, Mes et al. 1996, Mort et al. 2001). The large shrubby *Aeonium* species of East Africa are the result of dispersal events from Macaronesia back to the African continent (Mes et al. 1996).

Following the debate of the origin of the clade, there were contradicting views about the **woodiness** of the MCS species. For species of Atlantic islands woodiness was traditionally seen as relict status (Lems 1960, Bramwell 1976, Liu 1989) conflicting the theory of Carlquist (1962, 1974). Carlquist proposed that the large and woody taxa of otherwise predominately herbaceous species on oceanic islands evolved through an increased secondary “insular woodiness” (see Jorgensen and Frydenberg 1999).
Molecular and physiological data support Carlquist’s theory and suggest that the MCS derived from an herbaceous ancestor (Pilon-Smits et al. 1992, Mort et al. 2002) which is also supported by the terminal position of the woody African taxa in phylogenetic reconstructions (Mes et al. 1996). For species of the Pacific islands woodiness was always considered as being derived on islands by rapid evolution from ancestral herbaceous progenitors (Baldwin et al. 1998). Recent studies confirmed this evolutionary pathway as well as derived woodiness for species of the Atlantic islands, e.g., in the genera *Aeonium*, *Echium* or *Sonchus* (Böhle et al. 1996, Kim et al. 1996, Mes at al. 1996).

Based on **karyological** evidence, Uhl (1961) discussed the relationship between North African *Sedum* species and the MCS. Based on this information, Mes (1995) subsequently described the three herbaceous *Sedum* species, *S. jaccardianum*, *S. modestum*, and *S. surculosum* (*Sedum* series *Monanthoidea*) as well as *S. caeruleum* and *S. pubescens*, as basal to the MCS clade. This suggestion was confirmed by several studies using different nuclear DNA (nrDNA; the internal transcribed spacer region = nrITS) and chloroplast DNA (cpDNA) molecular markers (Mes and ‘t Hart 1994, Mes et al. 1996, van Ham and ‘t Hart 1998).

The respective *Sedum* species are diploid with exception of the tetraploid *S. surculosum* (for reference see Mes and 't Hart 1994). The **chromosome base number** of the species is variable: for *S. caeruleum* x = 12-13, for *S. pubescens* x = 11, and for *S. jaccardianum*, *S. modestum*, and *S. surculosum* x = 8.
The species of *Aeonium* and *Monanthes* have a strict chromosome base number of x = 18 and for the species of *Aichryson* x = 15, 16, and 17 is reported. All MCS species are exclusively polyploid and thus, colonization of the Canaries was accompanied by **polyploidization** (Mes 1995, Mes et al. 1996, Mort et al. 2001). The increase in chromosome number must have occurred either in North Africa in an extinct ancestral species or on the Canary Islands (Mes et al. 1996). Colonization connected with polyploidization and without further change of the chromosome base number is a well known phenomenon for radiations and island speciation (Mes et al. 1996 and references therein). However, within the MCS subsequent increase in chromosome numbers (tetraploid up to hexaploid species, e.g., *A. arboreum*, *A. haworthii*, *A. leucoblepharum*, *A. simsii*, *A. urbicum*, *Ai. pachycaulon*, *M. anagensis*, and *M. polyphylla*) are known (Mes 1995).

Beside polyploidization, **hybridization** is assumed to play a minor role in the MCS, and only few species of hybrid origin are known (Mes and 't Hart 1996, Mes et al. 1997, van Ham and 't Hart 1998, Jorgensen and Frydenberg 1999, Mort et al. 2002).
In general, island species and species of radiated clades are considered as highly cross compatible with few or no internal barriers for crossing (Whitkus 1998, Jorgensen and Olesen 2001). Even if, e.g., most *Aeonium* species have very restricted habitat requirements and show ecological and topographic isolation (Kull 1982, Jorgensen and Frydenberg 1999), almost all species of *Aeonium*, *Aichryson*, and *Monanthes* can hybridize when occurring in sympatry and with overlapping flowering times (Lems 1960, Jorgensen and Olesen 2001). Hybrids are viable and fertile even if parental species seem to be fitter and survive better (Liu 1989, Nyffeler 1992). It is possible that the rarity of hybrids in the field and the lack of extensive hybrid zones are caused by the absence of suitable habitats for hybrid species (Lems 1960). In addition, a pollinator study has shown that infraspecific pollen transfer prevails and interspecific pollen transfer is limited in nature (Esfeld et al. 2009).

Whereas the division in the three genera *Aeonium*, *Aichryson*, and *Monanthes* is widely accepted (Mes 1995, Mort et al. 2002), relationships within the group and especially within the genera and clades are rather unresolved and highly debated. For classification and phylogenetic reconstruction biochemical, cytological, molecular, morphological, and physiological data have been used (Lems 1960, Uhl 1961, Tenhunen et al. 1982, Liu 1989, Lösch 1990, Nyffeler 1992, Pilon-Smits et al. 1992, Mes 1995, Nyffeler 1995, Stevens 1995, Mes and 't Hart 1996, Mes et al. 1996, 1997, van Ham and 't Hart 1998, Jorgensen and Frydenberg 1999, Mort et al. 2001, 2002, Fairfield et al. 2004, Mort et al. 2007) and important aspects of the three genera are summarized below.

The genus ***Aeonium*** **Webb & Berthel.** (including *Greenovia* Webb & Berthel.) is formed of more than 40 species and is the largest radiation both on the Canary Islands and within the MCS. The species of this genus display a huge morphological and physiological variety. Their growth-forms range from small rosettes herbs to large woody plants including subshrubs, shrubs, candelabrum, and rosettes trees. Flower color varies between white, green, yellow, and rose nuances to dark red for *A. nobile*. The species colonize habitats in all vegetation zones of the Canary Islands and habitats are connected to physiological adaptation, e.g., strength of the C_3 and CAM activity (Pilon-Smits et al. 1992). The species are mainly distributed as single island endemics on all Canary Islands. Only six species are known outside this centre of diversity; they are distributed on Madeira (2 species), Cape Verde (1), in Morocco (1), and Eastern Africa (2; Mes 1995, Jorgensen and Frydenberg 1999).
Historically, the genus *Greenovia* was separated from the genus *Aeonium* according to the derived number of flower organs, the absence of nectariferous scales, and different placentation (Berger 1930, Liu 1989, Mes 1995). *Greenovia* species are characterized by small, hapaxanth rosettes and highly polymerous (16-35merous) flowers (Mes and 't Hart 1996). However, molecular data confirm the inclusion of the four *Greenovia* species in the genus *Aeonium* (Mes 1995, Mort et al. 2002). Additionally, this is supported by the chromosome base number and because hybridization is possible between species of *Greenovia* and *Aeonium* (Praeger 1932, see reference in Mort et al. 2002).
While classifying the *Aeonium* clade Lems (1960) largely followed the classification of Praeger (1932). He focused on several growth characteristics such as lignification

and length of the stem, branching pattern, inflorescence size, and pattern, type, and size of leaves. Section ***Holochrysa*** comprises crassicaulous, monocarpic shrubs with 8-12merous yellow flowers. In the sect. ***Megalonium*** *A. nobile*, a monocarpic shrub with 7-8merous dark red flowers, is the only species. Section ***Urbica*** comprises several crassicaulous or woody shrubs or subshrubs with white or pink 6-9merous flowers. Within sect. ***Canariensia*** branched or monocarpic rosette plants were found with yellow 7-13merous flowers. Finally, sect. ***Goochia*** is characterized by dwarf shrubs and stoloniferous rosette plants with yellow and 7-16merous flowers.
Liu (1989) focused on typical morphological characters (e.g., plant height, branching type, surface reticulation of the stem, length, width, and thickness of leaves, inflorescence position, and number of flower organs) for phenetic and cladistic analyses. In total, he used 39 characters to define seven sections, 31 taxa, and six varieties. Section ***Petrothamnium*** is represented by small twiggy subshrubs with yellow flowers (pinkish in *A. goochiae*). Section ***Chrysocome*** contains perennial terrestrial twiggy subshrubs without nectariferous glands and yellow flowers. In sect. ***Patinaria*** species are biennial to perennial herbs that form rosettes and have pale yellow to nearly white flowers. Section ***Aeonium*** comprises perennial terrestrial subshrubs with yellow flowers. The monotypic sect. ***Megalonium*** contains again only *A. nobile*. To sect. ***Pittonium*** belong perennial terrestrial subshrubs with yellow flowers and sect. ***Leuconium*** comprises perennial terrestrial subshrubs that have white petals, often with pink streaks on central regions.
Studies based on molecular markers were, e.g., done by Mes (1995), Mes and ‘t Hart (1996), Jorgensen and Frydenberg (1999), and Mort et al. (2002). Mes (1995) used chloroplast and nuclear spacer sequences to study phylogenetic relationships and inferred nine sections: ***Aeonium***, ***Canariensia***, ***Chrysocome***, ***Goochiae***, ***Greenovia***, ***Leuconium***, ***Patinaria***, ***Petrothamnium***, and ***Pittonium***. Mort et al. (2002) provided the most extensive study based on nrITS and cpDNA markers in the genera of MCS and described four *Aeonium* subclades. The infrageneric classification is summarized in table 15 (appendix). Figure 2 displays the phylogenetic relationships of the MCS species based on combined cpDNA and nrITS data of Mort et al. (2002).

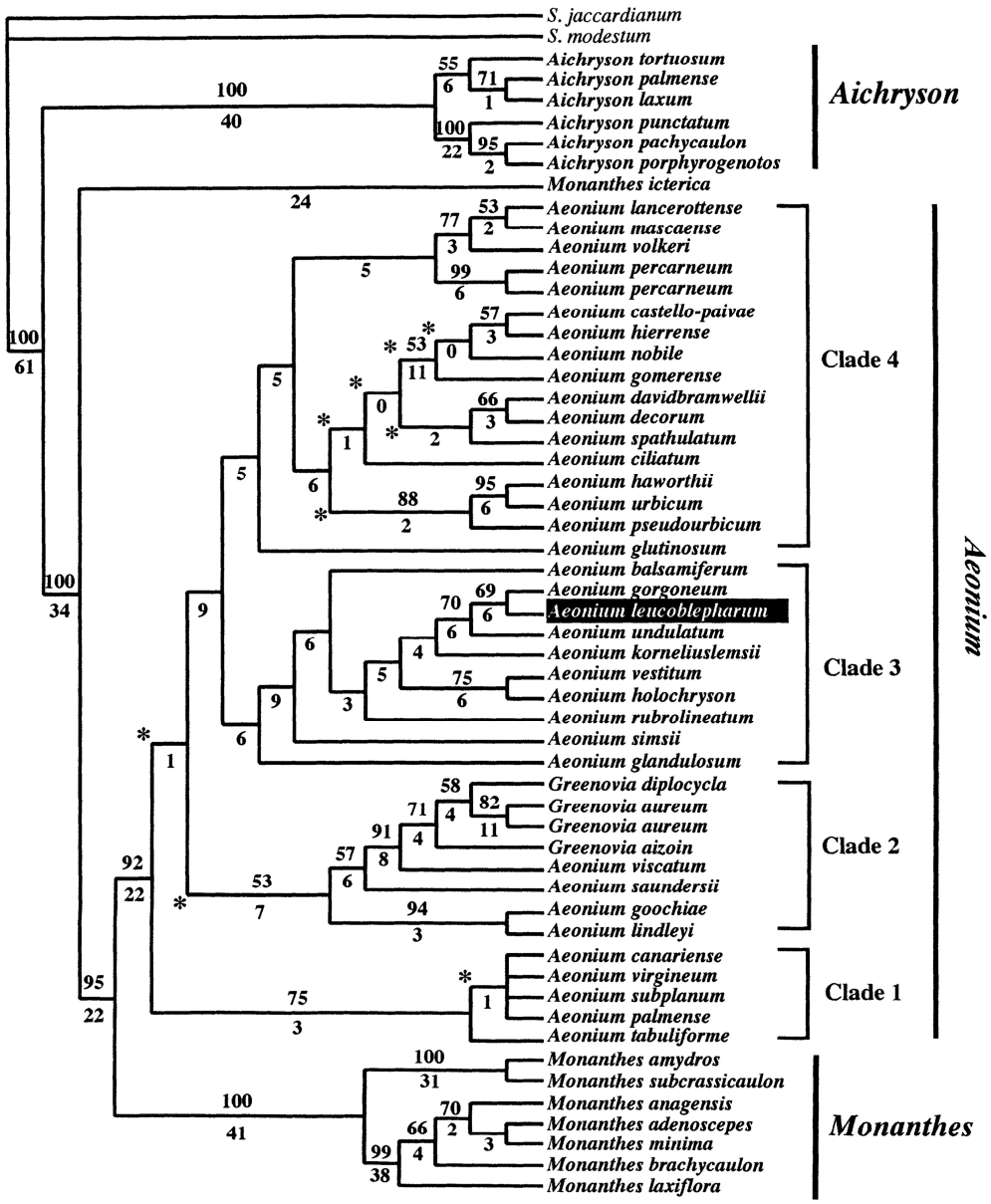


Fig. 2: Maximum parsimony phylogram of the MCS species based on cpDNA/nrITS data (from Mort et al. 2002).

The monophyletic genus ***Aichryson* Webb & Berthel.** contains about 13 species and is basal and sister to *Aeonium* and the perennial *Monanthes* species (Mort et al. 2002, Fairfield et al. 2004; fig. 2). The genus can be differentiated into six major clades and two lineages that correspond to habit and growth-form. One consists of the woody, perennial species *Ai. tortuosum* and *Ai. bethencourtianum*; the second comprises the herbaceous, annual members of the genus (Fairfield et al. 2004). The woody species are CAM species whereas the species of the second group mainly use the C_3 gas-exchange pathway (Lösch 1990). Species are generally highly branched, predominately hapaxanth, and up to 40 cm high (Mes 1995). They are

characterized by 6-12merous flowers with digitate nectariferous glands (Liu 1989, Fairfield et al. 2004).
Aichryson species colonize moist, shady habitats and are especially common in the laurel forest belt. Ten species are endemic to the Canary Islands whereas three, *Ai. villosum*, *Ai. divaricatum*, and *Ai. dumosum*, occur on Madeira and the Azores (Fairfield et al. 2004). At least one species of *Aichryson* can be found on each of the seven main Canary Islands. The centre of diversity is La Palma and only one species is found on Lanzarote. Three species are single-island endemics, *Ai. bethencourtianum* on Fuerteventura, *Ai. palmense* on La Palma, and *Ai. porphyrogennetos* on Gran Canaria. Fuerteventura and Lanzarote were the first colonized by species of *Aichryson*, and Madeira was subsequently populated from these two easternmost islands. Only once again this colonization pattern was found (genus *Crambe*; Francisco-Ortega et al. 2002) and contrasts biogeographic implications of other genera where a close biogeographic affinity between Madeira and the five western Canary Islands was detected (Fairfield et al. 2004).

Up to 13 species are known for the genus ***Monanthes*** **Haw.**. They are classified into three sections: *Monanthes*, *Monanthoidea*, and *Sedoidea* (Nyffeler 1992, Mes et al. 1997). They show a high level of different growth-forms comprising the dwarf annual herb *M. icterica*, perennial branched or unbranched herbaceous rosettes, and small branched shrublets. The large nectariferous scales are the most characteristic feature of *Monanthes*. Being the showiest part of the flowers, they take over the attracting function. Petals are very small, narrow, and decurved. Flowers are in general 6-8merous but the number of flower organs is not constant and can even vary within a single plant. Flower color is greenish or brownish and often variously variegated with red (Nyffeler 1992). Pollinators are flies that are attracted by segregated nectar and an intense musty scent in the evening hours (Nyffeler 1992). Species are self-compatible although temporal and spatial separation of male and female organs prevents self-pollination (Nyffeler 1992, 1995).
The species are distributed on all Canary Islands and the Salvage Islands with highest diversity on Tenerife. *Monanthes anagensis* and *M. minima* are single island endemics and only *M. laxiflora* and *M. polyphylla* are represented on more than two islands (Nyffeler 1992). Species of *Monanthes* colonize rather mesic habitats with a regular water supply. They grow in crevices of rocks and cliffs protected from direct

sunlight in north- or northeast-exposed locations. Thus, in physiological adaptation the C_3 gas-exchange prevails even if weak CAM activity is also known (Nyffeler 1992). In addition, the characteristic bladder cell-idioblasts at the margin of leaves are probably important for water storage (Nyffeler 1992, 1995).
Because different species with partly overlapping flowering times are often found in sympatry, hybridization is frequent and indicates lack of genetic isolation. Hybrids show intermediate habits, are restricted in their distribution, and usually occur only in small numbers (Nyffeler 1995). An exception is *M. icterica* for which no hybridization event is known (Mes et al. 1997). *Monanthes muralis* is of hybrid origin and the only known allotetraploid species (Mes et al. 1997). Further tetraploid species such as *M. laxiflora*, *M. pallens*, and *M. polyphylla* are known and *M. anagensis* is most probably hexaploid (2n = 108; Mes et al. 1997). In general, chromosomes are very small and the only exception is *M. icterica* which has large chromosomes and a chromosome base number of x = 10 compared to x = 18 for all other *Monanthes* species (Nyffeler 1992, Mes et al. 1997).

Based on the wide range of growth-forms, flower morphology, and habitats, the MCS were traditionally seen as an excellent example of an **adaptive radiation**. Lems (1960) even compared the species of *Aeonium* with the Darwin finches. However, molecular markers reveal a different picture. Inter-island colonization between similar ecological zones – together with adaptation and hybridization – is the main force driving speciation in the MCS (Jorgensen and Frydenberg 1999, Mort et al. 2002). Inter-island colonization is also known for other Macaronesian genera like *Argyranthemum*, *Bystropogon*, *Crambe*, and *Sonchus* (Francisco-Ortega et al. 1996, Kim et al. 1996, Juan et al. 2000, Francisco-Ortega et al. 2001, 2002, Trusty et al. 2005, Sanmartín et al. 2008). Exceptions to this rule are the two genera *Micromeria* (Meimberg et al. 2006) and *Sideritis* (Barber et al. 2000) in which adaptive radiation is the main explanation for speciation. For lineages of the Pacific archipelagos adaptive radiation is a common phenomenon and driving force for speciation with the outstanding example of the Hawaiian silversword alliance (HSA; Baldwin and Robichaux 1995, see also Barber et al. 2000).

Phylogenetic studies were done in many island species and radiations to understand relationships among and within genera or species, define their origin,

reveal phylogeographic and biogeographic patterns, and gain insight into speciation processes. In particular molecular markers have broadened the knowledge and shed new light on processes influencing radiation events (for review, e.g., Baldwin et al. 1998 or Emerson 2002).

Resolving relationships in recently diverged taxa is a huge problem in molecular systematics (Syring et al. 2005). Most common molecular markers, such as nrITS and cpDNA, are thought to be selectively nearly neutral and evolve slowly relative to speciation in radiations (Baldwin et al. 1998, Sang 2002, Small et al. 2004, Whittall et al. 2006). Thus, phylogenetic relationships were highly unresolved due to low genetic divergence and a lack of fixation of synapomorphic mutations (Mes et al. 1996, Baldwin et al. 1998). Subsequently, there is an ongoing debate that changes in coding genes could verify the diversity of radiated lineages and provide valuable tools for phylogenetic reconstructions (Baldwin et al. 1998, Sang 2002, Small et al. 2004).

Nuclear DNA (nrITS) and cpDNA markers are widely taken for phylogenetic reconstructions. Universal primers are available which can be used over a wide range of taxa (White et al. 1990, Taberlet et al. 1991, Blattner 1999). High copy numbers facilitate amplification and sequencing can be done without prior cloning. **Low-copy nuclear genes** are not extensively used until now (Bailey and Doyle 1999). Their main disadvantages are the development of species, gene or copy specific primers, cloning and intensive sequencing (Sang 2002, Small et al. 2004). Numerous nuclear coding genes exist in gene families and as multiple copies due to gene or genome duplications (Bailey and Doyle 1999, Small et al. 2004). Speciation events result in orthologous gene copies and duplication of genes within one species in paralogous copies (Litt and Irish 2003). Thus, orthologs and paralogs are frequently distributed in species and need to be distinguished since phylogenetic analyses depend strictly on comparisons of orthologous gene copies to deduce robust and correct relationships (Bailey and Doyle 1999, Litt and Irish 2003).

Advantages of nuclear coding genes are that they provide a nearly unlimited source of additional, independent, unlinked, and bi-parentally inherited phylogenetic information (Sang 2002, Small et al. 2004, Syring et al. 2005). They are alternative markers to explore hybridization and polyploidization and to resolve contrasting signals between nrITS and cpDNA data (Bailey and Doyle 1999, Small et al. 2004).

In most cases, nuclear coding genes show higher evolutionary rates than neutral evolving DNA regions and are therefore of special interest to resolve relationships at low taxonomic levels and within radiations (reviewed in Small et al. 2004). In addition, nuclear coding genes can be divided into four different parts that evolve differently. First the 5´-untranslated region (UTR) which comprises conserved promoter elements that are responsible for gene regulation. Highly variable parts may be used for phylogenetic reconstructions (Small et al. 2004). Second the conserved exon regions which contain the protein coding information. Within the 1st and 2nd codon positions a nucleotide change results in a nonsynonymous amino acid replacement. Nucleotide changes at the synonymous 3rd codon position result in the same amino acid and the nucleotides at this position typically diverge at similar rates as non-coding regions (Nei and Gojobori 1986, Purugganan et al. 1995, Nei and Kumar 2000, Small et al. 2004). Alignments of exon sequences are in most cases easy, especially at the amino acid level, even between genera or families. Thirdly and in contrast, introns are much more variable. This is particularly true for variations at the sequence level whereas the length of introns is often important for correct splicing. In several cases introns also contain important regulatory elements that are conserved (Small et al. 2004). Aligning intron regions is not trivial and in most cases only possible between closely related species (see, e.g., Fortune et al. 2007 or Zhang et al. 2008). The last region, the 3´-UTR, controls mRNA processing and the poly-A tail and is highly variable even among species of the same genus (Small et al. 2004, Whittall et al. 2006).

Molecular marker studies have surprisingly revealed that the high morphological divergence found in radiated lineages does not correlate with the molecular diversity of these species (see, e.g., Baldwin et al. 1998 or Purugganan and Robichaux 2005). King and Wilson (1975) postulated that changes in **regulatory genes** rather than changes in **structural genes** are the main factor for morphological variation of species. Structural genes encode proteins that directly fulfill their task in the organism. On the other hand, regulatory genes encode transcription factors that regulate the expression of other genes and therefore play a central role in eukaryotic development (Purugganan and Robichaux 2005). For example, in the adaptive radiation of the HSA accelerated substitution rates for homologs of the regulatory genes *APETALA1* (AS*AP1*) and *APETALA3* (AS*AP3*) compared to the structural

gene *CHLOROPHYLL A/B BINDING PROTEIN9* (*ASCAB9*) could be detected (Barrier et al. 2001).

Contributing to this ongoing debate on the emphasis of these two gene classes, the radiation of the MCS was studied. Physiological as well as morphological differences within the MCS indicate that several genes may have had impact on the speciation. Physiological differences are well known in adaptation to dry habitats and a limited availability of water. In response, CO_2 fixation is shifted from C_3 to CAM fixation in the MCS species (Lösch 1990). Here, fixation of CO_2 results first in phosphoenolpyruvate (PEP) that is converted to malic acid and stored in the vacuole during night. During daytime malate is decarboxylated and CO_2 re-assimilated via ribulose-1,5-bisphosphate carboxylase (RUBISCO). The fixation is characterized by changing pH-values during day and night and triggered by the enzyme **PEP carboxylase** (PEPC; Cushman and Bohnert 1999). PEPC is a multifunctional enzyme and ubiquitous in the plant kingdom with different isoforms and tissue-specific or specific physiological roles. This structural gene is encoded by a small multigene family and several gene copies per taxon are known (Gehrig et al. 1995). Protein alignments of the PEPC coding genes show highly conserved motifs. It is likely that these encode domains which are involved in the activity and regulation of the enzyme (Gehrig et al. 1995, Chollet 1996, Cushman and Bohnert 1999).
Considerable differences in activity and strength of CAM and C_3 for the species of the MCS were reported by Tenhunen et al. (1982). Lösch (1990) classified the species in strong, weak, and intermediate CAM and C_3 fixation types. Pilon-Smits et al. (1992) and Mort et al. (2007) refined these results and concentrated mainly on the impact of CAM in the evolution of the MCS. Thus, the structural gene which encodes for PEPC, afterwards named *MCS_PEPC*, is a good candidate for playing an important role in the speciation process.

Other candidate genes with a high impact on speciation might be found involved in reproduction. Mutations in genes that regulate floral color, reward or flower architecture can accelerate the diversification process and cause rapid isolation by influencing differential pollinator visitation (e.g., Baker and Baker 1983, Hoballah et al. 2007, Whittall and Hodges 2007).

Flower development is controlled by a large number of homeotic genes. Since the identity of floral organs strictly depends on the activity of these genes they may have an extraordinary influence in speciation (Theissen 2005, Erbar 2007). **Floral homeotic genes** are regulatory genes and almost all of them belong to the MADS-box gene family (Becker et al. 2000, Erbar 2007). As the name indicates, the **MADS** (**M**CM1 from budding yeast, **A**GAMOUS from *Arabidopsis*, **D**EFICIENS from snapdragon, and **S**RF from human) gene family exists in animals, fungi, and plants and control diverse developmental processes (Becker et al. 2000). In plants, most MADS-box genes display floral-specific expression. However, some are expressed in vegetative tissues and control, e.g., flowering time, inflorescence development and structure, leaf development, or determine cell specification (Purugganan et al. 1995, Baum 1998, Vergara-Silva et al. 2000, Saedler et al. 2001).

Plants possess type II MADS-box genes. The encoded proteins are of ~260 amino acids and characterized by the MIKC structure comprising the four conserved sequence regions: **M**ADS-box domain, short **I**ntervening region (I-region), **K**eratin-like domain (K-domain), and **C**-terminal region (Purugganan et al. 1995, Kramer and Hall 2005; fig. 3).

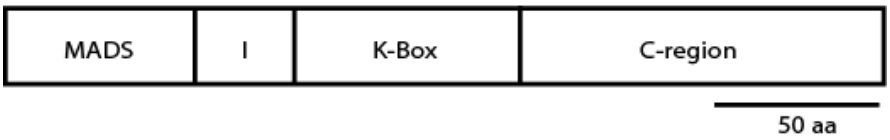


Fig. 3: MIKC-like structure of the plant MADS-box proteins (adapted after Purugganan et al. 1995).

The highly conserved MADS-box domain comprises ~57 amino acids and may act as a sequence specific DNA-binding motif (Coen and Meyerowitz 1991, Saedler et al. 2001). The I-region, which separates the MADS- and K-domain, shows considerable sequence variability (Purugganan et al. 1995, Saedler et al. 2001, Kramer and Hall 2005). According to Becker et al. (2000) this region is important for selective formation of DNA-binding heterodimers. The K-domain comprises ~70 conserved amino acids and might mediate protein-protein interactions (Coen and Meyerowitz 1991, Vergara-Silva et al. 2000, Saedler et al. 2001). The C-terminal region is most variable in sequence and length. It encodes a putative transactivation domain which is involved in the formation of heterodimers (Kramer et al. 1998, Becker et al. 2000, Vergara-Silva et al. 2000). MADS-box, I-region and the first part of the K-domain are known as the core region and are necessary for DNA-binding and dimerization

activity. The non-core region comprises the other part of the K-box and the C-terminal region and is important for strengthening the dimerization activities (Zhang et al. 2008).

MADS-box genes were first characterized in the model species *Arabidopsis thaliana* and *Antirrhinum majus* (for reference see, e.g., Erbar 2007) and defined as early and late acting genes in plant development (Purugganan et al. 1995). Multiple gene duplications have occurred and suggest that gene diversification via gene duplication plays an important role in this gene family (Aagaard et al. 2005 and references therein). At least three monophyletic MADS-box gene groups could be defined whose members share similar expression patterns and functional roles. Many species possess more than one locus from the three groups AGAMOUS, APETALA3/PISTILLATA, and APETALA1/AGL9 (Purugganan et al. 1995). The AGAMOUS-group consists of 10 genes such as *AGAMOUS* and *PLENA* which are inferred in the development of stamens and carpels (Kramer and Hall 2005). The APETALA3/PISTILLATA-group comprises 10 genes that are important for development of petals and stamens. The group can be subdivided into two distinct monophyletic subclades: the first comprises the orthologous genes *APETALA3* (*AP3*, from *A. thaliana*) and *DEFICIENS* (*DEF*, *A. majus*); the second *PISTILLATA* (*PI*, *A. thaliana*) and *GLOBOSA* (*GLO*, *A. majus*). Within one species *AP3* and *PI* as well as *DEF* and *GLO* are paralogs and both are required for gene activity. They form heterodimers consisting of *AP3/PI* (for *A. thaliana*) and *DEF/GLO* (*A. majus*) that bind to their own promoters to set-up self-regulation (Purugganan et al. 1995, Kramer et al. 1998, Saedler et al. 2001). Homologs of *AP3* also show an influence on the size of petals and stamens (Juenger et al. 2000, Lawton-Rauh et al. 2000). The third group, APETALA1/AGL9, contains 14 genes in three distinct monophyletic gene clades. The first clade, named APETALA1, consists of, e.g., *APETALA1* (*AP1*), *SQUAMOSA* (*SQUA*), *CAULIFLOWER* (*CAL*) or *AGL8*. The second clade AGL9 comprises *AGL9*, *TM5*, *FBP2*, while the third clade AGL6/AGL13 is represented by *AGL6* and *AGL13*. Some of these genes control floral meristem identity (*AP1* and *CAL*) whereas others (e.g., *AP1*) are important actors in floral organ identity of sepals and petals (Purugganan et al. 1995, Berbel et al. 2001). Furthermore, studies of quantitative trait loci (QTL) reported links between the *AP1* locus and flowering time as well as inflorescence branching patterns (Mandel and Yanofsky 1995).

Mechanisms controlling flower development and floral organ identity are highly conserved in evolution (Coen and Meyerowitz 1991). Flowers are organized in four concentric whorls of organs. The two outermost whorls comprise sepals and petals whereas the inner whorls consist of stamen and carpels as reproductive organs (Coen and Meyerowitz 1991). Floral architecture is determined by the overlapping activities of regulatory MADS-box genes and summarized in the **ABC(DE) model** (Coen and Meyerowitz 1991, Theissen et al. 2000, Theissen 2001, Lohmann and Weigel 2002, Erbar 2007; fig. 4).

A-function genes alone determine the organ identity of sepals whereas A- and B-function genes together determine petals. The B- and C-function genes are important for stamen organ identity while the C-function genes alone are responsible for carpels (Coen and Meyerowitz 1991). The classic ABC model was enlarged to the ABCDE model in 2001 when genes with D- and E-function were described (Vergara-Silva et al. 2000, Theissen 2001). D-function genes are important for the origin of ovules whereas E-function genes are crucial co-factors for the ABCD-function genes (Kramer and Hall 2005; fig. 4).

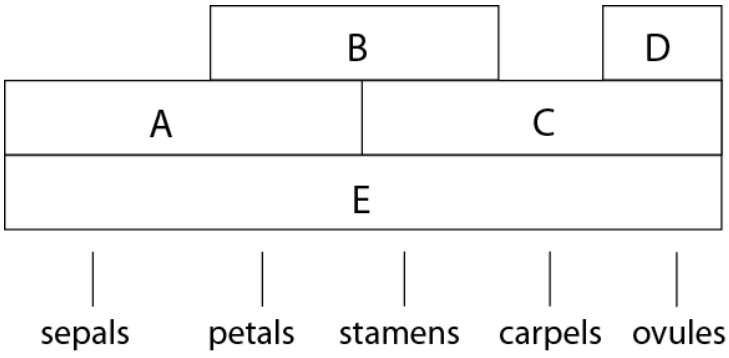


Fig. 4: The extended ABCDE model adapted after Erbar (2007).

In the present study, homologs of the A-function gene *AP1* and the B-function gene *AP3*, here afterwards named *MCS_AP1* and *MCS_AP3*, were selected to analyze their influence in the evolution of the MCS.

Since the selected genes exist in multigene families and the studied species are polyploid (e.g., Mes 1995, Lawton-Rauh 2003), **gene duplications** are expected to occur frequently in the MCS.

Gene duplications may provide the raw material for evolution (Ohno 1970, Baum 1998, Lynch and Conery 2000, Zhang 2003). They are common, frequent, and ongoing in organisms and without them, the plasticity of a genome or species in adaptation to changing environments would be limited (Lynch et al. 2001, Lawton-Rauh et al. 2003, Zhang 2003). They may arise via unequal crossing over, retroposition, or chromosome as well as genome duplications (Zhang 2003). Unequal crossing over usually generates tandem gene duplications. The duplicated region contains parts of a gene, an entire gene or several genes including the introns (Zhang 2003). If genes are duplicated by retroposition, messenger RNA (mRNA) is retrotranscribed to complementary DNA (cDNA) and then inserted into the genome. This process is connected with the loss of introns and regulatory sequences, the presence of poly A tracts and of flanking short direct repeats. The duplicated gene copies are usually unlinked to the original gene and often become pseudogenes immediately because they lack necessary transcription elements like promoters and regulatory sequences (Zhang 2003). Genome duplications are especially well known for plant species. 70-80% of the angiosperm species have undergone polyploidization at some point in their history (Moore and Purugganan 2005). The consequence of polyploidization is the duplication of the whole genome and thus, all genes are immediately doubled, which is in sharp contrast to single gene duplications.

Several scenarios and consequences could be assumed for duplicated genes: 1) nonfunctionalization, pseudogenization, and subsequently gene loss, 2) redundant maintenance, 3) subfunctionalization, and 4) neofunctionalization (Lynch and Conery 2000). According to Zhang (2003) many duplicated genes get lost. Furthermore, since gene duplications lead to functional redundancy, it is often not advantageous to have two identical copies (Baum 1998, Zhang 2003). It is likely that one gene copy acquires mutations that disrupt the structure and function of the copy which gradually becomes a pseudogene, whereas the other copy is involved in the proper function of the gene. Pseudogenization is believed to occur in the first million years after duplication if there are no other evolutionary processes acting on the duplicated gene copies (Lynch and Conery 2000, Zhang 2003).

If the presence of duplicates in the genome is beneficial gene conversion via concerted evolution or strong purifying selection can prevent duplicates from diverging or getting lost. The maintenance of two gene copies becomes more likely

when duplicates differ in some aspects of their functions (Duarte et al. 2006). This could be achieved through subfunctionalization where each gene copy takes part in the function of their ancestor gene. One important form, which seems to be a rule rather than an exception, is division of gene expression after duplication like for *AP1*/*CAL*/*SQUA* (see Duarte et al. 2006 or Shan et al. 2007). Another form can occur at the protein level where partitioning the tasks of the ancestral gene takes place and each copy is responsible for a unique set of subfunctions (Lynch et al. 2001, Zhang 2003). An example would be the *AP3*/*PI* homologs (Kramer et al. 1998).
One of the most important outcomes of gene duplication may be the origin of a novel function by relaxed purifying selection or positive selection (Zhang 2003). Relaxed purifying selection allows random mutations to become fixed in one gene copy and in a changing environment might induce an altered gene function. In contrast, if positive selection is involved in the development of a new gene function, accelerated fixation of advantageous mutation may occur (Zhang 2003). Moore and Purugganan (2003) showed that positive selection plays a key role in preserving gene copies and can act at early stages to maintain duplicates.

Selection pressure acting on genes might be different and may vary between retained orthologs and paralogs (see, e.g., Small and Wendel 2002 or Wang et al. 2007). In general, selection pressure acting on a gene could be distinguished into neutral, purifying, and positive selection and base on nucleotide differences. Nucleotide differences in the coding region have two different consequences. Either nucleotide alteration resulted in the same amino acid in the translated protein (synonymous or silent substitution) or in amino acid replacements (nonsynonymous substitutions). The latter case could have dramatic effects on the deduced proteins depending if very similar or quite different amino acids replace the original one (Hughes et al. 2000).
Based on the number of synonymous and nonsynonymous substitutions selection can be calculated. The selection pressure (ω) is defined as the ratio of Ka/Ks, where Ka denotes the pairwise nonsynonymous substitution per nonsynonymous site and Ks represents the number of synonymous substitution per synonymous site (Hurst 2002). If the ratio is nearly 1, neutral selection is expected indicating that chance alone determines whether a mutation will become fixed or not. Additionally, the likelihood of fixation in nonsynonymous substitutions is as high as those in

synonymous substitutions. When the ratio of Ka/Ks is less than 1, purifying selection is assumed that means selection against nonsynonymous nucleotide substitutions. Selection eliminates deleterious mutations and maintains the function of the protein. If the ratio of Ka/Ks is larger than 1, positive selection occurs where nucleotide substitutions are favored, fixed and changes in the amino acid composition lead to changes in the protein (Nekrutenko et al. 2001, Hurst 2002). In addition, if the value of Ka is significantly higher than that of Ks, the most probable assumption is that the gene has undergone adaptive evolution (Eyre-Walker 2006). So far, positive selection has been detected for genes involved in sexual reproduction (e.g., gamete recognition genes or self recognition genes in plants), for host-parasite interaction genes (e.g., plant resistance genes (R-genes), plant chitinases), for genes which encode for enzymes involved in energy metabolism (e.g., the pancreatic ribonuclease genes in leaf-eating monkeys), and for genes involved in adaptation to specific environments (e.g., regulatory genes involved in plant morphology; Bishop et al. 2000, Ford 2002, Barrier et al. 2003, Zhang 2003). Thus, the influence of genes, gene classes or gene copies in the speciation process could be estimated by calculating the selection pressure acting on a gene, gene region or gene copy.

Summarizing these paragraphs the following questions arise:

1.) Do the studied low-copy nuclear genes reflect the species relationships as inferred from other markers and are thus valuable tools to deduce and reconstruct phylogenetic relationships of the MCS and *Sedum* species?
2.) Can gene duplications and therefore different functional or orthologous and paralogous gene copies be detected for the MCS and/or *Sedum* species?
3.) If so, do gene copies evolve with same or different evolutionary rates and does the selection pressure acting on different gene copies vary?
4.) Do the selected regulatory or structural genes have different impact on the speciation process and what might have triggered the evolution of the MCS species?

The aim of this thesis is to find answers to these questions and to elucidate the difficult phylogenetic relationships within the MCS.

2. Materials and Methods

2.1. Study species

In the following section the studied species will be shortly introduced. After the species name sections referring to Lems (1960), Liu (1989), and Mes (1995), and the clade number corresponding to the combined cpDNA/nrITS phylogram of Mort et al. (2002) are mentioned for the *Aeonium* species, if possible (compare table 15 and fig. 2). Sections for the *Monanthes* species refer to Nyffeler (1992). For a better overview the most important characters are given in table 1. Information for the species are summarized from Lems (1960), Maire (1976), Kull (1982), Liu (1989), Lösch (1990), Nyffeler (1992), Hohenester and Welss (1993), Nyffeler (1995), Mes et al. (1997), Schönfelder and Schönfelder (1997), Mort et al. (2002), and Fairfield et al. (2004).

***Aeonium aureum* (C.Sm. ex Hornem.) T.Mes** (sect. *Greenovia* [Mes], clade 2) is a perennial terrestrial subshrub, 30-45 cm high and the most common species of the former genus *Greenovia*. Plants form rosettes and always daughter rosettes, leaves are glaucous and glabrous, and stems are unbranched with a lax branched inflorescence. Flowering time is between March and April and the golden yellow flowers are 20-35merous without nectaries. The species is endemic to the Canary Islands and grows on occasionally moist rocks, roofs, and on rocky surfaces in the pine region at 400-2000 m elevation.

***Aeonium canariense* Webb & Berthel.** (sect. *Canariensia*, *Patinaria*, *Canariensia*, clade 1) is a perennial terrestrial pilose herb which forms large rosettes. Stems are unbranched and up to 45 cm high, and the inflorescence up to 70 cm. Petals, 8-10, are very pale yellow-green to nearly white; flowers produce nectar and bloom between April and August. The species is found in the laurel forest on rocks, soil banks, and cliffs in fairly dry habitats up to 1300 m in the north of Tenerife.

***Aeonium cuneatum* Webb & Berthel.** (sect. *Canariensia*, *Patinaria*, *Canariensia*) is a typical representative of the Canarian laurel forest and is morphologically similar to *A. canariense*. It is a perennial terrestrial, but sometimes also epiphytic, unbranched herb which forms rosettes. Leaves are glabrous and glaucous. The inflorescence is 18-60 cm long; flowers are 8-9merous, bright or golden yellow and segregate nectar.

The species blooms between April and June and colonizes fairly moist habitats at elevations of 500-950 m. It is found on rocks and soil banks among bushes and occasionally on trees in the eastern and western part of Tenerife in the Anaga and Teno mountains.

***Aeonium goochiae* Webb & Berthel.** (sect. *Goochiae*, *Petrothamnium*, *Goochiae*, clade 2) is a perennial terrestrial viscid subshrub with rosettes and stems up to 40 cm, which are densely branched. Flowers are 7-8merous and the flower color is very pale yellow, rose or nearly white with a central portion of pink. The inflorescence comprises 10-45 flowers, which produce nectar and bloom between February and June. The plant is common on rocks, walls, and cliffs usually in the shadow of trees or rocks in fairly moist habitats. It is an endemic species of the humid north coast of La Palma at 100-700 m elevation.

***Aeonium nobile* (Praeger) Praeger** (sect. *Megalonium*, *Megalonium*, *Leuconium*, clade 4) is a monocarpic perennial terrestrial subshrub with stems up to 60 cm and large succulent leaves. It is rarely branching and the large inflorescence is flat-topped to broadly dome-shaped. Its dark red flowers are unique and the result of numerous reddish stripes on the 7-9 pale yellow petals. Flowers produce nectar and bloom between March and July. The species is endemic to La Palma where it can be found on oldest rock formations in dry slopes, soil banks, and cliffs in soil pockets and crevices, from the sea level up to 800 m.

***Aeonium rubrolineatum* Svent.** (sect. *Holochrysa*, *Aeonium*, *Aeonium*, clade 3) is a perennial terrestrial subshrub. Its stems are erect, branches often in groups, and the inflorescence is dense. Flowers are 9-11merous and yellow to pale yellow with reddish variegations caused by reddish veins or reddish bases and margins. Nectar is produced; plants bloom between May and November and loose all leaves during the flowering period. It colonizes soil banks and cliffs from 800-1200 m and is especially common in the SW sector of La Gomera.

***Aeonium saundersii* Bolle** (sect. *Goochiae*, *Petrothamnium*, *Petrothamnium*, clade 2) is a perennial terrestrial subshrub and unique in its balsam odor. Stems are up to 25 cm high and inflorescences have 5-70 flowers. These are 12-16merous, pale

yellow and do not segregate nectar. Flowering time is between April and June and the species is common on vertical rocks both in sun and shade. It is found in cervices where it grows rapidly, branching out and undergo a summer rest period with leaves forming rounded bud-like structures. It is endemic to La Gomera where it is found in the east at 150-800 m elevation.

***Aeonium smithii* Webb & Berthel.** (sect. *Goochiae*, *Chrysocome*, *Chrysocome*) is a perennial terrestrial unbranched herb with stems up to 60 cm. Plants are hirsute with multi-cellular trichomes. Number of sepals and petals is 8-12, mostly 10. The yellow flowers do not produce nectar and bloom between March and October. The species is common on rocks and cliffs in the pine forest zone and Cañadas between 150-2150 m. It is found in the south and east of Tenerife.

***Aichryson laxum* (Haw.) Bramwell** is an annual or biennial herb, pilose and often reddish. Stems are 15-30 cm high and regularly branched. Flowers are 8-12merous and bright yellow. The flowering time of the species is between March and June. The species colonizes moist crevices and walls and is occasionally epiphytic and endemic to the Canary Islands.

***Aichryson pachycaulon* Bolle** is an annual or biennial herb. Stems are thick, erect, up to 65 cm, and plants are glabrous. Flowers are pale to bright yellow, 7-8merous and bloom between October and April. *Aichryson pachycaulon* is classified into five distinct species or subspecies which cluster into three different clades (Fairfield et al. 2004). They show a range of morphological and cytological variation although they are generally characterized by the lack of pubescence. Each subspecies is a single island endemic and *Ai. pachycaulon* subsp. *immaculatum* occurs on Tenerife. The species colonizes mesic habitats like rocks in the laurel forest of the Canary Islands.

***Monanthes anagensis* Praeger** (sect. *Sedoidea*) is a slightly woody, diffusely branched shrublet up to 30 cm high. Leaves are ordered alternative and never covered with wax. The terminal inflorescence is regularly branched with 3-18 flowers. Flowers are 6-8merous, pale yellow, occasionally with brown-red stripes and with nectaries. The species is restricted to higher elevations of the Anaga Mountains on Tenerife from 600-900 m and locally common. It mostly occurs on moist rocks and

cliffs and sometimes in open and more arid places. It is a characteristic taxon of the laurel forest zone but partly occurring in the lowland xerophytic zone as well.

***Monanthes icterica* (Webb ex Bolle) Praeger** (sect. *Monanthoidea*) is a dwarf annual herb, reaches up to 6 cm height, is rather unbranched with alternate leaves and glabrous. It is the only known annual species of the otherwise perennial *Monanthes*. The inflorescence is terminal, regularly branched with 3-7 flowers. These are pale yellow, often slightly brown-red striped, and 6-7merous with nectaries. The species is distributed on Tenerife and La Gomera on ledges and in crevices of rocks and cliffs from 100-900 m.
The habit of *M. icterica* resembles species of the genus *Aichryson* but the enlarged nectaries, the entirely glabrous axes and leaves, and the bladder idioblasts place this species clearly in *Monanthes*.

***Sedum caeruleum* L.** is an herbaceous glabrous green plant with spots of red. The sepals often have a black spot at the base and the flowers are 5-9merous, normally 7merous. The single terminal inflorescence is not compact; flowers are star-shaped and azure or rarely white. They have very small white nectaries and flowers between March and June. Plants colonize stone fields, boulders, clearance with stones, and sometime banks of small rivers from sea level up to middle mountains. It is distributed in well watered regions in Southern Europe especially in Italy, Corsica, Sardinia, Sicily, and Malta.

***Sedum jaccardianum* Maire** is an herbaceous sticky plant. The petals are outside citron yellow and inside yellow with an orange zone in the upper part. The number of floral organs range between 6 and 10 and nectaries are very small and whitish. The species flowers between May and July and colonizes calcareous stones from 1600 m up to 2800 m. *Sedum jaccardianum* is endemic in Northern Africa.

***Sedum modestum* Ball** is a very small herb with rosettes. Its leaves are glabrous but hairs could be found at the back of the petals. Petals are golden yellow and later white with purple spots. Flowers are star-like in shape, 5-7merous and have nectaries. The inflorescence is terminal with one group of flowers and not compact. Flowering time is between April and June. The species colonizes stone fields, soil

filled crevices, and stumps from the sea level up to 2200 m in semiarid and well watered areas and is endemic in Northern Africa.

***Sedum pubescens* Vahl** is an herbaceous sticky plant. Flowers are bright yellow and have a star-like shape, number of floral organs is 5-6 and flowers have nectaries. The inflorescence is large, not compact, with only one flower at the end and flowering time is between May and July. Plants grow in forests, in low open scrublands with many evergreen shrubs (Garrigue), in understorey, field of stones, and in crevices of calcareous or silicate stones from 0-1300 m. The species occurs in semiarid and well watered regions of Northern Africa and Southern Europe.

***Sedum surculosum* Coss.** is an herbaceous perennial, glabrous plant with sessile loose flat rosettes. Flowers are 5-7merous and have a star-like shape. They are pale yellow with slight red-brown stripes and have small nectaries. Species flowers between July and August and colonizes damp rocks and cliffs. Two different species varieties can occur either on granitic and porphyritic rocks or on limestone (Nyffeler 1992). *Sedum surculosum* can be found along streams of high mountains from 2400-3800 m. The species is endemic to Northern Africa and distributed in Morocco in the Great Atlas and Anti-Atlas.

The taxonomic history of *S. surculosum* is quite interesting. In 1873 J. Ball described the species as *Monanthes atlantica* but in the same year a plant from the Moroccan Atlas Mountains was described as *S. surculosum* by E. Cosson. Berger (1930) retained it in the genus *Monanthes* whereas Praeger (1932) excluded it because of its broad yellow petals, which are more *Sedum* or *Aichryson*-like, and *S. jaccardianum* seems to connect *S. surculosum* with the *Sedum* species (Nyffeler 1992). Nevertheless, it was often still included in the genus *Monanthes* because of its large nectariferous scales and resemblances in habit and flower morphology (see Nyffeler 1992 or Mes et al. 1997). However molecular data strongly support the inclusion in the genus *Sedum* because it shares with *S. jaccardianum* a unique 70 base pair deletion in the cpDNA *trnL-trnF* intergenic spacer (Mes and ‘t Hart 1994).

Table 1: Overview of the main characteristics of the studied species.

taxon	distribution	CO_2 fixation type	floral organs	flower color	habit	habitat
A. aureum	GC,T,H	C3 - weak CAM	28-32	golden yellow	herbaceous	moist
A. canariense	T	C3 - weak CAM	8-10	pale yellow green	herbaceous	fairly dry
A. cuneatum	T	C3 - weak CAM	8-9	golden yellow	herbaceous	fairly moist
A. goochiae	P	C3 (CAM possible)	7-8	rose	subshrub	moist
A. nobile	P	strong CAM	7-9	dark red	subshrub	dry
A. rubrolineatum	G	C3 (CAM possible)	9-11	yellow	subshrub	dry
A. saundersii	G	C3	12-16	pale yellow	subshrub	mesic
A. smithii	T	strong CAM	8-12	yellow	herbaceous	dry
Ai. laxum	T,GC, P,H,G	C3	6-12	bright yellow	herbaceous	moist
Ai. pachycaulon	T,GC, P,F,G	C3	7-8	pale yellow	herbaceous	mesic
M. anagensis	T	C3	6-8	yellow	shrublet	moist
M. icterica	G,T	C3	6-7	pale yellow	herbaceous	mesic
S. caeruleum	S Europe	C3	5-9	azure, rose	herbaceous	mesic/moist
S. jaccardianum	MO	C3	6-10	citron yellow	herbaceous	mesic/moist
S. modestum	MO	C3	5-7	golden yellow	herbaceous	mesic/moist
S. pubescens	N Africa, S Europe	C3	5-6	bright yellow	herbaceous	mesic/moist
S. surculosum	MO	C3	5-7	yellow	herbaceous	moist

2.2. Molecular methods

2.2.1. RNA material

Extracted RNA was provided by Dr. M. Thiv (SMNS, Stuttgart). A combination of fresh bud material of different stages of several *Aeonium* species was collected in the Botanical Garden of the University Zurich and in the City Succulent Collection Zurich. RNA was extracted following the standard protocol for the Concert™ Plant RNA Reagent Kit (Invitrogen) and used for subsequent cDNA syntheses using Superscript II reverse transcriptase (Invitrogen).

2.2.2. DNA material

Genomic DNA was extracted using the DNeasy Plant Mini Extraction Kit (Qiagen). Living plants for most of the studied species were collected on Tenerife and La Gomera (MCS) or in Morocco (*Sedum*; table 16; appendix). Plants were cultivated at the State Museum of Natural History Stuttgart where vouchers of the species are also deposited. Material of *A. goochiae* and *A. nobile* was provided by Prof. Dr. R. Lösch (University Düsseldorf), material of *S. caeruleum* by Mrs. Lübenau-Nestle, and *S.*

pubescens by the City Succulent Collection Zurich (table 16; appendix). Leaves used for DNA extraction were harvested, grinded with liquid nitrogen and DNA subsequently extracted following the instructions of the supplier.

2.2.3. Amplification of the low-copy nuclear genes

Homologs of the floral homeotic regulatory genes *AP1* and *AP3* as well as of the structural gene *PEPC* were amplified comprising the following steps: 1) cDNA synthesis and subsequently the use of degenerated primers to amplify the respective gene region from cDNA using GoTaq polymerase (Promega). 2) The obtained sequences were aligned and blasted to check the sequence identity by similarity comparisons using the National Center for Biotechnology Information (NCBI) database. If the amplified sequences were confirmed as homologs of the target genes, 3) specific primers for each gene were deduced and used for genomic DNA amplification with proofreading enzymes (Pfu: Fermentas or Promega; Phusion: NEB). Proofreading enzymes correct incorporation of nucleotides in 3´ to 5´ direction and therefore reducing the number of PCR mistakes during template elongation.

All PCR reactions were performed in a final volume of 10 µl and components as indicated in table 2. PCR reactions were done in a T-Gradient cycler (Biometra, Göttingen) or GeneAmp 9700 PCR System (PE Biosystems, Foster City, CA). Amplified PCR products were separated on 1% ethidium bromide stained agarose gels to prove size and quality and were subsequently cloned and sequenced as described below.

MCS_PEPC

Direct amplification from genomic DNA was possible using the primers PEPC-F and PEPC-R (table 17; Gehrig et al. 1995). Initial PCR reactions were done using GoTaq (Promega) under the following conditions: initial treatment of 95°C for 3 min., 33 cycles of [95°C 45 sec., 52.5°C 50 sec., 72°C 2.5 min.] and a post treatment of 72°C for 7 min. before cooling.

For final reactions annealing temperature and elongation time were optimized and amplification from genomic DNA done with Pfu polymerase (Fermentas; table 2) under the following conditions: initial treatment of 95°C for 3 min., 33 cycles of [95°C 50 sec., 53°C 50 sec., 72°C 5 min.] and a post treatment of 72°C for 15 min. before cooling. For several species Pfu polymerase (Promega; table 2) was used and the

PCR was performed under the following conditions: initial treatment of 95°C for 3 min., 35 cycles of [95°C 45 sec., 52.5°C 50 sec., 72°C 5 min.] and a post treatment of 72°C for 15 min. before cooling.
For the *Sedum* species further primer optimizations were done based on conserved regions of the MCS species. For PCR reactions Phusion (NEB; table 2) was used, the primer combination PEPC_Sed_for_2/PEPC_Sed_rev (table 17), and reactions were performed under the following conditions: initial treatment of 98°C for 30 sec., 30 cycles of [98°C 10 sec., 64°C 30 sec., 72°C 2 min.] and a post treatment of 72°C for 10 min. before cooling.

MCS_AP1

cDNA synthesis with a specific poly(T)-AP1 primer (Litt and Irish 2003; table 17) was done following the instruction of the supplier (Invitrogen). For the first PCR primer AP1MDS3 (Litt and Irish 2003; table 17) and AP1-noT (table 17) were used and reactions were performed under the following conditions: initial treatment of 94°C for 3 min., 35 cycles of [94°C 50 sec., 50°C 50 sec., 72°C 2 min.] and a post treatment of 72°C for 10 min. before cooling. Amplified fragments were separated on a 1% ethidium bromide stained agarose gel, excised from the gel, and purified with GFX gel purification columns (GE Healthcare). Purified products were 1:25 diluted and used for nested PCR with the primer combination AP1MDS2/SQUAR (Litt and Irish 2003; table 17) as followed: initial treatment of 95°C for 3 min., 33 cycles of [95°C 50 sec., 55°C 50 sec., 72°C 2 min.], followed by post treatment of 72°C for 5 min. before cooling.
Finally the specific primers AP1-11F and AP1-704R (table 17) were used for genomic DNA amplification with Pfu (Promega; table 2) and the PCR program: initial treatment 95°C 3 min., 30 cycles [95°C 1 min., 54°C 50 sec., 72°C 5 min.] and post treatment of 72°C for 15 min. before cooling.
For the *Sedum* species further primer optimizations were done based on conserved regions of the MCS species. Primer combination AP1_Sed_for/AP1_Sed_rev (table 17) was used, Phusion (NEB; table 2), and PCR reactions were performed under the following conditions: initial treatment of 98°C for 30 sec., 30 cycles of [98°C 10 sec., 64°C 30 sec., 72°C 2 min.] and a post treatment of 72°C for 10 min. before cooling.

MCS_AP3

Two approaches were followed to develop specific primers for *AP3* homologs. For the first one cDNA synthesis was done with random primers as described by the supplier. Subsequently, first PCR reactions were performed with each time 1.1 µl cDNA and the primer combination ATG3/AP3-polydT (Kramer pers. communication; table 17) as followed: initial treatment of 95°C for 1 min., 5 cycles of [95°C 30 sec., 60°C 3 sec. with ramp of -0.2°C/sec. to 50°C, 72°C 50 sec.] followed by 35 cycles of [95°C 30 sec., 60°C 30 sec., 72°C 2 min.] and a post treatment of 72°C for 7 min. before cooling. Amplified fragments were separated on a 1% ethidium bromide stained agarose gel, excised from the gel between 500 and 1000 bp, and purified with GFX gel purification columns (GE Healthcare). Purified products were 1:1 diluted and nested PCR reactions with the primers MADS4 and AP3-polydT (Kramer pers. communication; table 17) were performed as followed: initial treatment of 95°C for 3 min., 33 cycles of [95°C 50 sec., 53°C 50 sec., 72°C 2 min.], followed by post treatment of 72°C for 5 min. before cooling.

Based on these sequences primers for the MCS species were improved and used in the second approach. Here, cDNA synthesis was done with the specific AP3-polydT primer (table 17) following the instruction of the supplier and subsequently the primer combinations MADS4-Aeo/AP3-724R, AP3-11F/AP3-724R, AP3-351F/AP3-polydT, and AP3-351F/AP3-724R (table 17) were tested using 1 µl diluted cDNA as indicated: initial treatment of 95°C for 3 min., 33 cycles of [95°C 50 sec., 53°C 50 sec., 72°C 2 min.], followed by post treatment of 72°C for 5 min. before cooling. In addition, another first PCR was performed with the primer combination ATG3/AP3-noT2 (table 17) as followed: initial treatment of 95°C for 3 min., 33 cycles of [95°C 50 sec., gradient of 50°C-60°C 50 sec., 72°C 2 min.], followed by post treatment of 72°C for 5 min. before cooling. Afterwards probes were 1:20 diluted and used for the second PCR with a successful amplification for the primer combination AP3-351F/AP3-724R (table 17) as followed: initial treatment of 95°C for 3 min., 33 cycles of [95°C 50 sec., 55°C 50 sec., 72°C 2 min.], followed by post treatment of 72°C for 5 min. before cooling.

Finally primers were optimized for genomic DNA and the combination PI-F-Aeo/AP3-724R (table 17) was used with Pfu (Promega; table 2) and the following PCR program: initial treatment 95°C 3 min., 30 cycles [95°C 1 min., 54°C 50 sec., 72°C 5 min.] and post treatment of 72°C for 15 min. before cooling.

For *Sedum* further primer optimizations were done based on exon regions of the MCS species. GoTaq (table 2) was used because amplification and/or subsequent cloning failed using proofreading enzymes. The primer combinations PI-F-Aeo/AP3-1766R (comprises the MCS specific forward primer) as well as AP3-82F/AP3-1766R (both *Sedum* specific; all table 17) were used and PCR cycling was performed under the following conditions: initial treatment 95°C 3 min., 30 cycles [95°C 50 sec., 55°C 45 sec., 72°C 2 min.] and post treatment of 72°C for 10 min. before cooling.

Table 2: Overview of the components used in PCR reactions for the different polymerases.

	GoTaq (Promega)	Pfu (Promega)	Pfu (Fermentas)	Phusion (NEB)
buffer	1x	1x	1x	1x
$MgSO_4$	-	-	2 mM	-
dNTPs	0.2 mM each	0.4 mM each	0.2 mM each	0.2 mM each
primer each	0.325 µM	1.0 µM	1.0 µM	0.5 µM
polymerase	0.05 U/µl	0.032-0.048 U/µl	0.025 U/µl	0.02 U/µl
DNA	30-50 ng	10-30 ng	10-50 ng	30-80 ng

Since no information concerning the copy number was available for MCS and *Sedum* species, one main question was if there are gene duplications within the MCS and/or *Sedum* species, and if orthologous and/or paralogous gene copies could be detected. Deduced primers were therefore specific for the gene region and species of interest but not for potential gene copies and allow a screening for copy number in the studied species groups.

2.2.4. Cloning of the low-copy nuclear genes

PCR products of the low-copy nuclear genes showed only a single band. However, used primers were just gene specific whereas species of the MCS are polyploid and the selected genes represented in multigene families. Thus, to prove and verify the number of amplicons, PCR products were cloned using the pGEM-T Easy vector system (Promega; T-cloning technique) or the Jet cloning kit (Fermentas; blunt-end cloning technique). Instructions of the supplier were followed but only half of the reaction mix containing buffer, vector, and ligase was used. PCR products resulting from amplification with GoTaq were cloned directly using the pGEM-T Easy vector.

GoTaq creates A-overhangs at PCR products which directly ligate to the T-overhang of the provided cloning vector. In contrast, PCR products of proofreading enzymes miss A-overhangs. Using the pGEM-T Easy vector system additional A-addition was done using GoTaq and following the instructions of the supplier, or PCR products were cloned directly into the Jet blunt-end cloning vector. Ligation was done for 1h at room temperature or over night at 4°C. The ligation reaction mix was transformed to 50 µl of competent cells (XL1 blue-cells) via heat-shock reaction for 60-90 sec. in a 42°C warm water bath or heat block and immediately returned on ice for 2 min. Transformed cells were grown in 700-900 µl SOC medium for 1.5 h at 37°C with shaking at ~300 rpm. 150-200 µl of the transformation culture were plated on LB/ampicillin/IPTG/X-Gal plates and grown over night at 37°C. Positive clones were selected by blue-white screening for the pGEM-T Easy vector system. Using Jet cloning only positive clones grow on ampicillin plates.

2.2.5. Colony PCR

Cloning was followed by colony PCR. In average, eight white clones per gene and taxon were selected randomly and amplified using the vector specific primers M13 (Promega; table 17) or Jet (Fermentas; table 17).

Reactions were performed in 20 µl final volume with one randomly selected white colony using Pfu (Promega) or GoTaq (Promega) with components as indicated in table 2, and the following program: initial treatment at 95°C for 3 min., 33 cycles of [95°C 50 sec., 55°C 50 sec., 72°C 12 min.], followed by a final elongation of 20 min. at 72°C before cooling. Using GoTaq (e.g., for *MCS_AP3* of *Sedum*) an elongation time of 3 min. was used and an annealing temperature of 57°C for the vector specific primer Jet.

Screening for successful amplification of the insert was done on 1% ethidium bromide stained agarose gels. Five successful amplified PCR products (equivalent to clones or colonies) were chosen randomly, purified using Nucleofast 96 PCR plates (Macherey-Nagel), and re-suspended in 20 µl TE buffer. Quantity and quality of purified PCR products were checked on a 1% ethidium bromide stained agarose gel and the amount of template product for the sequencing reaction estimated.

2.2.6. Amplification of nrITS and cpDNA regions

For the studied species the rDNA nrITS region and several cpDNA regions (*matK*, *trnL-trnF*, and *psbA-trnH* spacer regions) were amplified using primers and conditions as described by Mort et al. (2002). Forward and reverse sequences for each taxon were aligned. Obtained consensus sequences were used to prove species identity and to improve the nrITS alignment provided by Dr. M.E. Mort (University of Kansas) for subsequent phylogenetic species reconstructions.

2.2.7. Sequencing

Cycle sequencing was performed in a T-gradient cycler (Biometra) with the following cycling program: initial denaturation at 96°C for 2 min., 25 cycles of [96°C 10 sec., 50°C 5 sec., 60°C 4 min.] with a ramp of 1°C/sec. and immediately cooled to 4°C. For clones that were amplified with the primers Jet, a higher annealing temperature of 57°C was used while cycle sequencing to assure sequence quality.

Reactions were performed using 5 µl final volume with maximal 2.75 µl of the purified PCR product, 0.5 µM primer, 0.83 µl of 5x sequencing-buffer, and 0.33 µl BigDye Terminator (version 3.1., Applied Biosystem).

In addition to the vector specific primers M13 or Jet, several internal sequencing primers (table 17) were deduced for the low-copy nuclear genes to generate overlapping sequence regions. For the nrITS and cpDNA regions amplification primers were used.

Sequencing products were purified on 96-well Sephadex (GE Healthcare) plates and dried products were sent for sequencing. Sequencing was done on a capillary sequencer (ABI PRISM; Applied Biosystems) in the Institute of Systematic Botany, University of Zurich.

2.2.8. Sequence analysis

Obtained chromatograms of all sequences were checked and edited manually. Consensus sequences for each clone were obtained based on overlapping sequence regions using Sequencher 4.2.1. (Gene Codes).

Consensus sequences were aligned with help of Sequencher, manually improved, and nucleotide substitutions were carefully checked by comparison with the chromatogram of the original sequence.

Final alignments comprising all consensus sequences of respective clones were done using Sequencher and manually improved in Se-Al V2.0a11 (Rambaut 1996) and MEGA version 4 (Tamura et al. 2007).

2.2.9. Definition of the gene regions

Since amplification based on genomic DNA, coding and noncoding regions (exons and introns) could be defined following the GT/AG rule for excisions of introns. Comparisons were done with respective cDNA sequences of MCS species and/or with coding sequences of *Kalanchoe* species that were obtained from the NCBI database. *Kalanchoe gracilis* (AJ252946) and *K. pinnata* (AJ252919) were used for *MCS_PEPC*; *K. blossfeldiana* (DQ479358) for *MCS_AP3*.

Three datasets were obtained for each gene: a full-length dataset including all information of amplified templates (total amplified genomic DNA). Exclusion of exon regions will result in the intron dataset whereas via excluding the intron partitions the exon dataset is obtained. The exon alignment was carefully checked for the open reading frame (ORF) and adapted to it by translation into amino acid alignments in MEGA.

Furthermore, unambiguous alignments of intron sequences between genera are not trivial. Therefore, the *Sedum* sequences were removed, arising gaps deleted, and phylogenetic reconstructions done based on exclusive MCS intron regions.

2.2.10. Improvements of the datasets

Full-length nucleotide alignments for each specific taxon were compared and identical sequences (clones) excluded from datasets to decrease calculation time.

Additionally, datasets were checked for PCR mediated chimeric sequences. Occurrence of chimeric sequences is likely when similar templates are amplified within a single PCR reaction. Testing was first done by careful comparison of taxon specific sequences by eye, searching for recombination and break points. Afterwards checking was done with help of the Bellerophon web server that uses a partial tree building approach (Huber et al. 2004). Phylogenetic reconstructions are done with sequence fragments before and after an assumed break point and the topologies compared to infer incongruence that may indicate chimeric sequences. Here, the whole dataset was compared using four different correction methods (Huber-Hugenholtz, Kimura, Jukes-Cantor, and no correction). Aligned datasets and an open

window size of 300 were used and only adapted for the *MCS_AP1* dataset to an open window size of 200. Full-length and exon datasets were checked and the detected potential chimeric sequences again carefully proved by eye. Sequences resulting with every method and after final proof as chimers were excluded.
Next to redundant and chimeric sequences, sequences with single base pair deletions that lead to frameshift mutations and premature stop codons in the exon alignment were excluded based on comparisons of the exon and protein alignments.
Each final alignment was converted into nexus and phylip formatted files for phylogenetic reconstructions. MEGA formatted files were used for further analyses as indicated below.

2.2.11. Partition Homogeneity Tests

Conflicting signals in datasets could be estimated using the Partition Homogeneity Test (PHT) as implemented in PAUP* 4b10 (Swofford 2002). The reduced dataset of each of the three gene regions was used to estimate conflicting evolutionary signals concerning exon and intron regions. Datasets were changed by combining all exon and intron regions together and define them as data partitions. 1000 replicates were performed using the heuristic search modus. Starting trees were obtained by stepwise addition with random addition of sequences and 100 replicates. Number of trees hold in each step during random sequence addition was 1. Using TBR branch-swapping no more than 10 trees were saved. If the resulted p-value is significant ($p < 0.05$) there is a conflicting signal in the dataset between exon and intron sequence evolution. If the p-value is not significant no conflicting signal could be detected.

2.2.12. Phylogenetic reconstructions

The character based methods Bayesian Interference (BI) and Maximum likelihood (ML) were used to reconstruct the phylogenetic history of the studied genes and species. The advantage of both methods is the integration of a model of sequence evolution. The model of sequence evolution was first calculated using Modeltest 3.7 and the model defined by the Akaike information criterion (AIC) was selected (Posada and Crandall 1998). The obtained model was used to define the parameters for each BI analysis using MrBayes 3.1.2 (Huelsenbeck and Ronquist 2001, Ronquist and Huelsenbeck 2003) and for ML analyses using PHYML 3.0 (Guindon et al. 2005). Default settings (temperature 0.2, four chains, random starting tree, and

sampling every 100th generation) were used for BI. Only in one case, where after a reasonable high number of generations the standard deviation of the split frequencies of the four chains was not below 0.01, the analysis was stopped and restarted with the lower temperature of 0.1 (see table 9). If the standard deviation of the split frequencies for the chains was < 0.01 the analysis was finished. Based on the number of generations, posterior probabilities (pp) as well as trees were summarized using a burn-in of 30%.

For ML the web-based PHYML 3.0 program was used (Guindon and Gascuel 2003, Guindon et al. 2005). Based on the obtained model of sequence evolution all parameters needed for the ML analysis were specified, especially the model of sequence evolution and the discrete gamma-model, if possible. As tree topology search strategies NNI and SPR were used. A random BIONJ tree was generated as starting point and in total five replicates performed. Bootstrap support (bs) was obtained and the number of bootstrapped datasets range between 300 and 1000 depending on the computer resources and calculation time.

All trees were opened in TreeView 1.6.6 (Page 1996), rooted with the respective outgroup and ordered. Posterior probabilities and bootstrap support are indicated at the nodes and final trees were graphically improved in Adobe Illustrator 11.0.0.

2.2.13. Blast analyses

Obtained sequences were blasted against known sequences in the NCBI database to confirm results deduced from phylograms. The megablast search algorithm (highly similar sequences) was used and only if no results could be obtained the blastn search algorithm (somewhat similar sequences).

2.2.14. Neighbor-joining reconstructions for homologs of *PEPC*, *AP1*, and *AP3*

To infer the genealogy over a broader sampling and to confirm orthologous and paralogous copies for the studied genes and species, Neighbor-joining (NJ) analyses were done. Therefore, sequences that showed highest sequence similarity and randomly picked sequences from the NCBI database were added to the respective alignments. Alignments were done and improved based on translated amino acids and reverted nucleotide alignments subsequently used to infer the NJ phylogram. Alignments, translations, and analyses were done in MEGA with the following settings: all sites included, complete deletion of gaps, default model of “Maximum

Composite Likelihood", and 1,000 re-sampled pseudoreplicates to obtain bootstrap support.

2.2.15. Species phylogeny

To compare gene trees with the species phylogeny an existing nrITS dataset (M.E. Mort, University of Kansas) was used and own sequences added. Improvements of the given alignment were made and the resulting dataset used for a BI analysis.

The BI tree with branch lengths was furthermore used to construct an ultrametric tree using r8s v.1.70, a program for reconstructing divergence times and absolute rates of substitutions (Sanderson 2003). For this purpose a rate smoothing approach using penalized likelihood (PL) with the TN algorithm as well as nonparametric rate smoothing (NPRS) with Powell algorithm were used (Sanderson 1997). The age of the oldest island (Fuerteventura = 20.7 My) was used to calibrate the node of the Macaronesian taxa. The cross-validation procedure for PL was performed to determine the appropriate smoothing levels and was set to S = 1. Both methods resulted in highly similar ages. Therefore the ultrametric tree was constructed based on NPRS and Powell algorithm.

2.2.16. Nucleotide differences, replacements, and amino acid substitutions

Next to estimate phylogenetic relationships, exon and deduced amino acid alignments were used for further analyses. Based on exon phylograms orthologous and paralogous gene copies were distinguished and this information used for all further analyses.

Numbers of nucleotide differences were counted based on exon alignments and numbers of replacements based on deduced protein alignments. Total numbers of substitutions between all species-specific sequences were counted which corresponds mainly to paralogous gene copies (sequences between subclades). In a second approach it was distinguished in orthologous gene sequences (sequences within a subclade). Based on protein alignments amino acids were characterized and proved if the replaced amino acid was very different to the original one concerning charge, size, and further characteristics and if so, afterwards named and counted as quite different amino acid (Stryer 1995).

2.2.17. Relative Rate Tests

Differences in the evolutionary rates of duplicated gene copies were estimated with Tajima's relative rate test (RRT) as implemented in MEGA. Equality of evolutionary rates between two paralogous sequences in reference to an outgroup sequence is tested. P-values less than 0.05 are used to reject the null hypothesis of equal rates between lineages. Thus, if the null hypothesis is rejected, molecular clock can be rejected as well and sequences evolve differently.

2.2.18. Selection pressure

The selection pressure acting on the respective genes was estimated with three different methods. In advance, a further reduction of identical species-specific exon sequences was done. Furthermore, the *Kalanchoe* sequences were excluded from the *MCS_PEPC* and *MCS_AP3* datasets to calculate the selection pressure within the studied species of MCS and *Sedum*.

First, Ka/Ks-values were estimated that define the selection pressure acting on a gene with Ka/Ks = 1 indicating neutral selection, Ka/Ks < 1 purifying selection, and Ka/Ks > 1 positive selection. Ka- and Ks-values were obtained by calculating p-distances with the Nei-Gojobori method and complete deletion of gaps as implemented in MEGA. The Nei-Gojobori method calculates the numbers of synonymous and nonsynonymous substitutions and the numbers of potentially synonymous and nonsynonymous sites to define Ks and Ka, respectively (Nei and Gojobori 1986). Jukes-Cantor correction was used to correct the computed p-distances for multiple substitutions at the same site and standard errors for Ka and Ks were estimated from 1,000 bootstrap pseudoreplicates. Obtained Ka- and Ks-values were used with three decimal positions to calculate Ka/Ks-values (ω), mean Ka/Ks-values, and standard deviations in Excel 2003 (Microsoft). In addition, Ka/Ks-values were classified in specific ranges and presented in balk-diagrams as proportion of comparisons.

Two-tailed t-tests with unequal variances as implemented in Excel 2003 were done to prove for significant differences of Ka/Ks-values. Tests were done between 1) copy A and copy B of the MCS species; 2) gene copies of the MCS to sister- as well as outgroup species, and 3) regulatory and structural genes.

Next to the calculation of Ka/Ks-values, selection pressure was tested for each gene with the codon-based Z-test for large samples as implemented in MEGA. The null hypothesis of neutrality was compared with the alternative hypotheses of positive and purifying selection. Analyzed datasets were checked twice for each region, first, the whole dataset including MCS and *Sedum* species and second comprising only MCS species. Analyses were done with the Nei-Gojobori method with p-distance or Jukes-Cantor distance, and 1,000 bootstrap re-samples with complete deletion of gaps.

Verifying positive selection using estimation of Ka/Ks-values is difficult because an average over the whole protein sequence is done (Zhang 2003). Other methods allow estimation of positive selection acting on single amino acid sites like the one implemented in the Selecton web server (Doron-Faigenboim et al. 2005, Stern et al. 2007). Selecton enables to test the hypothesis of positive selection by comparing different evolutionary models with each other. One of the selected models (M8) allows for positive selection and is compared with the null model (M8a) which assumes that there is no positive selection by setting Ka/Ks = 1. Testing the null hypothesis of positive selection with the M8 model, the first result is a graphic indicating with different colors the selection pressure acting on each amino acid site. If there is evidence for positive selection the alternative model (M8a) is tested. Subsequent likelihood ratio tests (LRT) estimate which model explains best the analyzed data and indicate significance for positive selection.
Selecton enables the user to supply an accurate phylogeny to improve the results. Therefore, BI phylograms were implemented in the analyses. Testing positive selection within the *MCS_PEPC* gene region was done with two different datasets. The first excludes the *Kalanchoe* sequences and the four *Sedum* sequences S_pubescence_1c43 and 243 as well as S_surculosum_34341 and 54343, whereas these *Sedum* sequences were included in the second analysis. Two datasets of *MCS_AP1* were used to search for positive selection, one containing all analyzed sequences and the second excluded the highly unique sequences of A_aureum_13637 and 1b3637 as well as A_saundersii_33637. From the *MCS_AP3* dataset the *Kalanchoe* sequence was excluded and the test of positive selection performed.

3. Results

3.1. Datasets

Alignments of *MCS_PEPC*, *MCS_AP1*, and *MCS_AP3* were carefully checked for sequences of chimeric origin, with ambiguous sites, and single base pair (bp) deletions that result in frameshifts and premature stops in exon regions and translated proteins. Sequences showing these features were excluded from the datasets in a second approach although observation based on these sequences will be discussed.

For each studied species one individual plant was used to extract DNA and to amplify the gene of interest. Nevertheless, each species can be represented by several different cloned sequences (named as clones or sequences) which may represent different alleles or gene copies present in the species.

Identical clones are represented by a single sequence in the dataset and reduced datasets were used for the analyses.

MCS_PEPC

In total, 83 sequences of *MCS_PEPC* homologs were obtained (table 3).

Table 3: Number of obtained total and reduced cloned sequences for *MCS_PEPC*.

taxon	total number	reduced number
A. aureum	8	4
A. canariense	5	3
A. cuneatum	5	3
A. goochiae	5	5
A. nobile	5	4
A. rubrolineatum	6	4
A. saundersii	5	5
A. smithii	3	3
Ai. laxum	4	2
Ai. pachycaulon	5	3
M. anagensis	5	5
M. icterica	5	3
S. caeruleum	5	2
S. jaccardianum	5	5
S. modestum	0	0
S. pubescens	6	3
S. surculosum	6	5
	∑ 83	∑ 59

For *S. modestum* amplification failed and five sequences were excluded: Ai_pachycaulon_313T showed a chimeric sequence type and an indel of one nucleotide that will result in a frameshift mutation including several premature stop codons. Likewise of potential chimeric origin and excluded were A_cuneatum_402, A_nobile_361, A_canariense_41321, and M_icterica_472. Reduction of identical species-specific clones resulted in a dataset comprising 59 sequences (table 3).
In total, aligned sequences have a full-length of 1708 bp and three exons and two introns were detected (fig. 5, table 4).

Fig. 5: Schematic exon and intron structure of the *MCS_PEPC* gene sequences. Exons are shown as boxes and introns as lines. The relative length of the respective parts is given in table 4.

The exon alignment has a length of 1107 bp resulting in a translated protein of 369 amino acids. Within exon 3 up to 15 nucleotides were exclusive for the four unique sequences of *S. pubescens* and *S. surculosum* (compare below).

Table 4: Exon and intron positions and lengths of the *MCS_PEPC* sequences.

	position	length
exon 1	1 – 429	429
intron 1	430 – 799	370
exon 2	800 – 1186	387
intron 2	1187 – 1415	229
exon 3	1416 – 1708	293

The intron alignment comprises 599 bp but intron regions are quite divers and unambiguously alignable rather for taxa within the same genus. Therefore, MCS or *Sedum* species were removed from the alignments, respectively. Arising gaps were deleted but no further improvements of the MCS or *Sedum* intron alignments were done. The MCS intron alignment comprises 177 bp whereas the intron region of the *Sedum* species alone would result in 578 bp.
Deletion of the first intron was observed for *S. jaccardianum* and for S_surculosum_44318 and 44342.

MCS_AP1

The most incomplete dataset was obtained for *MCS_AP1*. No amplification products were obtained for *Ai. laxum*, *Ai. pachycaulon*, and *S. pubescens*. Sequencing failed partly for *S. caeruleum*, therefore, only four exon sequences are available.

Three sequences were excluded because of chimeric origin: A_aureum_32223, A_goochiae_12223, and A_nobile_3b3637. The first one showed additionally a premature stop codon in the translated protein alignment and was excluded together with further *A. aureum* sequences (12223, 23637, 33039, 4a2539) which shared this feature. Furthermore, the sequence of A_goochiae_23637 was excluded because of a 1 bp deletion that will result in a frameshift mutation and premature stop codon in the coding sequence. In addition, this sequence showed an indel of 56 amino acids between position 32 and 89 in the hypothetical protein. A frameshift mutation resulting in a premature stop codon was also detected for A_smithii_23637. In addition, this sequence had had an exclusive intron with a contrasting splice site. Instead of GT/AG for excision of the intron CA/AG was detected. The features of these sequences were confirmed by independent amplification. In total, 88 sequences were obtained and further reduction of identical species-specific clones resulted in an alignment comprising 42 sequences (table 5).

Table 5: Number of obtained total and reduced cloned sequences for *MCS_AP1*.

taxon	total number	reduced number
A. aureum	11	4
A. canariense	11	4
A. cuneatum	15	5
A. goochiae	7	1
A. nobile	5	3
A. rubrolineatum	4	1
A. saundersii	5	3
A. smithii	5	3
Ai. laxum	0	0
Ai. pachycaulon	0	0
M. anagensis	6	1
M. icterica	3	1
S. caeruleum	(4)	(3)
S. jaccardianum	4	4
S. modestum	8	5
S. pubescens	0	0
S. surculosum	8	4
	∑ 88	∑ 42

The full-length alignment comprises 3075 bp and eight exons and seven introns of highly variable length were observed (fig. 6, table 6).

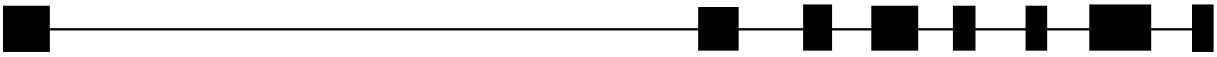


Fig. 6: Schematic exon and intron structure of the *MCS_AP1* gene sequences. Exons are shown as boxes and introns as lines. The relative length of the respective parts is given in table 6.

The exon alignment has a length of 651 bp (including *S. caeruleum*) which can be translated into 217 amino acids. Focusing on the translated protein alignment of this MADS-box gene, approx. amino acids 1-28 describe a part of the MADS-box domain, amino acids 29-69 the I-domain, amino acids 70-137 the K-domain, and amino acids 138-217 a part of the C-terminal domain.

Table 6: Exon and intron positions and lengths of the *MCS_AP1* sequences.

	position	length
exon 1	1 – 98	98
intron 1	99 – 1758	1660
exon 2	1759 – 1861	103
intron 2	1862 – 2020	159
exon 3	2021 – 2085	65
intron 3	2086 – 2216	131
exon 4	2217 – 2316	100
intron 4	2317 – 2424	108
exon 5	2425 – 2466	42
intron 5	2467 – 2603	137
exon 6	2604 – 2645	42
intron 6	2646 – 2764	119
exon 7	2765 – 2925	161
intron 7	2926 – 3037	112
exon 8	3038 – 3075	38

The intron alignment comprises 2426 bp but only 1758 bp if the dataset was reduced by the *Sedum* sequences. The intron alignment of the *Sedum* species comprised 1608 bp. Especially the first intron was long and extremely different between MCS

and *Sedum* sequences. It comprises 1660 aligned base pairs if both, MCS and *Sedum* species, were taken into account. However, only 522 bp were to some extent alignable and shared between the species. If the *Sedum* species were removed, the first intron comprises 1072 bp and 995 bp in a *Sedum* alignment.

MCS_AP3

For *MCS_AP3* 105 sequences of the studied species could be aligned (table 7). Four sequences showed a premature stop codon: A_aureum_263 and A_cuneatum_302b at the beginning, and A_rubrolineatum_26234 and A_rubrolineatum_305b at the end. In addition, A_aureum_263 seemed to be chimeric and was excluded; likewise also A_cuneatum_305b, A_rubrolineatum_301b, and A_saundersii_305. Reduction of identical species-specific clones resulted in an alignment of 75 sequences (table 7).

Table 7: Number of obtained total and reduced cloned sequences for *MCS_AP3*.

taxon	total number	reduced number
A. aureum	4	2
A. canariense	10	7
A. cuneatum	7	5
A. goochiae	5	3
A. nobile	12	3
A. rubrolineatum	7	4
A. saundersii	10	7
A. smithii	6	4
Ai. laxum	4	3
Ai. pachycaulon	7	5
M. anagensis	4	3
M. icterica	3	3
S. caeruleum	5	5
S. jaccardianum	4	4
S. modestum	6	6
S. pubescens	6	6
S. surculosum	5	5
	∑ 105	∑ 75

The full-length alignment resulted in a length of 2265 bp. For *K. blossfeldiana* (DQ479358) 30 nucleotides were missing at the beginning, for several *Sedum* sequences the first 42 bp, and for all these sequences the 3´-UTR region was not amplified.

Sequences comprising seven exons and six introns and with position 2180 the 3´-UTR region started (fig. 7; only amplified for MCS species). Exon and intron lengths are quite variable and summarized in fig. 7 and table 8.

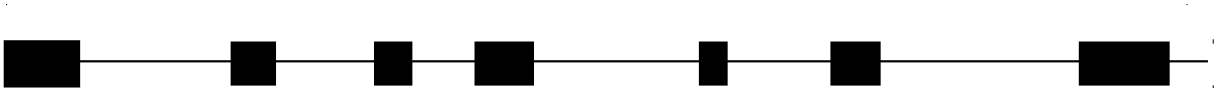


Fig. 7: Schematic exon and intron structure of the *MCS_AP3* gene sequences. Exons are shown as boxes and introns as lines. After the last exon box the 3´-UTR is indicated as line. The relative length of the respective parts is given in table 8.

The exon alignment comprises 669 nucleotides and resulted in a protein alignment of 223 amino acids. Translation in the hypothetical MADS-box protein would result in a part of the MADS-box (approx. amino acids 1-42), I-domain (43-71), K-domain (72-138), and C-terminal domain (139-223).

Table 8: Exon and intron positions and lengths for *MCS_AP3* sequences.

	position	length
exon 1	1 – 143	143
intron 1	144 – 424	281
exon 2	425 – 491	67
intron 2	492 – 699	208
exon 3	700 – 761	62
intron 3	762 – 886	125
exon 4	887 – 986	100
intron 4	987 – 1303	317
exon 5	1304 – 1345	42
intron 5	1346 – 1546	201
exon 6	1547 – 1627	81
intron 6	1628 – 2005	378
exon 7	2006 – 2179	174
3´-UTR region	2180 – 2265	86

The intron alignment consists of 1510 bp and was reduced to 1166 bp if only the MCS species were included. Introns of the *Sedum* species would result in 1318 bp. Intron 7 is unique because most parts correspond exclusively to several MCS species ordered in subclade B (see, e.g., fig. 14).

3.2. Phylogenetic reconstructions

In the present study full-length as well as exon and intron regions were used to study phylogenetic relationships. Combining exon and intron regions, which may evolve quite differently, in a full-length dataset reveals the problem that conserved (exon) and highly diverse (intron) regions are combined. Partition Homogeneity Tests indicate significant differences for *MCS_AP1* ($p < 0.05$) but not for *MCS_PEPC* ($p = 0.719$) and *MCS_AP3* ($p = 0.856$).

Phylogenies were reconstructed using BI and ML. The model of sequence evolution was first obtained by Modeltest and the respective model used to define the settings for the BI and ML analyses. In table 9 the model for each dataset as well as the settings and results of the BI analyzes are summarized.

Table 9: Selected model (AIC criterion) and results of the BI analyzes for different datasets. Default settings of MrBayes (four chains, sample every 100th generation, random starting tree) were used.

dataset	model	temperature	p-value	generations	burnin
MCS_PEPC_full-length	TVM+I+Γ	T = 0.2	0.005320	1,000,000	3,000
MCS_PEPC_exon	GTR+I+ Γ	T = 0.2	0.006626	1,000,000	3,000
MCS_PEPC_intron	HKY+I	T = 0.1	0.009362	2,000,000	6,000
MCS_PEPC_intron_Aeo	HKY+I	T = 0.2	0.008906	600,000	1,800
MCS_AP1_full-length	TVM+ Γ	T = 0.2	0.003558	1,000,000	3,000
MCS_ AP1_exon	TrN+ Γ	T = 0.2	0.009073	1,000,000	3,000
MCS_ AP1_intron	GTR+ Γ	T = 0.2	0.009856	100,000	300
MCS_ AP1_intron_Aeo	TVM+ Γ	T = 0.2	0.007175	100,000	300
MCS_AP3_full-length	HKY+ Γ	T = 0.2	0.009833	1,800,000	5,400
MCS_AP3_full-length_without_3´-UTR	HKY+ Γ	T = 0.2	0.009303	1,500,000	4,500
MCS_AP3_exon	TIM+I+ Γ	T = 0.2	0.008387	150,000	4,500
MCS_ AP3_intron	GTR+ Γ	T = 0.2	0.009091	1,000,000	3,000
MCS_ AP3_intron_Aeo	K81uf+ Γ	T = 0.2	0.004791	1,000,000	3,000

MCS_PEPC

Phylograms based on **full-length** alignments of *MCS_PEPC* were in agreement comparing BI (fig. 8) and ML (fig. 29; appendix). The BI phylogram provides a slightly better resolution. *Kalanchoe gracilis* (AJ252946) and *K. pinnata* (AJ252919) were used as outgroup. Several clones of *S. pubescens* (243 and 1c43) and *S. surculosum* (34341 and 54343) were with a very long branch clearly separated from the remaining *Sedum* and MCS species. These sequences are sister to the remaining analyzed sequences in the BI phylogram (pp = 1.00) but show an unresolved basal position for ML. Focusing on the main clade, S_pubescens_343 is basal to all other sequences (pp = 1.00, bs = 55%). Sequences of *S. surculosum* are found in different positions: S_surculosum_2c4391 basal to all other sequences (pp = 0.96, bs = 55%) and S_surculosum_44318 and 44342 as sister to *S. jaccardianum* (pp = 1.00, bs = 100%). *Sedum caeruleum* is sister to this latter group supported with a pp-value of 0.90 but without bootstrap support (43%) and these *Sedum* sequences are sister (pp = 1.00, bs = 53%) to the monophyletic MCS species. The sequences of the MCS species form two subclades supported with pp = 0.63 but without bootstrap support (28%). Subclade A contains sequences of *A. aureum*, *A. nobile*, *A. rubrolineatum*, *A. smithii*, *Ai. pachycaulon*, and *M. anagensis*. The *Aeonium* species are sister to *Ai. pachycaulon* and *M. anagensis* (pp = 1.00, bs = 75%) and the latter both species are also sister to each other (pp = 1.00, bs = 99%). *Aeonium rubrolineatum* and *A. nobile* are sister to each other (pp = 1.00, bs = 99%); *A. smithii* (pp = 1.00, bs = 95%) and *A. aureum* (pp = 1.00, bs = 100%) are successive basal to them. Subclade B contains duplicated sequences of *A. aureum*, *A. nobile*, *A. rubrolineatum*, *A. smithii*, and *Ai. pachycaulon*. Only *M. anagensis* shows no duplicates within subclade B. Next to the duplicated sequences the remaining MCS species, namely *A. canariense*, *A. cuneatum*, *A. goochiae*, *A. saundersii*, *Ai. laxum*, and *M. icterica*, could be found within subclade B. Basal and sister to the *Aeonium* species (pp = 1.00, bs = 100%) is a group comprising *Ai. laxum* and *Ai. pachycaulon* in a sister relationship to each other (pp = 1.00, bs = 99%) and to *M. icterica* (pp = 1.00, bs = 98%). Within subclade B *A. goochiae* is basal to all *Aeonium* species (pp = 1.00, bs = 95%). The remaining *Aeonium* sequences form two distinct groups (pp = 1.00, bs = 88%). The first comprises *A. aureum*, *A. saundersii*, and *A. smithii*. The latter two species are sister to each other (pp = 0.96, bs = 81%) and to *A. aureum* (pp = 0.95, bs = 86%). The second group comprises *A. canariense*, *A. cuneatum*, *A.*

nobile, and *A. rubrolineatum*. *Aeonium cuneatum* and *A. rubrolineatum* are sister to each other (pp = 1.00, bs = 80%) and A_canariense_411 is basal to them (pp = 1.00, bs = 81%). The remaining sequences of *A. canariense*, 412 and 41322, are sister to *A. nobile* (pp = 1.00, bs = 99%). Both species groups are sister to each other (pp = 1.00, bs = 98%).

If the unique sequences of *S. pubescens* and *S. surculosum* were removed, the same relationships were obtained. Also removing the *Kalanchoe* species and the unique *Sedum* sequences did not change the topologies but decrease resolution (data not shown).

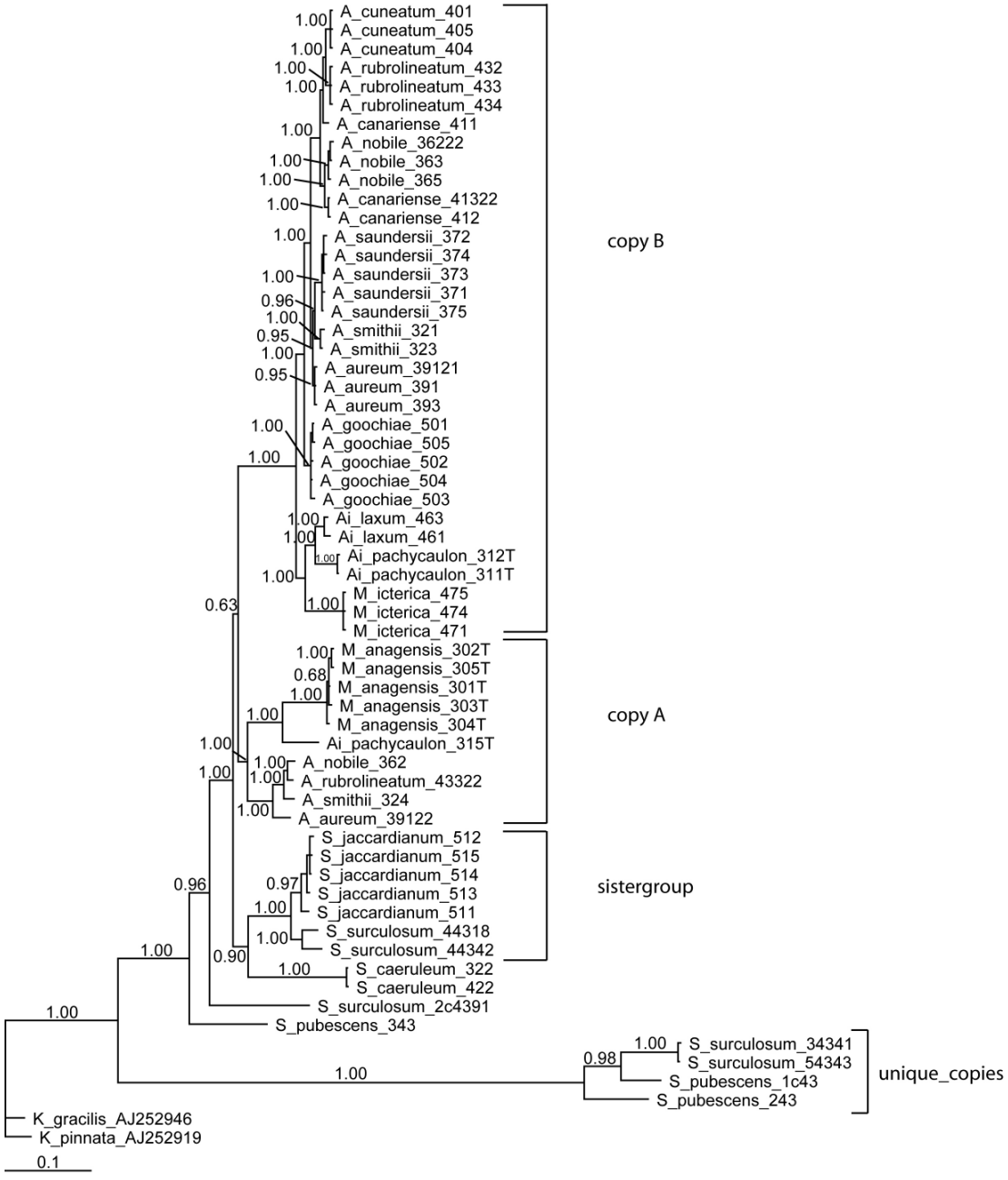


Fig. 8: BI phylogram based on the full-length *MCS_PEPC* data. Posterior probabilities are given at the nodes.

Also in the **exon** phylograms *K. gracilis* (AJ252946) and *K. pinnata* (AJ252919) were used as outgroup and both phylograms (BI fig. 9; ML fig. 30; appendix) were in agreement. Sedum_pubescens_243 and 1c43 as well as S_surculosum_34341 and 54343 are separated from the remaining sequences by a long branch in the BI phylogram (pp = 1.00). This relationship is unresolved in the ML phylogram. Sister to this group are all other species sequences. Sedum_pubescens_343 is basal of the main clade (pp = 0.99, bs = 76%). Positions of *S. surculosum* are divergent with S_surculosum_2c4391 basal to all remaining sequences (pp = 0.97, bs = 73%) and the other two *S. surculosum* sequences included in the *Sedum* subclade. *Sedum caeruleum* is sister to *S. jaccardianum* and *S. surculosum* (pp = 0.97, bs = 65%), which are also sister to each other (pp = 1.00, bs = 100%). The species of the MCS form a monophyletic clade, separated in two subclades. The position of subclade A is unresolved, comprising sequences of *A. aureum*, *A. nobile*, *A. rubrolineatum*, *A. smithii*, *Ai. pachycaulon*, and *M. anagensis*. Two subgroups could be described, one with *A. rubrolineatum* and *A. nobile* as sister to each other (pp = 0.88, bs = 87%) and *A. smithii* (pp = 1.00, bs = 100%) as well as *A. aureum* (pp = 1.00, bs = 96%) successive sister to them. This species group is sister (pp-value = 1.00, bs = 59%) to the sequences of *M. anagensis* and *Ai. pachycaulon*, which are also sister to each other (pp = 1.00, bs = 100%). Within subclade B duplicated sequences of *A. aureum*, *A. nobile*, *A. rubrolineatum*, *A. smithii*, and *Ai. pachycaulon* were found whereas sequences of *M. anagensis* were only detected within subclade A. Next to the duplicates, sequences of *A. canariense*, *A. cuneatum*, *A. goochiae*, *A. saundersii*, *Ai. laxum*, and *M. icterica* were found. Basal of subclade B are the sequences of *Ai. laxum* and *Ai. pachycaulon* in a sister relationship to each other (pp = 1.00, bs = 98%) and to *M. icterica* (pp = 1.00, bs = 99%). The split of this group is highly supported (pp = 1.00, bs = 100%) and *A. goochiae* is basal to all remaining *Aeonium* species of subclade B (pp = 1.00, bs = 96%). Within subclade B two groups are formed (pp = 1.00, bs = 84%). The first comprises *A. aureum*, *A. saundersii*, and *A. smithii*; the second *A. canariense*, *A. cuneatum*, *A. nobile*, and *A. rubrolineatum*. *Aeonium saundersii* and *A. smithii* are sister to each other (pp = 0.96, bs = 72%) with *A. aureum* basal to them (pp = 0.96, bs = 86%). In a sister relationship are also *A. canariense* and *A. nobile* (pp = 0.93, bs = 76%). Unique is the position of A_canariense_411 (pp = 1.00, bs = 77%) basal to the sister relationship of *A.*

cuneatum and *A. rubrolineatum* (pp 1.00 = bs = 75%). Both species groups are also sister to each other (pp = 1.00, bs = 94%).

Exclusion of the unique sequences of *S. pubescens* and *S. surculosum* did not change the topology. Also removing the *Kalanchoe* sequences and using S_pubescens_343 as outgroup resulted in the same relationships within the ingroup (data not shown).

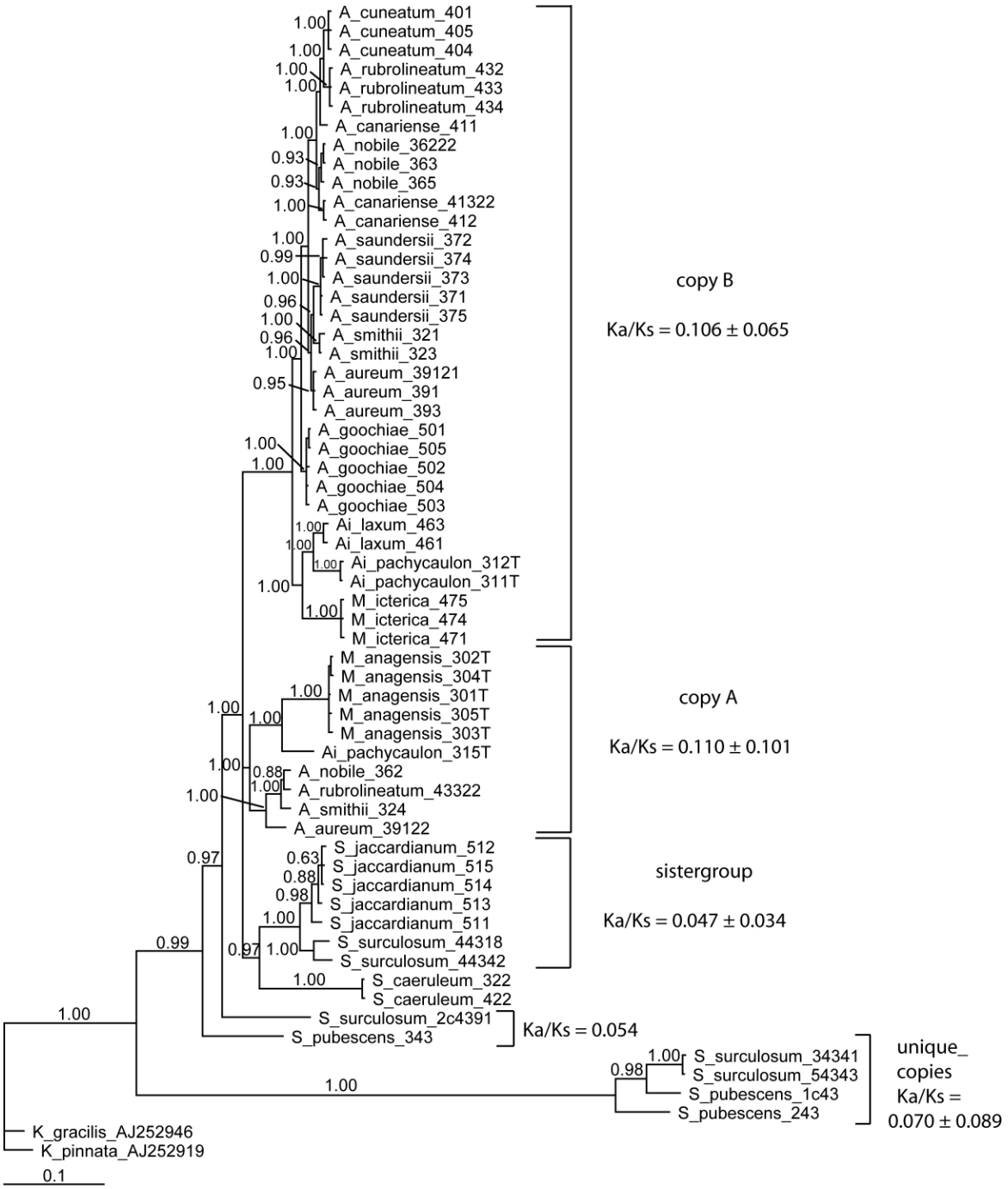


Fig. 9: BI phylogram based on the *MCS_PEPC* exon data. Posterior probabilities are given at the nodes.

The phylograms based on the **intron** data are highly unresolved. However, the ML phylogram (fig. 10) shows a better resolution compared to BI (fig. 31; appendix) and will be described below. The *Kalanchoe* sequences were excluded and S_pubescens_243 and 1c43 were used as outgroup. These sequences as well as S_surculosum_34341 and 54343 were unresolved at the basis and separated from the main clade by a long branch. *Sedum caeruleum* is sister (bs = 70%) to all remaining species and S_pubescens_343 and S_surculosum_2c4391 are basal to the main clade (bs = 40%). Further *S. surculosum* sequences are mixed with the MCS species in subclade B and were found together with *Ai. laxum*, *Ai. pachycaulon*, and *M. icterica*. Subclade B contains also the sequences of *S. jaccardianum* in unresolved positions. The separation of the main clade into two subclades obtains no bootstrap support and relationships are also highly unresolved in the BI phylogram. Subclade A comprises *A. aureum*, *A. nobile*, *A. rubrolineatum*, *A. smithii*, *Ai. pachycaulon*, and *M. anagensis* in two subgroups but the separation is not supported (bs = 25%). *Aeonium nobile* and *A. rubrolineatum* are sister to each other (bs = 60%) and *A. aureum* (bs = 66%) and *A. smithii* (bs = 78%) are successive sister to them. In the second subgroup Ai_pachycaulon_315T is sister to *M. anagensis* but without support. Basal of subclade B are *Ai. laxum* and *Ai. pachycaulon* in a sister relationship to each other and to *M. icterica* but no support was obtained for these relationships. Subclade B is furthermore separated into two species groups, one comprising *A. canariense* and *A. nobile* in a sister relationship to each other (bs = 70%) and *A. cuneatum* as well as *A. rubrolineatum* cluster to them. The second group is formed by *A. aureum*, *A. goochiae*, *A. saundersii*, and *A. smithii* sequences in unresolved relationships.

Given that unambiguous alignments of intron sequences between genera are not trivial *Sedum* sequences were removed, arising gaps deleted, and *Ai. laxum* used as outgroup. The ML phylogram shows a better resolution (fig. 33; appendix) but no conflicts were found in comparison to the BI phylogram (fig. 32; appendix). Aichryson_pachycaulon_311T and 312T cluster together unresolved basal of the respective phylograms and *M. icterica* is basal to all remaining species (pp = 0.97, bs = 92%). The separation of the main clade into two species groups is supported with a pp-value of 0.80 and a bootstrap of 66%. Within subclade A, Ai_pachycaulon_315T and *M. anagensis* are sister to each other (pp = 0.99, bs = 99%) and separated from the duplicated *Aeonium* species (pp = 1.00, bs = 100%). Here, *A. nobile* is sister to *A.*

rubrolineatum (pp = 0.96, bs = 71%) and both are successive sister to *A. aureum* (pp = 0.99, bs = 82%) and *A. smithii* (pp = 1.00, bs = 95%). Focusing on the ML phylogram, the remaining *Aeonium* sequences in subclade B form two subgroups but without support (40%). One subgroup comprises *A. canariense*, *A. cuneatum*, *A. nobile*, and *A. rubrolineatum* where *A. canariense* and *A. nobile* are sister to each other (bs = 95%) but excluding A_canariense_411. The second subgroup is formed by *A. aureum*, *A. goochiae*, *A. saundersii*, and *A. smithii* in unresolved relationships to each other.

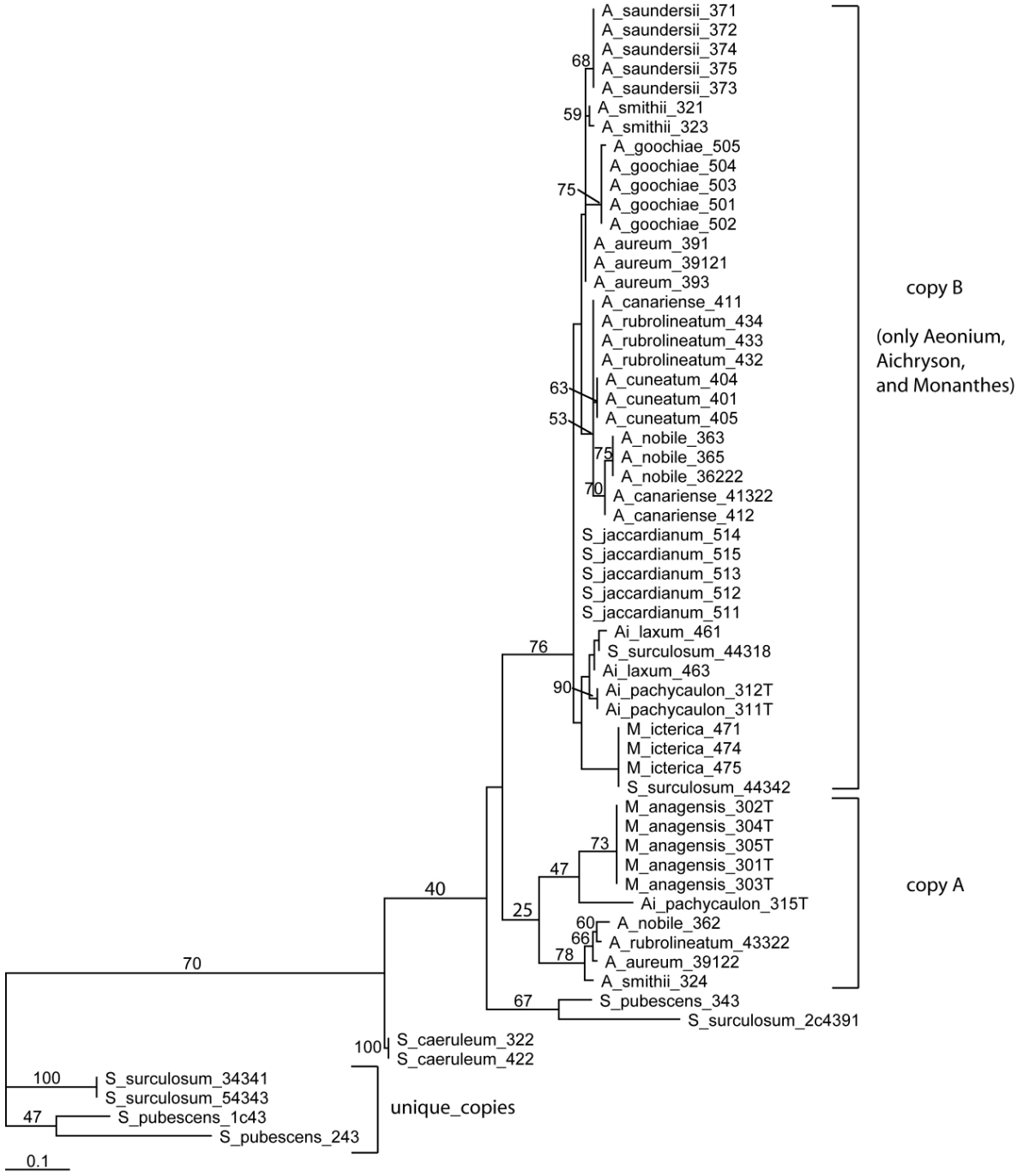


Fig. 10: ML phylogram based on the *MCS_PEPC* intron data. Bootstrap support is given at the nodes.

MCS_AP1

Full-length BI (fig. 34; appendix) and ML (fig. 11) phylogenies of *MCS_AP1* were in agreement with each other and relationships slightly better resolved in the ML phylogram. The analyzed *Sedum* species form a distinct group supported with a pp-value of 1.00 and bootstrap of 100%. *Sedum jaccardianum* and *S. surculosum* are sister to each other and to *S. modestum*; both relationships with highest support. Sedum_surculosum_3 seems to be paralogous to the remaining *S. surculosum* clones (pp = 1.00, bs = 78%). Nearly all sequences of the MCS species form one main clade. However, the species group is not monophyletic since A_aureum_13637 and 1b3637 as well as A_saundersii_33637 were detected in derived positions (pp = 1.00, bs = 100%). The two sequences of *A. aureum* cluster together but are separated from the additional *A. saundersii* sequence. The main clade consists of the MCS species that form two distinct subclades (A and B) in a sister relationship to each other (pp = 1.00, bs = 85%). Subclade B contains all studied and analyzed species whereas subclade A comprises only duplicated sequences of *A. aureum*, *A. canariense*, *A. cuneatum*, *A. nobile*, and *A. smithii*. Relationships within subclade A are rather unresolved. Only a sister relationship between *A. canariense* and *A. nobile* could be detected (pp = 1.00, bs = 86%) to which *A. cuneatum* is basal and sister (pp = 1.00, bs = 88%). In the ML phylogram *A. smithii* sequences were basal to this species group but without support (bs = 42%) and A_aureum_3b3637 basal to all of them (bs = 100%). Within subclade B relationships are better resolved. Both *Monanthes* species are basal to the *Aeonium* species even if their respective position is unresolved. *Aeonium goochiae* is basal to all other *Aeonium* species (pp = 1.00, bs = 90%) which form two distinct species groups (pp = 1.00, bs = 100%). Group one comprises *A. aureum* which is sister to *A. saundersii* but the support for this relationship is low (pp = 0.68, bs = 62%). *Aeonium smithii* is basal to them (pp = 0.99, bs = 82%) and *A. rubrolineatum* is basal of this whole subgroup (pp = 1.00, bs = 95%). The second species group consists of *A. canariense* and *A. nobile* as sister to each other (pp = 1.00, bs = 100%) and *A. cuneatum* basal of them (pp = 1.00, bs = 79%).

If the dataset is reduced by the first large intron relationships in the obtained BI phylogram are mainly the same. The only difference is the position of S_surculosum_3, which clusters separated from the remaining *S. surculosum* sequences as sister to *S. jaccardianum* (data not shown).

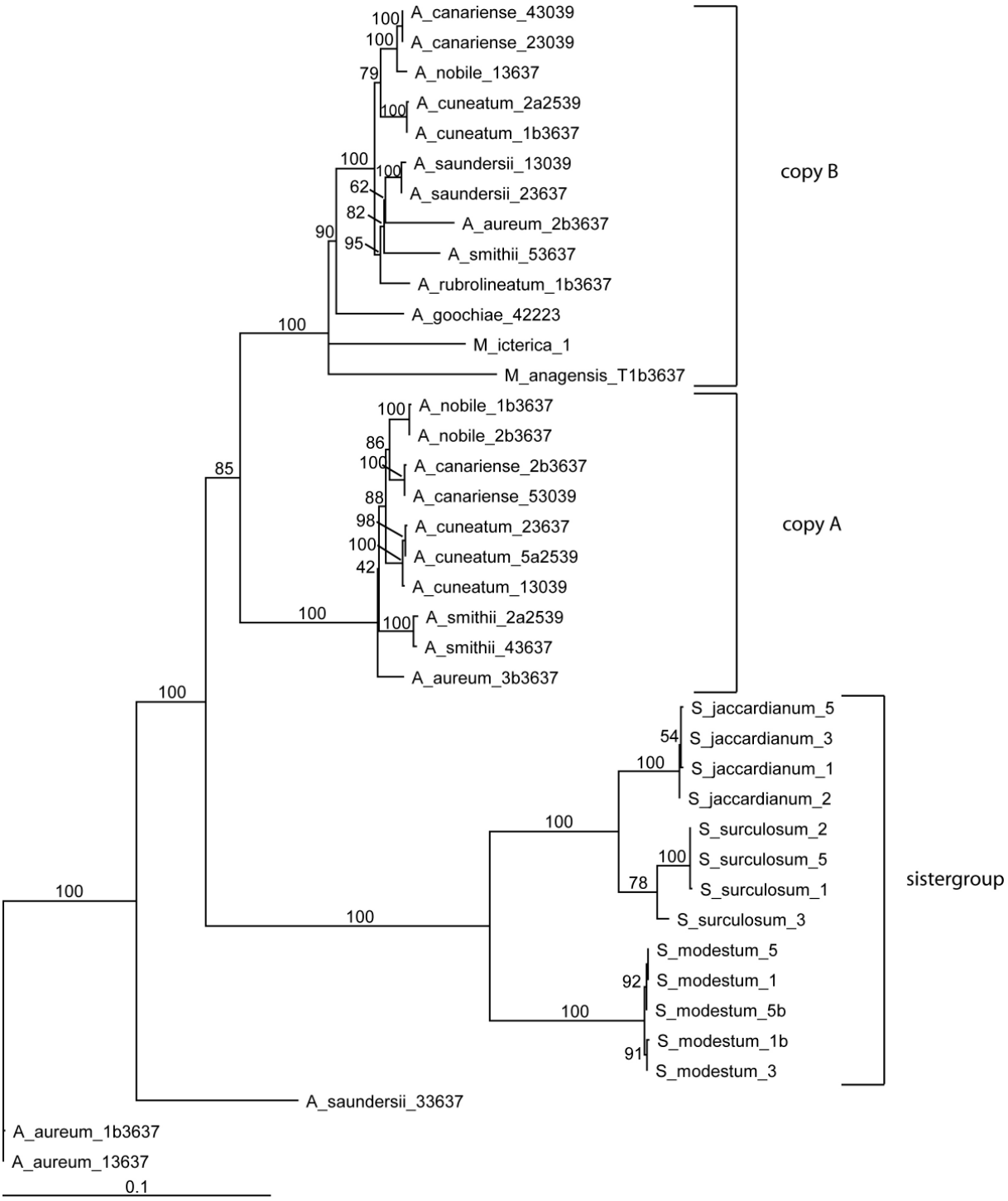


Fig. 11: ML phylogram based on the *MCS_AP1* full-length data. Bootstrap support is given at the nodes.

Phylograms of the **exon** sequences are highly unresolved but in agreement comparing BI (fig. 35; appendix) and ML (fig. 12) although the ML phylogram provides a slightly better resolution. *Sedum caeruleum* sequences could be obtained and were used as outgroup. The MCS species are not monophyletic because of the unique *A. aureum* and *A. saundersii* sequences. Aeonium_aureum_13637 and 1b3637 cluster unresolved at the basis of all other analyzed sequences. The sequence of A_saundersii_33637 is sister to the analyzed *Sedum* ser. *Monanthoidea* species but this relationship is only weakly supported by a pp-value of 0.63 (bs = 42%). Within the *Sedum* group *S. modestum* is sister (pp = 1.00, bs = 100%) to the sister species of *S. jaccardianum* and *S. surculosum*. Unique is again the position of S_surculosum_3 intermixed with *S. jaccardianum* (pp = 0.98, bs = 88%) whereas the remaining *S. surculosum* clones are sister to them (pp = 1.00, bs = 100%). The clade comprising the *Sedum* species and A_saundersii_33637 is sister to the main clade (pp = 0.51, bs = 54%) formed by the MCS species, which are distinguished into two subclades (pp = 0.81, bs = 61%). Within subclade A, relationships between *A. aureum*, *A. cuneatum*, and *A. smithii* are unresolved (pp = 0.69, bs = 57%). *Aeonium canariense* (pp = 0.81, bs = 58%) and *A. nobile* are basal to them. Subclade B comprises next to duplicated sequences of *A. aureum*, *A. canariense*, *A. cuneatum*, *A. nobile*, and *A. smithii* also the sequences of *A. goochiae*, *A. rubrolineatum*, *A. saundersii*, *M. anagensis*, and *M. icterica*. The latter two are unresolved at the base of the subclade focusing on the BI phylogram. In the ML phylogram, *M. anagensis* is basal of subclade B (bs = 84%) and *M. icterica* basal of all *Aeonium* species (bs = 45%). Basal of all other *Aeonium* species *A. goochiae* could be found (pp = 0.83, bs = 46%) and further relationships within subclade B are rather unresolved. Nevertheless, the sister relationship of *A. canariense* and *A. nobile* is observed (pp = 1.00, bs = 100%) and *A. cuneatum* is basal to them (pp = 0.66, bs = 53%). Focusing on the ML phylogram, also *A. aureum* and *A. smithii* (bs = 40%) are sister to each other.

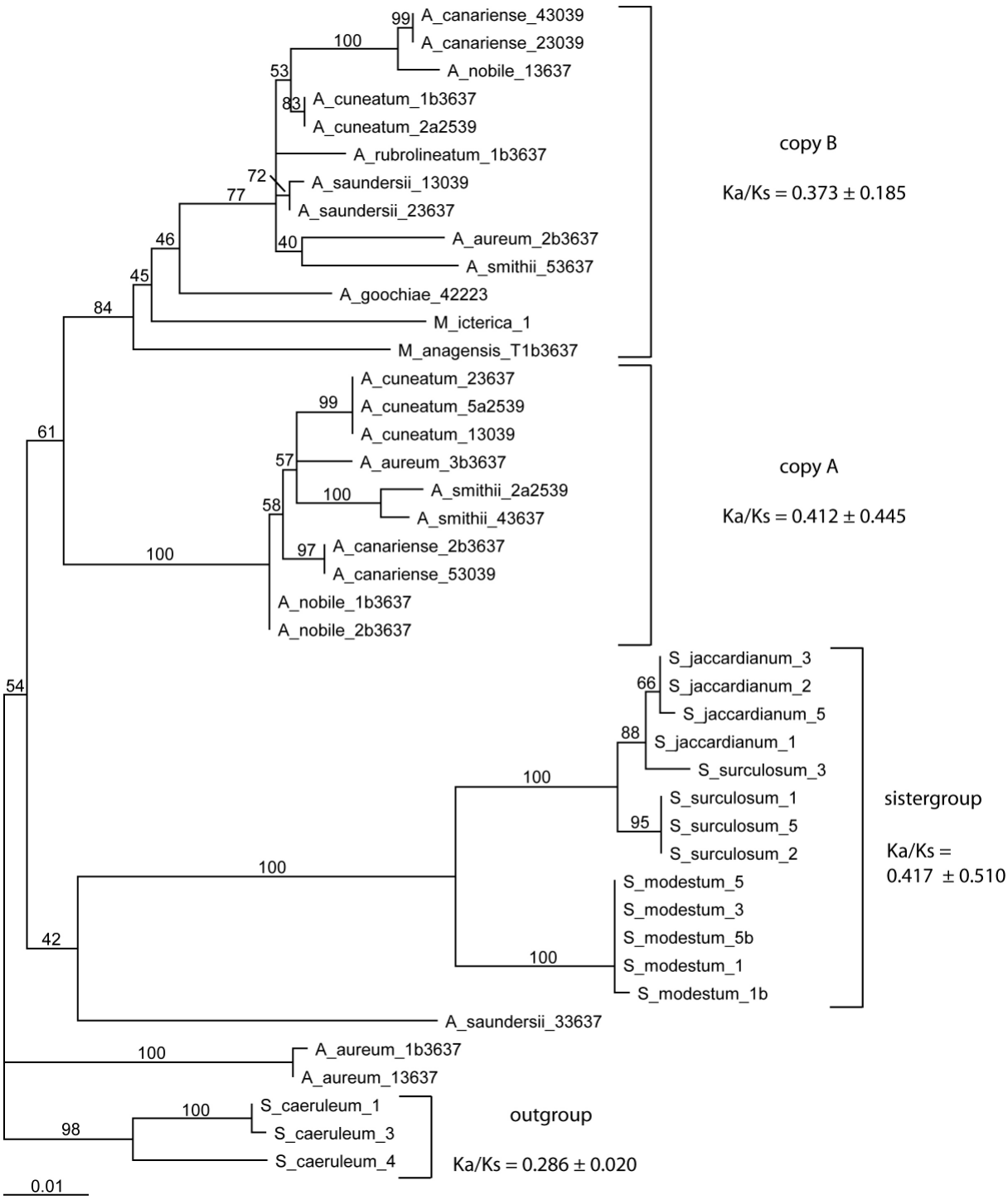


Fig. 12: ML phylogram based on the *MCS_AP1* exon data. Bootstrap support is given at the nodes.

Similar clades and relationships are obtained using exclusively the **intron** region of *MCS_AP1*. BI (fig. 13) and ML (fig. 36; appendix) phylograms were in major agreement and without conflicting signals. Excluding *S. caeruleum*, three groups of MCS species and one exclusive *Sedum* ser. *Monanthoidea* clade were obtained. A unique position was detected for A_aureum_13637 and 1b3637 as well as A_saundersii_33637. The first two sequences were selected as outgroup and

A_saundersii_33637 is basal to all remaining analyzed sequences in the BI phylogram (pp = 1.00). Again, due to these three sequences, the MCS species are not monophyletic. Within the *Sedum* clade *S. modestum* is sister to *S. jaccardianum* and *S. surculosum* (pp = 1.00, bs = 100%), which are also sister to each other (pp = 1.00, bs = 100%). Sedum_surculosum_3 showed again a unique position as sister to the other *S. surculosum* sequences (pp = 1.00, bs = 86%). The *Sedum* clade is basal (pp = 1.00, bs = 100%) to the main clade comprising the analyzed MCS. This main clade is divided into two distinct subclades (pp = 0.95, bs = 69%). Subclade A, comprising *A. aureum*, *A. canariense*, *A. cuneatum*, *A. nobile*, and *A. smithii*, separated into two species groups (pp = 1.00, bs = 100%). *Aeonium canariense* is sister to *A. nobile* (pp = 0.98, bs = 82%) and *A. cuneatum* is sister to them (pp = 1.00, bs = 97%). This species group is sister to the sister relationship of *A. aureum* and *A. smithii* (pp = 0.74, bs = 84%). The second subclade B contains next to the sequences of the above mentioned species the two *Monanthes* species and the sequences of *A. goochiae*, *A. rubrolineatum*, and *A. saundersii*. Focusing on the ML phylogram, both *Monanthes* species are sister to each other (bs = 54%) and basal to the species of *Aeonium* (bs = 100%). The relationship between the two *Monanthes* species is not resolved in the BI phylogram but the separation from the *Aeonium* species supported with a pp-value of 1.00. Basal of the *Aeonium* clade is *A. goochiae* (pp = 0.95, bs = 80%). Two species groups are formed (pp = 1.00, bs = 100%) where the first comprises *A. canariense* and *A. nobile* in a sister relationship to each other (pp = 1.00, bs = 100%) and to *A. cuneatum* (pp = 1.00, bs = 84%). The second species group is formed by *A. aureum* which is sister to *A. saundersii* (pp = 0.74, bs = 78%); *A. smithii* (pp = 0.98, bs = 83%) and *A. rubrolineatum* (pp = 1.00, bs = 93%) successive basal to them.

After removing the *Sedum* sequences, A_aureum_13637 and 1b3637 were used as outgroup (BI fig. 37 and ML fig. 38; both appendix). Aeonium_saundersii_ 33637 has a unique position basal to the remaining analyzed sequences and species (pp = 1.00, bs = 100%). The main clade is again divided into two subclades (pp = 1.00, bs = 100%) where exactly the same relationships as described above for the full intron dataset could be observed with high supports.

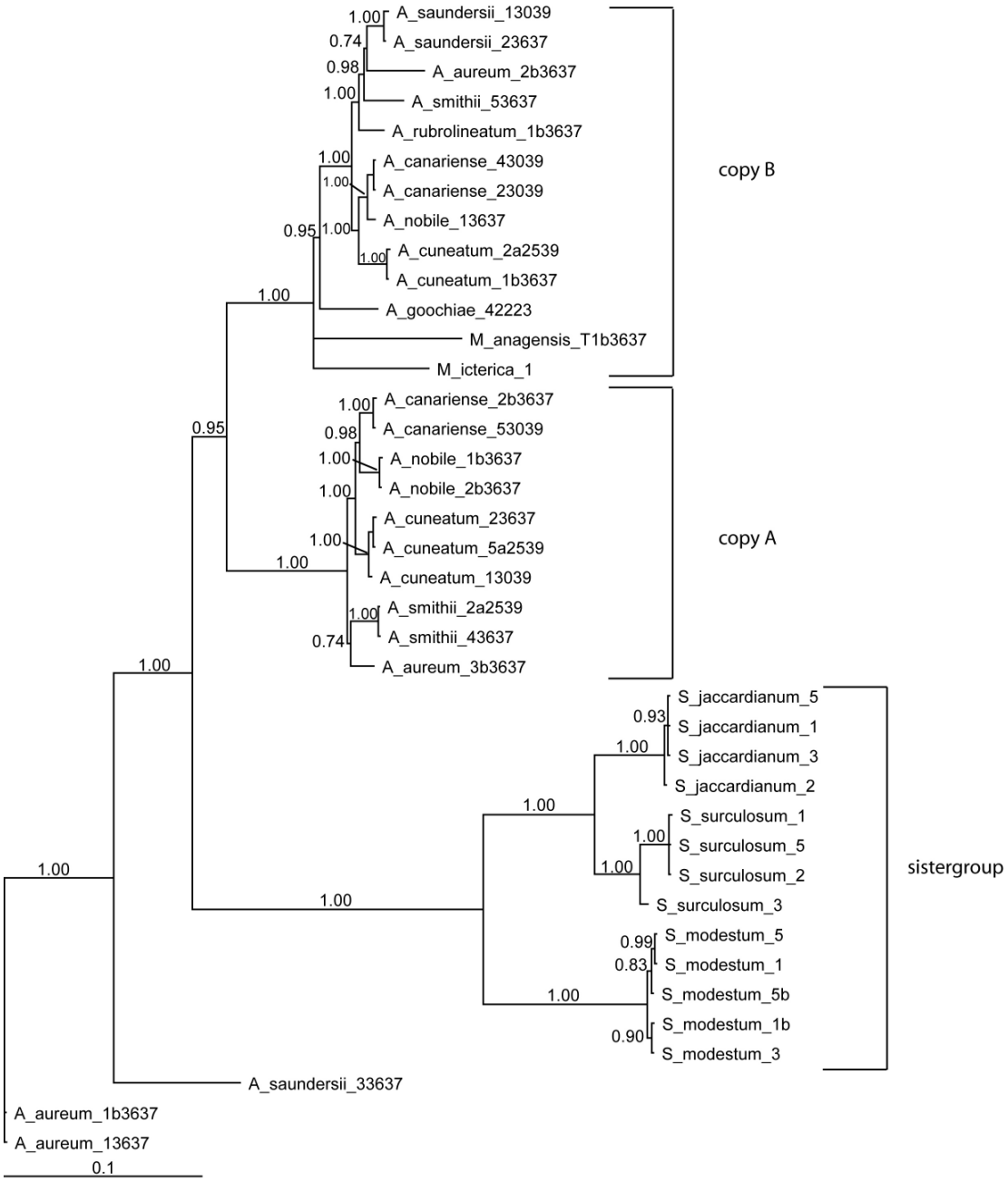


Fig. 13: BI phylogram based on the *MCS_AP1* intron data. Posterior probabilities are given at the nodes.

MCS_AP3

For phylogenetic reconstructions of the *MCS_AP3* sequences *K. blossfeldiana* (DQ479358) was used as outgroup in the **full-length** dataset. BI (fig. 39; appendix) and ML (fig. 14) phylograms are in major agreement and ML provides a slightly better resolution. Exclusion of the 3´-UTR results in the same topology and will be described below.

The sequences of *S. pubescens* form two distinct groups. One is unresolved at the base and the other sister and basal to all remaining species (pp = 0.99, bs = 69%). The remaining species form two major cluster (pp = 0.88, bs = 55%), one containing the *Sedum* species and the second monophyletic group is formed by the species of the MCS. *Sedum caeruleum* is sister to the *Sedum* ser. *Monanthoidea* species supported by a pp-value of 0.87 and bootstrap of 54%. *Sedum modestum* is sister to *S. jaccardianum* and *S. surculosum* (pp = 1.00, bs = 99%). The sequences of *S. jaccardianum* form a monotypic group and *S. surculosum* is sister to them. However, S_surculosum_1 and 8 are sister to *S. jaccardianum* (pp = 1.00, bs = 96%) whereas S_surculosum_2, 3, and 6 are sister and basal to this group of sequences (pp = 0.87, bs = 59%). The species of the MCS form two distinct subclades (pp-value = 1.00, bs = 86%). Subclade A contains exclusively *Aeonium* species (*A. aureum*, *A. canariense*, *A. cuneatum*, *A. rubrolineatum*, *A. smithii*) and subclade B comprises also the *Aichryson* and *Monanthes* species. Within subclade A two distinct species groups are formed (pp = 1.00, bs = 96%). Highly supported is the sister relationship between *A. canariense* and *A. cuneatum* (pp = 1.00, bs = 100%). These species are sister to the species group formed by *A. aureum*, *A. rubrolineatum*, and *A. smithii*. *Aeonium aureum* is basal (pp = 0.92, bs = 64%) to *A. rubrolineatum* and *A. smithii* which are sister to each other (pp = 1.00, bs = 100%). Within subclade B *Ai. pachycaulon* and *M. icterica* are intermixed (pp = 1.00, bs = 100%) with *Ai. laxum* as sister to them. Aichryson_laxum_26634 is in a derived position compared to the other two *Ai. laxum* sequences and shows differences between the BI and ML phylograms. This sequence is basal to the remaining *Ai. laxum*, *Ai. pachycaulon*, and *M. icterica* sequences in the ML phylogram with 100% bootstrap support. In contrast, in the BI phylogram the *Ai. laxum* clones form a monotypic species group (pp = 0.71) basal to *Ai. pachycaulon* and *M. icterica* (pp = 1.00) and with Ai_laxum_26634 as sister to the other *Ai. laxum* sequences. The group formed by *Ai. laxum*, *Ai. pachycaulon*, and *M. icterica* is basal to the remaining sequences of subclade B in the ML phylogram (bs = 99%). In the BI phylogram this relationship is unresolved. The position of *M. anagensis* is also unresolved in the BI phylogram, whereas this species is basal to all *Aeonium* species in the ML phylogram (bs = 53%). *Aeonium goochiae* is basal to the remaining *Aeonium* species (pp = 1.00, bs = 98%). Two distinct *Aeonium* species groups were formed (pp = 1.00, bs = 100%), one comprising *A. aureum*, *A. saundersii*, and *A. smithii* (pp = 0.97, bs = 68%). The second group combines *A.*

canariense, *A. cuneatum*, and *A. nobile* with *A. rubrolineatum* basal to them (pp = 0.65, bs = 48%). *Aeonium canariense* and *A. cuneatum* are sister to each other (pp = 0.57, bs = 55%) with *A. nobile* highly supported (pp = 1.00, bs = 100%) basal to them.

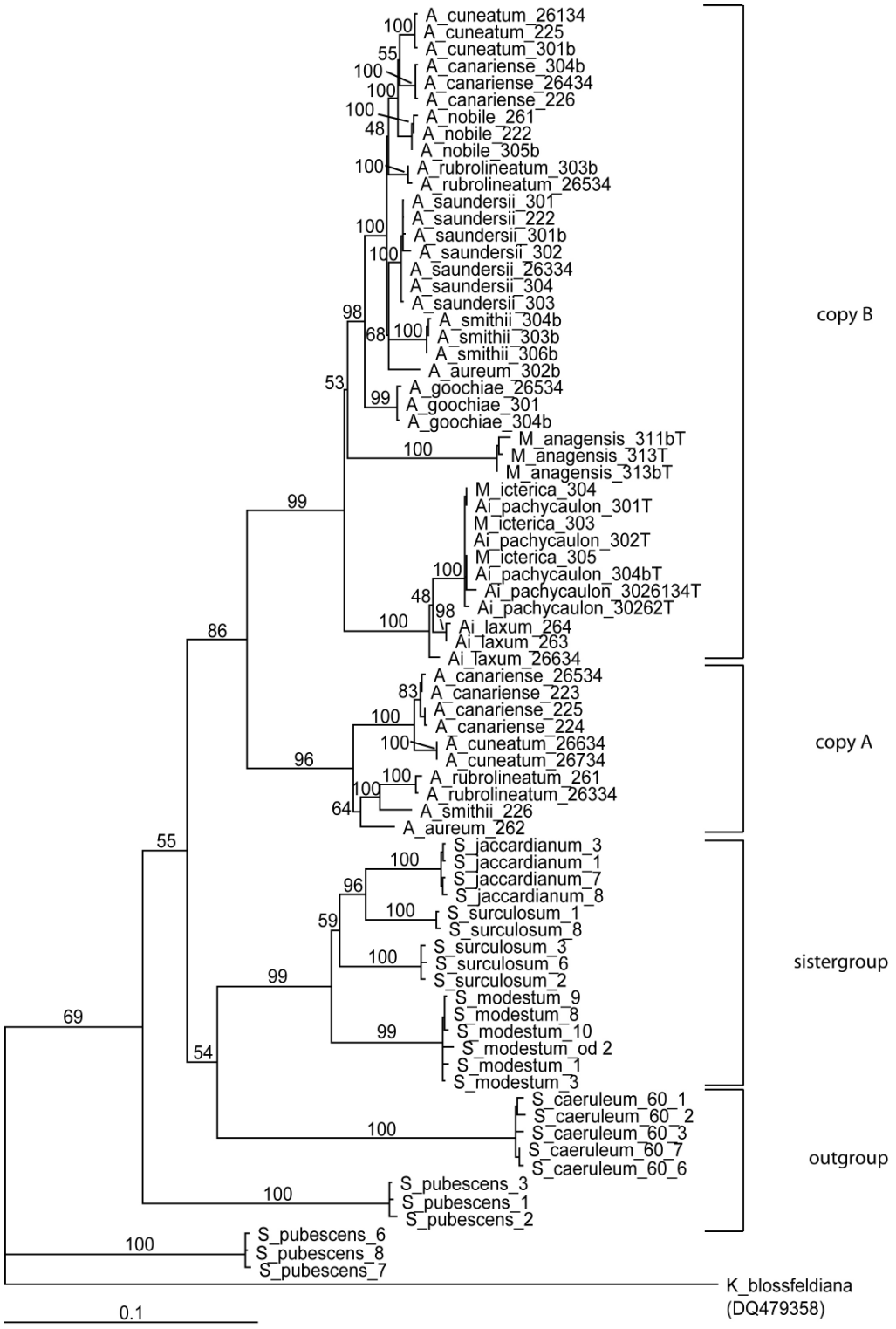


Fig. 14: ML phylogram based on the *MCS_AP3* full-length data. The 3´-UTR region is excluded. Bootstrap support is given at the nodes.

BI (fig. 15) and ML (fig. 40; appendix) phylograms of the **exon** region are in major agreement but some differences were observed in subclade A. *Kalanchoe blossfeldiana* (DQ479358) was used as outgroup and sequences of *S. pubescens* cluster into two distinct groups. One is unresolved at the base and one basal to all remaining species (pp = 0.96, bs = 71%). *Sedum caeruleum* is sister and basal to all further analyzed species (pp = 0.79, bs = 57%). *Sedum* ser. *Monanthoidea* species cluster together (pp = 1.00, bs = 97%) and *S. modestum* is sister to *S. jaccardianum* and *S. surculosum*. *Sedum jaccardianum* and *S. surculosum* are also sister to each other but with a unique pattern for *S. surculosum*. Sedum_surculosum_1 and 8 are sister to *S. jaccardianum* with a support of pp = 1.00 and bootstrap of 93%. In contrast, S_surculosum_2, 3, and 6 are basal and sister to this above mentioned group (pp = 0.98, bs = 65%). The *Sedum* species are separated from the species of the MCS with a pp-value of 0.88 but no bootstrap support is obtained (41%). Within the main clade of the MCS species two subclades could be distinguished (pp = 0.98, bs = 77%). Subclade A comprises only duplicated sequences of several *Aeonium* species whereas subclade B also contains the *Aichryson* and *Monanthes* species. Two species groups were detected within subclade A focusing on the BI phylogram and differences were observed between BI and ML phylograms. In the BI phylogram *A. rubrolineatum* and *A. smithii* are sister to each other (pp = 0.86) and *A. aureum* is basal to them (pp = 0.53). They are sister (pp = 1.00) to the second species group which is formed by *A. canariense* and *A. cuneatum* in a sister relationship to each other (pp = 1.00). In contrast, ML positioned A_smithii_226 basal of subclade A (bs = 97%); *A. canariense* and *A. cuneatum* are sister to each other (bs = 98%) with *A. aureum* basal to them (bs = 57%). The sequences of *A. rubrolineatum* are, without support, basal to this relationship. Relationships within subclade B are identical comparing BI and ML phylograms. Next to duplicated gene copies of several *Aeonium* species, subclade B also comprises the species of *Aichryson* and *Monanthes*, *A. goochiae*, *A. nobile*, and *A. saundersii*. *Aichryson laxum* is sister (pp = 1.00, bs = 100%) to the complete intermixed sequences of *Ai. pachycaulon* and *M. icterica* (pp = 0.99, bs = 69%). Aichryson_laxum_26634 is thereby basal to the both remaining *Ai. laxum* sequences (pp = 0.83, bs = 64%) and the whole species group is basal of subclade B (pp = 1.00, bs = 96%). *Monanthes anagensis* is basal to all *Aeonium* species (pp = 0.94, bs = 81%). *Aeonium goochiae* is basal and sister (pp = 0.88, bs = 68%) to the remaining *Aeonium* species which form two groups within

subclade B (pp = 1.00, bs = 84%). One subgroup (pp = 0.96, bs = 63%) comprises *A. aureum*, *A. saundersii*, and *A. smithii* in unresolved relationships. The second subgroup (pp = 0.97, bs = 58%) is formed by *A. canariense* and *A. cuneatum* which are sister to each other (pp = 0.99, bs = 60%) and cluster together with *A. nobile* and *A. rubrolineatum* in unresolved relationships.

The same relationships are also obtained if *Kalanchoe* is excluded and *S. pubescens* used as outgroup and also if the focus lies only on orthologous sequences (data not shown).

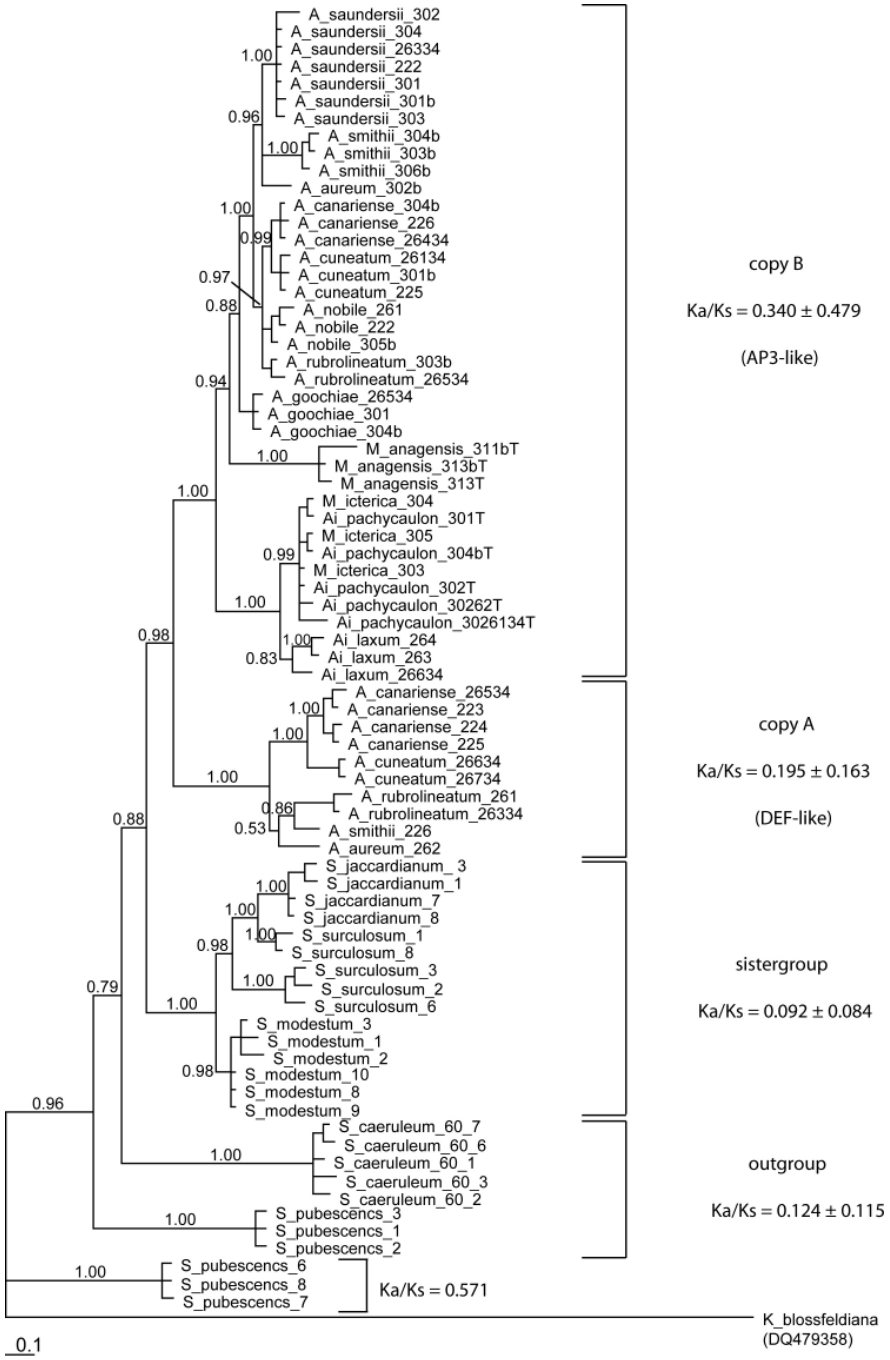


Fig. 15: BI phylogram based on the *MCS_AP3* exon data. Posterior probabilities are given at the nodes.

For the **intron** phylogram *Kalanchoe* was excluded and *S. pubescens* used as outgroup. Relationships inferred by BI (fig. 16) and ML (fig. 41; appendix) phylograms are in agreement focusing on the relationships of the MCS species but show conflicting signals for *Sedum*. Using S_pubescens_6, 7, and 8 as outgroup, the sequences S_pubescens_1, 2, 3 are unresolved at the base of all remaining species in the BI phylogram. In contrast, this species group is found basal to the main clade of the MCS species in the ML phylogram but lacks bootstrap support (38%). Within the BI phylogram, *S. caeruleum* is sister to the species of *Sedum* ser. *Monanthoidea* (pp = 0.53) and form together with them a unique *Sedum* clade in an unresolved position. *Sedum modestum* is sister and basal (pp = 1.00) to the sister relationship of *S. jaccardianum* and *S. surculosum*. The two sequences of S_surculosum_1 and 8 are sister to *S. jaccardianum* (pp = 0.99) whereas the three remaining *S. surculosum* sequences (2, 3, 6) are sister to this group (pp = 0.60). Relationships within the ML phylogram are differently resolved. The three species of *Sedum* ser. *Monanthoidea* form a unique clade in an unresolved position and *S. modestum* is sister and basal to *S. jaccardianum* and *S. surculosum* (bs = 100%). As in the BI phylogram, S_surculosum_1 and 8 are sister to *S. jaccardianum* (bs = 78%) whereas S_surculosum_2, 3, and 6 are sister to this group (bs = 69%). *Sedum caeruleum* is basal to the main MCS clade and S_pubescens_1, 2, and 3 but without support (37%). Relationships within the main clade comprising all studied *Aeonium*, *Aichryson*, and *Monanthes* species are identical comparing both methods. The main clade is separated into two subclades (pp = 1.00, bs = 95%). Subclade A comprising duplicated sequences of *A. aureum*, *A. canariense*, *A. cuneatum*, *A. rubrolineatum*, and *A. smithii*. *Aeonium canariense* and *A. cuneatum* are sister to each other (pp = 1.00, bs = 100%) and to the group formed by *A. aureum*, *A. rubrolineatum*, and *A. smithii* (pp = 1.00, bs = 100%). *Aeonium rubrolineatum* is sister to *A. smithii* (pp = 1.00, bs = 100%) and *A. aureum* basal to them (pp = 0.91, bs = 74%). Subclade B contains all remaining analyzed sequences of *Aeonium*, *Aichryson*, and *Monanthes*. *Monanthes anagensis* is, with highest support, basal to all other species. The two *Aichryson* species and *M. icterica* form a subclade basal to the *Aeonium* species (pp = 0.61, bs = 64%). Sequences of *Ai. pachycaulon* and *M. icterica* are intermixed (pp = 1.00, bs = 100%) and two sequences of *Ai. laxum* are sister to them (pp = 0.61, bs = 53%). Basal to this group is Ai_laxum_26634 (pp = 1.00, bs = 100%). *Aeonium goochiae* is basal to all other *Aeonium* species (pp = 1.00, bs = 99%) which could be

separated into two subgroups (pp = 1.00, bs = 100%). One subgroup describes the sister relationship of *A. canariense* and *A. nobile* to each other (pp = 0.97, bs = 88%) and *A. cuneatum* basal to them (pp = 1.00, bs = 100%). Relationships within the second subclade are unresolved containing *A. aureum*, *A. rubrolineatum*, *A. saundersii*, and *A. smithii*.

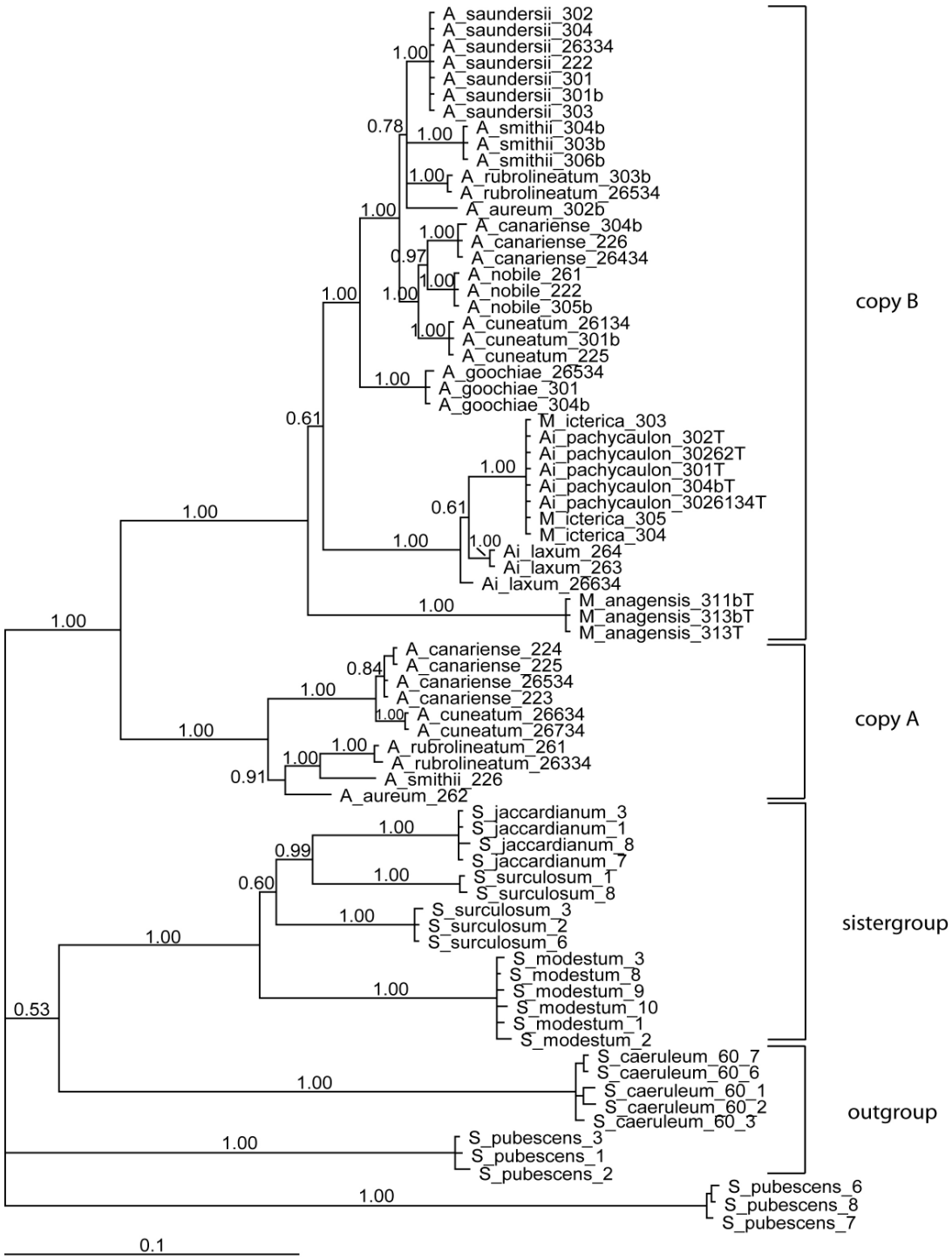


Fig. 16: BI phylogram based on the *MCS_AP3* intron data. Posterior probabilities are given at the nodes.

If only the MCS species were taken into account (BI fig. 42 and ML fig. 43; both appendix) and *Ai. laxum* is considered as outgroup, identical relationships are obtained in the BI and ML phylograms. *Monanthes icterica* is intermixed with *Ai. pachycaulon* (pp = 1.00, bs = 100%) in a sister relationship to *Ai. laxum* (pp = 0.66, bs = 55%) but the position of the whole subgroup is unresolved at the base of the respective phylograms. For the remaining species two subclades (pp = 1.00, bs = 100%) could be detected reflecting mainly the above mentioned relationships for the full intron dataset.

3.3. Blast and Neighbor-joining analyses

To confirm results deduced from the phylograms sequences were blasted. For further confirmation NJ analyses were done that help to classify the sequences and to distinguish more powerfully between orthologous and paralogous gene copies.

Interesting results were obtained for ***MCS_PEPC*** where the clear separation of the four unique *Sedum* sequences (S_pubsecens_243, 1c43 and S_surculosum_34341, 54343) is supported by Blast and NJ analyses. Whereas all other species clones show highest similarity to the same sequences – *PEPC* genes of several *Kalanchoe* species (e.g., AJ252917, X87819, X87818, AJ231288, AJ252946, AJ344052), *Cycas revoluta* (AJ312617) or *Euphorbia tirucalli* (AJ312660) – the separated *Sedum* show other similarities. For S_pubescens_1c43 highest similarity was as well found to *PEPC* genes of *Kalanchoe* species but all of them describe different isoforms as the above mentioned ones. High similarity was found to, e.g., isogenes 5 (AJ344056) and 6 (AJ344057) of *K. pinnata* and to *PEPC* genes of other species such as *Vitis vinifera* (AF236126). Sedum_pubescens_243 shows similarities to, e.g., the *K. pinnata PEPC* isogenes 5 (AJ344056), 6 (AJ344057), 7 (AJ244058), and to *PEPC* genes of *Crataegus* (EU500593), *Digitaria didactyla* (AM690213), *Lotus japonicus* (AB092820), *Lupinus luteus* (AM237200), and *Vicia faba* (AJ011303). Also S_surculosum_34341 and 54343 show highest sequence similarity to *PEPC* genes of different taxa like *Lupinus albus* (AY663388) or *Oryza sativa* (AY187619).

The clear separation of the unique sequences of *S. pubescens* and *S. surculosum* was also confirmed in the NJ phylogram (fig. 45; appendix). All *Aeonium*, *Aichryson*, *Monanthes*, and the remaining *Sedum* sequences are closely related to each other and in a sister relationship to the mentioned *PEPC* isogenes 1, 2, and 3 of diverse *Kalanchoe* species. In contrast, the unique *S. pubescens* and *S. surculosum*

sequences are closely related to the isogenes 5, 6, and 7 of several *Kalanchoe PEPC* genes.

For ***MCS_AP1*** most sequences show highest similarity to *AP1*-like sequences of *Corylopsis sinensis* (AY306146) and *Heuchera americana* (AY306148). Clade specific sequences could be observed for subclade B. Namely *A. goochiae*, *A. rubrolineatum*, *A. saundersii* (including the sequence 33637), and both *Monanthes* species show high sequence similarity to *AP1*-like genes of *Pyrus pyrifolia* (EF423915, EF423916), *Malus x domestica* (AB458503, AY071921, EU672877), *Prunus persica* (EU079377), and *Eriobotrya japonica* (AY880261, AY880262). Furthermore, all species of subclade B, with exception of the both *Monanthes*, show a high similarity to *Citrus sinensis AP1*-like sequences (AY338974, AY338975). Also the *Sedum* species have sequence similarity to the above mentioned species. Therefore, in general no convincing clade or subclade specific pattern was observed.

The NJ analysis of the enlarged *MCS_AP1* dataset revealed that all analyzed sequences of the MCS and *Sedum* species are in close relationship to each other and more distantly related to *AP1*-like sequences of other species. Even the three unique sequences of *A. aureum* and *A. saundersii* are imbedded. They show the closest relationship to *S. caeruleum* but belong clearly to the MCS/*Sedum* clade. Sister to *MCS_AP1* are *AP1*-like sequences of several genera such as *Citrus*, *Malus*, *Prunus*, and *Pyrus*. Basal in a sister relationship are homologs of *AGL8* and *FRUITFUL*-like sequences (data not shown).

A clear pattern was observed for the blasted ***MCS_AP3*** sequences of the studied species. All sequences show highest sequence similarity to the *AP3*-like protein of *K. blossfeldiana* (DQ479358). In addition, the sequences of subclade A show a high sequence similarity to *DEF*-like genes of *Pedicularis groenlandica* (AY524010) or in case of *A. rubrolineatum* to *DEF*-like genes of *Mazus reptans* (AY530538) and *Torenia fournieri* (AB359951). Exceptions are the sequences A_cuneatum_26734 and 26634 that show only similarity to the above mentioned *K. blossfeldiana AP3*-like gene (DQ479358). Nevertheless, a clear *AP3*-like subclade (B) and a *DEF*-like subclade (A) could be distinguished (e.g., fig. 15).

In the NJ analysis MCS and *Sedum* sequences are closely related to each other and cluster together in a sister relationship to *AP3*-like gene sequences of other species (data not shown).

3.4. Species phylogeny based on nrITS

One aim was to compare the genealogies of the low-copy nuclear coding genes with the species phylogeny based on neutral markers. The nrITS alignment was improved and the obtained BI phylogram (fig. 17) shows some differences to the nrITS Maximum parsimony (MP) phylogram of Mort et al. (2002). The MCS are monophyletic comprising all *Aeonium*, *Aichryson*, and *Monanthes* species. Using *S. caeruleum* as outgroup it is together with *S. pubescens* basal to the remaining species. *Sedum jaccardianum* and *S. surculosum* are sister to each other (pp = 1.00) and to *S. modestum* (pp = 1.00). *Sedum* ser. *Monanthoidea* species are in a sister relationship to the MCS species (pp = 1.00), which are distinguished in three main clades. Monophyletic groups are formed by *Aichryson* (pp = 1.00), the perennial *Monanthes* (pp = 1.00), and *Aeonium* (pp = 1.00). The position of the annual *M. icterica* is unresolved but still, the species is closer related to *Aichryson* than to the other *Monanthes* species. If the dataset of Mort et al. (2002) is used without optimization, the BI phylogram resolves *M. icterica* as sister to the *Aichryson* clade (data not shown). The species group comprising *M. icterica* and *Aichryson* is sister to *Aeonium* and to the perennial *Monanthes* species (p = 1.00). Mort et al. (2002) resolved this relationship differently with *M. icterica* basal and sister to *Aeonium* and to the remaining *Monanthes* species but not closely related to *Aichryson*. In the BI phylogram, the perennial *Monanthes* species are sister to *Aeonium* (pp = 1.00). Within the *Aeonium* clade relationships are mainly unresolved, nevertheless several subclades could be recognized. Basal of the *Aeonium* clade is the species group of *A. goochiae*, one studied species, and *A. lindleyi* (pp = 0.99). The position of *A. cuneatum* as further studied species is unresolved but seems to be rather basal to the remaining *Aeonium* species which form four further subclades. The second unresolved subclade B consists of five *Aeonium* species (pp = 0.94) including *A. canariense* as studied species. Subclade C comprises seven different species including *A. aureum*, *A. saundersii*, and *A. smithii* as studied species in sister relationships. *Aeonium smithii* is basal to the other species (pp = 0.70) and *A. saundersii* is sister (pp = 0.52) to the unresolved species group comprising *A. viscatum* and the former *Greenovia* species (pp = 1.00). Subclade D consists of 13 species with *A. nobile* as a further studied species and *A. glutinosum* and *A. glandulosum* basal (pp = 0.78) of this highly unresolved clade. Finally subclade E comprises 14 species, including the studied species *A. rubrolineatum* but

relationships are unresolved. The corresponding ultrametric tree is shown in fig. 46 (appendix).

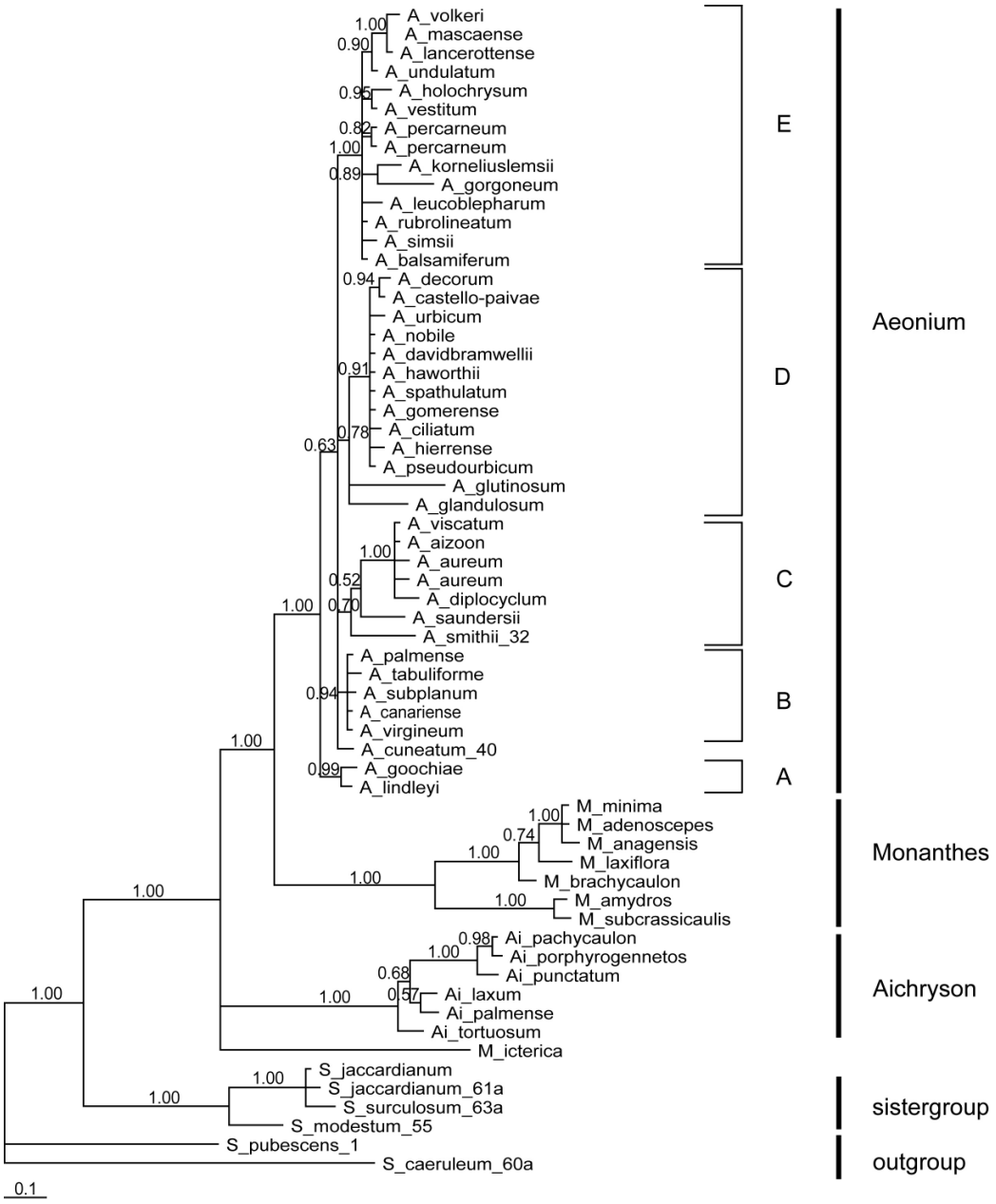


Fig. 17: BI phylogram based on the improved nrITS dataset. Own sequences are marked with numbers. Posterior probabilities are given at the nodes.

3.5. Gene duplications

Duplicated gene copies were detected based on exon phylograms for all three low-copy nuclear genes; for different species and distinguished into orthologs and paralogs. A general observation was the subclade specific duplication for the MCS species, a phenomenon which was not observed for the *Sedum* species. For MCS

species the main clade was separated into different subclades. The respective gene copies within one subclade are orthologous to each other and paralogous to the duplicated sequences (gene copies) of the other subclade. This pattern was mainly obtained for the *Aeonium* species with one exception: the duplication of *MCS_PEPC* for sequences of *Ai. pachycaulon*. Paralogous gene duplications were observed for several *Sedum* species but also for MCS species within the respective subclades. Summarizing the obtained pattern: species of the MCS, especially *Aeonium*, display two or more copies of *MCS_PEPC*, *MCS_AP1*, and *MCS_AP3* in their genomes whereas most *Sedum* species only possess one copy.

For ***MCS_PEPC*** duplicates were found for *A. aureum*, *A. nobile*, *A. rubrolineatum*, *A. smithii*, *Ai. pachycaulon*, *S. pubescens*, and *S. surculosum* (e.g., fig. 9). For the *Sedum* species the observed pattern is unique. Four of the duplicated *S. pubescens* and *S. surculosum* sequences are separated by a very long branch. Nevertheless, for *S. surculosum* a duplication was also detected within the main clade. The MCS species show a subclade specific duplication. The main clade represented by MCS could be separated into two subclades, which each comprises duplicated gene copies of the above mentioned *Aeonium* and *Aichryson* species. A unique pattern was observed for *A. canariense*. Duplicated genes copies were detected within subclade B where two clones are sister to *A. nobile* and A_canariense_411 is basal to the species group formed by *A. cuneatum* and *A. rubrolineatum*.

For ***MCS_AP1*** subclade specific duplications were detected for *A. aureum*, *A. canariense*, *A. cuneatum*, *A. nobile*, *A. saundersii*, and *A. smithii* (e.g., fig. 12). Sequences of *A. aureum* could be found in three different positions in the phylogram. In each subclade one copy of *A. aureum* was detected. Additionally, *A. aureum* forms a unique subclade unresolved at the basis of all analyzed sequences. Also the sequences of *A. saundersii* show a unique duplication pattern. Whereas A_saundersii_33637 is sister to the species of *Sedum* ser. *Monanthoidea*, further sequences were found in the main clade comprising all studied MCS species. Identical patterns of gene duplication were detected for *A. canariense*, *A. cuneatum*, *A. nobile*, and *A. smithii*, which each have a copy in both subclades. Paralogous duplications were detected for the *Sedum* species. Sedum_caeruleum_4 is in a derived position compared to the other *S. caeruleum* sequences. For *S. surculosum*: S_surculosum_3 clusters together with *S. jaccardianum* whereas further sequences form a unique clade as sister to them.

For ***MCS_AP3*** again two subclades were obtained for the MCS species (e.g., fig. 15). Duplications were detected for *A. aureum*, *A. canariense*, *A. cuneatum*, *A. rubrolineatum*, and *A. smithii*. These species are represented in each of the orthologous subclades A and B. Single paralogous gene duplications were detected within *Monanthes* and *Aichryson*. Further single duplication events resulting in paralogs were detected for other species such as *A. canariense* in subclade A or *A. nobile* and *A. smithii* in subclade B. Also for nearly all analyzed *Sedum* species single gene duplications were detected. *Sedum surculosum* shows hereby the strongest pattern. Sedum_surculosum_1 and 8 are sister to *S. jaccardianum* whereas S_surculosum_2, 3, and 6 are basal and sister to this relationship. Paralogs were also detected for *S. modestum* and *S. pubescens*.

In general, two or more subclade specific gene copies were detected for the species of the MCS but only one copy for the *Sedum* species. For *MCS_AP1* and *MCS_AP3* each time a main copy (one subclade) is detected comprising all studied species. In addition, a second (or more) subclades were observed comprising duplicated gene copies of *Aeonium* species. Single duplications were detected for several species within subclades and the deduced copies paralogous. *Sedum pubescens* and *S. surculosum* show duplications for all three studied gene regions. Focusing on the MCS species, *A. aureum* and *A. smithii* show for all three genes duplications; *A. canariense*, *A. cuneatum*, *A. nobile*, and *A. rubrolineatum* have gene duplications at least for two genes, and *A. saundersii* shows only a duplication for *MCS_AP1*. Several species like *A. goochiae*, *M. anagensis*, and *M. icterica* show never a gene duplication event.

3.6. Nucleotide differences, replacements, and amino acid substitutions

Nucleotide differences were counted between all obtained species-specific sequences and also distinguished into paralogous and orthologous species-specific sequences of subclade A and B, respectively. The number and percentage of differences is generally higher between paralogs. Even if synonymous and nonsynonymous substitutions were observed, only nonsynonymous replacements were counted because they may have an important impact on protein evolution. Additionally, it was recognized if the replaced amino acid was quite different compared to the original one.

For ***MCS_PEPC*** a unique pattern was observed. Four unique *Sedum* sequences are separated from the remaining sequences with a very long branch (e.g., fig. 9). Therefore, for *S. pubescens* and *S. surculosum* highest numbers and percentage of nucleotide differences and replacements were observed. They range between 24.9% and 27.6% for the nucleotide differences and between 15.7% and 16.8% for the replacements, respectively (table 10). For the remaining duplicates of *Sedum* 7.5% (*S. pubescens*) and 11.6% (*S. surculosum*) nucleotide differences could be observed. Those were clearly higher than that obtained for orthologs where values range between 0.1% and 0.6%; for *S. surculosum* a higher value of 2.2% was detected. For the MCS species comparisons of nucleotide differences for sequences from subclade A with sequences from subclade B range between 5.5% (*A. aureum*) and 9.4% (*Ai. pachycaulon*). Low values were observed for *A. canariense* (1.1%) and likewise for *S. jaccardianum* (1.4%). However, the relationships of these respective sequences have to be described as “paralogous alleles” since the compared species-specific sequences were found in the same subclade but in derived positions to each other. Nevertheless, their values are still different from those of true alleles or paralogs (for comparison, e.g., Fortune et al. 2007 or Zhang et al. 2008).

The values for replacements range for paralogous gene copies of MCS species between 4.1% (*A. aureum*) and 6.5% (*Ai. pachycaulon*). For orthologous sequences the values range between 0.3% and 0.8% with 0% for *Ai. pachycaulon*. For *Sedum* the values of replacements range between 0.8% and 5.7% for paralogs, and 0.3% and 0.5% for orthologous sequences (0% for *S. caeruleum*).

The values for quite different substituted amino acids (aa) range for paralogs between 54.2% (*Ai. pachycaulon*) and 75.8% (*S. surculosum*) with a mean of 66.9%. As already indicated in table 10 the replaced amino acids in orthologous gene copies are mostly quite similar to the original one.

Table 10: Number and percentage of nucleotide differences, replacements, and quite different amino acids (aa) between species-specific sequences of *MCS_PEPC* distinguished into orthologous and paralogous sequences.

species	paralogs						orthologs					
	nucleotide differences		replace-ments		quite different aa		nucleotide differences		replace-ments		quite different aa	
	no.	%	no.	%	no.	%	no.	%	no.	%	no.	%
A. aureum	61	5.5	15	4.1	10	66.7	2	0.2	1	0.3	0	0
A. canariense	12	1.1	4	1.1	3	75.0	1	0.1	1	0.3	0	0
A. cuneatum							3	0.3	1	0.3	1	50
A. goochiae							3	0.3	1	0.3	0	0
A. nobile	76	6.9	17	4.6	11	64.7	2	0.2	1	0.3	0	0
A. rubrolineatum	78	7.0	16	4.3	11	68.8	2	0.2	1	0.3	0	0
A. saundersii							5	0.5	2	0.5	1	50
A. smithii	82	7.4	18	4.9	12	66.7	3	0.3	2	0.5	1	50
Ai. laxum							5	0.5	1	0.3	0	0
Ai. pachycaulon	104	9.4	24	6.5	13	54.2	1	0.1	0	0.0	0	0
M. anagensis							5	0.5	3	0.8	2	66.7
M. icterica							3	0.3	1	0.3	0	0
S. caeruleum							1	0.1	0	0.0	0	0
S. jaccardianum	15	1.4	3	0.8	2	66.7	7	0.6	2	0.5	1	50
S. modestum	-		-		-		-		-		-	
S. pubescens	276	24.9	58	15.7	41	70.7						
S. pubescens	83	7.5	5	1.4	3	60.0						
S. surculosum	305	27.6	62	16.8	47	75.8	2	0.2	1	0.3	0	0
S. surculosum	128	11.6	21	5.7	14	66.7	24	2.2	2	0.5	1	50

Sequence divergence between paralogs of ***MCS_AP1*** range between 1.2% for *S. surculosum* and 8.3% for *A. aureum* (table 11). For the species of the genus *Aeonium* a mean of 5.8% could be estimated and a range between 4.3% and 8.3% (both *A. aureum*) observed. For *S. caeruleum* a value of 2.8% was found for the paralogous sequences. For orthologous sequences of the MCS species the obtained values are significantly lower than the one of the paralogs ranging between 0.2% (several species) and 0.8% for *A. smithii*. However, most species-specific orthologs showed no nucleotide differences at all. For the *Sedum* species the values range between 0.2% (*S. caeruleum* and *S. modestum*) and 0.3% (*S. jaccardianum*).

The mean value for replacements is 7.6%. Focusing on paralogous sequences the highest value of 12.9% was observed for *A. aureum* and the lowest value of 6.5% for *A. nobile* for the MCS species. For the *Sedum* species the values 2.8% (*S. surculosum*) and 4.6% (*S. caeruleum*) were observed.

Values for replaced amino acids with quite different characteristics were high (table 11) and ranged between 65.2% (*A. saundersii*) and 87.5% (*A. aureum*) for paralogous sequences with a mean of 74.7%.

Table 11: Number and percentage of nucleotide differences, replacements, and quite different amino acids (aa) between species-specific sequences of *MCS_AP1* distinguished into orthologous and paralogous sequences.

species	paralogs						orthologs					
	nucleotide differences		replace-ments		quite different aa		nucleotide differences		replace ments		quite different aa	
	no.	%	no.	%	no.	%	no.	%	no.	%	no.	%
A. aureum	54	8.3	28	12.9	23	82.1	1	0.2	1	0.5	1	100
A. aureum	28	4.3	16	7.4	14	87.5						
A. canariense	32	4.9	17	7.8	12	70.6	0	0	0	0	0	0
A. canariense							0	0	0	0	0	0
A. cuneatum	31	4.8	15	6.9	11	73.3	0	0	0	0	0	0
A. cuneatum							0	0	0	0	0	0
A. goochiae												
A. nobile	29	4.5	14	6.5	10	71.4	0	0	0	0	0	0
A. rubrolineatum												
A. saundersii	43	6.6	23	10.6	15	65.2	1	0.2	1	0.5	1	100
A. smithii	46	7.1	20	9.2	17	85.0	5	0.8	2	0.9	2	100
Ai. laxum	-		-		-		-		-		-	
Ai. pachycaulon	-		-		-		-		-		-	
M. anagensis												
M. icterica												
S. caeruleum	18	2.8	10	4.6	7	70.0	1	0.2	1	0.5	0	0
S. jaccardianum							2	0.3	0	0	0	0
S. modestum							1	0.2	0	0	0	0
S. pubescens	-		-		-		-		-		-	
S. surculosum	8	1.2	6	2.8	4	66.7	0	0	0	0	0	0

The mean value for nucleotide differences between paralogous sequences of ***MCS_AP3*** is 5.6%. The values range between 0.7% (*A. canariense* and *S. jaccardianum*), 1.2% for *Ai. laxum*, and 9.3% for *S. pubescens* (table 12). For *A. canariense*, *Ai. laxum*, and *S. surculosum* duplications were observed within one subclade and should thus rather be defined as “paralogous alleles”. In contrast, the sequences of *S. pubescens* cluster into two well defined clades. Paralogs of *Aeonium* species have a mean value of 5.3% and the observed values were quite similar (table 12). For orthologs lower values were observed that range between 0.1% (*A. canariense* and *S. jaccardianum*) and 1.3% (*M. anagensis* and *S. caeruleum*) with a mean of 0.6%.

Focusing again only on paralogs, the mean value for replacements is 4.5% and range between 0.9% for *A. canariense* and *S. jaccardianum* and 9% for *S. pubescens*.

Calculations of replacements with quite different amino acids resulted in a mean value of 53.6% and range between 30% (*S. pubescens*), 66.7% (*A. canariense* and *A. rubrolineatum*) and 100% for *A. canariense* and *S. jaccardianum*.

Table 12: Number and percentage of nucleotide differences, replacements, and quite different amino acids (aa) between species-specific sequences of *MCS_AP3* distinguished into orthologous and paralogous sequences.

species	paralogs						orthologs					
	nucleotide differences		replace-ments		quite different aa		nucleotide differences		replace-ments		quite different aa	
	no.	%	no.	%	no.	%	no.	%	no.	%	no.	%
A. aureum	41	6.1	9	4.0	5	55.6						
A. canariense	42	6.3	9	4.0	6	66.7	1	0.1	0	0.0	0	0
A. canariense	5	0.7	2	0.9	2	100	2	0.3	1	0.4	1	100
A. canariense							1	0.1	0	0.0	0	0
A. cuneatum	42	6.3	10	4.5	6	60.0	2	0.3	0	0.0	0	0
A. cuneatum							0	0.0	0	0.0	0	0
A. goochiae							2	0.3	0	0.0	0	0
A. nobile							3	0.4	1	0.4	1	100
A. rubrolineatum	42	6.3	12	5.4	8	66.7	2	0.3	2	0.9	2	100
A. rubrolineatum							3	0.4	1	0.4	1	100
A. saundersii							6	0.9	1	0.4	1	100
A. smithii	39	5.8	10	4.5	5	50.0	2	0.3	0	0.0	0	0
Ai. laxum	8	1.2	5	2.2	3	60.0	2	0.3	1	0.4	1	100
Ai. pachycaulon							7	1.0	2	0.9	2	100
M. anagensis							9	1.3	3	1.3	2	66.7
M. icterica							2	0.3	0	0.0	0	0
S. caeruleum							9	1.3	4	1.8	4	100
S. jaccardianum	5	0.7	2	0.9	2	100	1	0.1	0	0.0	0	0
S. jaccardianum							2	0.3	2	0.9	2	100
S. modestum							6	0.9	3	1.3	1	33.3
S. pubescens	62	9.3	20	9.0	6	30	3	0.4	1	0.4	0	0
S. pubescens							4	0.6	3	1.3	1	33.3
S. surculosum	22	3.3	5	2.2	2	40	2	0.3	1	0.4	0	0
S. surculosum							6	0.9	1	0.4	1	100

3.7. Relative Rate Tests

MCS_PEPC sequences revealed significant or highly significant rate differences for all analyzed paralogs. In dependence of the outgroup paralogs of *A. aureum*, *A. nobile*, *A. rubrolineatum*, *A. smithii*, and *Ai. pachycaulon* showed different rates. For *Ai. pachycaulon* this pattern of sequence evolution is only weakly supported compared to the *Aeonium* species.

For ***MCS_AP1*** relative rate tests did never reveal any differences in the evolutionary rate of the studied species and their paralogous gene copies.

For ***MCS_AP3*** rate differences were found for some of the studied species and their respective paralogous sequences. Several comparisons were made for *A. cuneatum* and *A. rubrolineatum* and in some cases significant p-values were observed. Results depend strongly on the selected outgroup species but were confirmed using different

outgroup references. *MCS_AP3* sequences of *A. aureum*, *A. canariense*, and *A. smithii* did not indicate any differences in their evolutionary rate.

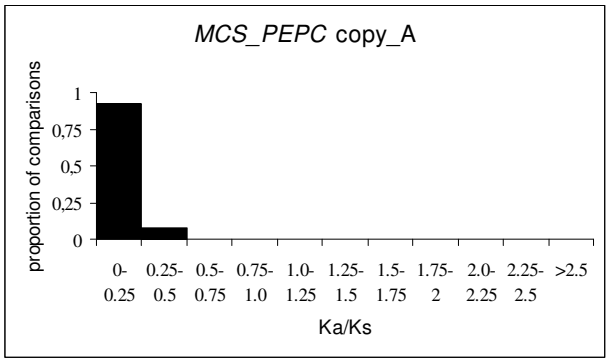
3.8. Ka/Ks-values

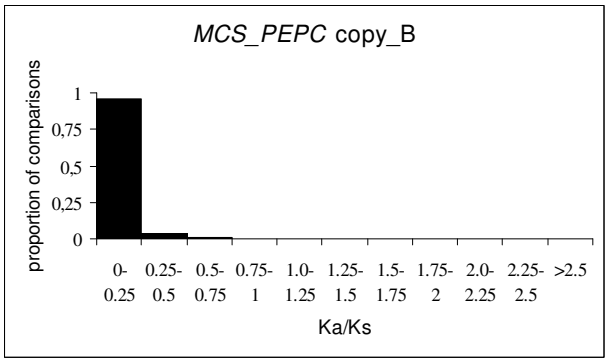
All estimated mean Ka/Ks-values were significantly lower than 1 and define purifying selection. Mean values for the MCS species were highest for *MCS_AP1* copy A and lowest for *MCS_PEPC* copy B. Overall the lowest mean value was detected for the sistergroup species of *Sedum* ser. *Monanthoidea* for *MCS_PEPC* and the highest for the *Sedum* ser. *Monanthoidea* species (sistergroup) for *MCS_AP1* (table 13).

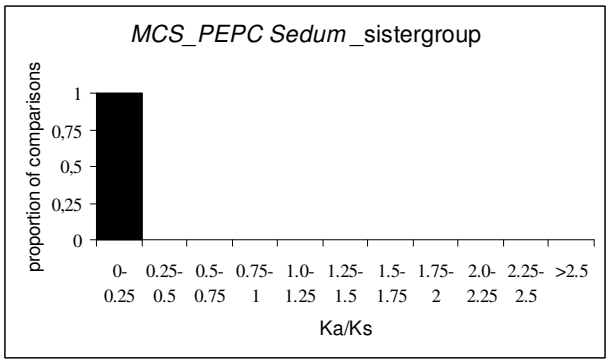
Table 13: Mean Ka/Ks-values. Indication of the copies follows the phylograms in fig. 8-16.

	copy A	copy B	*Sedum*_sister	*Sedum*_outgroup
MCS_PEPC	0.110 ± 0.101	0.106 ± 0.065	0.047 ± 0.034	
MCS_AP1	0.412 ± 0.445	0.373 ± 0.185	0.417 ± 0.510	0.286 ± 0.020
MCS_AP3	0.195 ± 0.163	0.340 ± 0.479	0.092 ± 0.084	0.124 ± 0.115

Figures 18-28 summarize the Ka/Ks-values. Most individual Ka/Ks-values were clearly below 1 but exception were detected for *MCS_AP1* copy A, for *MCS_AP1* of the *Sedum* ser. *Monanthoidea* species (sistergroup), and *MCS_AP3* copy B.







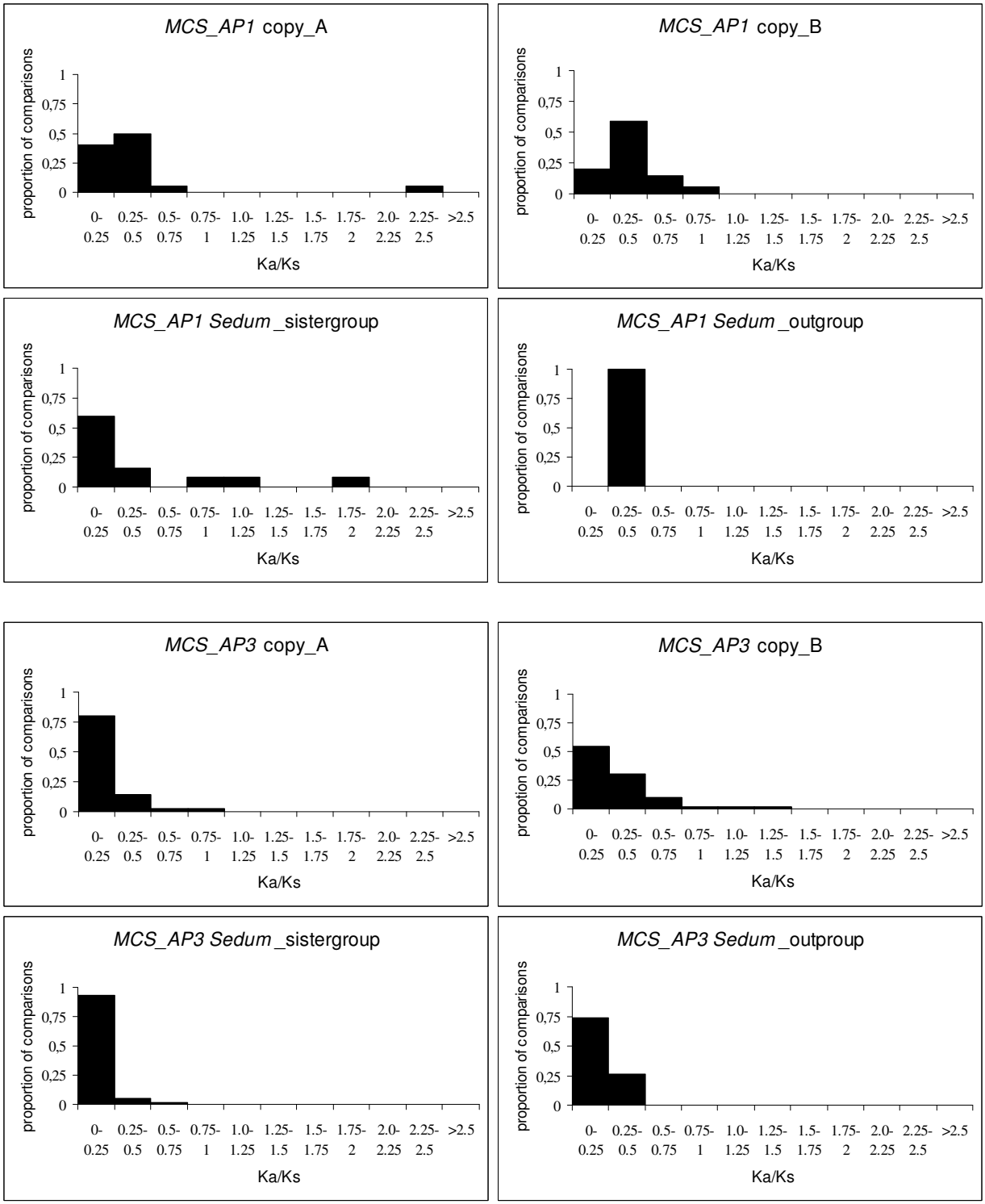


Fig. 18-28: Distribution of Ka/Ks-values for *MCS_PEPC*, *MCS_AP1*, and *MCS_AP3* genes for the MCS and the *Sedum* sister- as well as outgroup species. Indication of the copies follows the phylograms in fig. 8-16.

T-tests compare for each gene the Ka/Ks-values of 1) gene copy A versus copy B of the MCS species, 2) gene copies of the MCS vs. sistergroup and outgroup species, respectively and 3) regulatory versus structural genes (table 14).

Significant differences were observed for the two copies of *MCS_AP3* where the mean Ka/Ks-value of copy B is significantly higher than the one of copy A. For *MCS_PEPC* and *MCS_AP1* similar values were observed for each of the two copies (table 13 and 14).

In general, the gene copies of the MCS species have higher mean Ka/Ks-values compared to the *Sedum* specific copies. An exception is the copy of the sister species (*Sedum* ser. *Monanthoidea*) for *MCS_AP1* that has with 0.417 ± 0.510 a higher mean value than *MCS_AP1* copy A (0.412 ± 0.445) and copy B (0.373 ± 0.185); but the results are not significant (table 14).

The regulatory genes *MCS_AP1* and *MCS_AP3* have higher mean Ka/Ks-values as the structural gene *MCS_PEPC* and all these relationships are significant (table 14).

Table 14: Comparison of Ka/Ks-values for the respective genes for 1) copy A and B of the MCS species, 2) copies of the MCS vs. sister- and outgroup (OG) species, respectively, and 3) regulatory vs. structural genes based on t-tests (n.s. = not significant, * significant ($p < 0.05$) and ** highly significant ($p < 0.01$)).

	MCS_ PEPC _A	*MCS_ PEPC _B*	*MCS_ PEPC* _sister	*MCS_ AP1 _A*	*MCS_ AP1 _B*	*MCS_ AP1_* sister	*MCS_ AP1_* OG	*MCS_ AP3 _A*	*MCS_ AP3 _B*	*MCS_ AP3_* sister	*MCS_ AP3_* OG
MCS_ PEPC_A	-	n.s.	**	**	**			**	**		
MCS_ PEPC_B	n.s.	-	**	**	**			**	**		
MCS_ AP1_A	**	**		-	n.s.	n.s.	n.s.	*	n.s.		
MCS_ AP1_B	**	**		n.s.	-	n.s.	**	**	n.s.		
MCS_ AP3_A	**	**		*	**			-	**	**	*
MCS_ AP3_B	**	**		n.s.	n.s.			**	-	**	**

3.9. Selection pressure

No positive selection pressure was revealed for ***MCS_PEPC*** with the codon-based Z-test if the whole dataset is analyzed. Focusing only on the MCS species, positive selection was detected between M_anagensis_303T and 305T. In general, purifying selection is favored over strict neutrality and explains best the gene evolution.

For ***MCS_AP1***, the null hypothesis of strict neutral evolution was rejected favoring the alternative hypothesis of purifying selection. Positive selection was detected only between S_jaccardianum_1 and S_surculosum_3 (only for p-distance setting).

The analyzed sequences of ***MCS_AP3*** showed in most cases purifying selection but for several *Aeonium* sequences in subclade B the null hypothesis of strict neutrality could not be rejected. In addition, positive selection was indicated between the sequences of *A. nobile* and A_cuneatum_225, 26134, and 301b of the subclade B. Focusing only on the MCS species positive selection was additionally detected between M_anagensis_311T and 313T. Even if purifying selection is the main selection pressure, positive and especially neutral selection cannot be ruled out.

Selection pressure acting on a particular codon can differ and thus, analyses were done with Selecton to detect sites under positive selection. Sites under positive selection were detected for *MCS_PEPC*, *MCS_AP1*, and *MCS_AP3* but were only significant for the two regulatory genes.
Only one site under positive selection could be found for *MCS_PEPC*, 37 amino acids indicating positive selection for *MCS_AP1*, and three for *MCS_AP3*. For *MCS_AP3* all three amino acids were found within the MADS-box domain but must be handle with caution because missing data at the beginning of several *Sedum* sequences could influence the results. For *MCS_AP1* 17% of codons under positive selection were observed. Three codons indicating positive selection were found within the MADS-box (10.7%), six within the I-domain (14.6%), 10 within the K-domain (14.7%), and 18 in the C-terminal domain (22.5%).

4. Discussion

4.1. Phylogenetic reconstructions

Utilization of neutral markers for reconstructing phylogenetic relationships for species evolving within fast radiations often results in poorly resolved topologies or unresolved polytomies (Mes and ‘t Hart 1996, Baldwin et al. 1998). However, accurate knowledge of these relationships is critical for studying processes of speciation and evolution, and for the identification and proof of key innovations (Hodges 1997, Jorgensen and Frydenberg 1999, Jorgensen and Olesen 2001). Since low-copy genes can be involved in the determination of phenotypes and adaptation, a nuclear gene phylogeny might answer questions about morphological and physiological evolution influenced by the studied genes (Sang 2002). In addition, low-copy nuclear genes may provide more information to resolve relationships and thus, could overcome the problem of missing synapomorphic characters (reviewed in Sang 2002 or Small et al. 2004). Therefore, numerous studies have attempted to evaluate the use of low-copy nuclear genes to reconstruct phylogenies and to estimate their evolutionary impact (e.g., Bailey and Doyle 1999, Barrier et al. 2001, Malcomber 2002, Fan et al. 2004, Grob et al. 2004, Álvarez et al. 2005, Purugganan and Robichaux 2005, Syring et al. 2005, Janssens et al. 2007, Tu et al. 2008).

In the present study homologs of *PEPC*, *AP1*, and *AP3* genes were used to **reconstruct phylogenetic relationships** within the island radiation of the Macaronesian Crassulaceae Sempervivoideae (MCS). The results obtained from the full-length genic region, and exons and introns only of the studied low-copy nuclear genes, were very complex but some generalizations could be drawn especially by focusing on the main copies and clades.

The MCS and *Sedum* species are well separated. The studied *Sedum* species could be distinguished in the outgroup species *S. caeruleum* and *S. pubescens* and in the sistergroup species of *Sedum* ser. *Monanthoidea*. The species of the MCS cluster together in a main clade, and are separated into at least two subclades. The main copy (subclade) comprises in all but one case all studied MCS species and a separation between the three genera *Aeonium*, *Aichryson*, and *Monanthes* could be observed. The obtained relationships both support and contradict relationships inferred by other markers.

Within ***Aeonium*** a well supported relationship was found between *A. aureum*, *A. saundersii*, and *A. smithii*. This relationship is confirmed by morphological and molecular data (Mes and ‘t Hart 1996, Jorgensen and Frydenberg 1999, Mort et al. 2002, reanalyzed nrDNA ITS) even if Liu (1989) and Mes (1995) separated the species into different sections (see also table 15; appendix). These species share an increased number of flower organs that is highest for *A. aureum* (28-32), followed by *A. saundersii* (12-16), and *A. smithii* (8-12) compared to 6-11 for the remaining *Aeonium* species. Further common features are their yellow flower color (Liu 1989) and to some extent their CO_2 fixation pathway. For *A. aureum* C_3 fixation is reported; *A. saundersii* and *A. smithii* have an intermediated pathway including C_3 and CAM activity (Tenhunen et al. 1982, Pilon-Smits et al. 1992, Mort et al. 2007). Contradicting, Lösch (1990) reported strong CAM for *A. smithii*. Analysis of *MCS_AP1* places one additional species in this group: *A. rubrolineatum*. Like the other species, *A. rubrolineatum* has a slightly increased number of flower organs (9-11), one of the main characters encoded by homologs of *AP1*. In addition, it shares the yellow flower color with all species in the group and a subshrub habit with *A. saundersii* and *A. smithii* (Liu 1989).

Another close relationship, confirmed by all studied genes, was detected between *A. canariense*, *A. cuneatum*, and *A. nobile*. Analyses of both *MCS_PEPC* and *MCS_AP3* also include *A. rubrolineatum* in the group, contradicting the results of *MCS_AP1*. These relationships are surprising and generally not supported by other markers (Liu 1989, Mes 1995, Mes and ‘t Hart 1996, Jorgensen and Frydenberg 1999, Mort et al. 2002). Morphological data combine at least the herbaceous rosette plants *A. canariense* and *A. cuneatum*. Both belong to sections *Canariensia* (Lems 1960, Mes 1995) and *Patinaria* (Liu 1989), respectively. A sister relationship was also suggested by *MCS_AP3* but is not confirmed using neutral molecular markers (Mes and ‘t Hart 1996, Jorgensen and Frydenberg 1999, reanalyzed nrITS). That *MCS_AP3* combines both species could be due to the general similar habit and especially because of similar petals (8 to 10), a character encoded by *AP3* homologs. In the *MCS_PEPC* phylograms a sister relationship between *A. cuneatum* and *A. rubrolineatum* could be detected. For both species C_3 fixation prevails and weak CAM activity is possible. For *MCS_AP1*, *A. cuneatum* is basal to the sister relationship of *A. canariense* and *A. nobile*. Mes and ‘t Hart (1996) showed that *A. cuneatum* is closely related to *A. nobile* and *A. rubrolineatum* based on morphological

data and restriction site mutations of cpDNA, respectively. However, the latter relationship was not confirmed in the respective combined dataset (Mes and 't Hart 1996). No other markers support the relationship between *A. nobile* and *A. canariense* as suggested by *MCS_PEPC* and *MCS_AP1*, or that between *A. nobile* and *A. rubrolineatum* (*MCS_AP3*). Lems (1960) and Liu (1989) treated *A. nobile* as monotype of sect. *Megalonium*, and Mes (1995) ordered it to sect. *Leuconium* not closely related to any of the other above mentioned species. Mort et al. (2002) sorted it together with species such as *A. haworthii*, *A. pseudourbicum* or *A. urbicum* in clade 4. In contrast, *A. rubrolineatum* belongs to sect. *Holochrysa* (Lems 1960), Liu (1989) and Mes (1995) ordered it to sect. *Aeonium*, and Mort et al. (2002) in clade 3 (see also table 15 and fig. 2).

The close relationship of these four species is also surprising in the present study. Focusing on relationships inferred by *MCS_PEPC* data, the strong CAM species *A. nobile* is closely related to species where C_3 prevails (Lösch 1990). Concerning the number of flower organs the species are not very different (between 7 and 11; Liu 1989) and a close relationship might be assumed. In general the species show no strong morphological agreement. Species with pale yellow green flowers (*A. canariense*) are combined with yellow ones (*A. cuneatum*), yellow with red stripes (*A. rubrolineatum*), and dark red (*A. nobile*). *Aeonium canariense* and *A. cuneatum* are herbaceous species occurring in the laurel forest on Tenerife whereas *A. nobile* and *A. rubrolineatum* are subshrubs on La Palma and La Gomera (Liu 1989).

In all inferred phylograms and focusing on the main copy, ***A. goochiae*** is basal to all other *Aeonium* species. Its position is supported by morphological data (Liu 1989) and confirmed by nrITS but not when cpDNA and nrITS were combined (Mort et al. 2002). Jorgensen and Frydenberg (1999) reported a close relationship of *A. goochiae* with *A. lindleyi* and *A. viscatum* unresolved in a larger clade. Mort et al. (2002) confirmed the sister relationship to *A. lindleyi* in the combined cpDNA/nrITS phylogram. They ordered both species together with *A. aureum* and *A. saundersii* to clade 2. However, a relationship to both of the latter species is not detected in the present study. Liu (1989) described *A. goochiae* as a distinct species within sect. *Petrothamnium*. It can be distinguished by its pinkish flowers and very thin leaves whereas other characters suggest a distant relationship to *A. lindleyi*. *Aeonium goochiae* was suspected to connect the genus *Aeonium* with the genus *Aichryson*

(Lems 1960, Liu 1989). The present data clearly support this hypothesis given the basal position of *A. goochiae* in all phylograms.

Resolving relationships of ***Aichryson*** and ***Monanthes*** with the present data is difficult since only a low number of species was included. However, one noteworthy remark has to be made; as expected the *Monanthes* species show mostly diverging phylogenetic patterns. *Monanthes anagensis* is mainly found in sister relationships to the *Aeonium* species. A close relationship to *Ai. pachycaulon* is only resolved in the *MCS_PEPC* phylograms, but missing representatives of *M. anagensis* for the main copy may cover the pattern (e.g., fig. 9).

In contrast, ***M. icterica*** always shows a sister relationship to the *Aichryson* species (*MCS_PEPC*) or is even imbedded and intermixed with them (*MCS_AP3*). For *MCS_AP1* missing amplification of *Aichryson* makes a deduction of the relationships difficult. Most times the positions of both *Monanthes* are unresolved but they do not cluster as sister species. The position of *M. icterica* is traditionally highly debated. Nyffeler (1995) emphasis the *Aichryson*-like habit of *M. icterica* but the enlarged nectariferous scales and bladder cell-idioblasts order it unambiguously to *Monanthes*. Mes et al. (1997) discussed a sister relationship between *M. icterica* and *Aichryson* based on morphological and molecular data. Mort et al. (2002) reported, based on nrITS and combined cpDNA/nrITS data, that *M. icterica* is basal and sister to *Aeonium* and to the perennial *Monanthes* species. Chloroplast markers resolve *M. icterica* in a sister relationship to *Aichryson*. Reanalysis of the nrITS dataset also suggests this sister relationship. Based on the present study, the genus *Monanthes* is only monophyletic if *M. icterica* is separated from the perennial species.

As discussed above the uniqueness of *M. icterica* was shown before. The annual life form and the deviating chromosome base number of x = 10 distinguish this species. RAPD patterns and nrITS data differ significantly and support the assumption that *M. icterica* is genetically only distantly related to the perennial *Monanthes* (Mes et al. 1997). No hybrids have been reported involving *M. icterica*. Even if this phenomenon is assumed given the diverging chromosome base number, it is still interesting since hybridization in the genus is considered to be frequent and easily possible (Nyffeler 1995). The inclusion of *M. icterica* in the genus *Monanthes* based mainly on flower morphology but multiple origin of flowers with large nectariferous scales could be considered like in *S. surculosum* or *S. napiferum* (for reference see Mes et al. 1997).

Nevertheless, with respect to the studied low-copy nuclear genes *M. icterica* is clearly closely related to the *Aichryson* species.

Relationships and positions of the ***Sedum*** species are mainly confirmed with the analyzed low-copy nuclear genes as compared to existing morphological, chromosomal, and neutral molecular markers. *Sedum jaccardianum*, *S. modestum*, and *S. surculosum* represent the species of *Sedum* ser. *Monanthoidea* and cluster together as sistergroup species of the MCS. As predicted by other markers, *S. jaccardianum* and *S. surculosum* are thereby sister to each other and *S. modestum* is sister and basal to them. However, the pattern of *S. surculosum* for all studied genes is remarkable. For example, two *S. surculosum* sequences are sister to *S. jaccardianum* for the main copy of *MCS_PEPC* whereas the sequence S_surculosum_2c4391 is basal to the remaining species in the exon phylogram (e.g., fig. 9). In the exon phylograms of *MCS_AP1* S_surculosum_3 is mixed with *S. jaccardianum*, whereas the remaining sequences are sister to *S. jaccardianum* (e.g., fig. 12). Focusing on the coding region of *MCS_AP3* (e.g., fig. 15), *S. jaccardianum* and S_surculosum_1 and 8 are sister to each other. Sister to them are the remaining *S. surculosum* sequences (2, 3, 6) and these relationships were confirmed by the full-length data. A possible explanation could be hybridization, which is generally easy between *Sedum* species (for reference see van Ham and ‘t Hart 1998). Moreover, *S. surculosum* is tetraploid which may hint at an ancient introgression event via allopolyploidization. Given that in all cases some gene copies show a clear relationship to *S. jaccardianum*, this species may be involved in the evolution of *S. surculosum*.

Positions and relationships of both outgroup species, *S. caeruleum* and *S. pubescens*, were difficult to resolve because of partly missing amplification. However, they are basal to the remaining species and related to the analyzed species of *Sedum* ser. *Monanthoidea*.

In general the three **low-copy nuclear genes** support similar species groups and relationships; in several cases the species phylogeny based on morphological and other molecular markers is contrasted. Jorgensen (2002) discussed that radiation and morphological differentiation of *Aeonium* species is the result of adaptation to ecological conditions and not a reflection of the species phylogeny. Traits such as growth-form, plant height, leaf form, and inflorescence length (maybe partly

influenced by *MCS_AP1*) vary and are adaptive in response to ecological conditions. Thus, ecology may better predict the variance in a number of traits. If the characters are adaptive, selection may act on the genes that are involved in these above mentioned characters and would not evolve neutrally as phylogenetic markers like nrITS and cpDNA do (see also Hodges and Arnold 1994b). Selection pressure may fix mutations, which can be used as information for phylogenetic analyses (reviewed in Sang 2002). Predicted by Jorgensen and Olesen (2001), application of markers that may be related to these fitness factors could vary considerably and may better resolve relationships between the species (see also Mes and 't Hart 1996).

On the other hand, pubescence of floral organs (maybe related to CAM and thus, *MCS_PEPC*), flower size (petal length; *MCS_AP3*), number of ovaries, stamens per flower, and number of petals (the latter both maybe partly influenced by *MCS_AP3*) correlate with phylogeny (Jorgensen and Olesen 2001) suggesting that these genes may evolve neutrally and, thus, reflect the species phylogeny.

The assumption that *MCS_AP1* may have evolved in adaptation but *MCS_PEPC* and *MCS_AP3* under neutral selection could explain differences in the topologies inferred by these different genes. At least for *MCS_AP1* sites under positive selection could be found, even if patterns for *MCS_AP1* are quite diverse and difficult to explain (see below). On the other hand, for *MCS_AP3* evidence for neutral evolution was obtained. However, in general, strong purifying selection was detected acting on all analyzed genes and copies; substitution rates are restricted to avoid deleterious mutations within the coding sequence (Sang 2002, Janssens et al. 2007). Thus, conflicts between species and/or the respective gene topologies could maybe be explained by conservation and selection. Whereas conflicts between the genealogies may arise because different mutations are selectively fixed in the respective genes of the species, conflicts to the species phylogeny may arise because of the random accumulation of mutations in neutrally evolving nrITS and cpDNA regions.

Nevertheless, the obtained results must be handled with caution. The greatest challenge of working with low-copy nuclear genes is the differentiation between orthologous and paralogous gene copies since only the former resolve the species phylogeny (Litt and Irish 2003). Preliminary results of the present study suggest that the relationships inferred only from orthologs are similar to those of the whole dataset including paralogs and orthologs. However, inclusion of more species, improvement

of alignments comprising only orthologs, and subsequently further analyses are necessary to confirm the results and to enable more reliable conclusions.

Phylogenies based on low-copy nuclear genes have confirmed in several studies the relationships found for neutral evolving marker and/or other regulatory genes. Bailey and Doyle (1999) confirmed the species phylogeny of several Brassicaceae species using an intron region of the *PISTILLATA* gene. Koch et al. (2001) confirmed with their phylogeny based on *chalcone synthase* (*CHS*) and *AP3* gene promoter sequences relationships deduced by a combined *matK/CHS* analysis for Brassicaceae species. Analysis of *Atmyb2* flanking sequence data resulted in highly concordant phylogenies for *Arabidopsis* compared to a nrITS phylogeny (Beck et al. 2007). Durbin et al. (2003) found that relationships for *Ipomoea* based on *CHS* genes were quite similar to results based on nrITS and *waxy* gene sequences. The phylogeny of the tribe Andropogoneae (Poaceae) based on *FLORICAULA/LEAFY* genes agreed largely with previously published phylogenies using other nuclear genes (Bomblies and Doebley 2005) and the same was true for the genus *Amorphophallus* (Araceae) using the *FLORICAULA*/*LEAFY* second intron (Grob et al. 2004). Zhang et al. (2008) confirmed the species relationships of *Cornus* with their genealogy of a *PISTILLATA*-like gene copy and Fan et al. (2004) with the *Myc*-like anthocyanin regulatory gene. Álvarez et al. (2005) resolved congruent clades comparing previous phylogenies of *Gossypium* with one inferred by three low-copy nuclear genes and Janssens et al. (2007) found highly congruent topologies for *Impatiens* comparing the *AP3*/*DEF* K-domain with *atpB-rbcL* data.

In contrast to the multiplicity of the above mentioned studies is, e.g., the study of Malcomber (2002). He found that all used markers and methods were insufficient in constructing relationships of Rubiaceae and significant incongruent data partitions between nrITS and the *PEP*-large intron were revealed by a PHT test. Syring et al. (2005) found for four low-copy nuclear genes in *Pinus* that the individual loci do not uniformly support either the nrITS or cpDNA hypotheses. In some cases the low-copy nuclear genes produced even unique topologies. Fortune et al. (2007) detected incongruences between species and gene trees in the *waxy* gene phylogeny of the hexaploid *Spartina* species. Also phylogenetic analysis performed either on one or the other duplicated *waxy* gene copy did not alter the topologies. Tu et al. (2008) detected incongruences between cpDNA, *waxy*, and *LEAFY* gene phylogenies in the

genus *Nolana*, which may have been due to reticulate evolution, lineage sorting, and duplication of the *waxy* gene.

Just as the entire low-copy nuclear gene itself can alter the obtained topologies, the different **gene regions** may reveal conflicting signals as well. Exons are quite conserved and may not provide enough information to resolve relationships within closely related species. In contrast, introns are attractive for evaluating relationships among closely related taxa since they are diverse, fix mutations, and diverge at relatively rapid rates (Syring et al. 2005). Resolution power of full-length datasets may be simply explained by their length and amount of information that they provide even if they combine conserved and quite diverse data partitions.
In the present study, exon and intron regions were combined and PHT tests suggest significant differences between them for *MCS_AP1* ($p = 0.02$). Nevertheless, both datasets were combined due to the lack of well supported differences between the topologies inferred by the two independent datasets (see Grob et al. 2004). The full-length phylograms provide a better resolution and higher support values compared to exon phylograms, maybe because more phylogenetic information was provided. Conflicts were observed for the position of A_saundersii_33637. This sequence is basal to the main clade and the *Sedum* species in the full-length phylograms (e.g., fig. 11) but sister to the analyzed *Sedum* species in the exon phylograms (e.g., fig. 12). Aeonium_saundersii_33637 is an additional sequence and maybe gene copy of *MCS_AP1*. The observed conflicting position may be due to a diverging function. Therefore, different selection pressure and fixation may have acted on this gene copy and resulted in another position in the exon phylograms. Further conflicts were found for S_surculosum_3 that is intermixed with *S. jaccardianum* inferred by the exons but basal to the *S. surculosum* sequences in full-length phylograms. Hybridization and gene duplication could be reasons for the observed pattern. Relationships inferred by introns are similar to the full-length phylograms.
PHT tests revealed no differences between exon and intron regions for *MCS_PEPC* and *MCS_AP3* suggesting the same evolutionary rate for the data partitions. Relationships inferred in full-length and exon phylograms of *MCS_PEPC* were identical and comparisons to intron phylograms were difficult because of missing resolution.

For *MCS_AP3* minor differences could be observed between the three gene regions. Phylograms based on full-length data provided the best resolution. Conflicting positions were detected for the *Sedum* species. In the exon phylograms *S. caeruleum* is basal to the species of *Sedum* ser. *Monanthoidea* and to the main clade. In contrast, *S. caeruleum* is basal and sister to the species of *Sedum* ser. *Monanthoidea* in the full-length dataset and here, all *Sedum* species form a separated *Sedum* specific subclade. Exon phylograms based on conserved regions that were likely influenced by the gene functions may explain the classification of *S. caeruleum*. In the full-length datasets more information was used and the diverse intron data partitions may partly conflict the conserved exon partition. Intron phylograms show similar topologies as full-length phylograms except within subclade B where *A. rubrolineatum* and *A. cuneatum* showed conflicting positions (e.g., fig. 13).

In general, only few conflicts were found between intron, exon, and full-length phylograms. There are several explanations why intron and exon regions could evolve with similar rates. Low-copy nuclear genes could evolve in a neutral fashion and similar selection pressure and mutation rates could be assumed for the data partitions (see Beck et al. 2007). In accordance to Jorgensen and Olesen (2001) and Jorgensen (2002) neutral selection pressure may be assumed for *MCS_PEPC* and *MCS_AP3* and is at least partly confirmed for *MCS_AP3* in the present study. Furthermore, conserved structures in introns, like regulatory elements, could influence the divergence of this region (Bailey and Doyle 1999, Small et al. 2004) as well as hitchhiking where the evolution of intron regions follows the evolution of exon regions or concerted evolution may influence the diversification of intron and exon regions (e.g., Small and Wendel 2002).

4.2. Gene duplications

Polyploidy is a common phenomenon in plant evolution and duplication of genes in a genome is the most obvious molecular consequence (Kramer et al. 1998, Blanc and Wolfe 2004, Moore and Purugganan 2005). Polyploid species appear to vary from their diploid progenitors in a variety of ecologically important traits and genome duplication may provide a molecular mechanism for ecological diversification (Lawton-Rauh et al. 2003).

Gene duplication can also arise independently and randomly within species. MADS-box and *PEPC* genes are represented in gene families and therefore the studied low-copy nuclear genes are well-known members of multicopy gene families. Gene duplication could lead to increased diversity and functional innovation; recent studies described diverse potential outcomes: pseudogenization, functional redundancy, subfunctionalization, neofunctionalization or simple gene loss (Lynch and Conery 2000, Zhang 2003, Moore and Purugganan 2005).

Summarizing the results of the present study, **gene duplications** could be observed for all three analyzed genes and for different species. For *A. aureum* and *A. smithii* gene duplications were detected for all three analyzed genes. Other species such as *A. nobile*, *A. rubrolineatum*, and *A. saundersii* only show duplications for one or two of the genes. In addition, *A. goochiae*, both *Monanthes*, and *Ai. laxum* did not show any gene duplication. *Aichryson pachycaulon* showed duplication for *MCS_PEPC* but being tetraploid, detection of duplicates is likely. In contrast, *A. goochiae* or *M. anagensis*, tetra- or even hexaploid species (Mes et al. 1997, Jorgensen and Olesen 2001) never showed any duplicates. Subclade specific gene duplications for the polyploid MCS – especially *Aeonium* – was a common feature, whereas the diploid *Sedum* species showed duplications only within the *Sedum* specific clade. Thus, whereas for the MCS species at least two subclades comprising each orthologous gene copies could be obtained, single duplications that result in paralogous gene copies were observed for the *Sedum* species and additionally for some MCS species. Gene duplications within MCS species could be the result of two different events. Island colonization was connected with polyploidization and the duplication of the genome, and thus genes, is the most obvious consequence (Blanc and Wolfe 2004). On the other hand duplicates may have arisen randomly within species by single, independent duplication of a gene. In the present study both outcomes are suggested by the obtained topologies. The *MCS_PEPC* exon phylograms suggest that the duplication was the consequence of polyploidization. Two subclade specific gene copies were detected for the MCS species whereas only one CAM-specific gene copy was found for the *Sedum* species. Both MCS copies are under strong purifying selection, seem to be functional, and show so far no evidence for neofunctionalization. The main copy for the MCS species is copy B and five species exhibit the duplicated gene copy A (e.g., fig. 9). All species analyzed contain the

main copy B except *M. anagensis* which possesses only copy A. If redundant copies and thus gene expression is available, random silencing and/or loss of one copy can be tolerated in a species (Zhang 2003). Given that *M. anagensis* is tetraploid and of potential hybrid origin (Mort et al. 2002) additional copies should be detectable. However, if PCR failure is excluded, it must be assumed that one copy of *PEPC* is enough to fulfill the gene function and other copies were lost.

For *A. nobile and A. smithii* both *MCS_PEPC* copies were detected and the species are known for strong CAM activity (Lösch 1990). For these species, a duplication of the gene or retaining both copies may be advantageous. Harder to explain is the observed duplication for *A. aureum* and *A. rubrolineatum* because both are C_3 species with weak CAM activity. On the other hand, this observation may support the hypothesis that the duplicates are rather randomly retained or lost than gained. Since both gene copies are also under strong purifying selection further support for retention instead of gain is obtained. It is also noteworthy that *A. aureum* shows gene duplications for all three analyzed genes (see also below). *Aichryson pachycaulon* is another C_3 species that possesses both copies. In addition to polyploidization the tetraploid status (Uhl 1961) could be an explanation and that the duplicates may have arisen by hybridization or allopolyploidization. Still, the ordering of the duplicated copies in the two specific orthologous subclades suggests that polyploidization may explain this pattern better.

For *MCS_AP1* two main copies and two additional copies, for *A. aureum* and *A. saundersii*, were detected for the MCS species. For *Sedum* two copies were found; one copy being specific for the sistergroup species of *Sedum* ser. *Monanthoidea* and one specific for the outgroup species *S. caeruleum*. Concerning the gene duplications, the observed relationships seem to support both hypotheses: random and independent duplication within species and duplication as consequence of polyploidization. It could be imagined that the two copies in the main clade are the consequence of polyploidization whereas the additional copies of *A. aureum* and *A. saundersii* may result from single gene duplications. That *MCS_AP1* does not resolve the MCS species as monophyletic may support this suggestion. However, *AP1* is part of an extended gene family (Irish and Litt 2005) and *AP1*-like genes are not restricted to flower organs since they play broad roles in both vegetative and reproductive development (Shan et al. 2007). In the present study, amplification was based on genomic DNA and all potential copies could be amplified without restriction to specific

tissues. Additional copies could just represent different members of the gene family or tissue specific copies. Nevertheless, Blast and NJ analyses cluster the additional copies of *A. aureum* and *A. saundersii* with the other studied *MCS_AP1* sequences. Intensive estimation of sequence similarity, phylogenetic reconstruction, studies of cDNA sequences, predictions of expression patterns (paralogs may differ in timing or tissue), southern blotting, and genetic mapping would be powerful tools to understand the pattern better and in more detail and should be included in future studies. However, it is important to remember that *MCS_AP1* amplification for *Aichryson* and *S. pubescens* failed and may be an indication that the *MCS_AP1* gene family could be diverse and exist in several copies in the analyzed species. Also the specific gene copies of *Sedum* ser. *Monanthoidea* and *S. caeruleum* are in a derived relationship to each other and may represent different gene copies of *MCS_AP1* (e.g., fig. 12).

Focusing on the additional copies, whereas *A. aureum* shows duplications for all three analyzed genes, *MCS_AP1* is the only gene duplication detected for *A. saundersii*. This duplication is not subclade specific suggesting that maybe a derived copy was amplified. The *AP1* gene family is known for frequent major gene duplications such as *AP1* and *CAL* in *Arabidopsis* (Shan et al. 2007). In addition, recent duplications within one or few closely related species are extremely common. New genes are frequently recruited in the genome, however it is not clear whether recent duplicates will be functionally fixed in genomes or will eventually become pseudogenes (Shan et al. 2007). This phenomenon is also suggested in the present study and will be discussed in detail below.

Altogether, all observations – also the results of *MCS_AP3* with subclade specific duplications and an accelerated number of gene copies for the MCS compared to the *Sedum* species – rather suggest that the duplicated subclade specific gene copies may be the result or consequence of the polyploidization event that was connected with island colonization. This observation is in line with Barrier et al. (2001). Consistent with the allopolyploid origin of the studied HSA species (Barrier et al. 1999), these contain mostly two copies of the studied genes *ASAP1*, *ASAP3*, and *ASCAB9* whereas their recent common ancestor, the diploid North American Tarweeds, only possess one copy. Barrier et al. (2001) used copy specific primers to amplify the different target gene copies. A disadvantage of the present study is that no copy specific primers were used. Thus it is not possible to determine if missing

gene copies are the result of gene loss or PCR failure. However, in the study of Lawton-Rauh et al. (2003), a corresponding population based study to the one of Barrier et al. (2001), extensive tests revealed missing gene copies for some individuals that seemed to be deleted rather than not amplified despite evidence that both copies were expressed in the species.

Still, in the present study gene duplication could also have occurred later in the speciation process; after the split of *Aichryson* and *Monanthes* or after the separation of *A. goochiae* and *A. lindleyi* (compare indicated nodes 1 and 2 in fig. 46; appendix). However, chromosome base numbers do not support these hypotheses since they do not diverge for the species of *Aeonium* and *Monanthes* and given that also the *Aichryson* species have accelerated chromosome base numbers connected with polyploidization. Thus, random loss of gene copies in the polyploid species is a more likely explanation for the pattern of missing duplicated gene copies for several species, if PCR failure is excluded. In addition, the observed pattern of *Ai. pachycaulon* and *M. anagensis* for *MCS_PEPC* highly suggests that polyploidization may explain the general pattern of the duplicated genes. Independent single gene duplication or strong pseudogenization may play an additional role for *MCS_AP1*; however, more species have to be included in further studies to answer the question what triggers the gene duplication.

So far, all gene copies show purifying selection and no evidence for neofunctionalization. Single and independent duplication of each gene in the respective species is also rather unlikely since there was only a very weak trend that duplicated genes were connected with function. For example *A. nobile*, the species with the strongest CAM activity (Lösch 1990), had two *MCS_PEPC* copies that showed allelic differences. For *MCS_AP1* such evidence could be found for *A. saundersii* which has a accelerated number of flower organs and exhibits duplication only for this gene. Nevertheless, this gene duplication is not subclade specific and until now non-, neo-, or subfunctionalization could not be ruled out. Further studies have to be done to verify the results.

If genes were duplicated because of polyploidization one may assume that phylogenetic relationships within orthologs and between paralogs would be the same. Unfortunately, amplification of both copies was not successful to the same degree for both copies. Independent loss of one gene copy and/or different or preferential amplification may obscure patterns and make interpretation difficult (see

also, e.g., Baum et al. 2005 or Fortune et al. 2007). If missing amplification was the problem, gene copy specific primers or more extensive amplification, screening, and sequencing of clones may help. Mort and Crawford (2004) suggest to screen at least five clones to detect the potential gene copies of low-copy nuclear genes in a species. In addition, copy specific restriction enzyme digests might help to detect if different copies were amplified but by chance not cloned and sequenced. Nevertheless, Kramer et al. (1998) stated that if orthologs of both duplicated products could be found in more than one species, the duplication event must have occurred before the last common ancestor of the species in question splits. In the present study this would suggest duplications at the base of the MCS species, before the three genera separated. The basal duplication implies a possible connection with polyploidization and island colonization.

Still, random duplication may be an explanation for the observed duplicates. To prove if gene duplications in the present study are the consequence of single independent events or of polyploidization, more nuclear coding genes should be studied. Both, *PEPC* and the MADS-box genes are encoded by multigene families and duplicated within the angiosperms (Gehrig et al. 1995, Lawton-Rauh 2003). Other floral regulatory genes appear to be more conservative in copy number. If whole genome duplication was the driving force for gene duplication in the MCS, paralogs of those single-copy nuclear genes should be found in a corresponding study. For example, Aagaard et al. (2005) observed a gene duplication for *FLORICAULA/LEAFY* within the Lamiales even though this MADS-box gene was thought single-copy in all diploid angiosperms, and showing duplication only in tetraploid species. They discussed that their resolved pattern was consistent with an ancient duplication. Classification of the analyzed copies into gene clades and the same relative position of duplicated genes in the Lamiales phylogeny support a whole genome duplication theory.

For ***MCS_PEPC*** single **duplications** could be found for *Sedum* species. A duplication within the main clade was observed for the tetraploid *S. surculosum* that may explain this observed duplication. In addition, a clear separation between four sequences of *S. pubescens* and *S. surculosum* and all other *Sedum* and MCS sequences was observed in the genealogy and supported by a Blast analysis. Furthermore, the NJ analysis (fig. 45; appendix) reveals for nearly all sequences a close relationship to the mRNA sequences *PEPC Kb1* (X87818) or *Kb2* (X87819) of

K. blossfeldiana (Gehrig et al. 1995). In contrast, the four unique *Sedum* clones are closely related to *PEPC Kb3* (X87820) or *Kb4* (X87821). Gehrig et al. (1995) defined these isogenes. *Kb1*, *Kb2*, *Kb3*, and *Kb4* form two gene pairs, *Kb1*/*Kb2* and *Kb3*/*Kb4*, with 95-98% sequence homology within the pairs and only 75% between them. The orthologous sequence pair of *Kb1*/*Kb2* is attributed to the CAM state of *Mesembryanthemum crystallinum* whereas *Kb3* and *Kb4* are more closely related to C_3 specific PEPC isoforms. Thus, the unique pattern in the present study revealed a separation between C_3 and CAM specific *PEPC* gene copies. Most analyzed species possess the expected CAM specific gene copy whereas copies specific for the C_3 state were observed for two *Sedum* species. C_3 fixation prevails in *Sedum* and it is also the ancestral state for the evolution of the CAM biosynthetic pathway in the MCS species (Pilon-Smits et al. 1992, Mort et al. 2007).

As mentioned, two main and two additional copies were detected for ***MCS_AP1***. One copy was specific for *Aeonium* whereas the other also comprises the studied *Monanthes* species.

The most complex pattern could be detected for *A. aureum*. Four different gene copies were observed even though several were subsequently excluded because of premature stops. However, proteins that arise from sequences with premature stops may have an influence in evolution and should be considered. *Aeonium aureum* is the species with the most derived phenotype concerning the impact of *AP1* homologs. It has the highest number of sepals and petals representing partly one of the main functions encoded by *AP1*. Also *A. saundersii*, exhibiting 12-16merous flowers, shows a gene duplication. But this duplication was not subclade specific and there is no knowledge regarding the function of this additional copy. Within the MCS subclade gene duplications were observed for *A. canariense*, *A. cuneatum*, and *A. nobile* that have 7-10merous flowers. *Aeonium rubrolineatum*, with 9-11 flower organs, exhibited no duplicated gene copies and contrast the pattern. Other species like *A. goochiae*, *M. anagensis*, and *M. icterica* without duplicates have between 6 and 8 flower organs. If gene duplications within the main clade are indeed the result of polyploidization, independent gene loss in species such as *A. goochiae*, *A. rubrolineatum* or the *Monanthes* species with a low number of flower organs could be assumed; species with derived numbers of flower organs, like *A. aureum* and *A. saundersii*, simply retained duplicated copies. In general, Lynch and Conery (2000)

discussed that 30-50% of duplicated genes were preserved over periods of 10-100 million years following polyploidization. This number is exceptionally high and indeed the majority of duplicated genes, even without polyploidization, will be silenced rather than preserved and therefore are lost. The percentage of retained or lost duplicates varies between species depending on functions (Moore and Purugganan 2005).

For ***MCS_AP3*** MCS species showed subclade specific duplications compared to *Sedum* but a unique pattern could be found for *S. pubescens*. Two subgroups of sequences were detected, one positioned unresolved at the base of the phylogram and the other basal to all remaining analyzed sequences (e.g., fig. 15). It has to be assumed that this gene duplication arose within the species resulting in paralogous gene copies. This assumption is supported by the number of observed nucleotide differences (compare, e.g., studies of Baum et al. 2005, Fortune et al. 2007, Shan et al. 2007, Zhang et al. 2008) and the use of varying outgroups to root the obtained phylograms. Both copies showed significantly different Ka/Ks-values (data not shown). However the duplication does not correlate with the number of flower organs, and thus partly function of *AP3* homologs, since *S. pubescens* has 5-6merous flowers in agreement with the other *Sedum* species. Nevertheless, one has to keep in mind that two different primer combinations were used to amplify homologs of *MCS_AP3* in *Sedum*. For *S. pubescens*, sequences amplified with one or the other primer combination were clearly separated. However, this pattern was not found for the other *Sedum* species; their species-specific sequences cluster together. Nucleotide differences between these sequences were not primer specific and in the range of that for orthologous gene copies. Only *S. surculosum* showed a unique pattern and it was also the only studied *Sedum* species that had duplications in all three analyzed gene regions. Since *S. surculosum* is tetraploid additional gene copies were expected.

Homologs of *AP3* control petal and sepal organ identity and maybe size (Juenger et al. 2000, Lohmann and Weigel 2002). The number of flower organs for species with duplicated gene copies range between 7 and 32. Comparable to *MCS_AP1*, for species with the lowest number of flower organs no duplications were observed like for *A. goochiae* (7-8 petals), *A. nobile* (7-9), *M. anagensis* (6-8), and *M. icterica* (6-7). Again this pattern is not consistent because, e.g., *A. saundersii* has no duplicated gene copies but an accelerated number of flower organs (12-16). The high number of

analyzed clones, ten, makes it unlikely that additional copies were simply not detected. Nevertheless, the possibility of preferential amplification of a specific gene copy over other copies could not be ruled out. There were no specific nucleotide differences in the primer binding site between paralogs but primer affinity could also be related to differences in primary or secondary structure of the DNA at the potential target sites (Tu et al. 2008). This potential problem could be solved by designing gene copy specific primers for amplification, where one primer is designed to the conserved region and one to the variable region.

The subclade specific duplication of the MCS species with one main copy comprises all analyzed MCS species and a gene duplication for several *Aeonium* species (subclade A; e.g., fig. 15) was strongly support by Blast analyses. While the main copy showed the highest similarity to *AP3*-like sequence of *Kalanchoe*, the second copy showed the highest similarity to a *DEF*-like gene copy of *P. groenlandica*. Both copies are under strong purifying selection, indicating that they seem to be functional and no hint of neofunctionalization or pseudogenization could be detected. Kramer et al. (1998) showed that *DEF*, as *AP3* ortholog in *A. majus*, was able to largely replace the endogenous *AP3* function in *A. thaliana*. Thus it can be assumed that both gene copies fulfill their task as B-function MADS-box genes. The observed pattern suggests that divergence and subfunctionalization increase the chance of retention. Even if these phenomena could not be unequivocally proved with the present data, retention of duplicates is often connected with functional divergence. Different expression and/or subfunctionalization may be assumed for the *MCS_AP3* gene copies. Given that significant differences in the calculated mean Ka/Ks-values and in the evolutionary rates were observed (discussed below) the hypothesis of divergent functionalization is supported. Also the retention of one copy in all analyzed species and the loss of the other copy in a subgroup of the studied species suggest that different evolutionary fates may have acted on the gene copies (see Zhang et al. 2008).

Functional diversification of duplicated genes is an important feature in the long-term evolution of polyploids. Blanc and Wolfe (2004) showed that 57% of newly duplicated and 73% of old duplicated gene pairs of *A. thaliana* had divergent expression. Likewise Moore and Purugganan (2005) discussed that 57% of the duplicated genes had divergent expression patterns whereas only 20% had asymmetric rates of protein

evolution. Thus, a large majority of polyploid-derived duplicates in *Arabidopsis*, which remain duplicated, acquired divergent functions and became specialized.

One of the outcomes of the classic model for gene duplication is gene loss associated with **pseudogenization** (Ohno 1970). Since amplification in the present study was based on genomic DNA, it is not possible to distinguish between functional and nonfunctional gene copies but hints for pseudogenization were detected in some cases. Frameshift mutations and subsequent premature stops in the coding region may be signs of pseudogenization and were revealed in the present study for *MCS_PEPC* and *MCS_AP3* with weak patterns and a strong pattern for *MCS_AP1*. Several *MCS_AP1* sequences showed specific features and studying them could help to distinguish between potential outcomes of the gene duplication even if these sequences were finally excluded. Therefore, sequences of *A. aureum* with premature stops and two unique sequences, A_goochiae_23637 and A_smithii_23637, were included in a phylogenetic analysis that revealed the following pattern (fig. 44; appendix). Within the main clade, *A. aureum* sequences with premature stops clustered with two sequences of *A. aureum* without stops. The additional sequences A_aureum_13637 and 1b3637 clustered together with the unique sequence of A_goochiae_23637 that had an indel of 1 bp in the exon position 528 resulting in a frameshift mutation and two premature stop codons. In addition, A_goochiae_23637 had a deletion between base pair 96 and 265 resulting in 56 missing amino acids within the K-domain. Since the K-domain is quite conserved and is involved in protein-protein interactions it is unlikely that the hypothetical protein of this particular sequence would be functional. This assumption is in line with Durbin et al. (2003). They observed early stops for cDNA sequences of *Ipomoea* and for one transcript a 33 bp deletion that made it highly unlikely that the hypothetical protein was functional and pointed toward pseudogenization.
The other two *A. aureum* copies with stops clustered with A_saundersii_33637 and were not substituted by copies without premature stop codons. The unique sequence A_smithii_23637 also clustered in this above mentioned species group. Aeonium_smithii_23637 had an indel at position 635 and the resulting frameshift mutation yields in a premature stop. In addition, A_smithii_23637 had a unique intron with a divergent recognition site for the excision of the intron of CA/AG instead of GT/AG. Beside pseudogenization, this could also indicate alternative splicing since

premature stop codons were also observed for cDNA sequences of *MCS_AP1*. Pseudogenization or alternative splicing was also assumed for alcohol dehydrogenase (*Adh*) genes in cotton. For all sequences of the A-subgenome of *Gossypium barbadense* 67 bp were deleted from exon 3 and intron 4. Additionally, the first nucleotide of intron 6 in the D-subgenome of *G. hirsutum* was found to be polymorphic; 12 alleles having a guanine (G) and 32 alleles an adenine (A). Small and Wendel (2002) discussed that this could be an alternate splice recognition site that may also change the expression. Since diverging expression could be associated with subfunctionalization the phenomenon may also be assumed for the additional copies of *A. aureum* and *A. saundersii*, if they should be functional. This is especially interesting since A_saundersii_33637 showed a sister relationship to the *MCS_AP1* gene copy of the *Sedum* ser. *Monanthoidea* and likewise *A. aureum* to *S. caeruleum*. As already discussed above, both *Sedum* specific copies showed a derived relationship to each other which may represent different gene copies and functions. Both *Sedum* specific copies had Ka/Ks-values which indicate purifying selection, and no hints for pseudogenization or loss of activity could be detected. This relationship to potential functional gene copies weakens the assumption that the additional copies of *A. aureum* and *A. saundersii* are in the progress to become pseudogenes. Alternative splicing has been reported in several MADS-box genes and in most cases was believed to have negative consequences. On the other hand, occasionally it may have enhanced the function of a gene because novel transcripts may had the potential to function as different proteins (Shan et al. 2007).

Nevertheless Álvarez et al. (2005), who detected an unusual 3´ intron splice site for the *Adh*C sequences of several studied cotton accessions, discussed that this feature in addition to premature stops may indicate pseudogenization in some species. They discussed also an unusually high number of replacements for some *Ces*A1b sequences that also suggest pseudogenization. Baum et al. (2005) described a frameshift mutation alternating 16 amino acids of the exon 2 for amplified *LEAFY* sequences of some Brassicaceae species. If the same splice sites as in *Arabidopsis* are utilized, this frameshift mutation would result in a premature stop codon at the start of exon 3. However, the calculated pairwise Ka/Ks-value was 0.21 in contrast to 1 which would be predicted if neutral selection was acting (see, e.g., Zhang et al. 2008). This implies that purifying selection has continued to act after gene duplication and suggests that both loci are functional. In the present study another observation

was made. Calculated mean Ka/Ks-values of the above mentioned additional *A. aureum* and *A. saundersii* sequence groups vary between 0.819 ± 0.059 and 1.143 ± 0.436 (0.885 excluding the two *A. aureum* sequences; fig. 44; appendix), respectively. These values are significantly higher than those calculated for the main copies (discussed below) and values close to 1 indicate neutral selection associated with pseudogenization. Positive selection acting on a gene, and thus probably the development of a new function, could not be detected since calculated Ka/Ks-values were not significantly higher than 1. Even if neofunctionalization could not be ruled out, this result would rather support nonfunctionalization and formation of pseudogenes in the additional copies of *A. aureum* and *A. saundersii*. However, that contradicts the above discussed phenomenon of alternative splicing and the potential outcome of subfunctionalization. A reliable statement will only be possible after further analyses, e.g., after studying expression patterns. If diverging or maybe even tissue specific expression should be observed, neo- or subfunctionalization would have to taken into stronger account. Janssens et al. (2008) observed in *DEF*-like sequences of *Impatiens* a nucleotide deletion that caused a frameshift and resulted in a premature stop. However, this truncation could be detected for almost all *Impatiens* species suggesting that the derived protein may have a specific function throughout the genus. In general, frameshift mutations which mostly result in premature stop codons and truncated proteins were often considered as detrimental for protein function and therefore of little evolutionary significance. However, more recently, combinations of gene duplication and frameshift mutations were assumed to be an evolutionary important mechanism for the emergence of new biological functions and maybe important mechanisms driving speciation processes (Vandenbussche et al. 2003, Kramer et al. 2006, Janssens et al. 2008).

4.3. Selection pressure

Various evolutionary patterns have been found for different genes in plants. Selective diversification, balancing selection, positive selection, relaxation of selective constraints, and purifying selection could be estimated (e.g., Bishop et al. 2000, Barrier et al. 2001, Olsen et al. 2002, Chen et al. 2004, Purugganan and Robichaux 2005). In the present study all analyzed genes and copies were under strong purifying selection. Ka/Ks-values range between 0.106 and 0.110 for the structural

gene (*MCS_PEPC*) and between 0.195 and 0.412 for the regulatory genes (*MCS_AP1* and *MCS_AP3*).

In general, Purugganan et al. (1995) detected that plant MADS-box genes evolved more rapidly than typical eukaryotic loci but changed slower than other plant regulatory genes. Olsen et al. (2002) found diverging nucleotide diversity and contrasting selection pressure studying floral regulatory MADS-box genes of *A. thaliana* in detail. For example, coding regions of *TERMINAL FLOWER 1* (*TFL1*) and *LEAFY* displayed significant reduction in nucleotide variation suggesting a recent adaptive sweep. In contrast, coding regions of *AP3*, *PI*, *AP1*, and *CAL* showed similar levels of nucleotide diversity and no evidence of either positive or balancing selection, possibly because they control evolutionary conserved floral organ traits (Olsen et al. 2002). However, Kramer et al. (1998) discussed that genes encoding for A- and C-function such as *AP1* or *CAL* showed high levels of constraint due to their pleiotropic roles in floral meristem development and organ identity. B-class MADS-box genes such as *AP3* or *PI* evolved 20-40% faster than all other plant MADS-box genes since they are only involved in organ identity, with enormous morphological plasticity of petals and stamens (for reference see Kramer et al. 1998).

This general observation cannot be confirmed in the present study. The highest Ka/Ks-values were calculated for *MCS_AP1* with Ka/Ks = 0.412 (copy A) and 0.373 (copy B). Both copies of *MCS_AP3* had lower Ka/Ks-values. Also for the codon sites that indicate positive selection, *MCS_AP1* showed with 37 the highest number. The results of the present study are also in line with the one of Barrier et al. (2001) who detected the highest Ka/Ks-values for *ASAP1*, followed by *ASAP3* for species of the HSA. Barrier et al. (2001) as well as Lawton-Rauh et al. (2003; corresponding population study) discussed that even if no positive selection could be detected, both regulatory genes seemed to have a huge potential to influence the speciation process. The greater number of nonsynonymous over synonymous sites in most loci, and the potential therein to change the hypothetical protein in a drastic way, is consistent with the possibility of positive selection. However, homologs of *AP1* and *AP3* for HSA species showed clearly accelerated Ka/Ks-values (0.98 and 0.79, respectively) compared to the values inferred in the present study.

Still, the calculated mean Ka/Ks-values of *MCS_AP1* and *MCS_AP3* were higher than other estimated Ka/Ks-values for floral regulatory genes. Purugganan et al. (1995) estimated for such genes of *A. thaliana* similar levels of sequence constraints

and Ka/Ks-values ranging from 0.119 to 0.185. For the MCS species, only copy A of *MCS_AP3* showed a Ka/Ks-value (0.195) close to the above mentioned ones. This value is also similar to the one estimated by Hernández-Hernández et al. (2007) for all B-class MADS-box genes over a broad taxonomic range (Ka/Ks = 0.151). All other Ka/Ks-values of the present study were significantly higher, ranging between 0.340 and 0.412. This suggest a partial influence in the speciation process since these values are also much higher than mean Ka/Ks-values. Tiffin and Hahn (2002) estimated a mean Ka/Ks-value of 0.14 for 218 randomly analyzed sequences of Brassicaceae species. Lawton-Rauh et al. (1999) calculated a Ka/Ks-value of 0.18 for four genes compared between *A. thaliana* and *A. lyrata*. Barrier et al. (2003) calculated a mean Ka/Ks-value of 0.213 for *Arabidopsis* species. The values for the MCS species are also much higher than Ka/Ks-values estimated in other studies with comparable experimental designs to that of the present study. Baum et al. (2005) estimated for the transcription factor *LEAFY* in Brassicaceae a Ka/Ks-value of approximately 0.1. However different Ka/Ks-values were estimated for rosette-flowering lineages compared to inflorescence flowering lineages suggesting an impact on speciation. Guillet-Claude et al. (2004) estimated a Ka/Ks-value of 0.143 for conifer *knox*-I paralogs but further analyses suggested an influence of these genes in the evolution of conifers. Within the genus *Impatiens* Janssens et al. (2007) found a low Ka/Ks-value of 0.1 for *DEF*/*AP3*-like genes. This value is lower than the Ka/Ks-values for both copies of *MCS_AP3* in the present study indicating that homologs of *AP3* may be involved in the evolution of the MCS species. That homologs of *AP1* may also play a role in the speciation of MCS is suggested by higher Ka/Ks-values calculated for both *MCS_AP1* copies. The values were similar to the one estimated for the *Myc*-like anthocyanin regulatory gene in *Cornus*. Fan et al. (2004) estimated a mean value of 0.407 that exceeded the values reported for most structural and partly for regulatory genes.

Nevertheless, the present values are still below the levels seen for some other regulatory genes. For example, variable domains of the plant R-genes show Ka/Ks-values close to 1 (see Purugganan et al. 1995, Bishop et al. 2000). Even if the genes in the present study showed purifying selection, other studies concluded the influence of several genes in the speciation process based on the detection of positive selection. Positive selection, as fixation of advantageous mutations, has been an exciting topic to evolutionary biologist since adaptive changes in genes and

genomes are responsible for evolutionary innovations and species differences (Yang 2005). Nevertheless, beneficial mutations and positive selection are expected to be extremely rare and patterns of sequence variation generally support the view that many differences between and within species are nonadaptive (Ford 2002). In the present study, sites under positive selection could be detected using codon-based model tests but estimation over whole proteins suggested strong purifying selection acting on all analyzed genes and copies. So far, positive selection was detected for genes involved in sexual reproduction, for host-parasite interaction genes, for genes which encode for enzymes involved in energy metabolism, and for regulatory genes involved in plant morphology (Bishop et al. 2000, Ford 2002, Barrier et al. 2003, Zhang 2003, Purugganan and Robichaux 2005). Barrier et al. (2003) found for 304 orthologous loci compared between *A. thaliana* and *A. lyrata* 14 (5%) with an estimated Ka/Ks > 1, indicating positive selection and likely adaptive divergence between both species. 5% of genes under positive selection is an exceptionally high estimation, likely due to the fact that closely related species were studied. In contrast, Endo et al. (1996) found for 3595 analyzed genes only 17 (0.5%) with evidence for positive selection. However, their criterion, the estimation of Ka/Ks-values, averaged the selection pressure over entire genes and is not powerful for detection of positive selection. Also Tiffin and Hahn (2002) found no genes that indicated positive selection, maybe because only a small fraction of the approx. 26,000 genes of *A. thaliana* were analyzed, and the method of calculating Ka/Ks-values used had been insensitive. Positive selection acts in most cases only on a small region of a gene, may occur in an episodic fashion, and only in a narrow window of evolutionary time (Ford 2002, Zhang 2003). Thus, the signal of positive selection may be overwhelmed by purifying selection (Guillet-Claude et al. 2004, Zhang 2003). For example, the 14 genes under positive selection detected by Barrier et al. (2003) could not be confirmed in the corresponding population study since all Ka/Ks-values were lower than 1. An explanation may be heterogeneities in selective constraints across loci and the use of longer sequences.

Also several other authors discuss that substitution rates between protein coding regions are not equally distributed (e.g., Vergara-Silva et al. 2000, Ford 2002, Fan et al. 2004, Yang 2005). Most proteins have highly conserved regions where replacement mutations are not tolerated. For example, active sites are usually the subject of intense evolutionary constraint and thus are highly conserved to preserve

function (Bishop et al. 2000, Hughes et al. 2000, Koch et al. 2001). In the present study only three of the 37 amino acids indicating positive selection for *MCS_AP1* were found within the MADS-box but 18 amino acids indicated positive selection in the more variable C-terminal domain. In general, the MADS-box domain, encoding the putative DNA-binding region, has the lowest nonsynonymous substitution rate whereas the K-domain and C-terminal domain display greater amounts of protein sequence variation (Purugganan et al. 1995). Hernández-Hernández et al. (2007) observed positive selection mostly within the K-domain of some B-class MADS-box genes. These observations were confirmed in the present study where 14.7% (K-domain) and 22.5% (C-domain) of positive selected amino acids were found within these two domains respectively but only 11% within the partly amplified MADS-box domain.

Besides **differential selection pressure** acting on specific genes, duplicated gene copies may provide the raw material for evolution and species could evolve by **divergence of paralogs** (Ohno 1970, Zhang et al. 2002, Zhang 2003, Purugganan and Robichaux 2005). Many studies showed that selection pressure acting on orthologous and paralogous gene copies could be different. Gene duplications may result in redundancy of a gene function and allow the duplicated copies to evolve in different ways (Ohno 1970, Lynch and Conery 2000, Zhang 2003, Moore and Purugganan 2005). Conant and Wagner (2003) discussed that at least 20% of gene duplicates diverged asymmetrically in an average genome. In contrast, there is expanding literature demonstrating that gene duplications frequently showed evidence of purifying selection, that both copies evolved at similar rates, and are maintained long-term (see Bomblies and Doebley 2005). In the present study, results for *MCS_AP3* and partly for *MCS_PEPC* would support the first hypothesis whereas the results of *MCS_AP1* are in line with the second mentioned phenomenon. Both copies of *MCS_AP3* are under purifying selection but there are significant differences for selection pressure and evolutionary rates. *MCS_PEPC* paralogs showed significant differences in their evolutionary rates but purifying selection acts on both gene copies and was not significantly different. No evidence for differential selection pressure and evolutionary rates at all were estimated for *MCS_AP1* sequences.
No rate differences were observed in the study of Barrier et al. (2001) either, where paralogs of *ASAP1*, *ASAP3*, and *ASCAB9* evolved with similar rates. However,

Lawton-Rauh et al. (2003) detected different levels and patterns of variation for duplicated *ASAP3* genes, whereas in contrast similar levels of nucleotide diversity were detected for the duplicated *AP1* homologs in HSA species. Similar observations for the present *MCS_AP1* and *MCS_AP3* sequences may suggest an influence in the speciation of MCS as discussed by Barrier et al. (2001) and Lawton-Rauh et al. (2003) for their species.

Baum et al. (2005) discussed that purifying selection acts on the duplicated gene loci of the flower regulatory gene *LEAFY* in the Brassicaceae. However, duplicates showed a significant tendency to have elevated Ka/Ks-ratios (Baum et al. 2005). The paralogous loci *AtHVA22d* and *AtHVA22e* of *A. thaliana* studied by Chen et al. (2004) evolved under purifying selection. Nevertheless, replacement changes in the *AtHVA22d* locus were accelerated indicating relaxation of purifying selection after gene duplication in one copy. Partially non-overlapping modes of expression between the two functional paralogs suggest that subfunctionalization explain the maintenance of the duplicated loci (Chen et al. 2004). Similar patterns may be assumed for the homologs of *MCS_AP3* and should therefore be studied in detail in further analyses. Different substitution rates were found for duplicated genes in *Gossypium*. Small and Wendel (2002) detected varying evolutionary pressures acting on the two subgenomes (A and D). Nucleotide diversity was consistently higher for genes of the D-subgenome that evolved at a significantly faster rate than the A-subgenome sequences. The genetic redundancy caused by polyploidy or large gene families may have allowed relaxed selection in the D-subgenome whereas purifying selection was maintained in the A-subgenome sequences (Small and Wendel 2002). Zhang et al. (2008) detected two copies of a *PISTILLATA*-like gene in *Cornus*. Purifying selection dominated the evolution of this gene and the estimated mean Ka/Ks-values, 0.51 and 0.22, differed significantly and were larger than the one inferred for the outgroup taxa. The increased Ka/Ks-ratio was the result of accelerated nonsynonymous substitutions. Also Aagaard et al. (2005) showed that paralogs of *FLORICAULA* and *DEF* in Lamiales appeared to avoid silencing as typical fate of most gene duplicates. They found no evidence for adaptive divergence acting on duplicated copies but relaxed purifying selection followed the duplication in one or both copies. Purifying selection, despite known functional redundancy, was also detected for the genes *CYCLOIDEA* (*CYC*) and *DICHOTOMA* (*DICH*) in *A. majus* (Hileman and Baum 2003). *CYC* and *DICH* arise via a single duplication and

purifying selection acting on both loci is consistent with subfunctionalization. Both paralogs showed some differences in their expression pattern, a well-known mechanism for subfunctionalization and long-term maintenance of duplicated genes. In the present study, this phenomenon could be assumed for the redundant *AP3*- and *DEF*-like genes (compare, e.g., fig. 15). Huttley et al. (1997) found heterogeneity in the rate of substitution between members of the *CHS* gene family. The higher nonsynonymous substitution rate for one gene copy may indicate that it evolved a new function. Also Durbin et al. (2000) pointed out that new *CHS* genes were recruited in flowering plants and the nucleotide substitution rate was frequently accelerated for them. Different evolutionary rates were detected for the five *CHS* genes in *Ipomoea*. *CHS-A*, *B*, and *C* genes evolved about 2.7 times faster than *CHS-D* and *E* genes with a higher rate of replacement mutations in *CHS-A*, *B*, and *C* (Durbin et al. 2000). Wang et al. (2007) studied selective modes among *CHS* orthologs and paralogs in *A. thaliana* and *A. halleri* subsp. *gemmifera*. Purifying selection is the main source influencing the evolution of the *CHS* loci but diverging ratios of Ka/Ks-values imply different levels of functional constraints.

Conflicting to the above mentioned studies, and in line with the results observed for *MCS_AP1*, is the study of Fortune et al. (2007). The three divergent homologous *waxy* genes in the hexaploid *Spartina* species seem to have evolved under selective constraints and relative rate tests indicated no significant rate heterogeneity between the sequences. Zhang et al. (2002) found no evidence for positive selection and only little evidence that paralogs evolved at different rates among duplicated gene pairs of *A. thaliana*. Yang et al. (2002) studied duplicated *CHS* genes of the Asteraceae and in contrast to other studies found no significant rate differences comparing different copies. Altogether, the huge differences in selection pressure, in the evolutionary rates and the diverse patterns reflect the complexity of evolution for duplicated genes (Chen et al. 2004).

4.4. Regulatory versus structural genes

In adaptive radiations extensive studies were done to define major forces acting in the speciation process. Accelerated evolution of regulatory genes compared to structural genes has been proposed as main explanation for the decoupled rates of molecular and morphological evolution (King and Wilson 1975, Remington and Purugganan 2002, Durbin et al. 2003, Purugganan and Robichaux 2005). This rise

the question after major and candidate genes or gene classes responsible for rapid morphological evolution (e.g., Remington and Purugganan 2002). The present study contributes to this research work and the two regulatory genes studied showed significantly higher Ka/Ks-values when compared to the structural gene (table 13 and 14) even though all studied gene copies evolved under strong purifying selection.
The species of the MCS are extremely different concerning growth-form and flower morphology. This is also true for the species of the HSA, the classic example of an adaptive radiation in plants. Several genes such as the growth regulatory gene *GAI*, R-genes, and floral regulatory genes *ASAP1* and *ASAP3* were analyzed within the HSA to estimate their impact on the speciation process (Barrier et al. 2001, Remington and Purugganan 2002, Lawton-Rauh et al. 2003, Purugganan and Robichaux 2005). Selection pressures acting on these genes, gene copies, coding, and promoter regions, and therefore evolutionary consequences, were extremely different (Purugganan and Robichaux 2005). Barrier et al. (2001) found gene duplication and accelerated mutation rates in the HSA homologs of *AP1* and *AP3* compared to the North American Tarweed species. The estimated Ka/Ks-values for these homologs in the MCS species were significantly lower. To explain this observation it is important to take into account that the species of the MCS are not considered as an adaptive radiation (Mes 1995, Jorgensen and Frydenberg 1999, Mort et al. 2002). Strong selection by ecological factors is not the major force driving speciation in the MCS species. Instead, inter-island dispersal and subsequent speciation is the main impact confirmed, e.g., by the fact that the five major growth-forms arose only once and spread subsequently over different islands. Even if adaptation cannot be ruled out it likely plays only a minor role in the MCS (Mes and ‘t Hart 1996) whereas it is the driving force for speciation in the HSA (for reference see, e.g., Barrier et al. 1999).

Confirming the theory of King and Wilson (1975) both regulatory genes, *MCS_AP1* and *MCS_AP3*, evolved faster than the structural gene *MCS_PEPC*. Even if the Ka/Ks-values for the regulatory genes in the present study show no evidence for positive selection, they are higher than estimated mean Ka/Ks-values in other studies (see discussion above). This may indicate an impact of the studied regulatory genes in the speciation process even if maybe not as main driving forces. A pollination study, conducted on Tenerife, partly support this observation (Esfeld et al. 2009).

With help of a fluorescent dye powder (pollen analog) pollen transfer between sympatric *Aeonium* species with overlapping flowering times was estimated. Despite an overlapping spectrum of flower visitors, infraspecific pollen transfer clearly exceeded the interspecific pollen transfer. The degree of pollen transfer seems thereby be linked to flower morphology and biology. However, other characters than those encoded by *AP1* and *AP3* seem to influence reproductive isolation since highest pollen transfer rates were observed between species that had similarly colored flowers and the same reward of pollen and nectar. Therefore, it could be assumed that other genes such as genes encoding for flower color (*anthocyanin2*, e.g., Quattrocchio et al. 1999 or *CHS,* e.g., Durbin et al. 2003) and genes that have an impact on nectar or pollen may have had a higher influence in reproductive isolation and thus speciation of the MCS. In addition different flowering times leading to temporal isolation is discussed as the major reproductive barrier in *Aeonium*. Thus, flowering time genes should be studied to estimate their influence in the speciation process (Liu 1989, Jorgensen and Olesen 2001, Esfeld et al. 2009).

However, it is important to consider that significant differences in mean Ka/Ks-values were found between the MCS species and the respective sister- and outgroup *Sedum* species, suggesting an influence of the studied genes in the speciation process. For *MCS_PEPC* the species of *Sedum* ser. *Monanthoidea* had a significantly lower Ka/Ks-values than the MCS species. The same observation could be made for *MCS_AP3*, where both outgroup (*S. caeruleum* and *S. pubescens*) and sistergroup species (*S. jaccardianum*, *S. modestum*, and *S. surculosum*) showed a significant lower mean Ka/Ks-value compared to copy A and copy B of the MCS species. The difference between the mean Ka/Ks-values of the outgroup and sistergroup species was not significant (data not shown). In contrast the obtained pattern for *MCS_AP1* was different. Both copies of *MCS_AP1* for the MCS species showed no significant difference when compared to the mean Ka/Ks-values of the sister- and outgroup species. However, relationships between the amplified gene copies are more complicated (compare, e.g., fig. 12). The exon phylogram suggests a sister relationship between one additional *A. saundersii* sequence (33637) and the *Sedum* ser. *Monanthoidea* species. The outgroup species (*S. caeruleum*) could be found basal but the relationship is unresolved. Thus, it is possible that derived and different gene copies were amplified that do not reflect the true relationships and

should not be compared as sister- and outgroup relationships (see also discussion above).

In general not all regulatory genes are involved in morphological evolution or with the same strength in the speciation process (Olsen et al. 2002, Remington and Purugganan 2002). Despite high growth-form divergence selective constraints remain strong for duplicated *GAI* genes in the HSA. The constraint was somewhat relaxed in one of the two Hawaiian copies compared to the North American lineage but relative rate tests revealed no significant differences in the evolutionary rates of both gene copies. Remington and Purugganan (2002) failed to detect any evidence of positive selection and the calculated mean Ka/Ks-value was 0.37. Similar Ka/Ks-values were obtained in the present study for both copies of *MCS_AP1* and for copy B of *MCS_AP3*. This mild degree of relaxed selective constraint for the regulatory genes is in sharp contrast to the results of Barrier et al. (2001) who detected Ka/Ks-values of 0.98 (*ASAP1*) and 0.79 (*ASAP3*), respectively. However, the present Ka/Ks-values are slightly accelerated compared to that of the structural gene *ASCAB9* (Ka/Ks = 0.21; Barrier et al. 2001) pointing toward an impact of *MCS_AP1* and *MCS_AP3* in the speciation of MSC, although maybe a limited one.

Accelerated sequence evolution and positive selection is not a general phenomenon or only true for a small subset of the coding region of regulatory loci. For instance, Lukens and Doebley (2001) found no evidence for positive selection on transcription factors involved in growth regulation. Since many genes are involved in this process, overlaying effects may obscuring an existing pattern. This suggests that candidate genes or single speciation genes may not exist (see also Hodges 1997, Remington and Purugganan 2002, Dias et al. 2003).

In contrast, gene regulation due to changes in the coding sequence of regulatory loci or in cis acting promoter sequences can play a key role in morphological evolution (Baum 1998). Remington and Purugganan (2002) confirmed the importance of gene regulation in the speciation process. They found no evidence for positive selection acting on the coding region of the *GAI* gene in the HSA. In contrast, the ~900 bp upstream flanking region showed variable rates and patterns of evolution that may reflect positive selection. Further studies, e.g., the one of Doebley and Lukens (1998), have found higher rates of sequence polymorphism and divergence in functionally important promoter segments compared to the corresponding gene regions. Subsequently, changes in the level or expression pattern of genes were

observed that lead to novel phenotypes in maize. The influence of gene regulation was also extensively studied concerning the impact of flower color in the speciation process. Striking differences in flower color and morphology is associated with speciation, e.g., in *Aquilegia*, *Ipomoea*, and *Mimulus* (Hodges and Arnold 1994a, Schemske and Bradshaw 1999, Durbin et al. 2003). Plant pigments influencing flower color and hence pollinator attraction. To figure out their impact on the speciation process, many studies focused on the structural gene *CHS*, a starting point in the flavonoid synthesis (Clegg and Durbin 2000). However, most species differences found so far were associated with changes in regulation of gene expression. Durbin et al. (2003) found only one case where phenotypic differences in pigmentation patterns could be explained by mutation of the structural gene. There are only a limited number of ways that a mutation in the structural gene can lead to a change in flower color or this change may rather result in the total loss of pigmentation. Thus, changes in a structural gene will not have the same impact in terms of adaptive evolution. Changes in regulatory genes on the other hand often result in novel patterns of pigmentation that can influence pollinator behavior or even attract new pollinators highly influencing the speciation process (Durbin et al. 2003).

Thus, there are many more ways in which mutations in gene regulation can occur considering the complexity and many different types of genes involved in regulation (Purugganan and Robichaux 2005). Mutations in regulatory genes can provide a more rapid response to the environment and a larger adaptive success whereas in contrast, structural gene evolution may not be the key to rapid and extensive diversification in organismal form (Barrier et al. 2001, Durbin et al. 2003, Purugganan and Robichaux 2005).

5. Summary

Evolution and speciation belong to the most exciting topics in biology. The diversity of species evident in radiations provides a valuable tool to study these processes. Numerous studies have shown that the morphological variation of radiated species does not correlate with their molecular diversity and that regulatory genes evolve faster than structural genes.

However, the determination of candidate genes involved in speciation is difficult. Based on results of Barrier et al. (2001) the present study contributes to the ongoing debate about what triggers speciation and evolution. The two regulatory genes, homologs of the homeotic floral genes *APETALA1* and *APETALA3*, as well as the structural gene encoding for phosphoenolpyruvate carboxylase (*PEPC*) were studied in the radiation of the Macaronesian Crassulaceae Sempervivoideae (MCS).

Robust phylogenies are needed to determine the influence of traits in speciation, and their lack is a well known problem for radiations and fast species evolution. However, comparisons of species phylogenies and genealogies may enable the observation of major impact factors since low-copy genes can be involved in the determination of phenotypes and adaptation. The studied low-copy nuclear genes both support and contradict relationships observed for the MCS species based on morphological and molecular markers. Several relationships like the one between *A. aureum*, *A. saundersii*, and *A. smithii* or between the *Sedum* species are highly supported. On the other hand, conflicting relationships were observed for *A. canariense*, *A. cuneatum*, *A. nobile*, and *A. rubrolineatum*. In addition, the low-copy nuclear genes shed light on highly debated relationships like the one of *A. goochiae* at the base of all *Aeonium* species and *M. icterica* closely related to the *Aichryson* clade. *MCS_AP3* is the most promising phylogenetic marker and may also have been involved in the speciation as it showed significant differences in the estimated Ka/Ks-values and in the evolutionary rates between the amplified gene copies.

An important observation concerns the copy number of the studied genes. Subclade specific gene duplications were observed for the polyploid species of the MCS island radiation but not for the diploid sister- and outgroup species of *Sedum*. The obtained pattern strongly suggests that the gene duplication was connected with polyploidization and island colonization. Non-species specific gene loss or retention could be observed as well as potential random gain and pseudogenization.

Gene duplications may provide the raw material for evolution and in general, the influence of genes and gene copies in speciation could be defined by the selection pressure. Positive selection connected with the development of a new function would provide a strong evidence. However, purifying selection was observed for all studied genes and copies. Nevertheless, the estimated Ka/Ks-values, defining the selection pressure, for the regulatory genes were higher than mean Ka/Ks-values and also higher than values observed in other studies with a comparable study design. *MCS_AP1* showed the highest Ka/Ks-values followed by the second regulatory gene *MCS_AP3*. As predicted by the hypothesis of King and Wilson (1975), the structural gene *MCS_PEPC* had the lowest Ka/Ks-values.

However, overall a limited influence of the analyzed genes in the speciation process was suggested. Therefore, ongoing studies should focus on additional genes as well as on gene regulation (promoter sequences) and should include more species of the MCS for reliable conclusions. Following the results and observations of the present study, e.g., genes influencing growth-form, flowering time, flower color, position of inflorescence, number of floral organs, and genes influencing woodiness should be studied to gain insights into the speciation process of the Macaronesian Crassulaceae Sempervivoideae.

Literature

Aagaard, J.E., Olmstead, R.G., Willis, J.H. and Phillips, P.C. (2005): Duplication of floral regulatory genes in the Lamiales. American Journal of Botany: 92, pp. 1284-1293.

Álvarez, I., Cronn, R. and Wendel, J.F. (2005): Phylogeny of the New World diploid cottons (*Gossypium* L., Malvaceae) based on sequences of three low-copy nuclear genes. Plant Systematics and Evolution: 252, pp. 199-214.

Andrus, N., Trusty, J., Santos-Guerra, A., Jansen, R.K. and Francisco-Ortega, J. (2004): Using molecular phylogenies to test phytogeographical links between East/South Africa-Southern Arabia and the Macaronesian islands – a review, and the case of *Vierea* and *Pulicaria* section *Vieraeopsis* (Asteraceae). Taxon: 53, pp. 333-346.

Arnedo, M.A., Oromí, P. and Ribera, C. (1996): Radiation of the genus *Dysdera* (Araneae, Haplogynae, Dysderidae) in the Canary Islands: the western islands. Zoologica Scripta: 25, pp. 241-274.

Bailey, C.D. and Doyle, J.J. (1999): Potential phylogenetic utility of the low-copy nuclear gene *pistillata* in dicotyledonous plants: comparison to nrDNA ITS and *trnL* intron in *Sphaerocardamum* and other Brassicaceae. Molecular Phylogenetics and Evolution: 13, pp. 20-30.

Baker, H.G. and Baker, I. (1983): A brief historical review of the chemistry of floral nectar. In: Bentley, B. and Elias, T. (eds): The biology of nectaries. Columbia University Press, New York.

Baldwin, B.G. and Robichaux, R.H. (1995): Historical biogeography and ecology of the Hawaiian silversword alliance: new molecular phylogenetic perspectives. In: Wagner, W.L. and Funk, V.A. (eds): Hawaiian biogeography: evolution on a hot spot archipelago. Smithsonian Institution Press, Washington, D.C.

Baldwin, B.G., Crawford, D.J., Francisco-Ortega, J., Kim, S.-C., Sang, T. and Stuessy, T.F. (1998): Molecular phylogenetic insights on the origin and evolution of oceanic island plants. In: Soltis, D.E., Soltis, P.S. and Doyle, J.J. (eds): Molecular Evolution and Plant Systematics.

Barber, J.C., Francisco-Ortega, J., Santos-Guerra, A., Marrero, A. and Jansen, R.K. (2000): Evolution of endemic *Sideritis* (Lamiaceae) in Macaronesia: insights from a chloroplast DNA restriction site analysis. Systematic Botany: 25, pp. 633-647.

Barrier, M., Baldwin, B.G., Robichaux, R.H. and Purugganan, M.D. (1999): Interspecific hybrid ancestry of a plant adaptive radiation: allopolyploidy of the Hawaiian silversword alliance (Asteraceae) inferred from floral homeotic gene duplications. Molecular Biology and Evolution: 16, pp. 1105-1113.

Barrier, M., Robichaux, R.H. and Purugganan, M.D. (2001): Accelerated regulatory gene evolution in an adaptive radiation. Proceedings of the National Academic Society of the USA: 98, pp. 10208-10213.

Barrier, M., Bustamante, C.D., Yu, J. and Purugganan, M.D. (2003): Selection on rapidly evolving proteins in the *Arabidopsis* genome. Genetics: 163, pp. 723-733.

Baum, D.A. (1998): The evolution of plant development. Current Opinion in Plant Biology: 1, pp. 79-86.

Baum, D.A., Yoon, H.-S. and Oldham, R.L. (2005): Molecular evolution of the transcription factor *LEAFY* in Brassicaceae. Molecular Phylogenetics and Evolution: 37, pp. 1-14.

Beck, J.B., Al-Shehbaz, I.A., O´Kane Jr., S.L. and Schaal, B.A. (2007): Further insights into the phylogeny of *Arabidopsis* (Brassicaceae) from nuclear *Atmyb2* flanking sequence. Molecular Phylogenetics and Evolution: 42, pp. 122-130.

Becker, A., Winter, K.-U., Meyer, B., Saedler, H. and Theissen, G. (2000): MADS-box gene diversity in seed plants 300 million years ago. Molecular Biology and Evolution: 17, pp. 1425-1434.

Berbel, A., Navarro, C., Ferrándiz, C., Cañas, L.A., Madueño, F. and Beltrán, J.P. (2001): Analysis of *PEAM4*, the pea *AP1* functional homologue, supports a model for *AP1*-like genes controlling both floral meristem and floral organ identity in different plant species. The Plant Journal: 25, pp. 441-451.

Berger, A. (1930): Crassulaceae. In: Engler, A. and Prantl, K. (eds): Die natürlichen Pflanzenfamilien. 2nd ed. 18a. W. Engelmann, Leipzig.

Bishop, J.G., Dean, A.M. and Mitchell-Olds, T. (2000): Rapid evolution in plant chitinases: Molecular targets of selection in plant-pathogen coevolution. Proceedings of the National Academic Society of the USA: 97, pp. 5322-5327.

Blanc, G. and Wolfe, K.H. (2004): Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. The Plant Cell: 16, pp. 1679-1691.

Blattner, F.R. (1999): Direct amplification of the entire ITS region from poorly preserved plant material using recombinant PCR. BioTechniques: 27, pp. 173-178.

Böhle, U.-R., Hilger, H.H. and Martin, W.F. (1996): Island colonization and evolution of the insular woody habit in *Echium* L. (Boraginaceae). Proceedings of the National Academic Society of the USA: 93, pp. 11740-11745.

Bomblies, K. and Doebley, J.F. (2005): Molecular evolution of *FLORICAULA/LEAFY* orthologs in the Andropogoneae (Poaceae). Molecular Biology and Evolution: 22, pp. 1082-1094.

Bramwell, D. (1976): The endemic flora of the Canary Islands. In: Kunkel, G. (ed): Biogeography and ecology in the Canary Islands. Junk Publishers, The Hague.

Carlquist, S. (1962): A theory of paedomorphosis in dicotyledonous woods. Phytomorphology: 12, pp. 30-45.

Carlquist, S. (1974): Island Biology. Columbia University Press, New York & London.

Carracedo, J.C. (1994): The Canary Islands: an example of structural control on the growth of large oceanic-island volcanoes. Journal of Volcanology and Geothermal Research: 60, pp. 225-241.

Chen, C.-N., Chiang, Y.-C., Ho, T.-H.D., Schaal, B.A. and Chiang, T.-Y. (2004): Coalescent processes and relaxation of selective constraints leading to contrasting genetic diversity at paralogs *AtHVA22d* and *AtHVA22e* in *Arabidopsis thaliana*. Molecular Phylogenetics and Evolution: 32, pp. 616-626.

Chollet, R. (1996): Phosphoenolpyruvate carboxylase: a ubiquitous, highly regulated enzyme in plants. Annual Review of Plant Physiology and Plant Molecular Biology: 47, pp. 273-298.

Clegg, M.T. and Durbin, M.L. (2000): Flower color variation: a model for the experimental study of evolution. Proceedings of the National Academic Society of the USA: 97, pp. 7016-7023.

Coen, E.S. and Meyerowitz, E.M. (1991): The war of the whorls: genetic interactions controlling flower development. Nature: 353, pp. 31-37.

Conant, G.C. and Wagner, A. (2003): Asymmetric sequence divergence of duplicate genes. Genome Research: 13, pp. 2052-2058.

Cushman, J.C. and Bohnert, H.J. (1999): Crassulacean acid metabolism: Molecular genetics. Annual Review of Plant Physiology and Plant Molecular Biology: 50, pp. 305-332.

Dias, A.P., Braun, E.L., McMullen, M.D. and Grotewold, E. (2003): Recently duplicated maize *R2R3 Myb* genes provide evidence for distinct mechanisms of evolutionary divergence after duplication. Plant Physiology: 131, pp. 610-620.

Doebley, J. and Lukens, L. (1998): Transcriptional regulators and the evolution of plant form. The Plant Cell: 10, pp. 1075-1082.

Doron-Faigenboim, A., Stern, A., Mayrose, I., Bacharach, E. and Pupko, T. (2005): Selecton: a server for detecting evolutionary forces at a single amino-acid site. Bioinformatics Applications Note: 21, pp. 2101-2103.

Duarte, J.M., Cui, L., Wall, P.K., Zhang, Q., Zhang, X., Leebens-Mack, J., Ma, H., Altman, N. and dePamphilis, C.W. (2006): Expression pattern shifts following duplication indicative of subfunctionalization and neofunctionalization in regulatory genes of *Arabidopsis*. Molecular Biology and Evolution: 23, pp. 469-478.

Durbin, M.L., McCaig, B. and Clegg, M.T. (2000): Molecular evolution of the chalcone synthase multigene family in the morning glory genome. Plant Molecular Biology: 42, pp. 79-92.

Durbin, M.L, Lundy, K.E., Morrell, P.L., Torres-Martinez, C.L. and Clegg, M.T. (2003): Genes that determine flower color: the role of regulatory changes in the evolution of phenotypic adaptations. Molecular Phylogenetics and Evolution: 29, pp. 507-518.

Emerson, B.C. (2002): Evolution on oceanic islands: molecular phylogenetic approaches to understanding pattern and process. Molecular Ecology: 11, pp. 951-966.

Endo, T., Ikeo, K. and Gojobori, T. (1996): Large-scale search for genes on which positive selection may operate. Molecular Biology and Evolution: 13, pp. 685-690.

Erbar, C. (2007): Current opinions in flower development and the evo-devo approach in plant phylogeny. Plant Systematics and Evolution: 269, pp. 107-132.

Esfeld, K., Koch, M.A., van der Niet, T., Seifan, M. and Thiv, M. (2009): Little interspecific pollen transfer despite overlap in pollinators between sympatric *Aeonium* (Crassulaceae) species pairs. Flora. doi:10.1016/j.flora.2008.10.002.

Eyre-Walker, A. (2006): The genomic rate of adaptive evolution. Trends in Ecology and Evolution: 21, pp. 569-575.

Fairfield, K.N., Mort, M.E. and Santos-Guerra, A. (2004): Phylogenetics and evolution of the Macaronesian members of the genus *Aichryson* (Crassulaceae) inferred from nuclear and chloroplast sequence data. Plant Systematics and Evolution: 248, pp. 71-83.

Fan, C., Purugganan, M.D., Thomas, D.T., Wiegmann, B.M. and Xiang, Q.-Y. (2004): Heterogeneous evolution of the *Myc-like* anthocyanin regulatory gene and its phylogenetic utility in *Cornus* L. (Cornaceae). Molecular Phylogenetics and Evolution: 33, pp. 580-594.

Fishbein, M., Hibsch-Jetter, C., Soltis, D.E. and Hufford, L. (2001): Phylogeny of Saxifragales (Angiosperms, Eudicots): analysis of a rapid, ancient radiation. Systematic Biology: 50, pp. 817-847.

Ford, M.J. (2002): Applications of selective neutrality tests to molecular ecology. Molecular Ecology: 11, pp. 1245-1262.

Fortune, P.M., Schierenbeck, K.A., Ainouche, A.K., Jacquemin, J., Wendel, J.F. and Ainouche, M.L. (2007): Evolutionary dynamics of *Waxy* and the origin of hexaploid *Spartina* species (Poaceae). Molecular Phylogenetics and Evolution: 43, pp. 1040-1055.

Francisco-Ortega, J., Jansen, R.K. and Santos-Guerra, A. (1996): Chloroplast DNA evidence of colonization, adaptive radiation, and hybridization in the evolution of the Macaronesian flora. Proceedings of the National Academic Society of the USA: 93, pp. 4085-4090.

Francisco-Ortega, J., Santos-Guerra, A., Kim, S.-C. and Crawford, D.J. (2000): Plant genetic diversity in the Canary Islands: a conservation perspective. American Journal of Botany: 87, pp. 909-919.

Francisco-Ortega, J., Barber, J.C., Santos-Guerra, A., Febles-Hernández, R. and Jansen, R.K. (2001): Origin and evolution of the endemic genera of Gonosperminae (Asteraceae: Anthemideae) from the Canary Islands: evidence from nucleotide sequences of the internal transcribed spacers of the nuclear ribosomal DNA. American Journal of Botany: 88, pp. 161-169.

Francisco-Ortega, J., Fuertes-Aguilar, J., Kim, S.-C., Santos-Guerra, A., Crawford, D.J. and Jansen, R.K. (2002): Phylogeny of the Macaronesian endemic *Crambe* section *Dendrocrambe* (Brassicaceae) based on internal transcribed spacer sequences of nuclear ribosomal DNA. American Journal of Botany: 89, pp. 1984-1990.

Gehrig, H., Taybi, T., Kluge, M. and Brulfert, J. (1995): Identification of multiple PEPC isogenes in leaves of the facultative Crassulacean Acid Metabolism (CAM) plant *Kalanchoe blossfeldiana* Poelln. cv. Tom Thumb. FEBS Letters: 377, pp. 399-402.

Grob, G.B.J., Gravendeel, B. and Eurlings, M.C.M. (2004): Potential phylogenetic utility of the nuclear *FLORICAULA/LEAFY* second intron: comparison with three chloroplast DNA regions in *Amorphophallus* (Araceae). Molecular Phylogenetics and Evolution: 30, pp. 13-23.

Guillet-Claude, C., Isabel, N., Pelgas, B. and Bousquet, J. (2004): The evolutionary implications of *knox-I* gene duplications in conifers: correlated evidence from phylogeny, gene mapping, and analysis of functional divergence. Molecular Biology and Evolution: 21, pp. 2232-2245.

Guindon, S. and Gascuel, O. (2003): A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic Biology: 52, pp. 696-704.

Guindon, S., Lethiec, F., Duroux, P. and Gascuel, O. (2005): PHYML Online – a web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Research: 33, pp. W557-W559.

Hernández-Hernández, T., Martínez-Castilla, L.P. and Alvarez-Buylla, E.R. (2007): Functional diversification of B MADS-box homeotic regulators of flower development: adaptive evolution in protein-protein interaction domains after major gene duplication events. Molecular Biology and Evolution: 24, pp. 465-481.

Hileman, L.C. and Baum, D.A. (2003): Why do paralogs persist? Molecular evolution of *CYCLOIDAE* and related floral symmetry genes in Antirrhineae (Veronicaceae). Molecular Biology and Evolution: 20, pp. 591-600.

Hoballah, M.E., Gübitz, T., Stuurman, J., Broger, L., Barone, M., Mandel, T., Dell'Olivo, A., Arnold, M. and Kuhlemeier, C. (2007): Single gene-mediated shift in pollinator attraction in *Petunia*. The Plant Cell: 19, pp. 779-790.

Hodges, S.A. and Arnold, M.L. (1994a): Columbines: a geographically widespread species flock. Proceedings of the National Academic Society of the USA: 91, pp. 5129-5132.

Hodges, S.A. and Arnold, M.L. (1994b): Floral and ecological isolation between *Aquilegia formosa* and *Aquilegia pubescens*. Proceedings of the National Academic Society of the USA: 91, pp. 2493-2496.

Hodges, S.A. (1997): Floral nectar spurs and diversification. International Journal of Plant Science: 158, pp. S81-S88.

Hohenester, A. and Welss, W. (1993): Exkursionsflora für die Kanarischen Inseln. Ulmer, Stuttgart.

Huber, T., Faulkner, G. and Hugenholtz, P. (2004): Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. Bioinformatics: 20, pp. 2317-2319.

Huelsenbeck, J.P. and Ronquist, F. (2001): MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics: 17, pp. 754-755.

Hughes, A.L., Green, J.A., Garbayo, J.M. and Roberts, R.M. (2000): Adaptive diversification within a large family of recently duplicated, placentally expressed genes. Proceedings of the National Academic Society of the USA: 97, pp. 3319-3323.

Hurst, L.D. (2002): The *Ka/Ks* ratio: diagnosing the form of sequence evolution. TRENDS in Genetics: 18, pp. 486-487.

Huttley, G.A., Durbin, M.L., Glover, D.E. and Clegg, M.T. (1997): Nucleotide polymorphism in the chalcone synthase-A locus and evolution of the chalcone synthase multigene family of common morning glory *Ipomoea purpurea*. Molecular Ecology: 6, pp. 549-558.

Irish, V.F. and Litt, A. (2005): Flower development and evolution: gene duplication, and diversification and redeployment. Current Opinion in Genetics and Development: 15, pp. 454-460.

Janssens, B.S., Geuten, K., Viaene, T., Yuan, Y.-M., Song, Y. and Smets, E. (2007): Phylogenetic utility of the *AP3/DEF* K-domain and its molecular evolution in *Impatiens* (Balsaminiacea). Molecular Phylogenetics and Evolution: 43, pp. 225-239.

Janssens, B.S., Viaene, T., Huysmans, S., Smets, E. and Geuten, K.P. (2008): Selection on length mutations after frameshift can explain the origin and retention of the *AP3/DEF*-like paralogues in *Impatiens*. Journal of Molecular Evolution: 66, pp. 424-435.

Jorgensen, T.H. and Frydenberg, J. (1999): Diversification in insular plants: inferring the phylogenetic relationship in *Aeonium* (Crassulaceae) using ITS sequences of nuclear ribosomal DNA. Nordic Journal of Botany: 19, pp. 613-621.

Jorgensen, T.H. and Olesen, J.M. (2001): Adaptive radiation of island plants: evidence from *Aeonium* (Crassulaceae) of the Canary Islands. Perspectives in Plant Ecology, Evolution and Systematics: 4, pp. 29-42.

Jorgensen, T.H. (2002): The importance of phylogeny and ecology in microgeographical variation in the morphology of four Canarian species of *Aeonium* (Crassulaceae). Biological Journal of the Linnean Society: 76, pp. 521-533.

Juan, C., Emerson, B.C., Oromí, P. and Hewitt, G.M. (2000): Colonization and diversification: towards a phylogeographic synthesis for the Canary Islands. TRENDS in Ecology and Evolution: 15, pp. 104-109.

Juenger, T., Purugganan, M.D. and Mackay, T.F.C. (2000): Quantitative trait loci for floral morphology in *Arabidopsis thaliana*. Genetics: 156, pp. 1379-1392.

Kim, S.-C., Crawford, D.J., Francisco-Ortega, J. and Santos-Guerra, A. (1996): A common origin for woody *Sonchus* and five related genera in the Macaronesian islands: molecular evidence for extensive radiation. Proceedings of the National Academic Society of the USA: 93, pp. 7743-7748.

King, M.C. and Wilson, A.C. (1975): Evolution at two levels in humans and chimpanzees. Science: 188, pp. 107-116.

Koch, M.A., Weisshaar, B., Kroymann, J., Haubold, B. and Mitchell-Olds, T. (2001): Comparative genomics and regulatory evolution: conservation and function of the *Chs* and *Apetala3* promoters. Molecular Biology and Evolution: 18, pp. 1882-1891.

Kramer, E.M., Dorit, R.L. and Irish, V.F. (1998): Molecular evolution of genes controlling petal and stamen development: duplication and divergence within the *APETALA3* and *PISTILLATA* MADS-box gene lineages. Genetics: 149, pp. 765-783.

Kramer, E.M. and Hall, J.C. (2005): Evolutionary dynamics of genes controlling floral development. Current Opinion in Plant Biology: 8, pp. 13-18.

Kramer, E.M., Su, H.-J., Wu, C.-C. and Hu, J.-H. (2006): A simplified explanation for the frameshift mutation that created a novel C-terminal motif in the *APETALA3* gene lineage. BMC Evolution and Biology: 6, pp. 30-47.

Kull, U. (1982): Artbildung durch geographische Isolation bei Pflanzen – die Gattung *Aeonium* auf Teneriffa. Natur und Museum: 112, pp. 33-40.

Lawton-Rauh, A., Buckler IV, E.S. and Purugganan, M.D. (1999): Patterns of molecular evolution among paralogous floral homeotic genes. Molecular Biology and Evolution: 16, pp. 1037-1045.

Lawton-Rauh, A., Alvarez-Buylla, E.R. and Purugganan, M.D. (2000): Molecular evolution of flower development. TRENDS in Ecology and Evolution: 15, pp. 144-149.

Lawton-Rauh, A. (2003): Evolutionary dynamics of duplicated genes in plants. Molecular Phylogenetics and Evolution: 29, pp. 396-409.

Lawton-Rauh, A., Robichaux, R.H. and Purugganan, M.D. (2003): Patterns of nucleotide variation in homoeologous regulatory genes in the allotetraploid Hawaiian silversword alliance (Asteraceae). Molecular Ecology: 12, pp. 1301-1313.

Lems, K. (1960): Botanical notes on the Canary Islands. II. The evolution of plant forms in the islands: *Aeonium*. Ecology: 41, pp. 1-17.

Litt, A. and Irish, V.F. (2003): Duplication and diversification in the *APETALA1*/*FRUITFUL* floral homeotic gene lineage: implications for the evolution of floral development. Genetics: 165, pp. 821-833.

Liu, H.-Y. (1989): Systematics of *Aeonium* (Crassulaceae). National Museum of Natural Science, Taichung.

Lohmann, J.U. and Weigel, D. (2002): Building Beauty: the genetic control of floral patterning. Developmental Cell: 2, pp. 135-142.

Lösch, R. (1990): Funktionelle Voraussetzungen der adaptiven Nischenbesetzung in der Evolution der makaronesischen Semperviven. Dissertationes Botanicae: 146. Cramer, Berlin.

Lukens, L. and Doebley, J. (2001): Molecular evolution of the *teosinte branched* gene among maize and related grasses. Molecular Biology and Evolution: 18, pp. 627-638.

Lynch, M. and Conery, J.S. (2000): The evolutionary fate and consequences of duplicate genes. Science: 290, pp. 1151-1155.

Lynch, M., O'Hely, M., Walsh, B. and Force, A. (2001): The probability of preservation of a newly arisen gene duplicate. Genetics: 159, pp. 1789-1804.

MacArthur, R.H. and Wilson, E.O. (1967): The theory of island biography. Princeton University Press, Princeton.

Maire, R. (1976): Flore de l'Afrique du Nord. Vol. 14. Lechevallier, Paris.

Malcomber, S.T. (2002): Phylogeny of *Gaertnera* Lam. (Rubiaceae) based on multiple DNA markers: evidence of a rapid radiation in a widespread, morphologically diverse genus. Evolution: 56, pp. 42-57.

Mandel, M.A. and Yanofsky, M.F. (1995): A gene triggering flower formation in *Arabidopsis*. Nature: 377, pp. 522-524.

Meimberg, H., Abele, T., Bräuchler, C., McKay, J.K., Pérez de Paz, P.L. and Heubl, G. (2006): Molecular evidence for adaptive radiation of *Micromeria* Benth. (Lamiaceae) on the Canary Islands as inferred from chloroplast and nuclear DNA sequences and ISSR fingerprint data. Molecular Phylogenetics and Evolution: 41, pp. 566-578.

Mes, T.H.M. and 't Hart, H. (1994): *Sedum surculosum* and *S. jaccardianum* (Crassulaceae) share a unique 70 bp deletion in the chloroplast DNA trnL (UUA) - trnF (GAA) intergenic spacer. Plant Systematics and Evolution: 193, pp. 213-221.

Mes, T.H.M. (1995): Phylogenetic and systematic implications of chloroplast and nuclear spacer sequence variation in the Macaronesian Sempervivoideae and related Sedoideae. In: 't Hart, H. and Eggli, U. (eds): Evolution and Systematics of Crassulaceae. Backhuis Publishers, Leiden.

Mes, T.H.M. and 't Hart, H. (1996): The evolution of growth-forms in the Macaronesian genus *Aeonium* (Crassulaceae) inferred from chloroplast DNA RFLPs and morphology. Molecular Ecology: 5, pp. 351-363.

Mes, T.H.M., van Brederode, J. and ‘t Hart, H. (1996): Origin of the woody Macaronesian Sempervivoideae and the phylogenetic position of the East African species of *Aeonium*. Botanica Acta: 109, pp. 477-491.

Mes, T.H.M., Wijers, G.-J. and ‘t Hart, H. (1997): Phylogenetic relationships in *Monanthes* (Crassulaceae) based on morphological, chloroplast and nuclear DNA variation. Journal of Evolutionary Biology: 10, pp. 193-216.

Moore, R.C. and Purugganan, M.D. (2003): The early stages of duplicate gene evolution. Proceedings of the National Academic Society of the USA: 100, pp. 15682-15687.

Moore, R.C. and Purugganan, M.D. (2005): The evolutionary dynamics of plant duplicate genes. Current Opinion in Plant Biology: 8, pp. 122-128.

Mort, M.E., Soltis, D.E., Soltis, P.S., Francisco-Ortega, J. and Santos-Guerra, A. (2001): Phylogenetic relationships and evolution of Crassulaceae inferred from matK sequence data. American Journal of Botany: 88, pp. 76-91.

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

Mort, M.E. and Crawford, D.J. (2004): The continuing search: low-copy nuclear sequences for lower-level plant molecular phylogenetic studies. Taxon: 53, pp. 257-261.

Mort, M.E., Soltis, D.E., Soltis, P.S., Santos-Guerra, A. and Francisco-Ortega, J. (2007): Physiological evolution and association between physiology and growth form in *Aeonium* (Crassulaceae). Taxon: 56, pp. 453-464.

Nei, M. and Gojobori, T. (1986): Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Molecular Biology and Evolution: 3, pp. 418-426.

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

Nekrutenko, A., Makova, K.D. and Li, W.-H. (2001): The Ka/Ks ratio test for assessing the protein-coding potential of genomic regions: an empirical and simulation study. Genome Research: 12, pp. 198-202.

Nyffeler, R. (1992): A taxonomic revision of the genus *Monanthes* Haworth (Crassulaceae). Bradleya: 10, pp. 49-82.

Nyffeler, R. (1995): Die Gattung *Monanthes* Haworth. Kakteen und andere Sukkulenten: 46, pp. 156-165.

Ohno, S. (1970): Evolution by gene duplication. Springer-Verlag, Berlin.

Olsen, K.M., Womack, A., Garrett, A.R., Suddith, J.I. and Purugganan, M.D. (2002): Contrasting evolutionary forces in the *Arabidopsis thaliana* floral developmental pathway. Genetics: 160, pp. 1641-1650.

Page, R.D. (1996): TreeView: an application to display phylogenetic trees on personal computers. Computer Applications in the Biosciences: 12, pp. 357-358.

Pilon-Smits, E.A.H., 't Hart, H., Maas, J.W., Meesterburrie, J.A.N., Kreuler, R. and van Brederode, J. (1992): The evolution of the crassulacean acid metabolism in *Aeonium* inferred from carbon isotope composition and enzyme activities. Oecologia: 91, p. 548-553.

Posada, D. and Crandall, K.A. (1998): Modeltest: testing the model of DNA substitution. Bioinformatics: 14, pp. 817-818.

Praeger, R.L. (1932): An account of the Sempervivum group. Royal Horticulturae Society London.

Purugganan, M.D., Rounsley, S.D., Schmidt, R.J. and Yanofsky, M.F. (1995): Molecular evolution of flower develpoment: diversification of the plant MADS-box regulatory gene family. Genetics: 140, pp. 345-356.

Purugganan, M.D. and Robichaux, R.H. (2005): Adaptive radiation and regulatory gene evolution in the Hawaiian silversword alliance (Asteraceae). Missouri Botanical Garden: 92, pp. 28-35.

Quattrocchio, F., Wing, J., van der Woude, K., Souer, E., de Vetten, N., Mol, J. and Koes, R. (1999): Molecular analysis of the *anthocyanin2* gene of *Petunia* and its role in the evolution of flower color. The Plant Cell: 11, pp. 1433-1444.

Rambaut, A. (1996): Se-Al: Sequence Alignment Editor. Available from: http://tree.bio.ed.ac.uk/software/seal/

Remington, D.L. and Purugganan, M.D. (2002): *GAI* homologues in the Hawaiian silversword alliance (Asteraceae-Madiinae): molecular evolution of growth regulators in a rapidly diversifying plant lineage. Molecular Biology and Evolution: 19, pp. 1563-1574.

Ronquist, F. and Huelsenbeck, J.P. (2003): MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics: 19, pp. 1572-1574.

Saedler, H., Becker, A., Winter, K.-U., Kirchner, C. and Theissen, G. (2001): MADS-box genes are involved in floral development and evolution. Acta Biochimica Polonica: 48, pp. 351-358.

Sanderson, M.J. (1997): A nonparametric approach to estimating divergence times in the absence of rate constancy. Molecular Biology and Evolution: 14, pp. 1218-1231.

Sanderson, M.J. (2003): r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics: 19, pp. 301-302.

Sang, T. (2002): Utility of low-copy nuclear gene sequences in plant phylogenetics. Critical Reviews in Biochemistry and Molecular Biology: 37, pp. 121-147.

Sanmartín, I., van der Mark, P. and Ronquist, F. (2008): Inferring dispersal: a Bayesian approach to phylogeny-based island biogeography, with special reference to the Canary Islands. Journal of Biogeography: 35, pp. 428-449.

Schemske, D.W. and Bradshaw Jr., H.D. (1999): Pollinator preference and the evolution of floral traits in monkeyflowers (*Mimulus*). Proceedings of the National Academic Society of the USA: 96, pp. 11910-11915.

Schönfelder, P. and Schönfelder, I. (1997): Die Kosmos-Kanarenflora. Franckh-Kosmos, Stuttgart.

Shan, H., Zhang, N., Liu, C., Xu, G., Zhang, J., Chen, Z. and Kong, H. (2007): Patterns of gene duplication and functional diversification during the evolution of the *AP1/SQUA* subfamily of plant MADS-box genes. Molecular Phylogenetics and Evolution: 44, pp. 26-41.

Small, R.L. and Wendel, J.F. (2002): Differential evolutionary dynamics of duplicated paralogous *Adh* loci in allotetraploid cotton (*Gossypium*). Molecular Biology and Evolution: 19, pp. 597-607.

Small, R.L., Cronn, R.C. and Wendel, J.F. (2004): Use of nuclear genes for phylogeny reconstruction in plants. Australian Systematic Botany: 17, pp. 145-170.

Stern, A., Doron-Faigenboim, A., Erez, E., Martz, E., Bacharach, E. and Pupko, T. (2007): Selecton 2007: advanced models for detecting positive and purifying selection using a Bayesian inference approach. Nucleic Acids Research: 35, W506-W511.

Stevens, J.F. (1995): Chemotaxonomy of the Eurasian Sedoideae and Sempervivoideae. In: ‘t Hart, H. and Eggli, U. (eds): Evolution and Systematics of Crassulaceae. Backhuis Publishers, Leiden.

Stryer, L. (1995): Biochemistry. 4.ed. Freeman, New York.

Swofford, D.L. (2002): PAUP*. Phylogenetic analysis using parsimony (*and other methods), version 4. Sinauer, Sunderland.

Syring, J., Willyard, A., Cronn, R. and Liston, A. (2005): Evolutionary relationships among *Pinus* (Pinaceae) subsections inferred from multiple low-copy nuclear loci. American Journal of Botany: 92, pp. 2086-2100.

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

Tamura, K., Dudley, J., Nei, M. and Kumar, S. (2007): MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution: 24, pp. 1596-1599.

Tenhunen, J.D., Tenhunen, L.C., Ziegler, H., Stichler, W. and Lange, O.L. (1982): Variation in carbon isotope ratios of Sempervivoideae species from different habitats of Tenerife in the spring. Oecologia: 55, pp. 217-224.

Theissen, G., Becker, A., di Rosa, A., Kanno, A., Kim, J.T., Münster, T., Winter, K.-U. and Saedler, H. (2000): A short history of MADS-box genes in plants. Plant Molecular Biology: 42, pp. 115-149.

Theissen, G. (2001): Development of floral organ identity: stories from the MADS house. Current Opinion in Plant Biology: 4, pp. 75-85.

Theissen, G. (2005): On the origin of flowers: molecular views. Nova Acta Leopoldonina: 342, pp. 209-219.

Tiffin, P. and Hahn, M.W. (2002): Coding sequence divergence between two closely related plant species: *Arabidopsis thaliana* and *Brassica rapa* ssp. *pekinensis*. Journal of Molecular Evolution: 54, pp. 746-753.

Trusty, J.L., Olmstead, R.G., Santos-Guerra, A., Sá-Fontinha, S. and Francisco-Ortega, J. (2005): Molecular phylogenetics of the Macaronesian-endemic genus *Bystropogon* (Lamiaceae): palaeo-islands, ecological shifts and interisland colonizations. Molecular Ecology: 14, pp. 1177-1189.

Tu, T., Dillon, M.O., Sun, H. and Wen, J. (2008): Phylogeny of *Nolana* (Solanaceae) of the Atacama and Peruvian deserts inferred from sequences of four plastid markers and the nuclear *LEAFY* second intron. Molecular Phylogenetics and Evolution: 49, pp. 561-573.

Uhl, C.H. (1961): The chromosomes of the Sempervivoideae (Crassulaceae). American Journal of Botany: 48, pp. 114-123.

Vandenbussche, M., Theissen, G., van der Peer, Y. and Gerats, T. (2003): Structural diversification and neo-functionalization during floral MADS-box gene evolution by C-terminal frameshift mutations. Nucleic Acids Research: 31, pp. 4401-4409.

van Ham, R.C.H.J. and 't Hart, H. (1998): Phylogenetic relationships in the Crassulaceae inferred from chloroplast DNA restriction-site variation. American Journal of Botany: 85, pp. 123-134.

Vergara-Silva, F., Martínez-Castilla, L. and Alvarez-Buylla, E.R. (2000): MADS-box genes: development and evolution of plant body plans. Journal of Phycology: 36, pp. 803-812.

Wang, W.-K., Schaal, B.A., Chiou, Y.-M., Murakami, N., Ge, X.-J., Huang, C.-C. and Chiang, T.-Y. (2007): Diverse selective modes among orthologs/paralogs of the chalcone synthase (*Chs*) gene family of *Arabidopsis thaliana* and its relative *A. halleri* ssp. *gemmifera*. Molecular Phylogenetics and Evolution: 44, pp. 503-520.

White, T.J., Bruns, T., Lee, S. and Taylor, J. (1990): Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis, M.A., Gelfand, D.H., Sninsky, J.J. and White, T.J. (eds): PCR Protocols. A Guide to Methods and Applications. Academic Press, San Diego.

Whitkus, R. (1998): Genetics of adaptive radiation in Hawaiian and Cook Islands species of *Tetramolopium* (Asteraceae). II. Genetic linkage map and its implications for interspecific breeding barriers. Genetics: 150, pp. 1209–1216.

Whittall, J.B., Medina-Marino, A., Zimmer, E.A. and Hodges, S.A. (2006): Generating single-copy nuclear gene data for a recent adaptive radiation. Molecular Phylogenetics and Evolution: 39, pp. 124-134.

Whittall, J.B. and Hodges, S.A. (2007): Pollinator shifts drive increasingly long nectar spurs in columbine flowers. Nature: 447, pp. 706-710.

Yang, J., Huang, J., Gu., H., Zhong, Y. and Yang, Z. (2002): Duplication and adaptive evolution of the chalcone synthase genes of *Dendranthema* (Asteraceae). Molecular Biology and Evolution: 19, pp. 1752-1759.

Yang, Z. (2005): The power of phylogenetic comparison in revealing protein function. Proceedings of the National Academy of Sciences of the USA: 102, pp. 3179-3180.

Zhang, L., Vision, T.J. and Gaut, B.S. (2002): Patterns of nucleotide substitution among simultaneously duplicated gene pairs in *Arabidopsis thaliana*. Molecular Phylogenetics and Evolution: 19, pp. 1464-1473.

Zhang, J. (2003): Evolution by gene duplication: an update. TRENDS in Ecology and Evolution: 18, pp. 292-298.

Zhang, W., Xiang, Q.-Y., Thomas, D.T., Wiegmann, B.M., Frohlich, M.W. and Soltis, D.E. (2008): Molecular evolution of *PISTILLATA*-like genes in the dogwood genus *Cornus* (Cornaceae). Molecular Phylogenetics and Evolution: 47, pp. 175-195.

Websites:

1) http://www.ncbi.nlm.nih.gov/ (last access: 09/2009)
2) http://selecton.tau.ac.il/ (last access: 09/2009)
3) http://foo.maths.uq.edu.au/~huber/bellerophon.pl (last access: 09/2009)
4) http://www.atgc-montpellier.fr/phyml/(last access: 09/2009)

Abbreviations

aa	amino acid
Adh	alcohol dehydrogenase
AIC	Akaike information criterion
AP1	*APETALA1*
AP3	*APETALA3*
AtpB	membrane-bound ATP synthase
At	*Arabidopsis thaliana*
BI	Bayesian Interference
bp	base pair
bs	bootstrap support
CAB9	CHLOROPHYLL A/B BINDING PROTEIN9
CAL	*CAULIFLOWER*
CAM	Crassulacean Acid Metabolism
cDNA	complementary DNA
Ces	carboxylesterase
CHS	chalcone synthase
cpDNA	chloroplast DNA
CYC	*CYCLOIDEA*
DEF	*DEFICIENS*
DICH	*DICHOTOMA*
DNA	deoxyribonucleic acid
F	Fuerteventura
G	La Gomera
GAI	gibberellic acid insensitive
GC	Gran Canaria
GLO	*GLOBOSA*
GTR	General Time Reversible Model
H	El Hierro
HKY	Hasegawa-Kishino-Yano-85
HSA	Hawaiian silversword alliance
I	invariate
IPTG	isopropyl-beta-D-thiogalactopyranoside
ITS	internal transcribed spacer
K81uf	Kimura-3-Parameter with unequal frequencies
knox	knotted related homeobox
LB	Luria-Bertani
LRT	likelihood ratio test
matK	maturase K
MCS	Macaronesian Crassulaceae Sempervivoideae
ML	Maximum likelihood
MO	Morocco
MP	Maximum parsimony
mRNA	messenger RNA
My	million year

NCBI	National Center for Biotechnology Information
NEB	New England Biolab
NJ	Neighbor-joining
NNI	Nearest Neighbor Interchange
NPRS	Nonparametric rate smoothing
nr	nuclear
OG	outgroup
ORF	open reading frame
P	La Palma
PCR	polymerase chain reaction
PEP	phosphoenolpyruvate
PEPC	phosphoenolpyruvate carboxylase
PHT	Partition Homogeneity Test
PI	*PISTILLATA*
PL	Penalized likelihood
pp	posterior probabilities
QTL	quantitative trait loci
RbcL	RUBISCO large (subunit)
rDNA	ribosomal DNA
R-gene	resistence gene
RNA	ribonucleic acid
RRT	Relative Rate Test
RUBISCO	ribulose-1,5-bisphosphate
sect	section
ser	series
SMNS	State Museum of Natural History Stuttgart
SPR	Subtree Pruning and Regrafting
SQUA	*SQUAMOSA*
T	Tenerife
TBR	tree bisection reconnection
TE	Tris-EDTA (ethylenediaminetetraacetic acid)
TFL	*TERMINAL FLOWER*
TIM	Transitional Model
trnF	transfer RNA for Phenylalanine
trnL	transfer RNA for Leucine
TrN	Tamura-Nei
TVM	Transversional Model
UTR	untranslated region
ω	omega = Ka/Ks
X-Gal	bromo-chloro-indolyl-galactopyranoside

Overview of scientific contributions

Talks:

Esfeld, K. and Thiv, M. (2004): Studying adaptive radiation at the molecular level: a case study in the Macaronesian Crassulaceae-Sempervivoideae. 7. Annual Meeting of the GfBS. Stuttgart.

Esfeld, K. (2006): The process of speciation: evidence from an adaptive radiation. Vavilov-Seminar, IPK Gatersleben.

Esfeld, K., Thiv, M. and Koch, M. (2008): The use of nuclear coding genes for phylogenetic reconstructions in an adaptive radiation. Systematics. Göttingen.

Thiv, M., Esfeld, K. and Koch, M. (2009): The impact of floral genes in the evolution of an adaptive radiation. Systematics 2009, Leiden.

Poster:

Esfeld, K. and Thiv, M. (2005): A comparison of floral morphology and physiology genes in an adaptive radiation. XVII International Botanical Congress. Vienna.

Esfeld, K. and Thiv, M. (2006): The evolution of the Macaronesian Crassulaceae-Sempervivoideae: Studying adaptive radiation at the molecular level. 17th International Symposium Biodiversity and Evolutionary Biology, Bonn.

Publications:

Esfeld, K., Koch, M.A., van der Niet, T., Seifan, M. and Thiv, M. (2009): Little interspecific pollen transfer despite overlap in pollinators between sympatric *Aeonium* (Crassulaceae) species pairs. Flora. doi:10.1016/j.flora.2008.10.002.

Thiv, M., Esfeld, K. and Koch, M. (2009, in press): Studying “adaptive radiation” at the molecular level: a case study in the Macaronesian Crassulaceae-Sempervivoideae. In Glaubrecht, M. and Schneider, H. (eds.): Evolution in Action - Adaptive Radiations and the Origins of Biodiversity. Springer, Hamburg.

Other contributions:

Esfeld, K., Hensen, I., Wesche, K., Jakob, S.S., Tischew, S. and Blattner, F.R. (2008): Molecular data indicate multiple independent colonizations of former lignite mining areas in Eastern Germany by *Epipactis palustris* (Orchidaceae). Biodiversity and Conservation: 17, pp. 2441-2453.

Tadele, Z. and Esfeld, K. (2008): Applications of TILLING to the understudied crops from Africa: the case of tef. FAO/IAEA - International Symposium on Induced Mutations in Plants, Vienna. Poster.

Esfeld, K., Plaza, S. and Tadele, Z. (2009): Bringing high-throughput techniques to orphan crop of Africa: Highlights from the Tef TILLING Project. Gene Conserve: 8, pp: 783-788.

Tadele, Z., Esfeld, K. and Plaza, S. (2009): Applications of high-throughput techniques to the understudied crops of Africa. Conference on Agriculture: Africa's "Engine for Growth" - Plant science & biotechnology hold the key, Rothamsted Research, Harpenden, Herts.

Tadele, Z., Esfeld, K. and Plaza, S. (2009): Employing green revolution genes to improve orphan crop tef. Tadele, Z. (ed.): New approaches to plant breeding of orphan crops in Africa: Proceedings of an International Conference, 19-21 September 2007, Bern, Switzerland.

Appendices

1.) Datasets and alignments on CD

A) PEPC_full_length_Diss.fas
B) PEPC_full_length_Diss.nex
C) PEPC_exon_Diss.fas
D) PEPC_exon_Diss.nex
E) AP1_full_length_with_cDNA_Diss.fas
F) AP1_full_length_with_cDNA_Diss.nex
G) AP1_full_length_without_cDNA_Diss.fas
H) AP1_full_length_without_cDNA_Diss.nex
I) AP1_exon_Diss.fas
J) AP1_exon_Diss.nex
K) AP3_full_length_Diss.fas
L) AP3_full_length_Diss.nex
M) AP3_exon_Diss.fas
N) AP3_exon_Diss.nex

O) Thesis PDF

2.) Comparison of infrageneric classification (table 15).

3.) Location of the studied species (table 16).

4.) Overview of the used primers (table 17).

5.) Fig. 29-43: BI and ML phylograms based on various datasets.

6.) Fig. 44: BI phylogram based on the exon dataset of *MCS_AP1* including all amplified unique sequences.

7.) Fig. 45: NJ phylogram based on the enlarged exon *MCS_PEPC* dataset.

8.) Fig. 46: Ultrametric tree.

Table 15: Comparison of the infrageneric classification of *Aeonium*: Lems (1960), Liu (1989), Mes (1995), and Mort et al. (2002). Studied species are indicated in bold.

	Lems (1960)	Liu (1989)	Mes (1995)	Mort et al. (2002; cpDNA/nrITS)
A. aizoon			Sect. *Greenovia*	Clade 2
A. aureum			**Sect. *Greenovia***	**Clade 2**
A. arboreum	Sect. *Holochrysa*	Sect. *Aeonium*	Sect. *Aeonium*	
A. balsamiferum	Sect. *Holochrysa*	Sect. *Aeonium*	Sect. *Aeonium*	Clade 3
A. canariense	**Sect. *Canariensia***	**Sect. *Patinaria***	**Sect. *Canariensia***	**Clade 1**
A. castello-paivae	Sect. *Urbica*	Sect. *Leuconium*	Sect. *Leuconium*	Clade 4
A. ciliatum	Sect. *Urbica*	Sect. *Leuconium*	Sect. *Leuconium*	Clade 4
A. cuneatum	**Sect. *Canariensia***	**Sect. *Patinaria***	**Sect. *Canariensia***	
A. davidbramwellii	Sect. *Urbica*	Sect. *Leuconium*	Sect. *Leuconium*	Clade 4
A. decorum	Sect. *Urbica*	Sect. *Leuconium*	Sect. *Leuconium*	Clade 4
A. diplocyclum			Sect. *Greenovia*	Clade 2
A. dodrantale			Sect. *Greenovia*	
A. glandulosum	Sect. *Canariensia*	Sect. *Patinaria*	Sect. *Patinaria*	Clade 3
A. glutinosum	Sect. *Canariensia*	Sect. *Pittonium*	Sect. *Pittonium*	Clade 4
A. gomerense	Sect. *Urbica*	Sect. *Leuconium*	Sect. *Leuconium*	Clade 4
A. goochiae	**Sect. *Goochiae***	**Sect. *Petrothamnium***	**Sect. *Goochiae***	**Clade 2**
A. gorgoneum	Sect. *Holochrysa*	Sect. *Pittonium*	Sect. *Aeonium*	Clade 3
A. haworthii	Sect. *Urbica*	Sect. *Leuconium*	Sect. *Leuconium*	Clade 4
A. hierrense	Sect. *Urbica*	Sect. *Leuconium*	Sect. *Leuconium*	Clade 4
A. holochrysum	Sect. *Holochrysa*	Sect. *Aeonium*	Sect. *Aeonium*	Clade 3
A. korneliuslemsii	Sect. *Holochrysa*	Sect. *Aeonium*	Sect. *Aeonium*	Clade 3
A. lancerottense	Sect. *Urbica*	Sect. *Leuconium*	Sect. *Leuconium*	Clade 4
A. leucoblepharum	Sect. *Holochrysa*	Sect. *Pittonium*	Sect. *Aeonium*	Clade 3
A. lindleyi	Sect. *Goochiae*	Sect. *Petrothamnium*	Sect. *Goochiae*	Clade 2
A. mascaense				Clade 4
A. nobile	**Sect. *Megalonium***	**Sect. *Megalonium***	**Sect. *Leuconium***	**Clade 4**
A. palmense	Sect. *Canariensia*	Sect. *Patinaria*	Sect. *Canariensia*	Clade 1
A. percarneum	Sect. *Urbica*	Sect. *Leuconium*	Sect. *Leuconium*	Clade 4
A. pseudourbicum				Clade 4
A. rubrolineatum	**Sect. *Holochrysa***	**Sect. *Aeonium***	**Sect. *Aeonium***	**Clade 3**
A. saundersii	**Sect. *Goochiae***	**Sect. *Petrothamnium***	**Sect. *Petrothamnium***	**Clade 2**
A. sedifolium	Sect. *Goochiae*	Sect. *Petrothamnium*	Sect. *Petrothamnium*	
A. simsii	Sect. *Goochiae*	Sect. *Chrysocome*	Sect. *Aeonium*	Clade 3
A. smithii	**Sect. *Goochiae***	**Sect. *Chrysocome***	**Sect. *Chrysocome***	
A. spathulatum	Sect. *Goochiae*	Sect. *Chrysocome*	Sect. *Chrysocome*	Clade 4
A. stuessyi	Sect. *Holochrysa*	Sect. *Pittonium*	Sect. *Aeonium*	
A. subplanum	Sect. *Canariensia*	Sect. *Patinaria*	Sect. *Canariensia*	Clade 1
A. tabuliforme	Sect. *Canariensia*	Sect. *Patinaria*	Sect. *Canariensia*	Clade 1
A. urbicum	Sect. *Urbica*	Sect. *Leuconium*	Sect. *Leuconium*	Clade 4
A. undulatum	Sect. *Holochrysa*	Sect. *Aeonium*	Sect. *Aeonium*	Clade 3
A. valverdense	Sect. *Urbica*	Sect. *Leuconium*	Sect. *Leuconium*	
A. vestitum				Clade 3
A. virgineum	Sect. *Canariensia*	Sect. *Patinaria*	Sect. *Canariensia*	Clade 1
A. viscatum	Sect. *Goochiae*	Sect. *Petrothamnium*	Sect. *Goochiae*	Clade 2
A. volkeri				Clade 4

Table 16: Location of the collection sites of the studied species. Indicated are area and name of the collection site as well as coordinates in Degrees Minutes and Seconds.

species	area	location	N	W	remarks
A. aureum	T	Los Carrizales	28° 11' 27.47"N	16° 30' 40.82"W	
A. canariense	T	San Andres	28° 18' 48.18"N	16° 06' 51.23"W	
A. cuneatum	T	El Bailadero	28° 19' 28.63"N	16° 07' 23.83"W	
A. goochiae	P	Los Tilos			provided by R. Lösch
A. nobile	P	Barranco de las Angustias			provided by R. Lösch
A. rubrolineatum	G	Vallehermoso	28° 05' 41.63"N	17° 10' 26.96"W	
A. saundersii	G	Lomito	28° 03' 57.06"N	17° 05' 39.25"W	
A. smithii	T	Vila Flor	28° 05' 41.65"N	16° 22' 53.63"W	
Ai. laxum	T	Genoves by Icod	28° 12' 57.22"N	16° 26' 39.54"W	
Ai. pachycaulon	T	Taborno	28° 19' 18.77"N	16° 09' 38.57"W	
M. anagensis	T	Taganana	28° 19' 49.02"N	16° 07' 19.43"W	
M. icterica	T	Tabaiba	28° 11' 34.95"N	16° 30' 36.48"W	
S. caeruleum	Sardinia	Arzana	39°55' N	9° 31' O	provided by R. Lübenau-Nestle
S. jaccardianum	MO	Bekrite	33° 00' 55.25"N	5° 08' 30.73"W	two extractions
S. modestum	MO	Taddert	31° 24' 20.47"N	4° 06' 18.29"W	two extractions, cultivated from collected seeds
S. pubescens	Tunesia	City Succulent Collection Zurich			RS-TAV 996096/0
S. surculosum	MO	Oukaimeden	31° 06' 15.37"W	7° 30' 40.97"W	two extractions

Table 17: Primers used to amplify the respective gene regions. Indicated are name, gene region, sequence, and application.

primer name	region	sequence	applied for
M13_F (Promega)	vector	GTAAAACGACGGCCAG	vector
M13_R (Promega)	vector	CAGGAAACAGCTATGAC	vector
Jet_F (Fermentas)	vector	GCCTGAACACCATATCCATCC	vector
Jet_R (Fermentas)	vector	GCAGCTGAGAATATTGTAGGAGATC	vector
PEPC-F	*PEPC*	TCWGATTCAGGAAAAGATGC	amplification
PEPC-R	*PEPC*	GCAGCGATRCCCYTCATTGT	amplification
PEPC_Sed_for_2	*PEPC*	TCWGATTCAGGAAAAGATGCWGG	amplification
PEPC_Sed_rev	*PEPC*	GCAGCGATRCCCYTCATTGTCAA	amplification
PEPC_446F	*PEPC*	TGTCGCCACAGAGCAMTATC	sequencing
PEPC_968R	*PEPC*	GTATGCATCACGCAGACGMA	sequencing
PEPC_1331R	*PEPC*	GCTTGGCATACATTTAGTGT	sequencing
PEPC_342F	*PEPC*	CCTCATGGATGAAATGGC	sequencing
PEPC_293F_Sed	*PEPC*	AGCATGGYATGAATCCWCCT	sequencing
poly(T)-AP1	*AP1*	GACTCGAGTCGACATCGATTTTTTTTTTTTTTTT	cDNA
AP1MDS2	*AP1*	TGGNYTKNTSAAGAARGCTCATGA	cDNA
AP1MDS3	*AP1*	GTNCARYTNARRMGNATNGARAAYAAGAT	cDNA
SQUAR	*AP1*	GCAAAGCATCCMAKATGGCATG	cDNA
AP1-noT	*AP1*	GACTCGAGTCGACATCGA	cDNA
AP1-11F	*AP1*	AAGAAGGCTCATGAGATCTC	amplification
AP1-704R	*AP1*	TGCTCATTCTCCCTTCATCTTC	amplification
AP1_Sed_for	*AP1*	GAGATCTCYGTCTTGTGTGATGC	amplification
AP1_Sed_rev	*AP1*	CTCCCTTCATCTTCTCCCTGGTAAG	amplification
AP1_238F	*AP1*	GGCCAAGCTTGATCTCTTGCAGA	sequencing
AP1_286R	*AP1*	TGTAACTCTCTCATGCTCAATGC	sequencing
AP1_1631R	*AP1*	AAGCTCAGAAATTGACTCATGC	sequencing
AP1_379R	*AP1*	GCTCAGAAATTGACTCATGCATA	sequencing
AP1_482R	*AP1*	CACTGTGCTTCGTCTTGCTGCAC	sequencing
AP1_515	*AP1*	GCACTTAAGGCTCAATGCTT	sequencing
AP1_5567R	*AP1*	TTIATSCAGCAAAGCATCCAAG	sequencing
AP1_251F	*AP1*	AAAATCACAGGCATTACTTGGG	sequencing
AP1_559F	*AP1*	CCGCTTAGCTGTAGAAACTGAG	sequencing
AP1_375SeqF	*AP1*	TGCACTTTGTTTTTGGACTTG	sequencing
Seq_Sed_AP1_R	*AP1*	AACATGGCTGTGATMTTTACACA	sequencing
AP3-polydT	*AP3*	CCGGATCCTCTAGAGCGGCCGCTTTTTTTTTTTTTTTT	cDNA
ATG3	*AP3*	ATGGSIMGIMMIAARATISARAT	cDNA
MADS4	*AP3*	AAYMGRCARGTIACITWYAARMGRMG	cDNA
MADS4-Aeo	*AP3*	AATAGGCAGGTGACGTTYAAGAG	cDNA
AP3-11F	*AP3*	TGACGTTTAAGAGGCGGAAC	cDNA
AP3-noT2	*AP3*	ATCCTCTAGAGCGGCCGC	cDNA
AP3-351F	*AP3*	AGCTATTCGTGCTCGCAAGT	cDNA
PI-F-Aeo	*AP3*	AATAGGCAGGTGACGTTYTCGAAGCGGAG	amplification
AP3-724R	*AP3*	TTGGCAAAAACAACGAAACA	amplification
AP3-82F	*AP3*	AARAAGGCAGAGGAGCTYAC	amplification
AP3-1766R	*AP3*	TCAGTTACCACTCATGAGRGTGTAA	amplification
AP3_399R	*AP3*	TCAGTTCTGCTTGAAATTGCTT	sequencing
AP3_570F	*AP3*	GCAAGAAACTTTGAGGAAAGTGA	sequencing
AP3_586SeqF	*AP3*	AATGCAAGAAACTTTGAGGAA	sequencing
AP3_1109SeqR	*AP3*	TCKGCTTGAAATTGCTTTG	sequencing

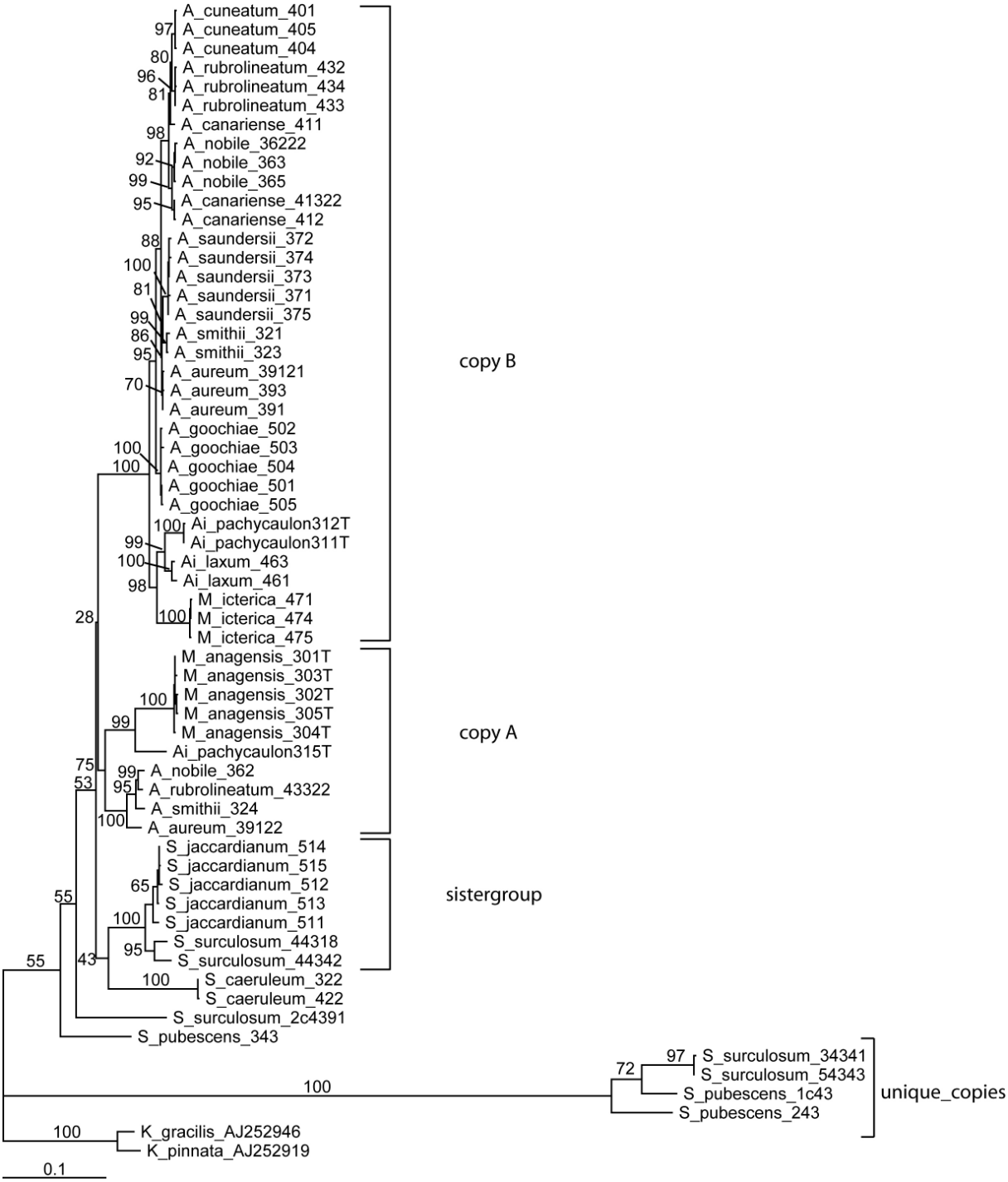


Fig. 29: ML phylogram based on the *MCS_PEPC* full-length data. Bootstrap support is given at the nodes.

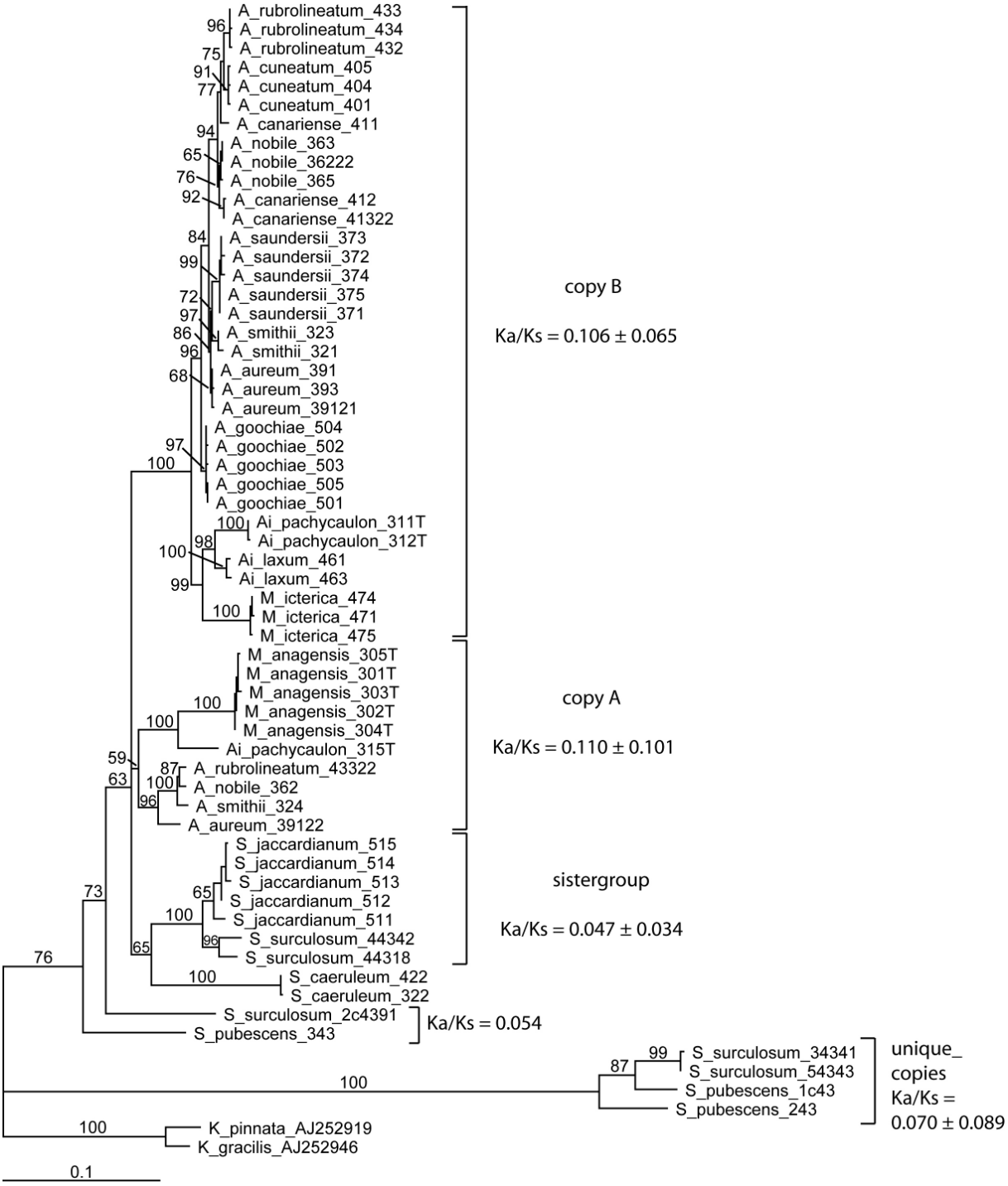


Fig. 30: ML phylogram based on the *MCS_PEPC* exon data. Bootstrap support is given at the nodes.

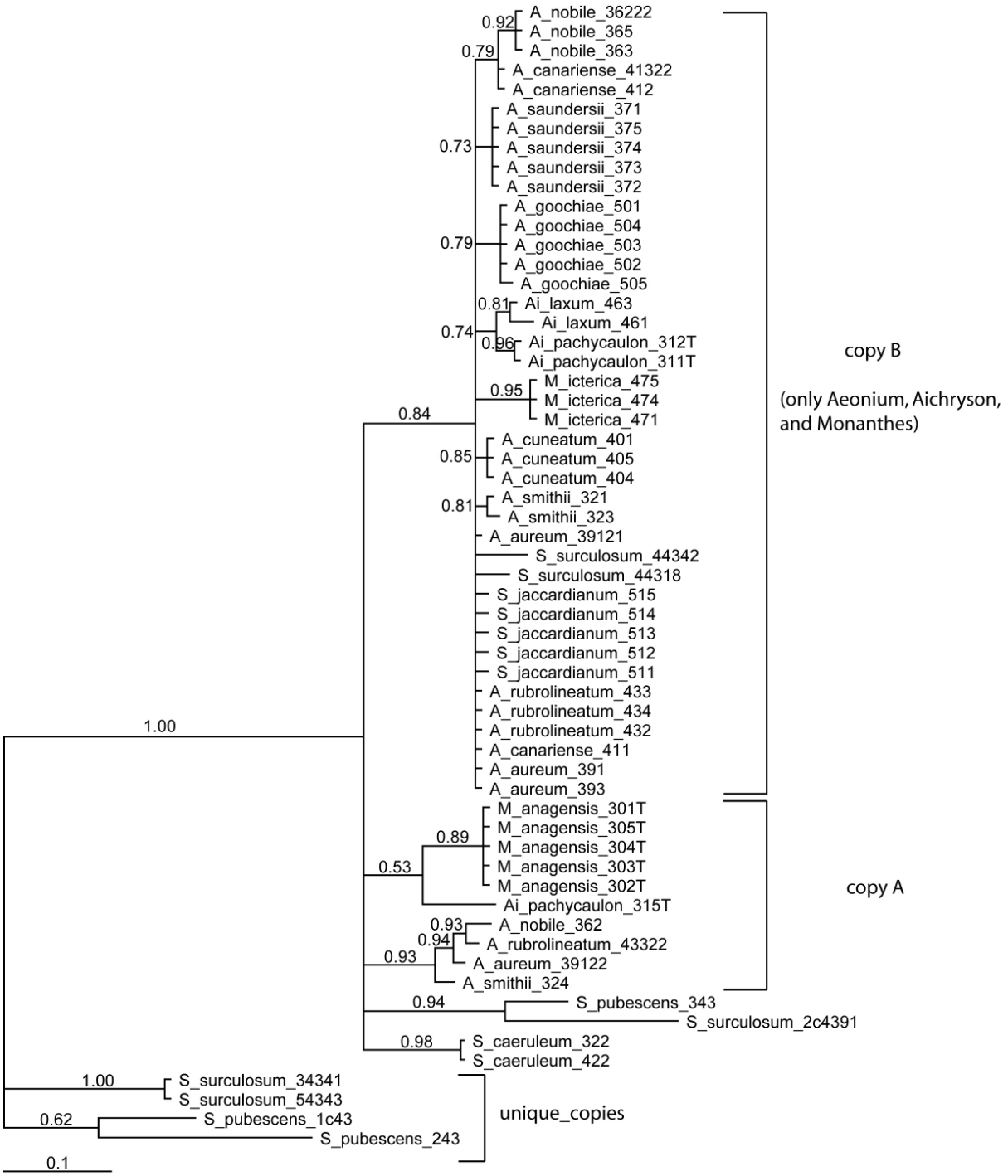


Fig. 31: BI phylogram based on the *MCS_PEPC* intron data. Posterior probabilities are given at the nodes.

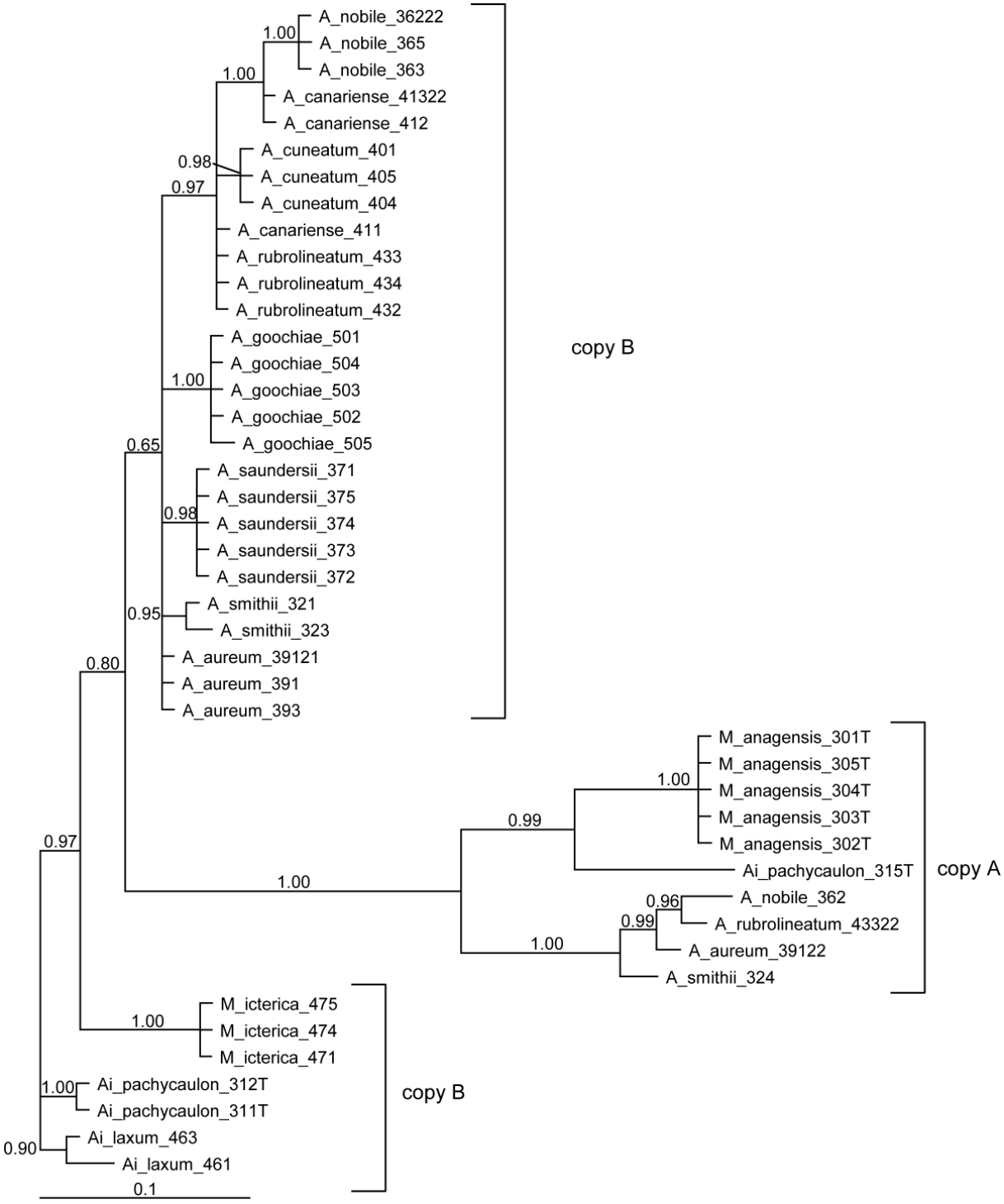


Fig. 32: BI phylogram based on the *MCS_PEPC* MCS intron data (*Sedum* sequences were excluded). Posterior probabilities are given at the nodes.

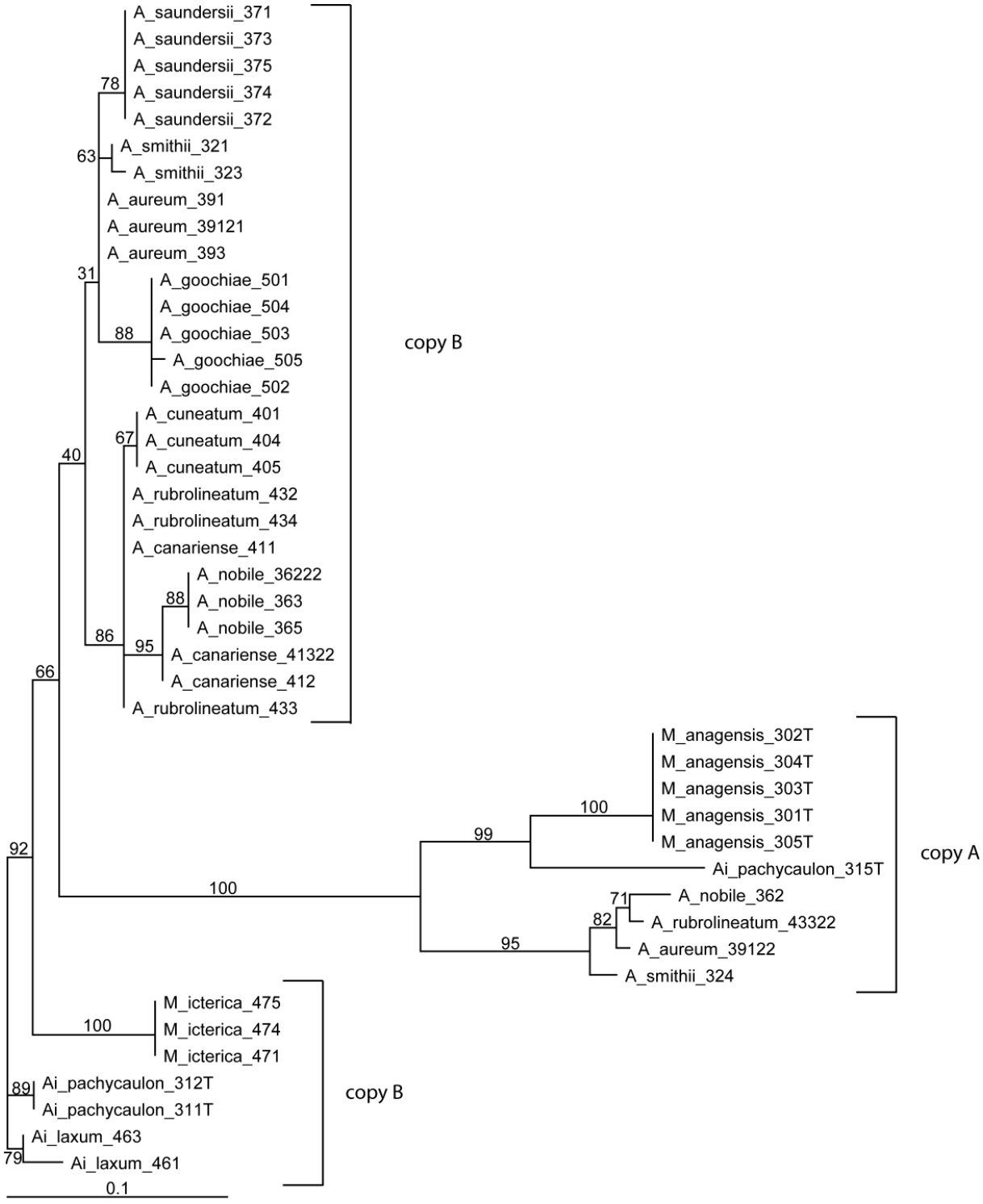


Fig. 33: ML phylogram based on the *MCS_PEPC* MCS intron data (*Sedum* sequences were excluded). Bootstrap support is given at the nodes.

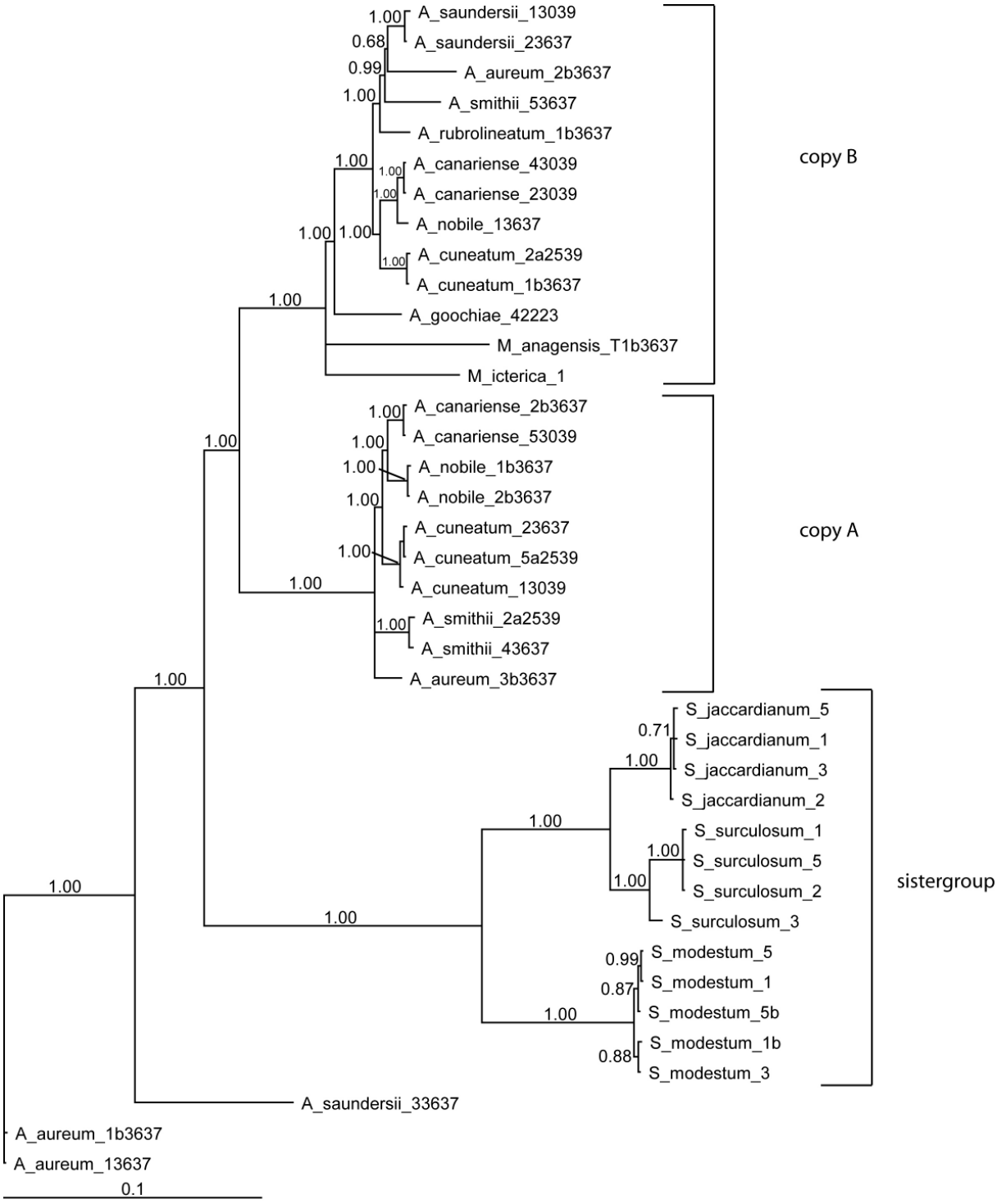


Fig. 34: BI phylogram based on the *MCS_AP1* full-length data. Posterior probabilities are given at the nodes.

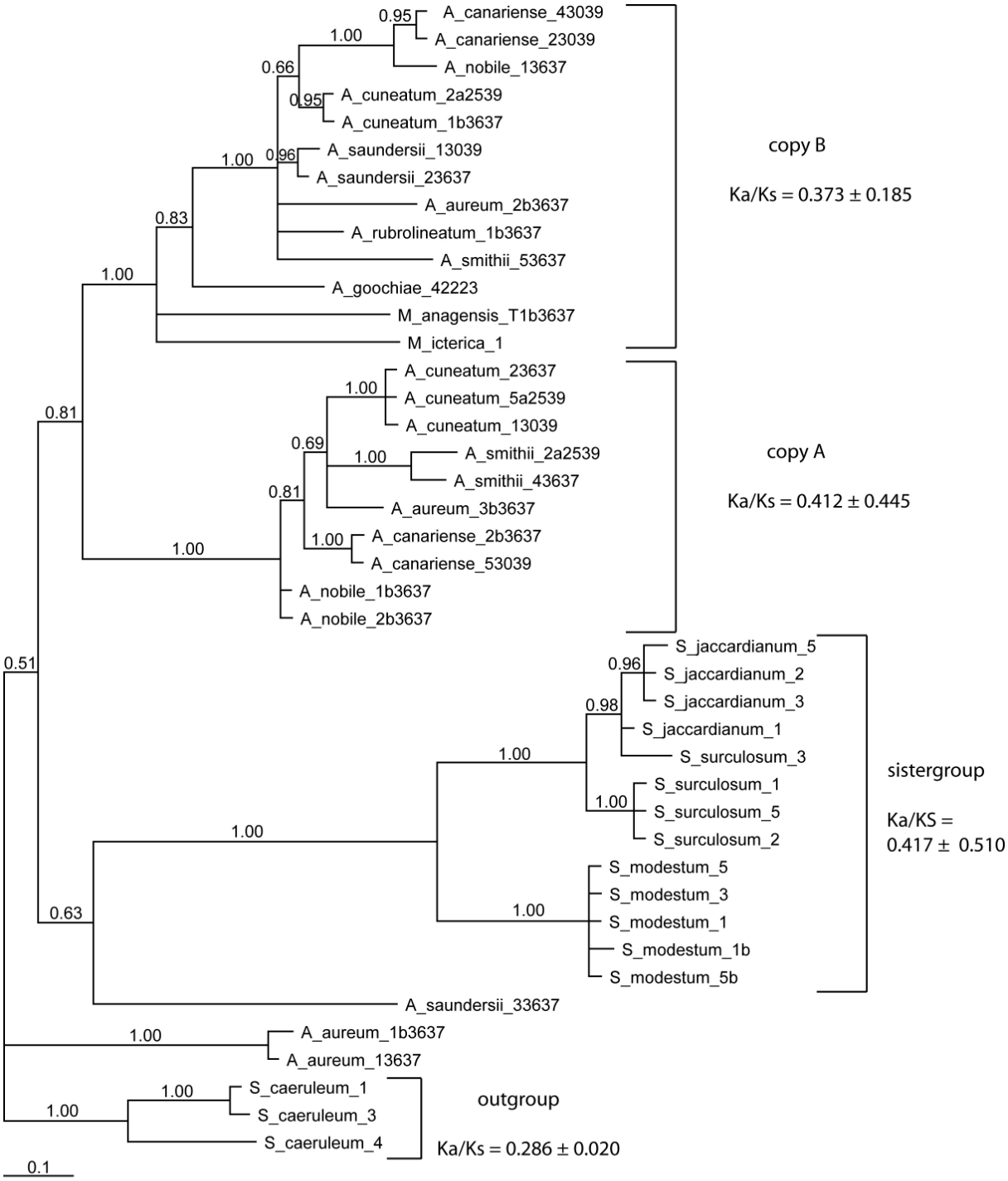


Fig. 35: BI phylogram based on the *MCS_AP1* exon data. Posterior probabilities are given at the nodes.

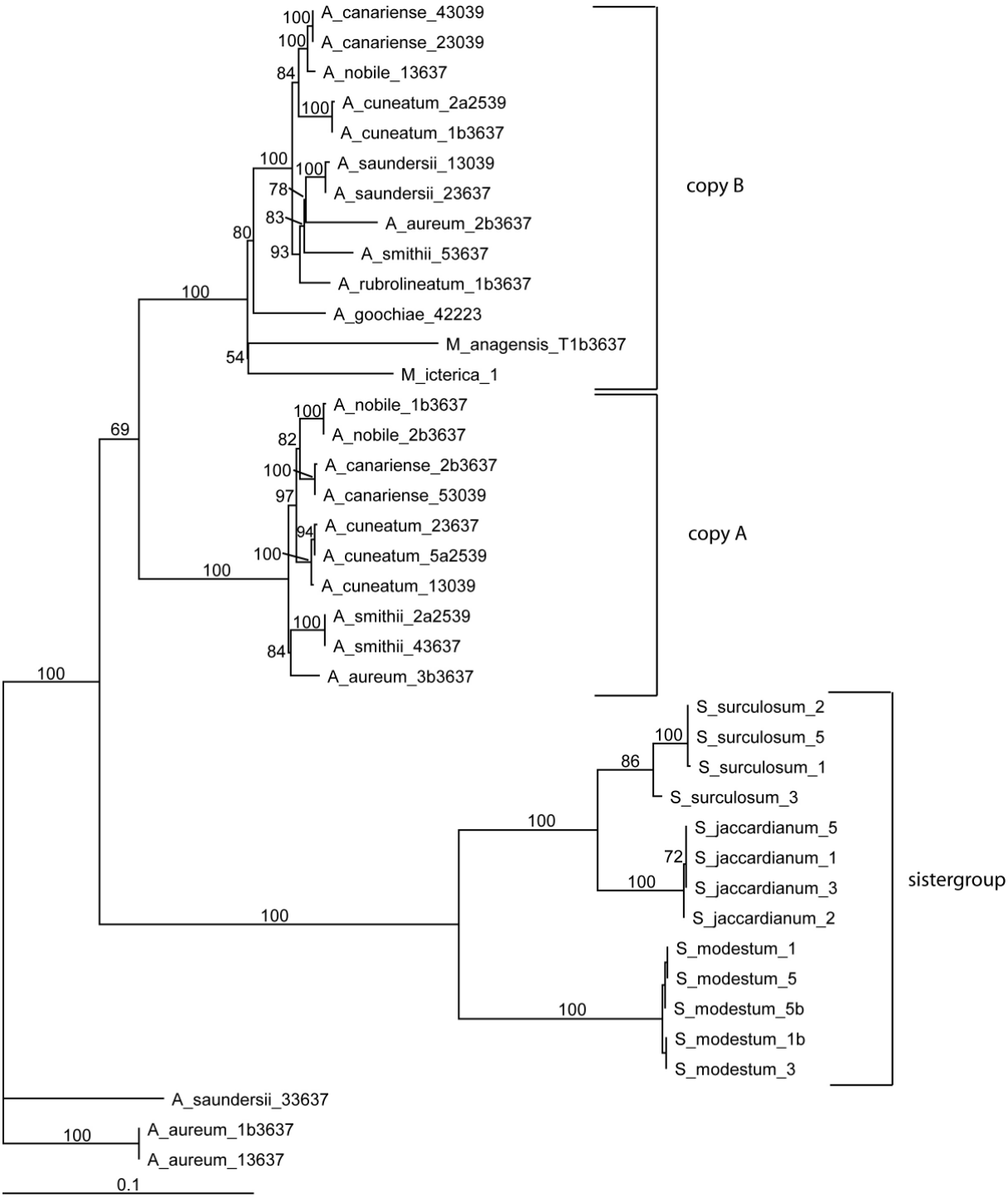


Fig. 36: ML phylogram based on the *MCS_AP1* intron data. Bootstrap support is given at the nodes.

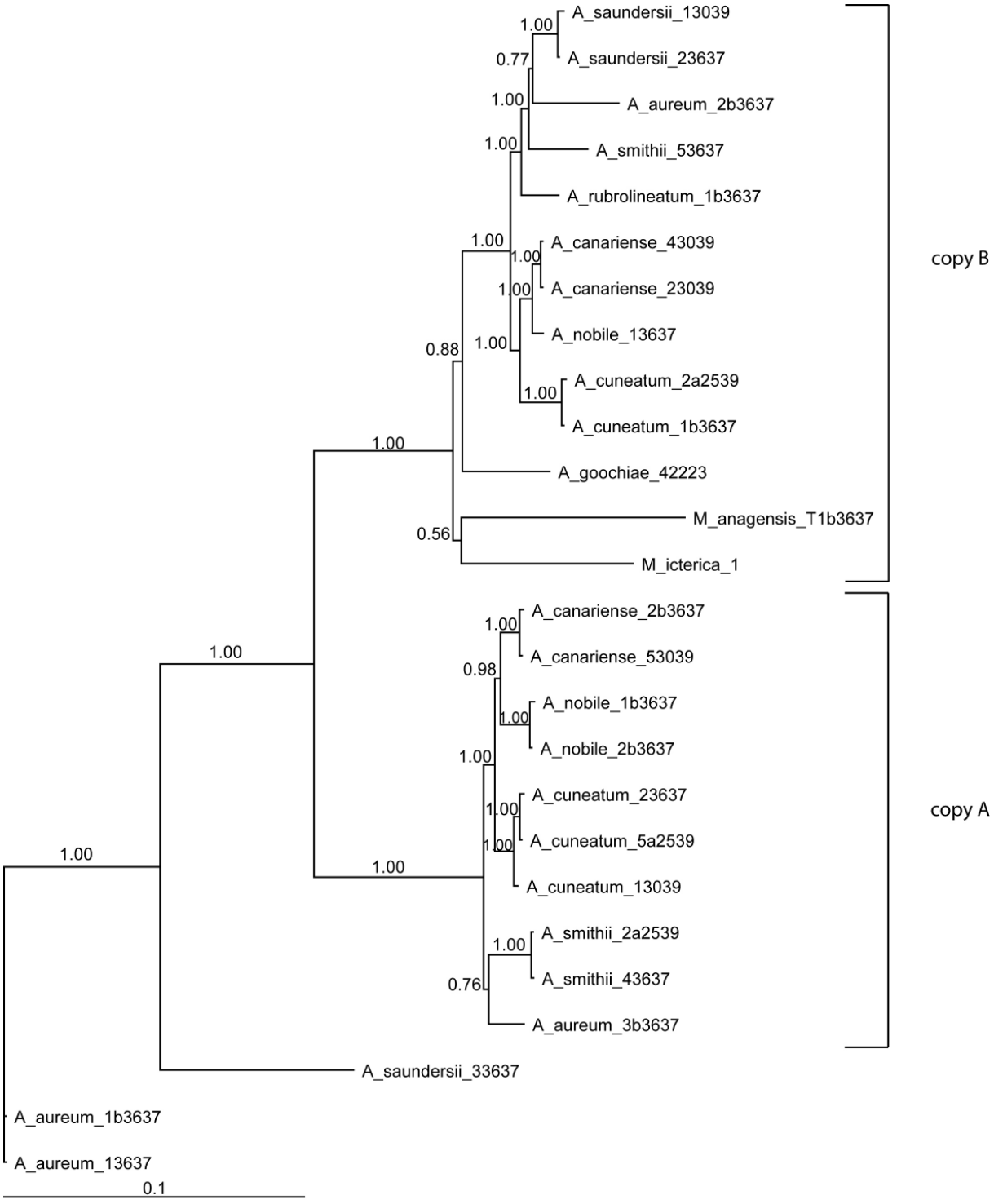


Fig. 37: BI phylogram based on the *MCS_AP1* MCS intron data (*Sedum* sequences were excluded). Posterior probabilities are given at the nodes.

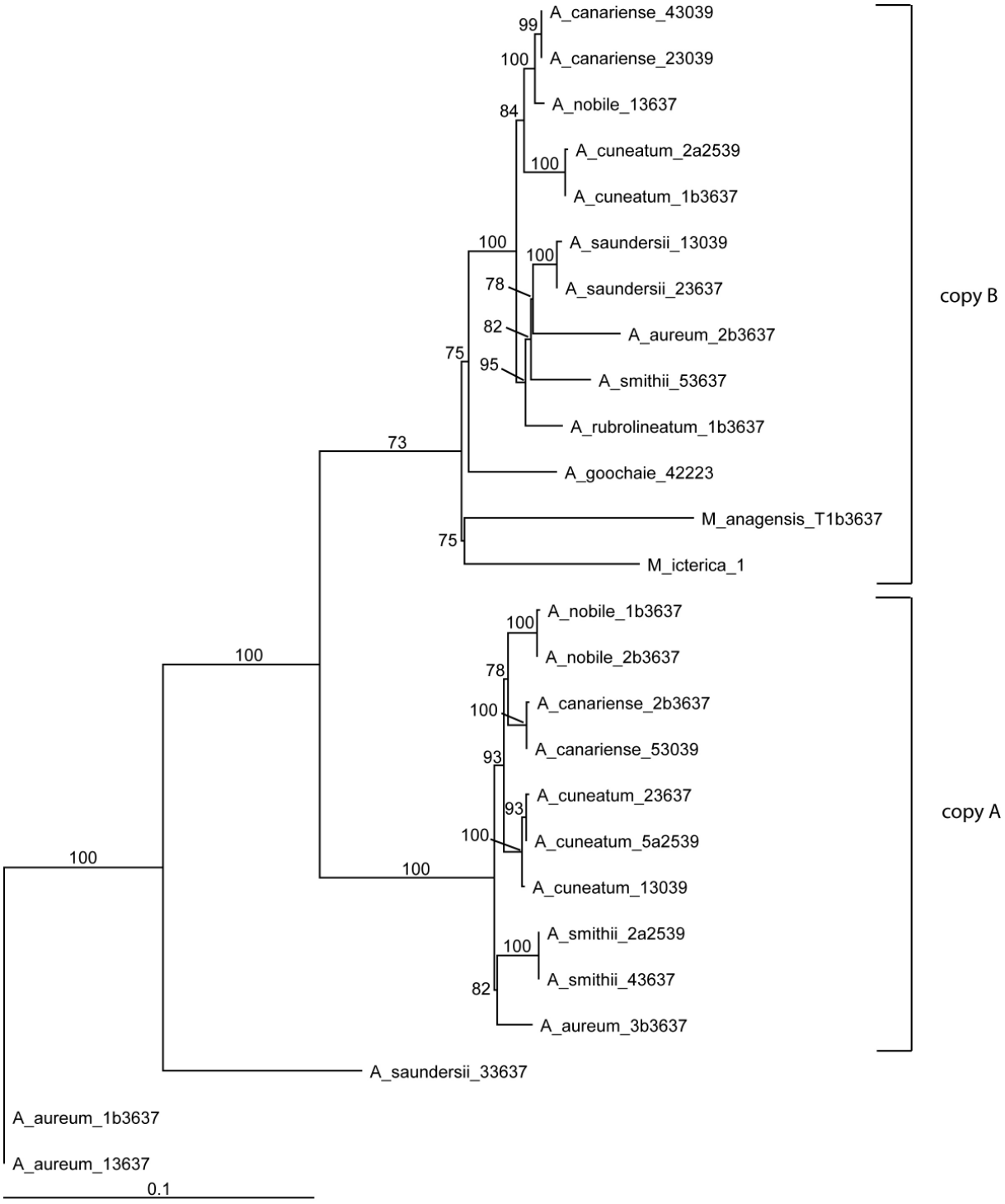


Fig. 38: ML phylogram based on the *MCS_AP1* MCS intron data (*Sedum* sequences were excluded). Bootstrap support is given at the nodes.

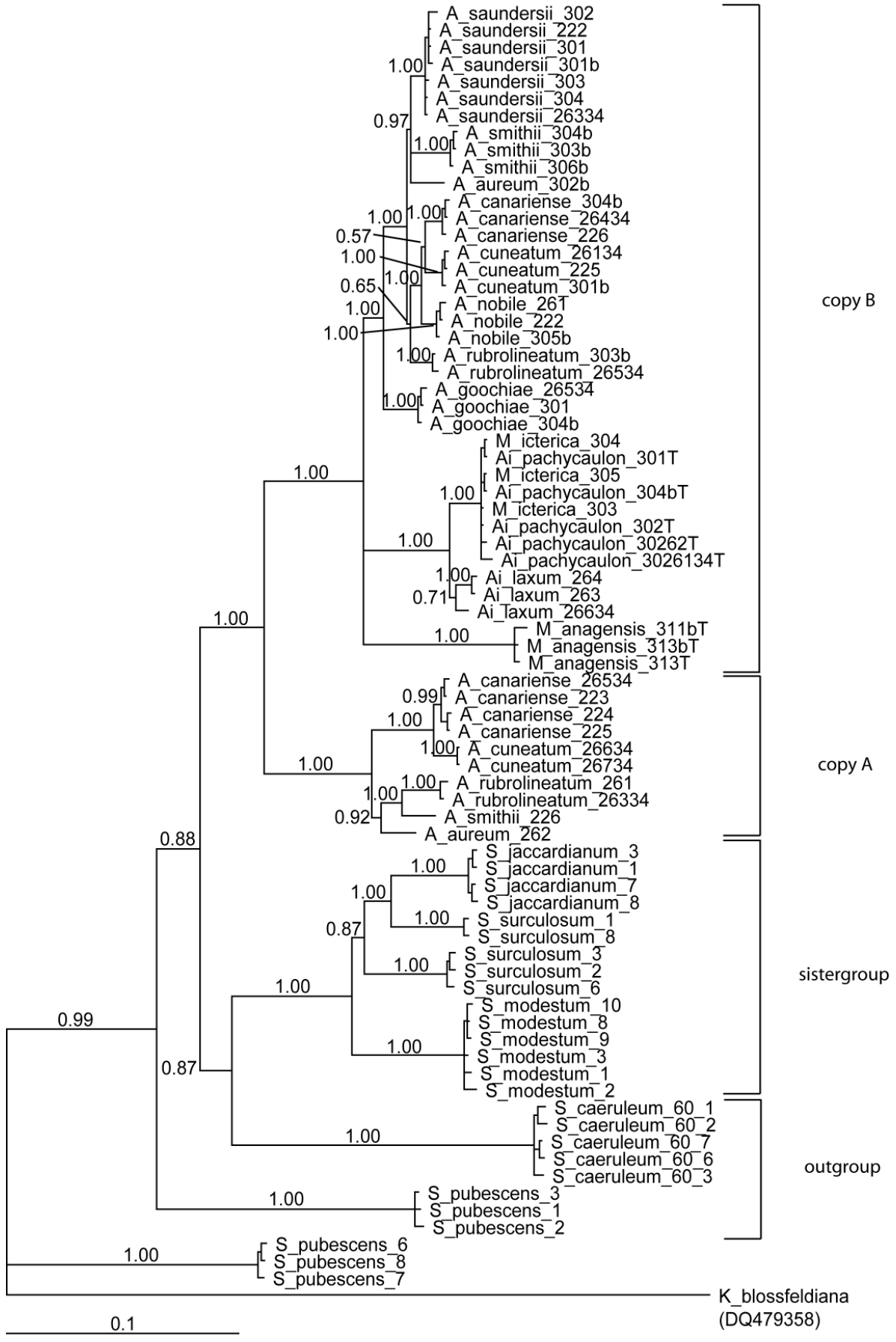


Fig. 39: BI phylogram based on the *MCS_AP3* full-length data. The 3´-UTR-region is excluded. Posterior probabilities are given at the nodes.

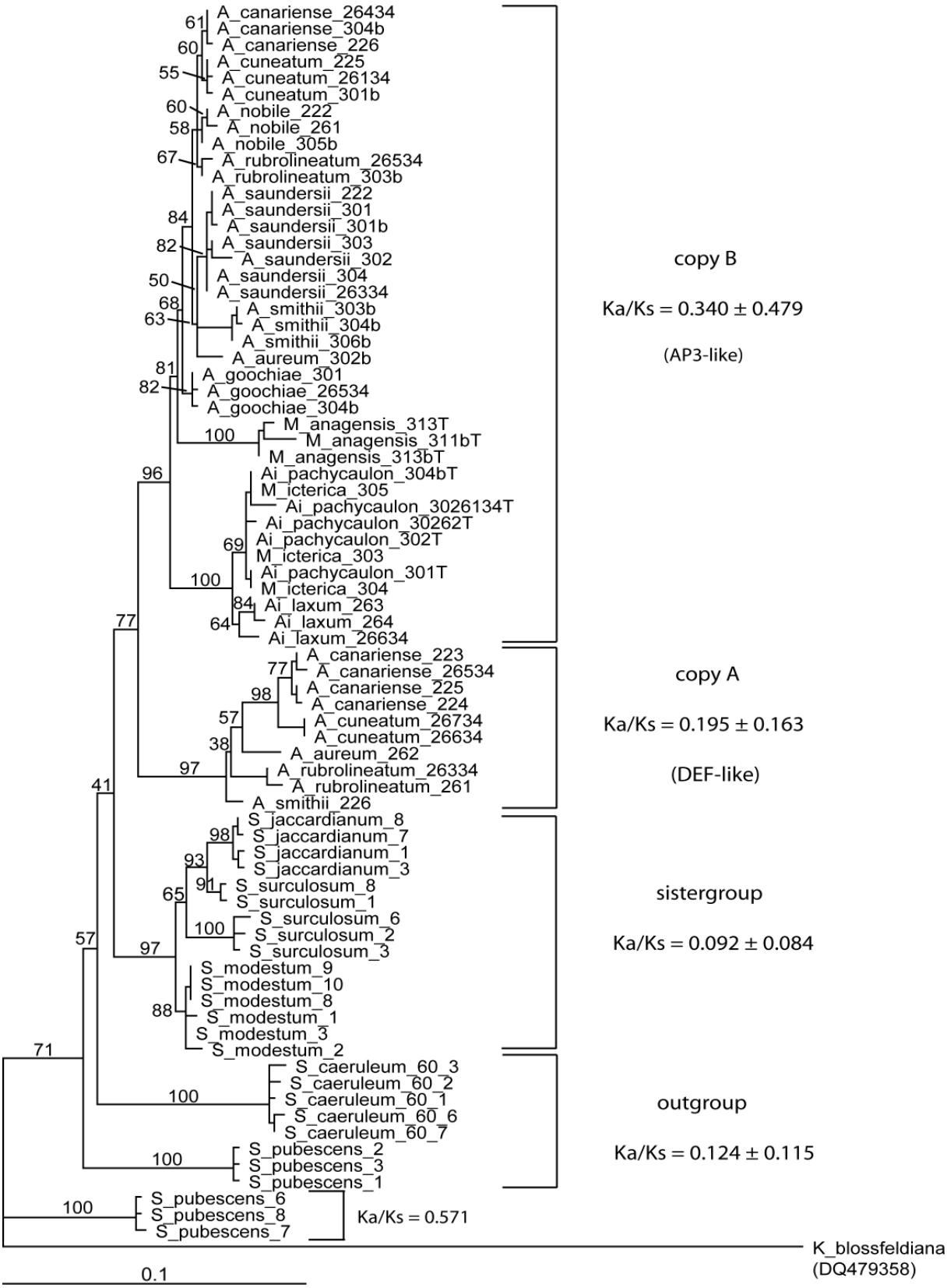


Fig. 40: ML phylogram based on the *MCS_AP3* exon data. Bootstrap support is given at the nodes.

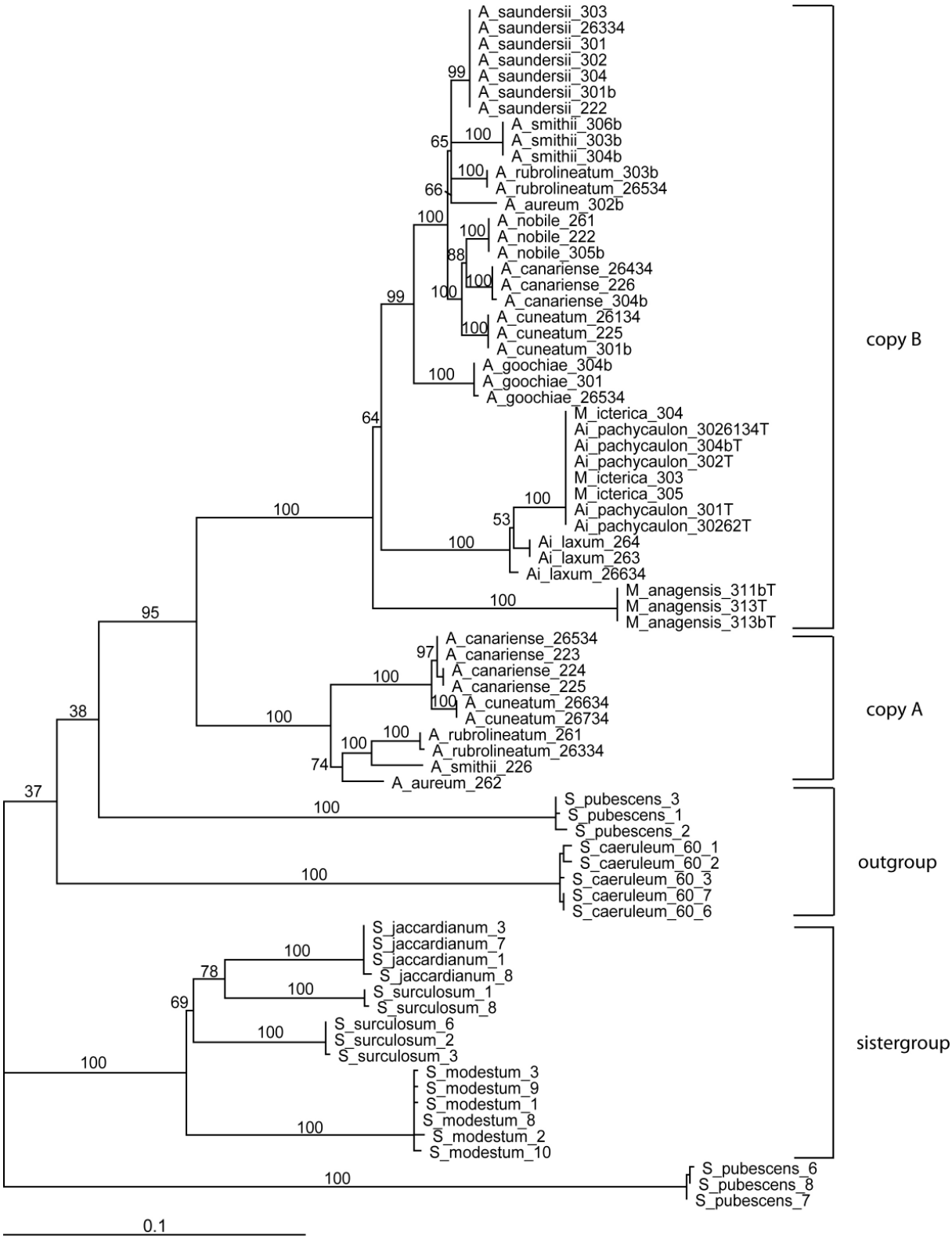


Fig. 41: ML phylogram based on the *MCS_AP3* intron data. Bootstrap support is given at the nodes.

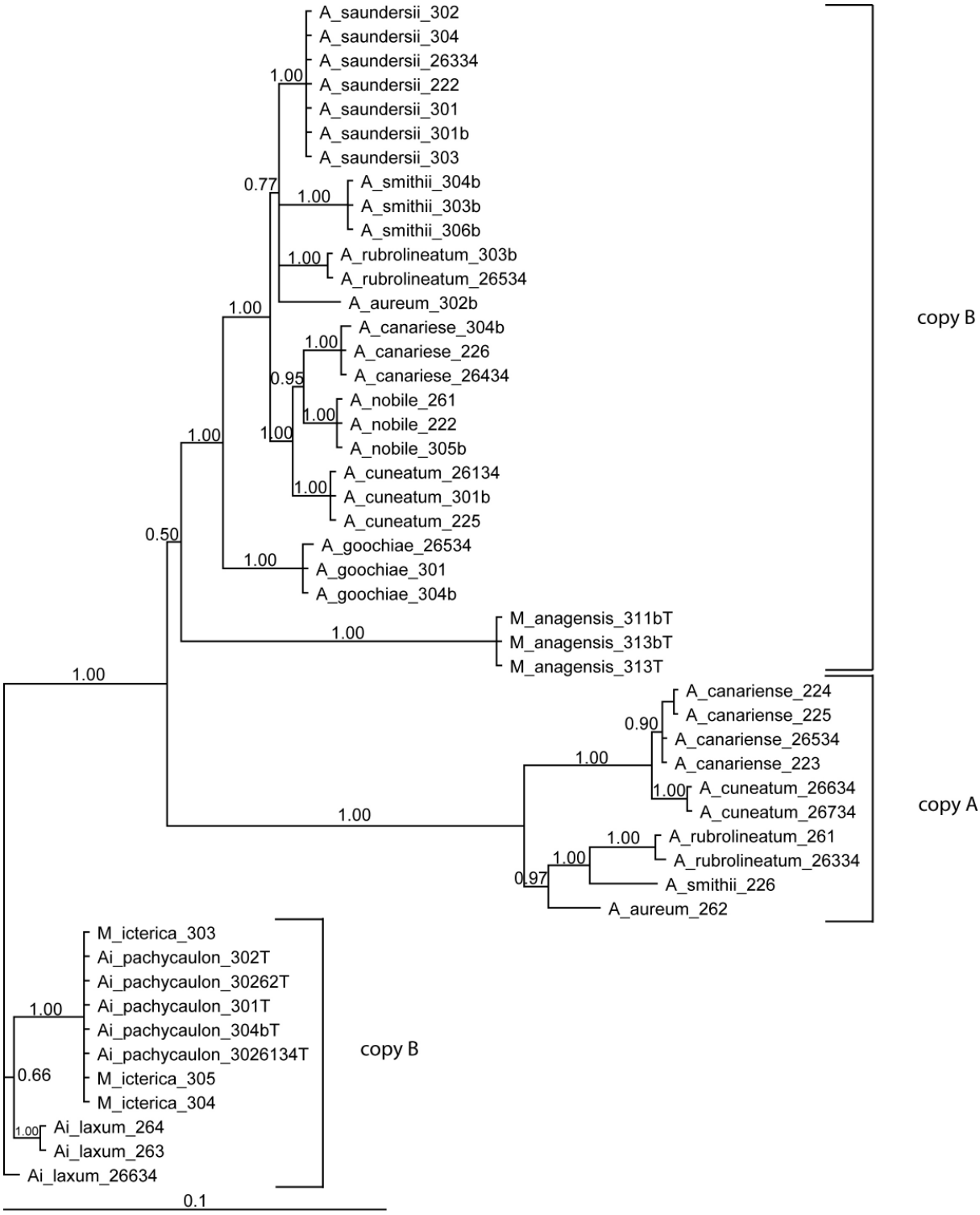


Fig. 42: BI phylogram based on the *MCS_AP3* MCS intron data (*Sedum* sequences were excluded). Posterior probabilities are given at the nodes.

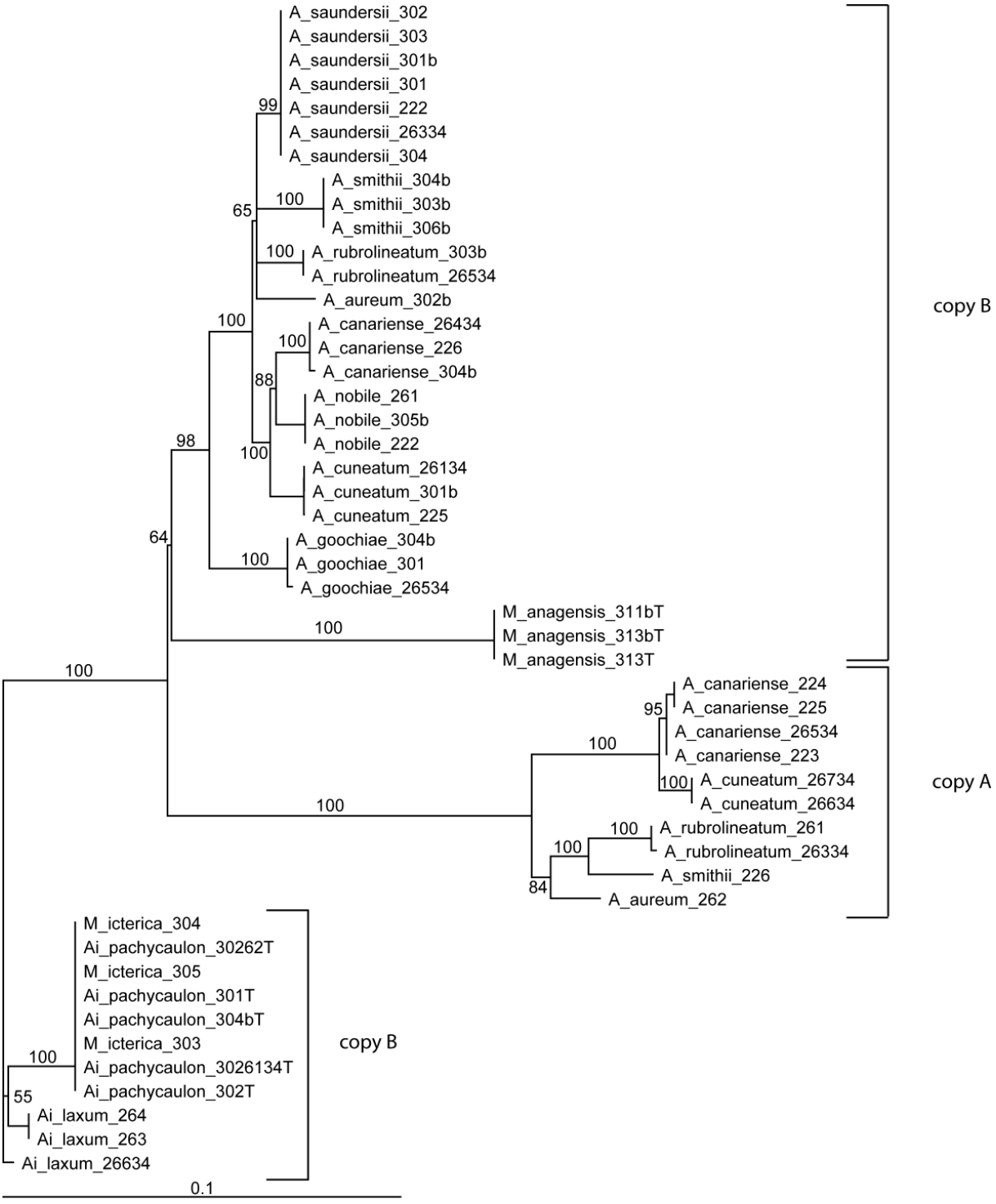


Fig. 43: ML phylogram based on the *MCS_AP3* MCS intron data (*Sedum* sequences were excluded). Bootstrap support is given at the nodes.

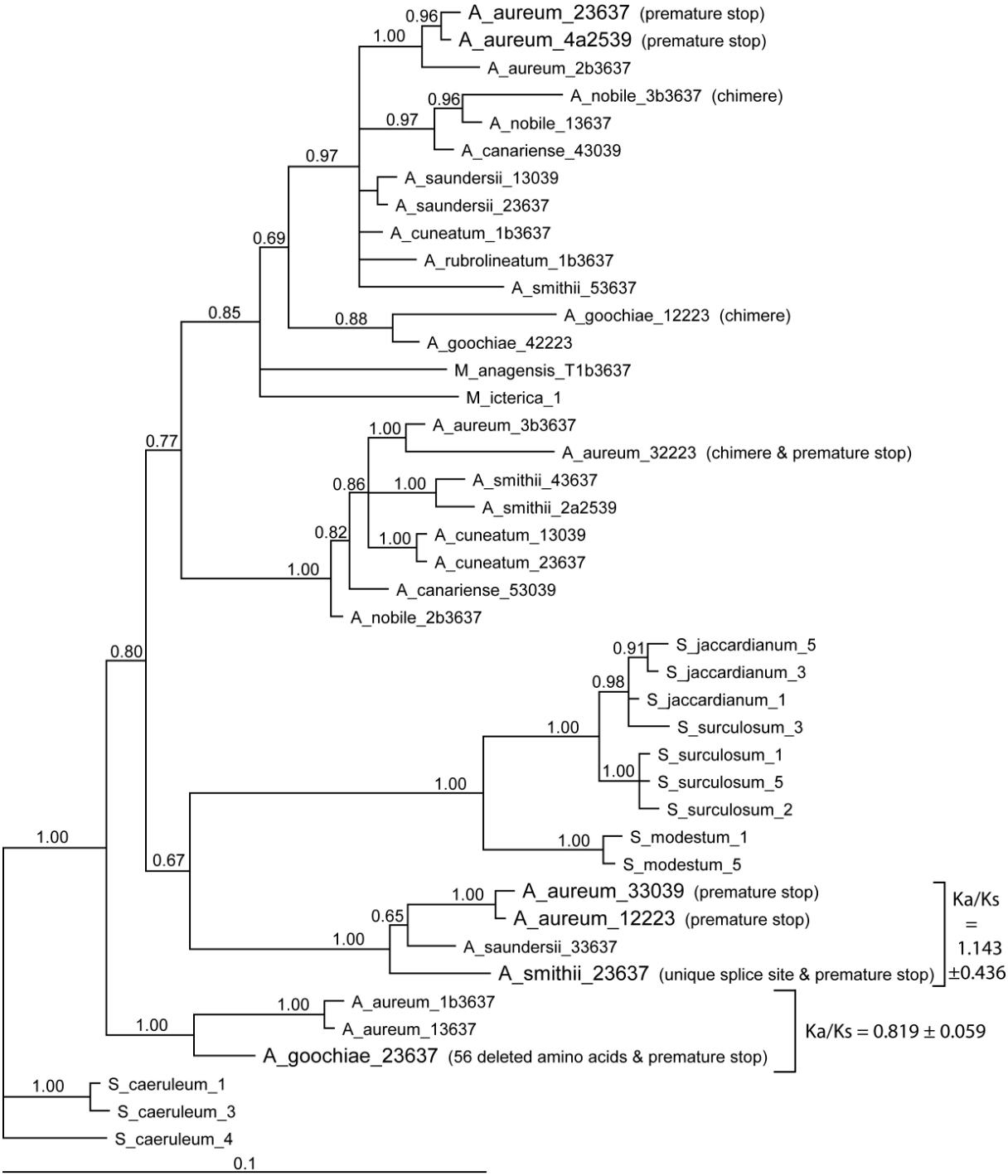


Fig. 44: BI phylogram based on exon data of *MCS_AP1* including all amplified sequences (chimers, sequences with unique splice sites or introns, and with premature stops caused by frameshift mutations). Posterior probabilities are given at the nodes.

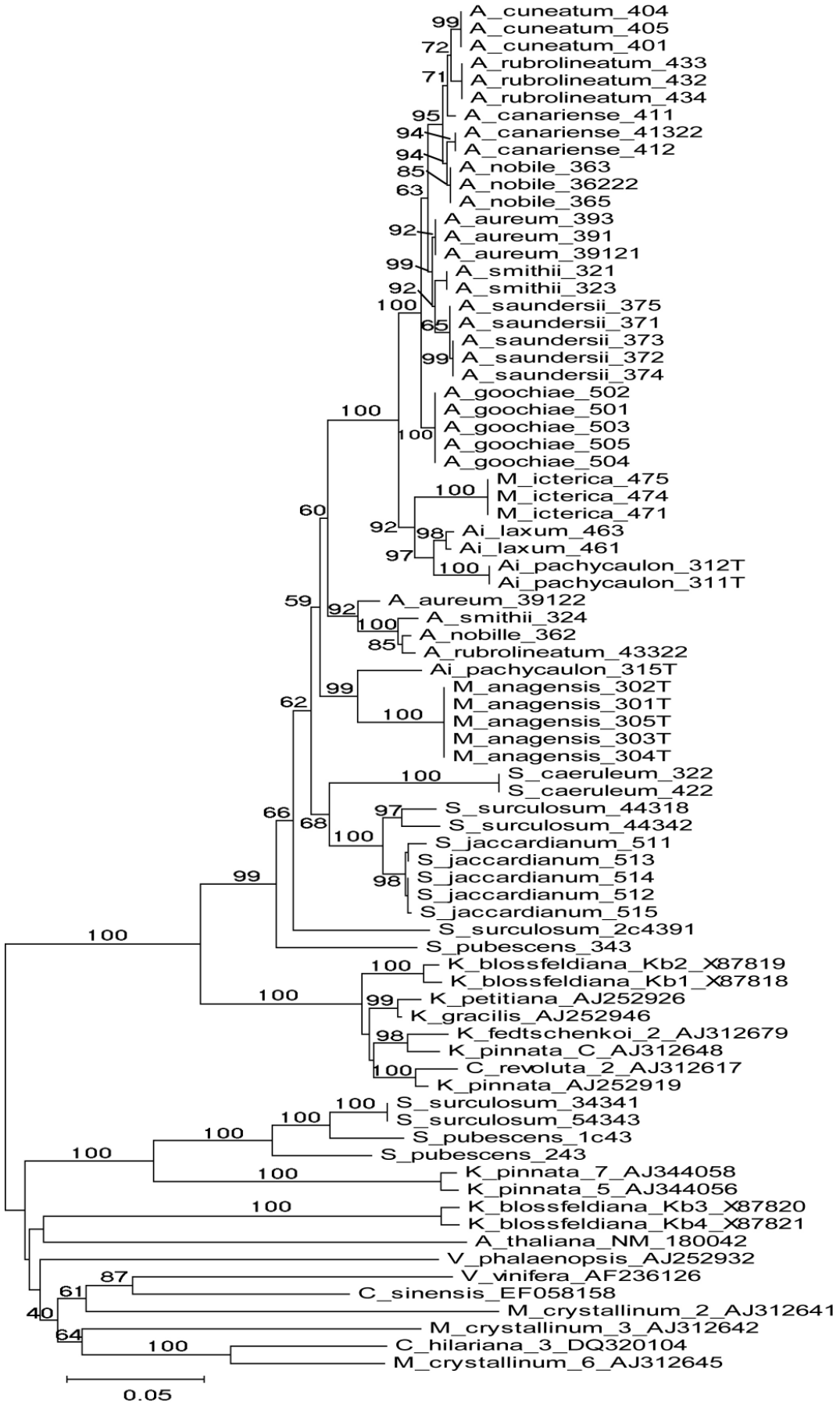


Fig. 45: NJ phylogram of the enlarged *MCS_PEPC* dataset. Bootstrap support is given at the nodes.

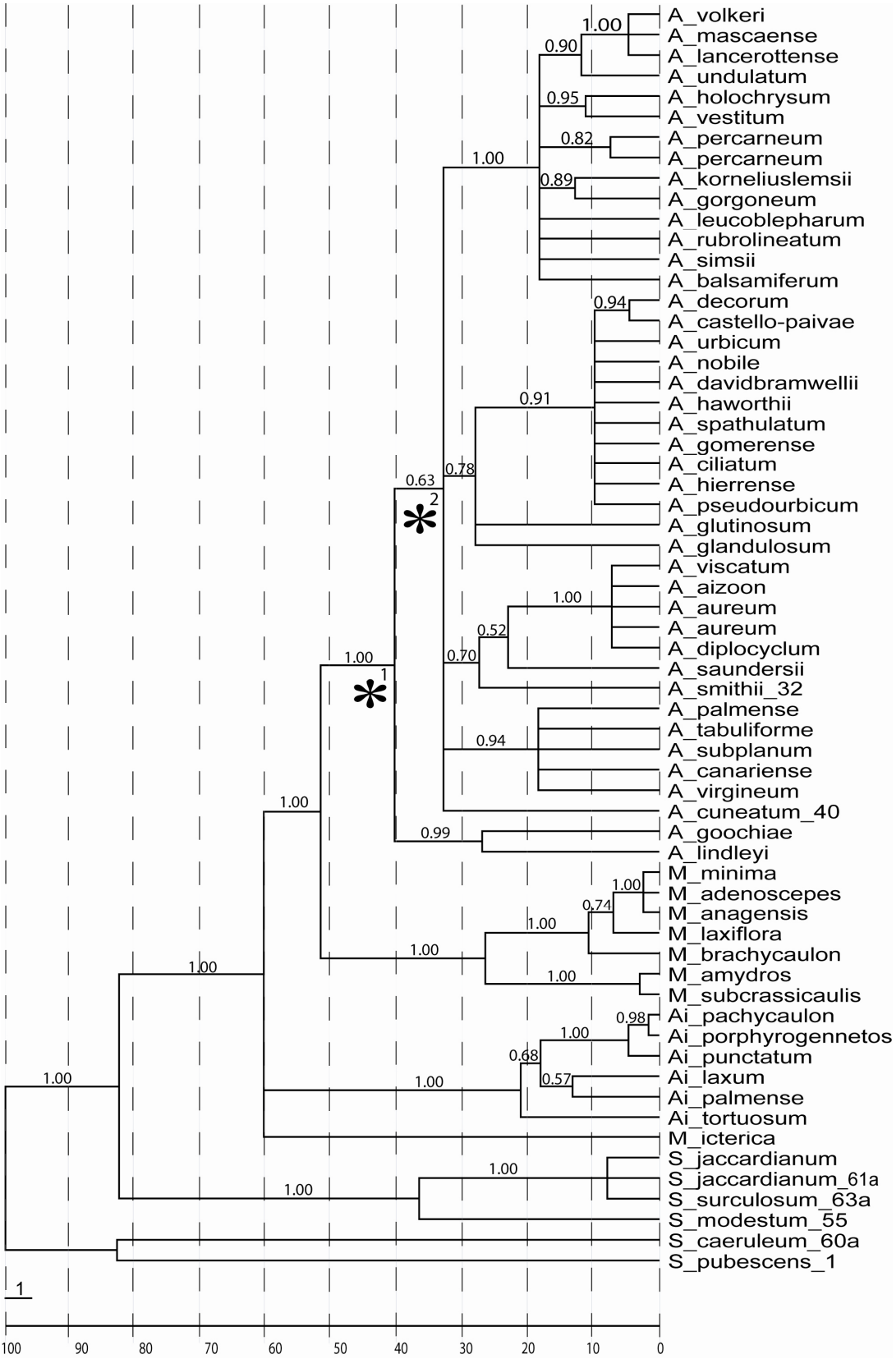


Fig. 46: Ultrametric tree. Posterior probabilities are given at the nodes. Asterisks mark the positions of potential alternative duplication events.

Acknowledgment

Mein erster Dank gilt Herrn Prof. Dr. Marcus Koch, der die Betreuung der Arbeit übernommen und damit die Promotion möglich gemacht hat. Ganz besonders herzlich danke ich ihm für seine Geduld während der Fertigstellung der Arbeit sowie für die Diskussionen und Denkanstösse.

Auch bei Frau Prof. Dr. Claudia Erbar bedanke ich mich sehr herzlich für die unkomplizierte Zusammenarbeit und für die freundliche Übernahme der Zweitkorrektur.

Bei Dr. Mike Thiv bedanke ich mich für die Überlassung des Themas, für die Bereitstellung des Labor- und des Arbeitsplatzes am Naturkundemuseum Stuttgart. Ich danke ihm für zahlreiche Diskussionen und das Korrekturlesen der Arbeit sowie die eingebrachten Verbesserungsvorschläge.

Mein ganz besonders herzlicher Dank gilt Dr. Frank Blattner – ohne ihn wäre die Arbeit nie beendet worden. Ich danke ihm sehr für die vielen hilfreichen und intensiven Diskussionen und Denkanstösse, die er mir unermüdlich während der Entstehung der Arbeit gab. Er war für meine Sorgen und Probleme ansprechbar und hat damit wesentlich zum Gelingen der Arbeit beigetragen. Auch danke ich ihm von ganzem Herzen für das Asyl, dass er mir immer mal wieder in seiner Arbeitsgruppe gegeben hat. Auch das ausdauernde Korrekturlesen war von Nöten wie Willkommen.

Ganz besonders bedanke ich mich auch bei Dr. Zerihun Tadele, meinem „neuen“ Chef, der immer ein Auge zugedrückt hat wenn auf dem Rechner mal wieder eine Analyse lief. Damit und auf vielfältige andere Art und Weise motivierte und unterstützte er mich.

Mein herzlichster Dank gilt meinen Kollegen aus dem Labor in Hohenheim, meinen „Sequenzierknechten“ Timo, Aelys, Philip, Valentin und Chloe in Zürich und der ETX in Gatersleben: Petra und Birgit für eure Hilfe und Unterstützung im Labor während meiner Zeit in Gatersleben; Maia, Nico, Enoch habt Dank für die vielen Diskussionen und das ihr mir MrBayes und PHYML näher gebracht habt, Maia auch lieben Dank für die Unterstützung bei r8s. Inga, Martin, Elke, Dominik und Markus für die Zeit am Museum. Und all meinen Kollegen an der Uni Bern, die mir im letzten Jahr soviel beigebracht haben. Ohne euch alle wäre die vorliegende Arbeit nicht möglich gewesen.

Mein ganz besonderer Dank gilt meinen Freunden, die mich in der Zeit unterstützt, motiviert und immer mal wieder aufgefangen haben: Kerstin, Sabine, Frank, Micha, Timo, Nicole, Beate; Anna, Katrin und Siobhan habt Dank für das Korrekturlesen und das ihr Bern zu einer tollen Stadt für mich macht. Katrin, Matti, Christiane und Thomas vielen lieben Dank für eure Freundschaft und alles was dazu gehört. Dafür sei auch euch, Anke und Jule, von ganzem Herzen gedankt.

Und von ganzem Herzen danke ich auch meiner Familie für die Unterstützung, dass ihr mich aufgefangen und ertragen habt, wenn es mal wieder nötig war! Vielen, vielen lieben Dank für Alles, was ihr für mich tut!!

Selbständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angeführten Quellen und Hilfsmittel angefertigt habe. Wörtlich oder inhaltlich entnommene Stellen benutzter Werke wurden als solche gekennzeichnet.

Ferner erkläre ich, dass ich an keiner anderen Stelle ein Prüfungsverfahren beantragt bzw. die Dissertation in dieser oder anderer Form bereits an einer anderen Fakultät zur Prüfung vorgelegt habe.

Bern, den 27.09.09