History of plant colonization, speciation and genetic diversity in the afro-alpine environment: Evidence from molecular data

TIGIST WONDIMU KIFILEMICHAEL

A Thesis submitted to

The department of Plant Biology and Biodiversity Management

Presented in Fulfillment of the Requirement of the Degree of Doctor of Philosophy in Biology (Botanical Science)

Addis Ababa University

Addis Ababa, Ethiopia

July 2012

Addis Ababa University
School of Graduate Studies

This is to certify that the thesis prepared by Tigist Wondimu Kifilemichael, entitled "*History of plant colonization, speciation and genetic diversity in the afro-alpine environment: Evidence from molecular data"* and submitted in fulfillment of the requirements for the Degree of Doctor of Philosophy complies the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the Examining Committee:

Examiner ____________________ Signature______________Date_____________

Examiner ____________________ Signature______________Date_____________

Advisor Prof. Sileshi Nemomissa__ Signature______________Date______________

Advisor Prof. Christian Brochmann Signature______________Date_____________

ABSTRACT

History of plant colonization, speciation and genetic diversity in the afro-alpine environment: Evidence from molecular data

Tigist Wondimu

Addis Ababa University, 2012

Recent phylogenetic and phylogeographic studies of some afro-alpine plants suggested repeated long-distance migrations of plants into Africa mainly from Eurasia. Such repeated migrations may have played an important role in the plant species composition of the isolated afro-alpine 'sky island' flora. The current study used African *Swertia* spp., *Trifolium cryptopodium*, *Geranium arabicum*, a putative new species of *Geranium*, *Carduus schimperi* and *Umbilicus botryoides* to assess the history of plant colonization and speciation in the high mountains of Ethiopia and East Africa (Kenya, Tanzania and Uganda). The African species of *Swertia* were used in phylogenetic and biogeographic analyses while the remaining species were used to explore phylogeographic history and the level of their genetic diversity across the afro-alpine environment of Eastern Africa. Sequences of two non-coding plastid regions [the *trn*L-F region (both the intron and the intergenic spacer)] and the *trn*S-fM intergenic spacer, one coding plastid region (*mat*K) and the nuclear ribosomal internal transcribed spacer (ITS) were used to investigate the phylogenetic relationships, biogeography and divergence time of African *Swertia*. The African species formed two distantly related clades, each closely related to different Himalayan species of the genus. Thus, this study revealed two episodes of colonization of Africa by different lineages of *Swertia* from the Himalayas. Molecular divergence data analysis of these taxa indicated that the earliest migration of *Swertia* into Africa took place in the

Miocene Epoch. In the phylogeographic studies of the other taxa, two or three primer combinations were selected for AFLP analysis. The phylogeographic analyses of the AFLP data indicated several long-distance migrations among the high mountains in Ethiopia and East Africa. In two species, *Carduus schimperi* and *Trifolium cryptopodium*, the populations from Bale Mountains in Ethiopia were genetically more similar to populations from the East African Mountains than to those from other Ethiopian mountains. The putative new species of *Geranium* (listed as *Geranium sp. nov. = G. S. Miehe 3002* in Flora of Ethiopia and Eritrea as endemic to the Bale Mountains) was not found to be genetically distinct from the widespread and genetically variable *G. arabicum*, suggesting that the characters used to distinguish it merely represent phenotypical plasticity. In contrast, the previous division of *Carduus schimperi* into three subspecies based on morphology was supported by the genetic data, except that the material from the Bale Mountains which previously had been referred subsp. *schimperi* grouped with the East African subsp. *nanus*.

Key words: Afro-alpine flora, AFLP, Biogeography, Divergence time, *In situ* radiation, Phylogeny, Phylogeography, *mat*K, *trn*L-F, *trn*S-fM, ITS.

Acknowledgements

The following institutions are very much acknowledged for providing me the study opportunity and offering to use their facilities: School of Graduate studies in Addis Ababa University; College of Natural Sciences and Plant Biology and Biodiversity Management Program Unit; and University of Oslo, the Natural History Museum. I am also very grateful to the office of Graduate studies, Addis Ababa University for covering my travel costs to Prague that gave me an opportunity to attend an international conference and a workshop.

I would like to express my deepest and sincere gratitude to my supervisors, Professor Sileshi Nemomissa (Addis Ababa University), Professor Christian Brochmann (University of Oslo, Natural History Museum) and Dr. Magnus Popp (University of Oslo, Natural History Museum) without the help of who this Ph. D. would not be possible. They generously shared their wide knowledge to me from which I learned a lot. They have been encouraging and guiding me throughout the study period that provided an excellent basis of knowledge to me.

I wish to express my warm and sincere thanks to Dr. Charlotte Bjorå (University of Oslo, Natural History Museum), who joined my project with warm willing as a supervisor, for her kind and wholehearted follow up during my study period in Oslo. She has contributed a lot to make this study successful by providing me with her own resources to generate additional data. Her understanding, encouraging and personal guidance has made a great input to my study and gave me untiring help during my difficult moments. I am very much thankful to Professor Jan Suda (Charles University, Prague) for offering me the opportunity to attend the international conference and waived me the training fee on techniques of Flow Cytometry.

My deepest gratitude goes to Abel Gizaw who has been persistent throughout the study period providing me support with academics, social and boosting my liveliness in all the difficult times. His priceless advice and support has made me to achieve. I warmly thank him and my other colleagues in Afroalp II, Felly Tussime, Catherine Masao and Ahmed Abdi with whom I successfully carried out the field work. Their ideals and concepts have had a remarkable influence on my study. During this work I have collaborated with many colleagues for whom I have great regard, and wish to extend my warmest thanks to all those who have helped me.

Many thanks to all the field assistants, drivers and all other contributors in all the countries (Ethiopia, Kenya, Tanzania, Uganda and Madagascar) without whom the field work would not be possible. Ato Wege Abebe and Ato Assefa Hailu are among all others who contributed most in assisting me to successfully achieve the field work and preparation of the specimens.

I owe my best sincere gratitude to Dr. Gurja Belay for his timeless advice and encouragement. I warmly thank Dr. Tamirat Bekele for his valuable advice, allowing me to use his resources and warm welcoming me to share his office where I worked the final part of the study contentedly.

My warm gratitude goes to Dr. Rahel G/wold and her family who made my life in Oslo superb. Her extensive discussions around my work and interesting advices have been very helpful to pursue my study with potency. My heartfelt gratitude to Dr. Manuel Pimentel and Dr. Sabrina Mazzonii for their sincere support and priceless advises. Virginia Mirré, Lisbeth Thorbek and Dr. Mika Bendiksby are warmly appreciated for their help in the process of generating and analyzing the data. I am very much grateful to Dr. Zufan Abera, Dr. Kidist Bobosha, W/o Kelemua Debebe,

W/t Shewangizhiw Lema and many other friends, their kind support and guidance have been great value in this study. W/t Etaferahu, secretary of the Plant Biology and Biodiversity Management Program Unit is also highly appreciated for her kind support whenever needed.

I owe my loving thanks to my daughter Mihiret Shimelis, for her patience and tolerance compromising all her needs during my departures. My warmest thanks go to my parents without their encouragement and understanding it would have been impossible for me to finish this work. My special gratitude goes to my mother for her enduring and loving support. She always inspire me for not to give up and hold on. My younger sister Emahoy Tsedalemariam and Emahoy Ametemariam are warmly appreciated for their prayers and support. I also pass my gratitude to Getachew Worku, Cheramlak Wondimu and all other family members who have in one way or another contributed to my success.

This study is part of the project „AFROALP-II - Afro-alpine „sky islands‟: genetic versus taxonomic biodiversity, climate change, and conservation‟ funded by The Norwegian Programme for Development, Research and Higher Education (NUFU project 2007/1058) to S. Nemomissa and C. Brochmann.

Table of Contents

List of Figures

List of Tables

List of Appendix

1. Introduction

The unique plant diversity of the high mountains of Ethiopia and East Africa has fascinated scholars of natural history for a long time (e.g. Fries, 1923; Hedberg, 1951; 1957; 1961; 1965; 1969; 1970; 1995; Hedberg and Hedberg, 1979). These mountains are scattered along the eastern and western arc of the Great Rift Valley in Africa. Simen Mountains and Mt. Choke (in Ethiopia), the Ruwenzori, Mt. Muhavura and Mt. Gahinga (in Uganda), and Mt. Elgon (in Kenya) lie in the Western arc of the Rift, while Bale Mountains and Mt. Garamuleta (in Ethiopia), Mt. Aberdares and Mt. Kenya (in Kenya), and Mt. Kilimanjaro and Mt. Meru (in Tanzania) lie in the Eastern arc of the Rift. These mountains can be characterized biologically as isolated „sky islands in a sea of lowlands‟.

The climate in the afro-alpine environment is harsh with small seasonal variation and large fluctuations on a daily basis, seemingly "summer every day and winter every night" (Hedberg, 1957; 1964). This remarks the intense sunshine during the day time and the nightly frosts all the year round that are experienced in most of the mountains. In most of the mountains the weather condition involves clear weather from sunrise until mid morning; fog and increasing cloud cover in the late morning; rain, sleet, and mist periodically through the afternoon, with clearing in late afternoon or early evening (Smith & Young, 1987). Daytime air temperatures at 1 m above the ground may reach 10 to 15^0 C during clear periods; nighttime air temperatures may reach -5 to -10^0 C just before sunrise on a clear night.

The afro-alpine ecosystem exhibits three major vegetation zones: the lowermost afro-montane forest zone, the mid-altitudinal ericaceous zone, and the uppermost afro-alpine zone (Hedberg, 1951). Different species of trees, for example *Hagenia abyssinica* J.F.Gmel. and *Juniperus procera* Hochst. ex. Endl., are common in the afro-montane forest zone while *Erica* shrubs and trees are dominant in the ericaceous zone (Hedberg, 1951; White, 1983; Wesche *et al*., 2000). The afro-alpine zone is dominated by tussock grasses and sedges, small-leaved shrubs (*Alchemill*a and *Helichrysum*) and the giant rosette plants of *Lobelia* and *Dendrosenecio* (Hedberg, 1951). More than 80% of the plant species in this zone are endemic, with their supposedly closest relatives in temperate alpine areas in other parts of the world, and several of the remaining afro-alpine species also occur in other temperate areas (Hedberg, 1961; 1965).

The responses of afro-alpine flora to the Pleistocene climate oscillations (glacial and interglacial cycles) are little understood in an African context. However, a few recent studies have yielded some useful information (Assefa *et al*., 2007; Ehrich, *et al*., 2007, Kebede *et al*. 2007). Although the current study may not provide the final, complete view concerning the responses of afro-alpine flora to Pleistocene climate oscillations, it will contribute to knowledge of the dynamics of plant biodiversity in terms of genetic structure and gene flow during this period. The time of origin and sources of the current afro-alpine flora have been a point of debate (Hedberg, 1961; 1965; 1969; Koch *et al*., 2006; Ehrich, 2007; Gizaw *et al*, unpublished data). We have used *Swertia* species to shed light on the sources of afro-alpine flora and also to suggest the time of origin of this genus in Africa. Most African *Swertia* are distributed in afro-alpine environment and Nemomissa (1994) has made detailed morphological study of the Northeast African Swertia. Hence, the African *Swertia* are good models to be considered in this study. Moreover, the study included *Carduus schimperi*, *Trifolium cryptopodium*, *Geranium arabicum* subsp. *arabicum*, *Geranium sp. nova*., *Swertia abyssinica* and *Umbilicus botryoides* to investigtate genetic diversity and population structuring to infer their phylogeographic history.

2. Literature Review

2.1 Molecular Phylogeny, Biogeography and Phylogeography

Phylogeny refers to the evolutionary relationship among extant taxa. The focus of phylogenetic studies is on shared derived characters and variation between taxa resulting from evolution. Evolution results in descent with modifications resulting largely from genetic changes caused by mutation (Wiley *et al*. 1991; Kitching *et al*., 1998; Simpson, 2006). As a consequence the accumulation of different mutations, isolated populations may evolve into different independent lineages. Members of each new population are more similar to each other than to members of the other population or to the ancestral population (Wiley *et al.,* 1991). These relationships between taxa can be traced back through shared ancestors, and can be expressed as a tree, which is referred to as a phylogenetic tree (Kitching *et al*., 1998; Simpson, 2006). The study of evolutionary relationships within and between groups of living organisms is called phylogenetics (McNeill, 1979; cited in Stuessy, 2009; Kitching *et al*., 1998).

In phylogenetic studies, the aim is to discover a group which is considered to be natural or monophyletic (Kitching *et al*., 1998; Simpson, 2006). The phylogenetic method attempts to find critical evidence for delimiting taxa as a monophyletic group. This evidence relies upon derived characters which are common to all descendants of a common ancestory. The relatives of an extant species must be determined by close examination of the species under study, particularly with regard to characters that are believed to be heritable (Simpson, 2006: Judd *et al*. 2007). Such evolutionary inherited characteristics are called homologous characters, and it is only these features that clearly show the genealogic relationships of taxa (Kitching *et al*., 1998; Mayr and Bock, 2002; Simpson, 2006; Stuessy, 2009). In contrast to this, related taxa may share similar characters that are not inherited from their common ancestor. These characters are called analogous (homoplastic) and usually result from convergent evolution (Mayr and Bock, 2002; Simpson, 2006; Stuessy, 2009). The latter cannot be used for assessing evolutionary relationships among taxa. Whereas a pre-existing (ancestral) character state changed into a new (derived) character state, the newly derived character state is called an apomorphy while the ancestral character state is termed as plesiomorphy (Kitching *et al*., 1998; Simpson, 2006).

In phylogenetic systematics, an accepted natural group must share a common derived character state among the descendants and with the recent common ancestor. The shared derived character state is known as a synapomorphy (Kitching *et al*., 1998; Simpson, 2006). Through time synapomorphies may become plesiomorphies (shared ancestral character states) as evolution is a continuous process (Kitching *et al*., 1998). Synapomorphies play the greatest role in reconstructing phylogenetic histories of organisms (Kitching *et al*., 1998; Simpson, 2006; Stuessy, 2009).

The construction of a Molecular phylogeny encompasses a series of approaches in which phylogenetic relationships are inferred using information from macromolecules and micromolecules of the organisms under study (Simpson, 2006; Stuessy, 2009). This approach is more reliable in determining whether similar character states are due to homology or homoplasy. The types of molecular data that can be used in such studies are obtained from DNA sequences, allozymes, VNTRs (Variable Number Tandem Repeats; macrosatellite, minisatellite and/or microsatellite DNA), RFLPs

(Restriction Fragment Length Polymorphism), RAPDs (Random Amplified Polymorphic DNA), and AFLPs (Amplified Fragment Length Polymorphism).

Biogeography is the discipline of comparative biology primarily concerned with the history of organisms in geographical space. It involves a range of scientific disciplines that includes geography, geology and biology. Some biogeographers have recognized two research traditions, to which may be attached labels such as ecological biogeography and historical biogeography (Myers and Giller, 1988). Ecological biogeography is concerned with ecological processes occuring over short temporal and small spatial scales, whereas at the other end, historical biogeography is concerned with evolutionary processes over millions of years on a large, often global scale.

Historical biogeography is in the midst of its evolution whereby most of its techniques have been developed in the last two decades (Morrone and Crisci, 1995). This discipline is going through an extraordinary revolution concerning its foundations, basic concepts, methods, and relationships to other disciplines of comparative biology. The approach in modern historical biogeography mainly focuses on special analyses. It deals with formal models of spatial organization, assumes knowledge of spatial arrangements and space-time processes (Gatrell, 1983). Further explanation of these themes was given by Crisci (2001); models of special organization refers to the spatial postdictions and predictions; past and future biogeographical events, respectively. Spatial postdictions deal with the establishment of past biogeography of either a taxon or an area while spatial prediction seeks to model or forecast future biogeographies. The space-time process assumes extinction, duplication (sympatric diversity), vicariance and dispersal as natural processes that modify the diversity of organisms. Spatial arrangement emphasizes the value of geographical distribution in biogeography.

Hence, the whole practice of biogeography is to understand the event-based (temporal) and geographical (spatial) distribution of organisms. The inclusion of temporal information is crucial to assess the historical relationship of organisms. Otherwise there would be no reliable inference about the true relationships of organisms under study (Donoghue and Moore, 2003). Absolute dating in phylogenetic and biogeographic studies enables this to be obtained. Absolute dating or estimation of the divergence date of a group of organisms can be done once a rooted phylogenetic tree with estimated branch lengths has been generated and geological and fossil information is at hand.

Phylogeography deals with the history of spatial and temporal patterns of species evolution, diversification and migration. Phylogeographic studies unravel the factors (of the current and past climatic conditions) that determine the contemporary distribution of species and their genes (Hewitt, 1996; 1999). The term phylogeography was introduced by Avise *et al*. (1987) and is defined as a field of study concerned with the principles and processes governing the geographical distribution of genealogical lineages, especially those operating within and among closely related species (Avise, 2000). In phylogeographic studies, phylogenies are related to patterns of geographical distribution. The approach investigates temporal and spatial histories (origin, diversification and implicitly extinction, and migration) and hence, is mainly concerned with the distributional history of populations, and the subsequent history of floras and faunas in different areas. Methodologically, phylogeography combines elements of population genetics, phylogeny and historical biogeography. Such studies are very efficient for investigating the historical assembly of particular taxa (e.g. Riddle *et al*, 2000). Phylogeographic studies have become increasingly important in taxonomic research as well as in the field of conservation biology. In most cases, molecular data employed in phylogeographic studies include DNA sequences, VNTRs and AFLPs.

2.2 DNA sequencing and AFLPs: methods and applications

It is obvious that closely related organisms show a higher degree of concurrence in the molecular structure of DNA than those organisms that are distantly related and can provide robust information. Thus, they are very useful in the study of relationships between organisms at higher levels of the hierarchy (Hamby and Zimmer, 1992). The use of molecular techniques in plant systematics started with protein-based techniques and proceeded to DNA-based techniques. In this section of the Thesis, a brief overview of DNA-based techniques used for this study will be provided.

2.2.1 DNA sequencing

DNA sequencing methods have become key tools in the study of the histories and relationships of organisms. Sequences of DNA are expected to accumulate mutations over time. Molecular phylogenetics uses such data to build a "relationship tree" that shows the probable evolution of various organisms. The most common approach is the comparison of sequences for homogenous genes using sequence alignment techniques to identify similarity. DNA sequence data basically refer to the sequence of nucleotides in a particular region of the DNA of a given taxon. Comparisons of homologous regions of DNA among the taxa under study yield the characters and character-states that are used to infer relationships in phylogenetic analyses.

DNA sequencing has become very useful in investigating the pattern of relationships of plants at different levels. The application of this method in plant systematic studies has provided useful evidence during the last two decades. Therefore, several novelties have been communicated within this short period of time. For example, DNA sequences have been used to investigate evolutionary relationships within the green algae (Wilcox *et al*. 1992); bryophytes (Hedderson *et al*., 1996; Capesius and Stech, 1997); ferns (Kranz and Huss, 1996); conifers (Stefanović *et al*., 1998); and angiosperms (APG III, 2009).

DNA sequencing is considered a superior method for examining evolutionary relationships since the actions of evolution are ultimately reflected in those sequences. At present it is still a long and expensive process to sequence the entire DNA of an organism (its genome), and this has been done for only a few species. However it is quite feasible to determine the sequence of a defined area of a particular chromosome.

Methods for generating and analysing DNA sequence data

The first step in DNA sequencing is to identify a particular region of DNA to be compared between species.

A. Polymerase chain reaction (PCR)

After a gene sequence of interest is identified, the DNA from a given plant sample is isolated and purified by various chemical procedures. Following this, the sequences of interest are amplified using the polymerase chain reaction (PCR). The invention of this technology was essential to modern DNA sequencing, as it permits rapid and efficient DNA amplification by replicating thousands of copies of DNA (Barlett and Stirling, 2003). The polymerase chain reaction works as follows. Prior research establishes the occurrence of relatively short regions of DNA that flank (occur at each end of) the gene or DNA sequence of interest, and which are both unique (not occurring elsewhere in the genome) and conserved (i.e., invariable) in all taxa to be investigated. These short, conserved, flanking regions are used as a template for the synthesis of multiple, complementary copies, known as primers (Cheng *et al*., 1994). Primers ideally are constructed such that they do not bind with one another.

In the polymerase chain reaction, a solution is prepared, made up of 1) the isolated and purified DNA of a sample; 2) multiple copies of primers; 3) free nucleotides; 4) DNA polymerase molecules (typically TAQ polymerase, which can tolerate heat); and 5) buffer and salts. This solution is heated to a point at which the sample DNA denatures, whereby the two strands of DNA separate from one another. Once the sample DNA denatures, the primers in the solution may bind with the corresponding, complementary DNA of the sample. Following binding of the primer to the sample DNA, individual nucleotides in the solution attach to the 3“ end of the primer, with the sample DNA acting as a template; the DNA polymerase catalyzes this reaction. A second primer, at the opposite end of the DNA sequence of importance, is used for the complementary, denatured DNA strand. Thus, the two denatured strands of DNA are replicated. After replication, the solution is cooled to allow for annealing of the replicated DNA with the complementary DNA single strands. This is followed by heating to the point that the DNA denatures to repeat the process. A typical PCR can produce more than a million copies of DNA in a matter of hours (Fig. 1). PCR products are subjected to a 0.1 % agarose gel medium to test whether the DNA has amplified or not (Fig. 2).

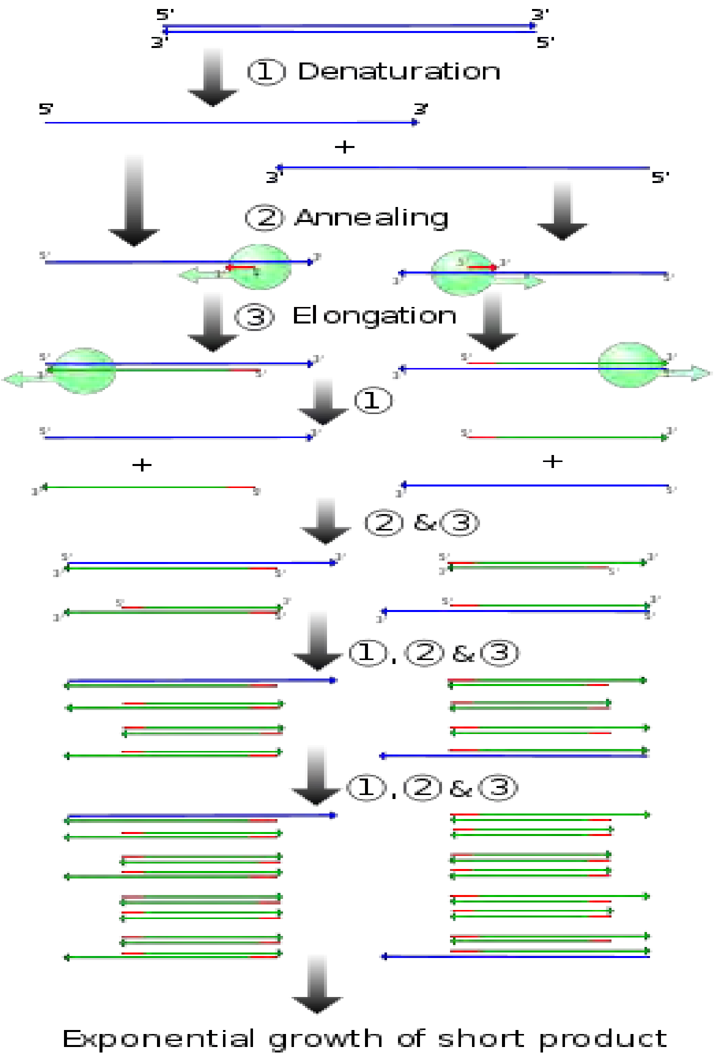


Fig. 1 Schematic drawing of the PCR cycle: (1) Denaturing at 94–96 C. (2) Annealing at 65 C (3) Elongation at 72 C. Four cycles are shown here. The blue lines represent the DNA template to which primers (red arrows) anneal that are extended by the DNA polymerase (light green circles), to give shorter DNA products (green lines), which themselves are used as templates as PCR progresses. (http://en.wikipedia.org/wiki/File:PCR.svg, 2008)

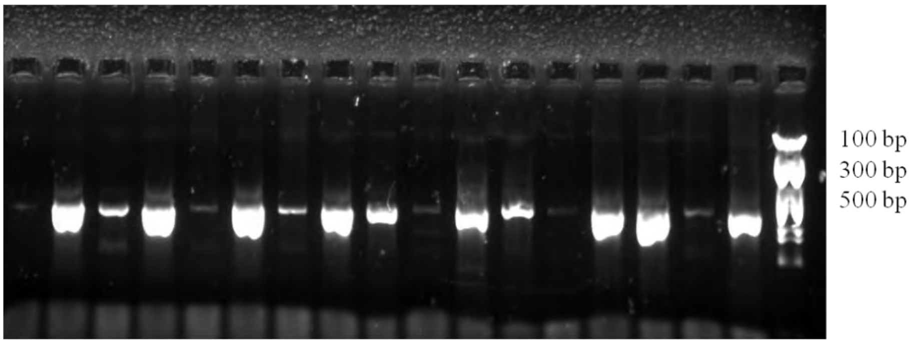


Fig. 2 Electrophoretic gel picture showing amplified plastid DNA (*trn*S-fM region), for 17 samples of African *Swertia*. DNA Ladder at the end of the gel (right hand side)

B. DNA sequencing reaction

Following the development of dye-based sequencing methods with automated analysis, DNA sequencing has become easier. The sequencing reaction follows the PCR by which the DNA is replicated.

Though DNA sequencing may be done in several ways, the method developed by Sanger *et al.* (1977) became the initial method of choice. However the method utilized radioisotopes and required four lanes to run one sample. Later, Smith *et al.* (1986) developed a set of four different fluorescent dyes that allowed all four reactions to be separated in a single lane. This led to the innovation of „automated sequencing‟ rather than „manual sequencing‟ as was done by Sanger‟s method. The automated sequencing method allows the exposure of the whole gel-slab to a photographic plate for a fixed time and post-analysis detection (Griffin and Griffin, 1993; Adams *et al.* 1996). Hence, the sequencing method advanced with more sophisticated separation procedures and enabled 96 samples to be handled at a time. The most commonly used machine reads fluorescent dyes with a laser detector.

The modern procedure for DNA sequencing is as follows. A purified DNA polymerase enzyme is used to synthesize DNA chains of varying lengths. The production of dye-labelled DNA is very similar to DNA replication using the PCR. The replicated DNA is placed into solution with DNA polymerase, primers, free nucleotides, and a small concentration of synthesized compounds called dideoxynucleotides that are each attached to a different type of fluorescent dye. As in the polymerase chain reaction, the sample DNA is heated until the double helix unwinds and the two complementary DNA chains separate. At this point, a primer attaches to a conserved region of one of the strands of DNA, and free nucleotides in solution join to the 3‟ end of the primer, using the sample DNA as a template and catalyzed by the DNA polymerase. Thus, a replicated copy of the DNA strand begins to form. However, at some point a dideoxynucleotide joins to the new strand instead of a nucleotide doing so. The dideoxynucleotides resemble the four nucleotides, except that they lack a hydroxyl group. Once a dideoxynucleotide is joined to the chain, absence of the hydroxyl group prevents the DNA polymerase from joining it to anything else. Thus, with the addition of a dideoxynucleotide, synthesis of the new DNA strand terminates.

The ratio of dideoxynucleotides to nucleotides in the reaction mixture is carefully set and is such that the concentration of dideoxynucleotides is always much smaller than that of normal nucleotides. Thus, the dideoxynucleotides may terminate the new DNA strand at any point along the gene being replicated. For example, some of the new DNA strands will be the length of the primer plus one additional base (in this case the dideoxynucleotide); some will be the primer length plus two bases (a nucleotide and the terminal dideoxynucleotide); some will be the primer length plus three bases (two nucleotides and the terminal dideoxynucleotide); etc. There are many thousands, if not millions, of copies of the sample DNA. Thus, there will be an equivalent number of newly replicated DNA strands, of all different lengths.

The final step of DNA sequencing entails subjecting the DNA strands to electrophoresis, in which the DNA is loaded onto a flat gel plate or in a thin capillary subjected to an electric current. Because the phosphate components of nucleic acids give DNA a net negative charge, the molecules are attracted to the positive pole. The DNA strands migrate through the medium over time, the amount of migration inversely proportional to the molecular weight of the strand (i.e., lighter strands migrate further). Each strand is terminated with a dideoxynucleotide to which a flurescent dye is attached; each of the four dideoxynucleotides has a different type of fluorescent dye, which (upon excitation) emits light of a different that is detected and recorded as a peak, which measures the light intensity. Because a given emitted wavelength (“color”) is determined by one of the four dideoxynucleotides, the corresponding nucleotide can be inferred and its position identified by the timing of migration of the DNA strands. In this way, the sequence of nucleotides of the DNA strand can be inferred (Fig. 3).

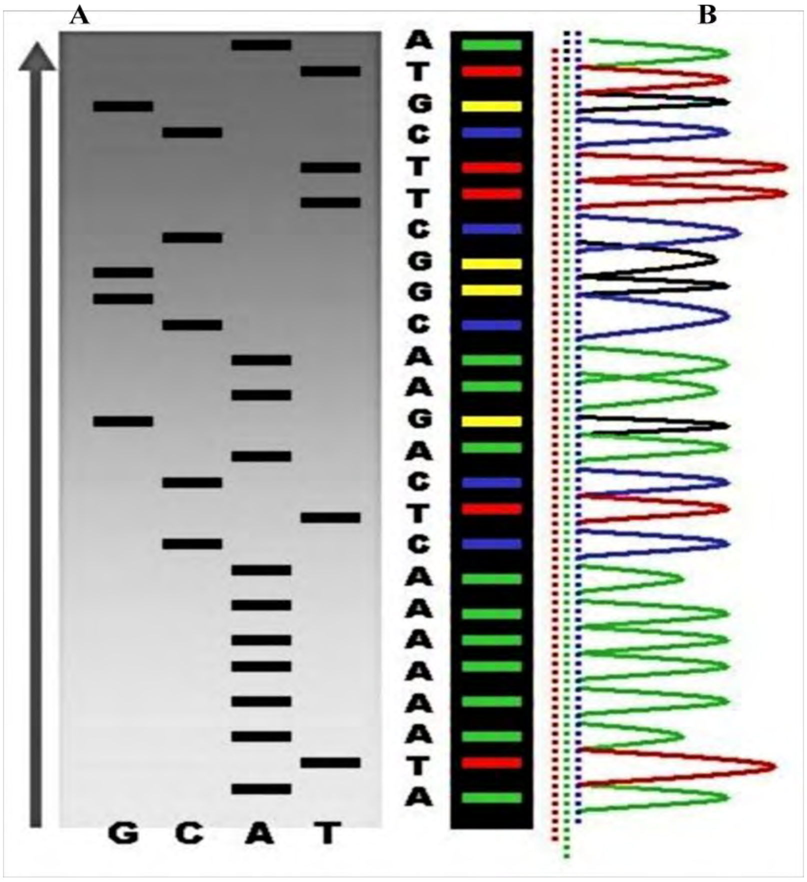


Fig. 3 Sequencing and visualising sequences. A) Manual (radioactive) sequencing, B) Automated (Fluorescent) sequencing. (Abizar at Wikipedia, http://en.wikipedia.org/wiki/DNA_sequencing, 2008)

A more recently and an improved technology of the DNA sequencing method is known as the **Next-generation DNA sequencing**. This technique has the potential to dramatically accelerate biological and biomedical research, by enabling the comprehensive analysis of genomes, transcriptomes and interactomes to become inexpensive, routine and widespread, rather than requiring significant production-scale efforts (Schendure and Hanlee, 2008). Plant systematics also benefits from this development of DNA sequencing that includes better resolved phylogenies, potential for dealing with the difficulties posed by polyploidy and hybridisation, and new options for studying species boundaries and species relationships in recent radiations (Harison *et al*., 2011).

One of the next-generation sequencing techniques is the single-nucleotide addition (SNA) method of DNA sequencing called pyrosequencing (Hyman, 1988) or 454 sequencing (after the name of the Roche-owned company that developed it). In pyrosequencing, the number of individual nucleotides is limited to the point at which DNA synthesis pauses. Then, unlike with the Sanger method, chain elongation can be resumed with the addition of nucleotides. A tiny amount of visible light is generated by enzymatic action as each nucleotide is added to a growing chain; this light is recorded as a series of peaks called a pyrogram, which corresponds to the order of lettered nucleotides that are added and ultimately reveals the underlying DNA sequence (Metzker, 2005). Thus, by correlating when a sample flashes with the nucleotide that is present at that time, researchers can sequence a stretch of DNA. 454 sequencing can read up to 20 million bases per run by applying the pyrosequencing technique on picotiter plates that facilitate sequencing of large amounts of DNA at low cost compared to earlier methods (Adams, 2008).

C. Analysis of DNA sequence data

DNA sequence data are converted to characters and character-states to be used in phylogenetic analyses. First, the sequences of a given length of DNA are aligned, in which homologous nucleotide positions are arranged in corresponding columns (Fig. 4). For some genes that are relatively conserved, alignement is straightforward, as all taxa have the same number of nucleotides per gene. For other genes or DNA segments, some taxa may have one or more additions, deletions, inversions, or translocations relative to other taxa. The occurrence of these mutations, and/or the occurrence of considerable homoplasy among taxa, can make alignment of DNA sequences difficult. In addition, multiple copies of a gene can make homology assessment difficult. Various computer algorithms can be used to automatically align sequences of the taxa being studied, but these have assumptions that must be carefully assessed.

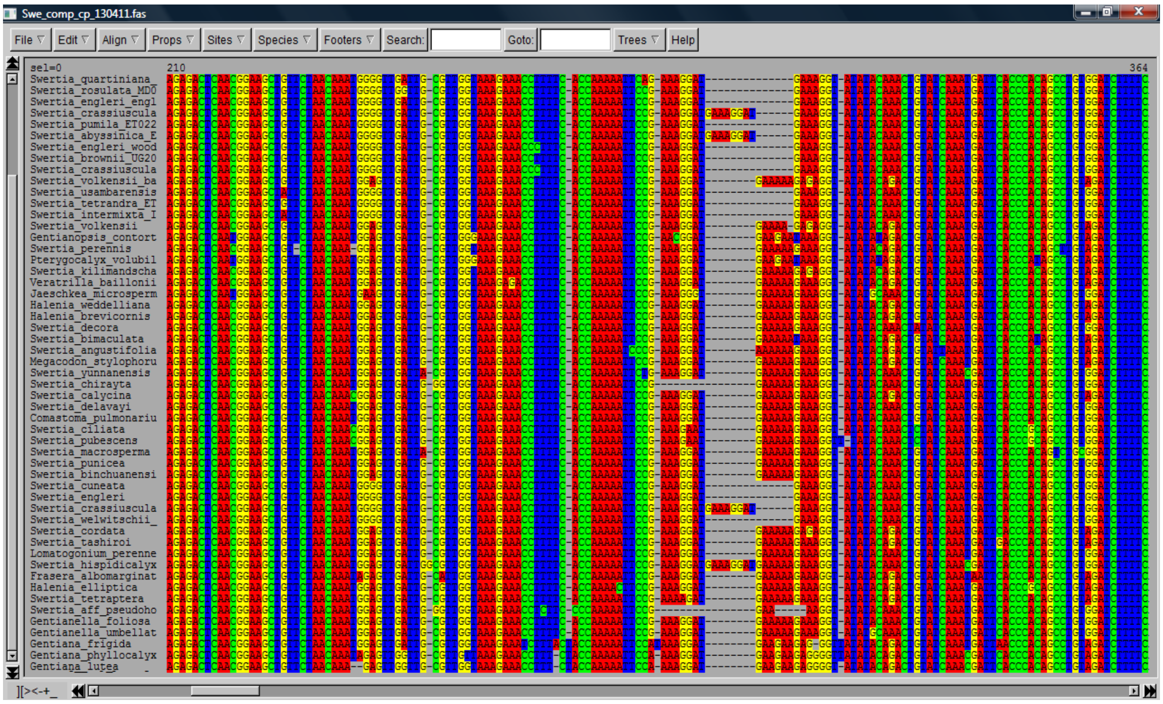


Fig. 4 Screen print image of aligned plastid DNA (*trn*L-F) sequences of 59 and 3 species from the subtribe Swertiinae and Gentianinae, respectively.

Generally, in using DNA sequence data in a phylogenetic analysis, a character is equivalent to the nucleotide position, and a character state of that character is the specific nucleotide at that position (there being four possible character-states, corresponding to the four nucleotides). A large number of nucleotide positions are generally invariable among taxa, and some of the variable ones are often uninformative by being autapomorphic for a given taxon; thus, relatively few sites are informative and therefore useful in phylogenetic reconstruction.

However, a major addition, deletion, inversion, or translocation can in itself be identified as an evolutionary novelty (apomorphy), used in grouping lineages together. For example, members of the Faboideae (of the Fabaceae) lack, by deletion, one of the inverted repeats found in the chloroplast of most angiosperms. Major mutations such as these may be coded separately from single base differences and may be given a relatively greater weight in inferring relationships.

Several types of weighting schemes may be used with molecular data. For protein encoding genes, the codon position may be differentially weighted. For example, because of redundancy of the genetic code, the third codon position is generally more liable (a change more likely to have occurred randomly) than the second, and the second may be more liable than the first. Thus, the first and second codon positions may be given relatively greater weight, respectively (such as a weight of 10 for the first codon position, 5 for the second position, and 1 for the third position). The logic here is that a change in codon position 1 or 2 is less likely to have occurred at random within a taxon and more likely represents evolutionary novelties that are shared among taxa. Weighting by codon position may be based on empirical data. For a given data set, the number of changes occurring for codon positions 1, 2, and 3 may be used (inversely) to establish the relative weights.

Another weighting parameter that may be used with DNA sequence data concerns transitions versus transversions. Transitions are evolutionary changes from one purine to another purine (A to G or G to A) or from one pyrimidine to another pyrimidine (C to T or T to C). Transversions are evolutionary changes from a purine to a pyrimidine (A to C, A to T, G to C, or G to T) or from a pyrimidine to a purine (C to A, C to G, T to A, or T to G). Weighting using transitions versus transversions may be based on empirical data. For a given data set, the relative frequency of transitions versus transversions may be used (inversely) to establish the relative weights. For example, for a given group under study, if transitions occur 5x more frequently than transversions, the latter may be given a weight of 5 and the former a weight of 1. On the other hand, some characters could have evolved convergently (homoplasy). In order to get a better supported tree, it is rational to give less weight to such characters during reconstruction of a phylogenetic tree (Goloboff *et al*., 2008).

DNA sequence data can also be used to evaluate the secondary structure of a molecule. Thus, nucleotide differences that result in major changes in the conformation of the product (whether ribosomal RNA or protein) may have a much greater physiological effect than those that do not and might receive a higher weight. Computer algorithms can evaluate this to some degree. Parsimony, maximum likelihood, and Bayesian methods are commonly used to infer phylogenetic relationships using DNA sequence data. The most robust hypotheses of relationships have arisen from studies using sequence data from multiple, ideally unlinked, genes.

Types of DNA sequence data

For plants, the three basic types of DNA sequence data stem from the three major sources of DNA: nuclear (nDNA), chloroplast (cpDNA), and mitochondrial (mtDNA). Nuclear DNA is transmitted from parent(s) to offspring by nuclear division (meiosis or mitosis) via sexual or asexual (somatic) reproduction. Chloroplasts and mitochondria, however, replicate and divide independently of the nucleus and may be transmitted to offspring in a different fashion. For example, in angiosperms these organelles are (with some exceptions) sexually transmitted only maternally, being retained in the ovule but excluded in pollen cells. In conifers, interestingly, chloroplast DNA is transmitted paternally (Powell *et al*., 1995; Du et al., 2009).

The use of sequence data from the DNA of chloroplasts have proven to be highly useful in elucidating both lower and higher level relationships. Like all organelle and prokaryotic DNA, it is circular. Curiously, most angiosperms have a region of chloroplast DNA known as the inverted repeat (Kolodner and Tewari, 1979; Graham and Olmstead, 2000), which is the mirror image of the corresponding region.

Nuclear DNA sequencing has been used to a lesser degree in plant systematics. One of the more useful types of nuclear DNA sequences has been the internal transcribed spacer (ITS) region, which occurs as multiple copies (as opposed to single copies found in many protein-coding genes). The ITS region lies between the 18S and 26S of the nuclear ribosomal DNA (nrDNA). ITS sequence data have been most valuable for inferring phylogenetic relationships at a lower level, e.g., between closely related species. However, it has also been used for elucidating higher level relationships (Baldwin, 1992).

2.2.2 AFLP (Amplified Fragment Length Polymorphism)

AFLP-PCR was originally described by Zabeau and Vos (1993). The method is based on the detection of restriction fragments by PCR amplification, and therefore combines restriction digestion and PCR (Vos *et al*., 1995). AFLPs are fragments of DNA that have been amplified using directed primers from restriction digested genomic DNA.

AFLP markers are advantageous in several aspects. They have good taxonomic applicability in that AFLP markers can be generated for any organism with DNA while no prior knowledge about the genomic makeup is needed (Mueller and Wolfenbarger, 1999). The production of a large number of polymorphisms allows researchers to apply the AFLP technique to study intraspecific genetic diversity and phylogenies among closely related species. Since AFLP markers are distributed across the genome they have a high multiplex ratio (Myburg, *et al*., 2001) that allows analyzing a large number of different genetic loci simultaneously. AFLP data are equally applicable to most species using a minimal amount of template DNA and are more reproducible than are RAPDs (Vos *et al.* 1995; Mueller and Wolfenbarger, 1999). Thus, they can be used to identify genetic differences among individuals using high number of DNA loci.

In spite of the above mentioned advantages of using AFLPs, some limitations are encountered in this technique. AFLP markers are dominant, with polymorphisms detected as either band presence or absence. Dominant markers are not as efficient as co-dominant markers for population genetic studies (Mueller and Wolfenbarger, 1999). In some cases too many fragments may be generated making it difficult to distinguish them on an electrophoretic gel. However, a slight modification of the primers used may limit the number of fragments that are amplified, enabling them to be more easily identified.

Several papers have been published on the phylogeography and population genetics of organisms using AFLPs (e.g. Gaudeul *et al*., 2000; Schönswetter *et al*., 2002; Tribsch *et al*., 2002; Campbell *et al*., 2003; Wang *et al*., 2003; Skrede *et al*., 2006; Piñeiro *et al*., 2007; Kebede *et al*., 2007, Misra *et al*., 2010, etc.). A number of researches have also employed AFLP in phylogenetic studies of closely related species and even, in some cases, for higher-level relationships (Kardouls *et al*., 1998; Despres *et al*., 2003; Brout *et al*., 2004; Tremetsberger, *et al*., 2006, etc.).

Methods for generating and analysing AFLPs data

Generating AFLP data can be achieved in four successive steps: 1) Digestion of total cellular DNA with two restriction enzymes and ligation of the restriction site specific adaptors to all restriction fragments (RL: Restriction-Ligation). Normally a four base restriction enzyme, *Mse*I, is used as a frequent cutter and a six base restriction enzyme, *EcoR*I, as a rare cutter. The restriction digestion is followed by ligation of the fragments by specific double stranded synthetic oligonucleotide adapters at each restriction site. The adapters are designed in such a way to prevent restoration of the initial restriction site, and hence allow the possibility of undertaking both reactions (the restriction and ligation) simultaneously (Janssen *et al*., 1996).

2) Pre-selective amplification of the fragments (PS: Pre-selective PCR). This reaction involves oligonucleotide primers that are complementary to the adapters and the restriction sites. The 5‟ end of the primers is complementary to the adapters while the 3‟ end is extended in to the restriction site by few arbitrarily selected nucleotides. This allows the primer to amplify only a subset of fragments.

3) Selective amplification of some of these fragments with two PCR primer combinations that have corresponding adaptor and restriction site specific sequences (SL: Selective PCR). Selective amplification is carried out using the oligonucleotide primers used in PS but with two extra nucleotides. Note that different taxa respond in different way for a particular primer. Therefore, successive primer test must be performed prior to the actual AFLP run.

4) Electrophoretic separation of amplicons on a gel matrix, followed by visualisation of the band pattern. In this step the subset of fragments are separated by denaturing polyacrylamide gel electrophoresis or capillary gel electrophoresis to generate a fingerprint of DNA bands (Weising *et al*., 2005). Radioisotopes, fluorescent dyes or silver staining can be used to detect polymorphic alleles. However, loading fluorescent primers as of three, each labeled with different colored dye, into the same gel lane maximizes the number of polymorphic loci ensuring high reliability of data gathering. The schematic summary of the steps are presented in Fig. 5.

Once the data are generated in the lab, the results will be visualized for analysis using specialized softwares such as Genemapper and Genograph. In the AFLP profile, presence of a fluorescence peak corresponds to the presence of an amplified restriction fragment, and absence of this peak corresponds to the absence of an amplified restriction fragment. A presence (1) and absence (0) matrix is produced for further analysis. Fig. 6 presents a screen print image of AFLP data profiles.

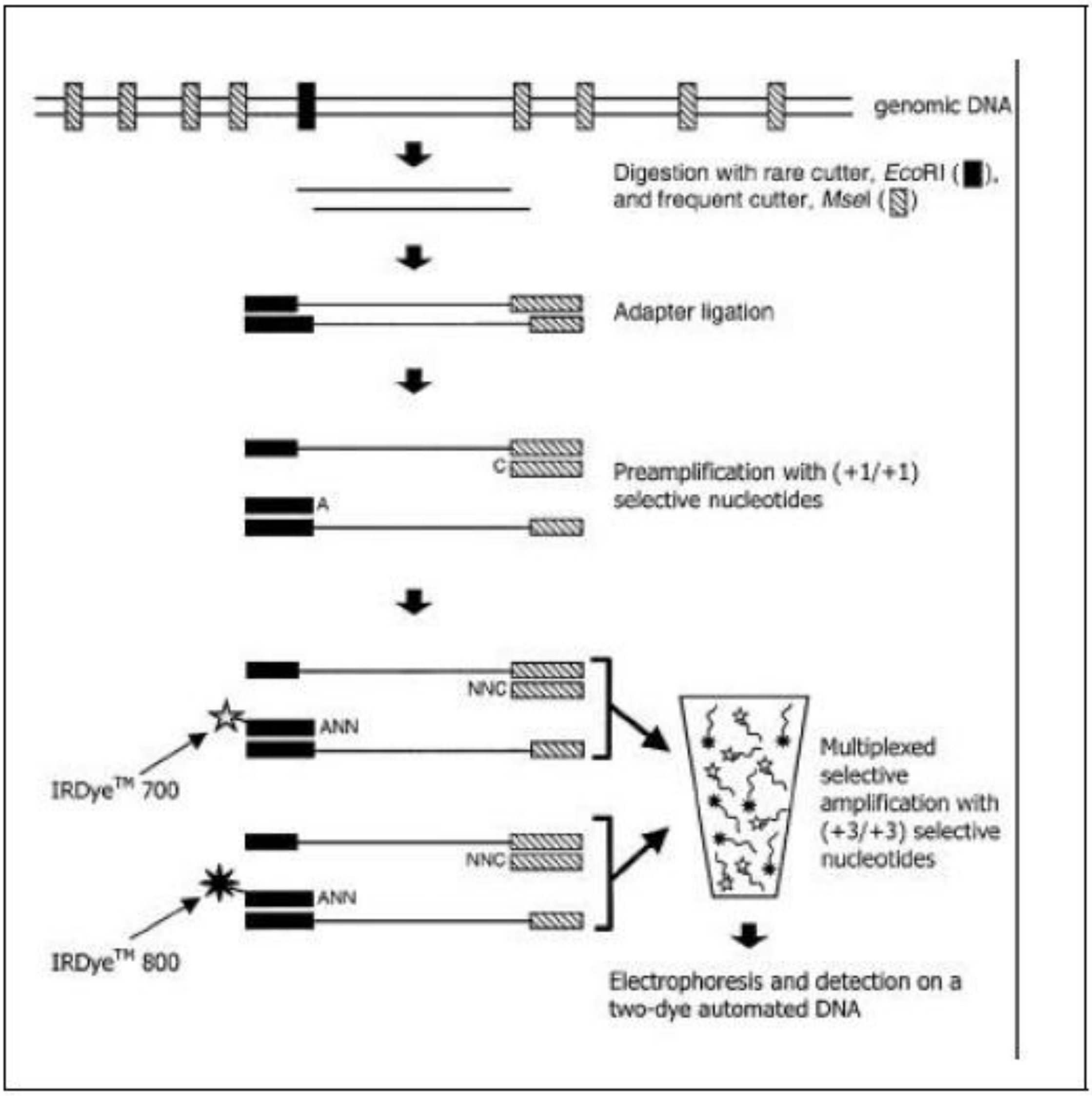


Fig. 5 Summary of AFLP data generation in the lab.: adopted from Myburg, *et al.* (2001)

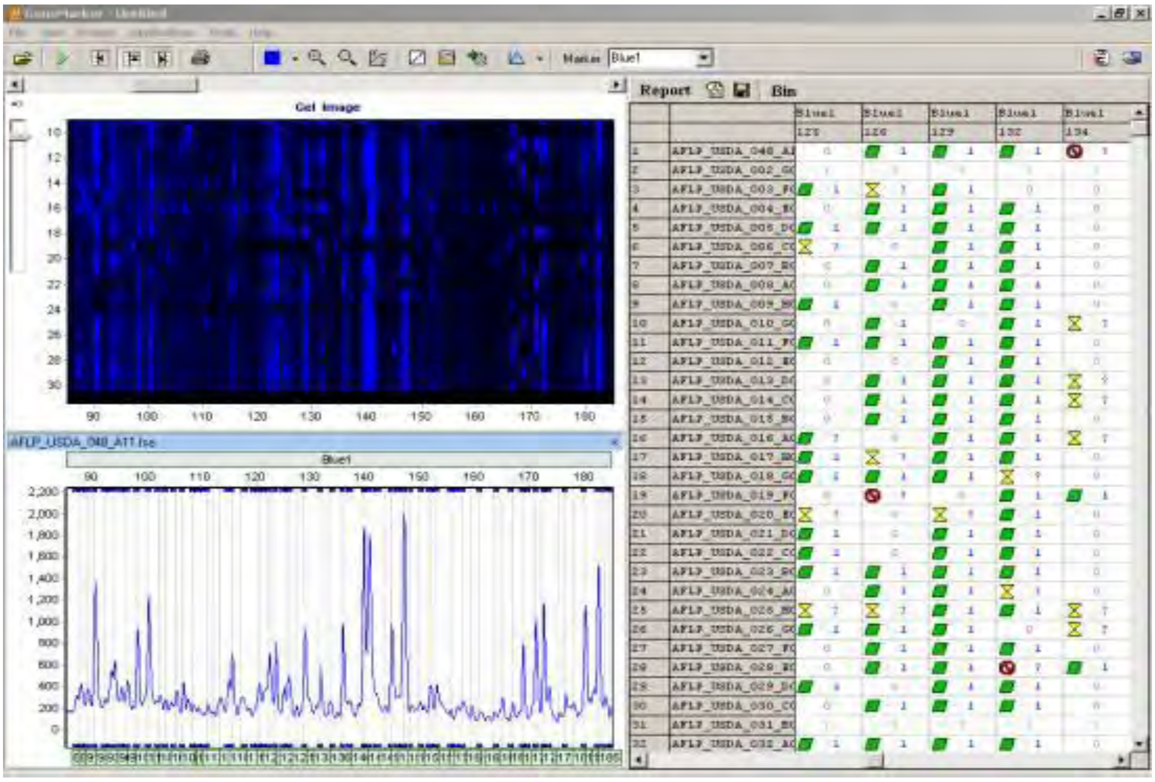


Fig. 6 Screen print image showing data visualization of AFLP data set.

3 Study background and Justification

3.1 Phylogenetic and biogeographic study

Recent phylogenetic and phylogeographic studies of some afro-alpine plants, e.g. *Arabis, Ranunculus, Carex, Alchemilla* and *Lychnis,* suggested that they migrated to Africa mainly from Eurasia, in some cases even from Asia 4000-5000 km away (Koch *et al*., 2006; Assefa *et al*., 2007; Popp *et al*., 2008; Gehrke & Linder, 2009). Interestingly, there is evidence for multiple colonization of Africa in several plant groups. Thus, repeated long-distance dispersal may have played an important role for the species composition of the isolated afro-alpine „sky island" flora (Assefa *et al*., 2007, Gehrke & Linder, 2009), as also reported for the flora of isolated arctic islands (Alsos *et al*., 2007).

A total of 24 taxa of the African *Swertia* L. were used to study the source of origin of these taxa and time of immigration into Africa. Fries (1923) classified the African *Swertia* (Gentianaceae) into two informal groups mainly based on the number of foveae (nectaries) as the Binectariate and the Uninectariate groups (Table 1). Hedberg"s (1957) classification recognized three informal groups, the Crassiuscula, Kilimandscharica and Volkensii groups based on seed and pollen characters, the number and nature of the margins of foveae, chromosome number and duration of the life cycle. More exhaustive sampling was done in northeast Africa and Tropical East Africa by Nemomissa (1994). He used 34 characters in a phenetic analysis, revealing two distinct groups of species best distinguished by four morphological characters (number of nectaries per corolla lobe, seed surface, seed shape and surface of the exine). Group 1 has corolla lobes with two nectaries, ovoid seeds with smooth surface and spinulose exines, whereas Group 2 has one nectary per corolla lobe, ellipsoid seeds with ridge-like outgrowths and striato-reticulate exines (Table 1, see also Fig. 7, 8).

Group 1 includes the Crassiuscula group of Hedberg (1957) and five of the species from the Binectariate group of Fries (1923). Group 2 combines the Kilimandscharica and Volkensii groups of Hedberg (1957). This group also includes all species with one nectary per corolla lobe from the Uninectariate and the Binectariate groups of Fries (1923). Four of the species from the Binectariate group of Fries (1923) were proved to possess one nectary at the base of their corolla lobes by Nemomissa (1994). See Table 1.

Nemomissa (1994) also compared African *Swertia* with Asian *Swertia* based on gross morphological characters and suggested a close relationship between them. He hypothesized, for example, that the African *S. kilimandscharica* is more closely related to the Himalayan *S. pseudohookeri* H. Smith than to any other African *Swertia*, that the African *S. macrosepala* is most closely related to the Himalayan *S. multicaulis* D. Don, and that the African *S. crassiuscula* is most closely related to the Himalayan *S. alternifolia* Royle.

Table 1 Informal grouping of African *Swertia* according to Fries (1923), Hedberg (1957) and Nemomissa (1994)

Fries (1923)		Hedberg (1957)			Nemomissa (1994)	
Binectariate	Uninectariate	Crassiuscula	Kilimandscharica	Volkensii	Group 1	Group 2
S. brownii	*S. schimperi*				*S. brownii*	*S. schimperi*
S. crassiuscula	*S. kilimandscharica*	*S. crassiuscula*	*S. kilimandscharica*		*S. crassiuscula*	*S. kilimandscharica*
S. engleri	*S. scandens*				*S. engleri*	*S. scandens*
S. pumila					*S. pumila*	*S. adolfi-friderici*
S. quartiniana					*S. quartiniana*	*S. squamigera*
S. macrosepala				*S. macrosepala*	*S. abyssinica*	*S. macrosepala*
S. subnivalis				*S. subnivalis*	*S. eminii*	*S. subnivalis*
S. uniflora				*S. uniflora*	*S. intermixta*	*S. uniflora*
S. volkensii				*S. volkensii*	*S. schliebenii*	*S. volkensii*
					S. scotii	
					S. tetrandra	
					S. usambarensis	
					S. welwitschii	

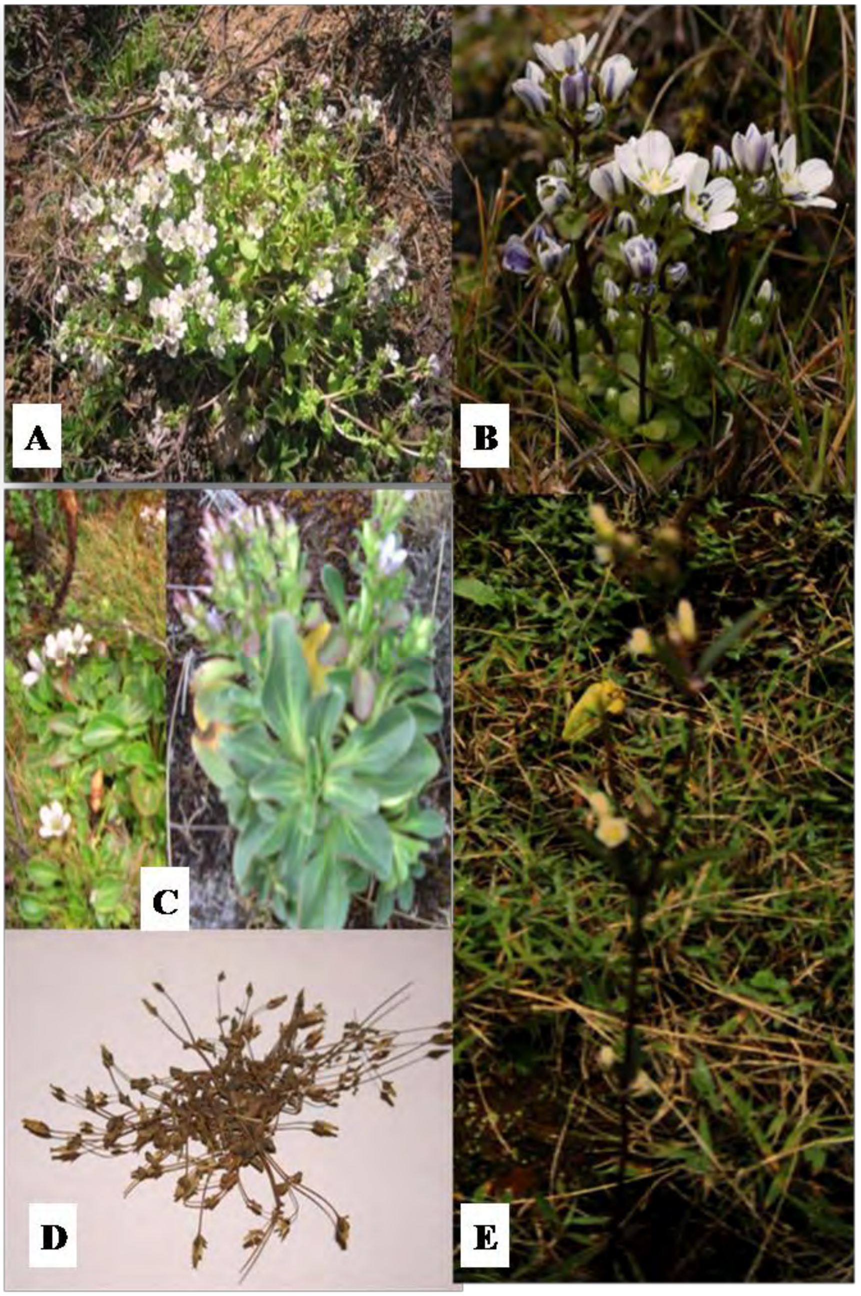


Fig. 7 Flowers and fruits of some *Swertia* species that grouped in the informal Group 1. A) & B) *S. abyssinica*, C) *S. crassiuscula* subsp. *crassiuscula*, D) *S.* aff. *abyssinica*, E) *S. rosulata*

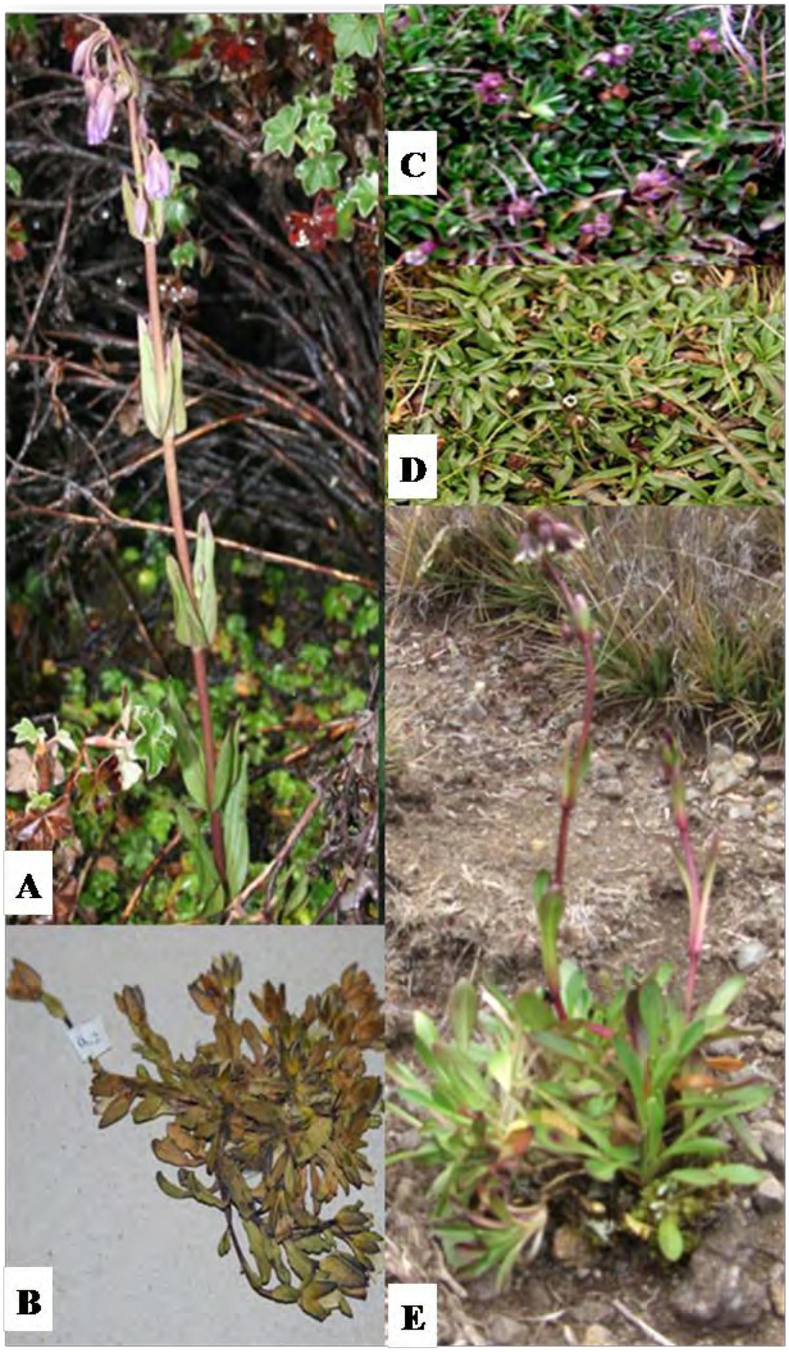


Fig. 8 Flowers and fruits of some *Swertia* species that grouped in the informal Group 2. A) *S. kilimandscharica*, B) *S. volkensii* var. *baleensis*, C) *S. subnivalis*, D) S. macrosepala subsp. microsperma E) *S. volkensii* var. *volkensii*

A phylogenetic study of the *Gentianella* lineage based on chloroplast and nuclear (ITS) DNA sequences indicated that the genus *Swertia* is highly paraphyletic (Chassot *et al*., 2001). This study only included four African species of *Swertia*, and these were placed into two distantly related clades. This result supports Nemomissa‟s (1994) morphology-based hypothesis of more than one migration of *Swertia* into Africa, and the two clades corresponded to his informal grouping of the African species. A recent study of pollen morphology in subtribe Swertiinae supported a similar subdivision of African *Swertia* as well as suggesting close relationships between African and Asian *Swertia* (Chassot & von Hagen, 2008).

However, the sampling in these previous studies represented only a small fraction of the diversity of African *Swertia*. Here we assess their phylogeny and biogeography and re-address earlier hypotheses based on chloroplast and nuclear DNA variation among 21 of the 25 African species, most of which were collected during our recent field trips. We selected the same two plastid DNA regions and one nrDNA region used by Chassot *et al*. (2001) to be able to include the 39 plastid DNA and nrDNA sequences of Gielly and Taberlet (1996) and Chassot *et al*. (2001) in extended phylogenetic analyses (Table 3).

Little is known about the time of plant colonization of the afro-alpine environment. Among the number of phylogenetic and phylogeographic studies of the afro-alpine plants, very few have dealt with time of colonization (Koch *et al*., 2006; Gizaw *et al*., unpublished). Divergence dating estimated a post-Gondwana divergence of the Gentianaceae about 50 myr ago (Yuan *et al*., 2003) and the subtribe Swertiinae at about 15 myr ago (Yuan *et al*, 2005; Favre *et al*., 2010). In this study, we used three plastid DNA (*trn*L-F, *trn*S-fM, & *mat*K) regions and one nrDNA (ITS) region to estimate date of divergence of the African *Swertia*. We sequenced a protein coding region (*mat*K) and sampled six additional sequences that belong to the lineage in which the fossil evidence is found. These additional sequences were used to reconstruct the phylogeny of tribes and subtribes in the family Gentianaceae by Struwe *et al*. (2002).

Therefore, here, we will address 1) whether the phylogenetic relationships inferred from the molecular data are consistent with the morphology-based grouping of the African taxa; 2) How many times did *Swertia* immigrate into Africa?; 3) Whether different Himalayan species form sister-group relationships with African clades and 4) when did the African *Swertia* diverge from their closest relatives?

3.2 Comparative Phylogeography and investigation of genetic diversity

Several phylogeographic studies have provided impressive molecular evidence about the history and evolution of plants, for example the arctic flora (Tremblay and Schoen, 1999; Abbott, *et al.*, 2000; Bronken *et al.*, 2001; Hewitt, 2004; Alsos *et al.*, 2005). These studies have shown how past climatic changes determined the current geographic structuring of the genetic diversity in plants in the Arctic. Other studies of arctic-alpine plants have described the spatial and temporal patterns of species evolution, diversification and migration in relation to past climatic changes (Koch, 2002; Stehlik, *et al.*, 2002; Stehlik, 2003; Reisch *et al.*, 2003; Brochmann *et al.*, 2003; Abbott and Brochmann, 2003; Comes and Kadereit, 2003; Schönswetter, 2002, 2003; Tribsch and Schönswetter, 2003).

Despite the increasing number of plant phylogeographic studies, there are few such studies that have explored the evolution, diversification and genetic structure of afro-alpine plants (Koch *et al.*, 2006; Assefa *et al.*, 2007, Ehrich *et al.*, 2007, Kebede *et al.*, 2007). The research presented in this thesis includes phylogeographic analyses of six taxa, i.e., *Carduus schimperi* Sch.Bip. (Asteraceae), *Trifolium cryptopodium* Steud.ex A. Rich. (Fabaceae), *Geranium arabicum* Forssk. (Geraniaceae), *Geranium sp. nov. = G. & S*. Miehe 3002, *Swertia abyssinica* Hochst. and *Umbilicus botryoides* Hoscht. ex A. Rich.(Crassulaceae). These taxa are from the afro-alpine environment and the analyses have involved an assessment of their genetic structure and diversity across their geographical ranges based on AFLP data. Taxa that have different evolutionary histories and are unrelated are good models for comparative phylogeographic studies (Morris *et al*., 2005). Apart from the two *Geranium* species, the taxa chosen for this study are all taxonomically distantly related.

3.3 Objective(s) of the study

General objective(s)

- To infer the history of plant colonization and diversification in the afro-alpine environment
- To explore genetic diversity in the afro-alpine environment

Specific objective(s)

- To reconstruct the phylogenetic tree of African species of *Swertia*
- To infer the biogeographic history of African species of *Swertia*
- To estimate the divergence date of African species of *Swertia*
- To assess the comparative phylogeography of some selected species from the afro-alpine environments; i.e. *Carduus schimperi* and *Trifolium cryptopodium*
- To investigate the genetic relationships and diversity of *Geranium arabicum* and *Geranium sp. nov.= G. & S. Miehe 3002* and assess the taxonomic validity of the new species
- To study genetic diversity among populations of *S. abyssinica* from high mountains of Ethiopia
- To examine and compare genetic diversity of *Umbilicus botryoides* between the populations from the Simen and Bale Mountains

4 Materials and Methods

4.1 Study areas

The current study was carried out exclusively in the afro-alpine environment except for the collection of some essential samples outside this area (Table 2 and Fig. 9). The high mountains of Ethiopia and East Africa were the main study sites. Extensive sampling was done from four of the high mountains and mountain systems of Ethiopia (Simen Mountains, Bale Mountains, Mt. Choke and Gara Muleta); four in Kenya (Mt. Elgon, Mt. Aberdare and Mt. Kenya); two in Tanzania (Mt. Kilimanjaro and Mt. Meru); and three in Uganda (Mt. Ruwenzori, Mt. Muhavura and Mt. Gahinga). Special collections were also made to ensure the inclusion of as many species as possible for this study. For example, separate collections were made from Ethiopia (Mt. Entoto, Mt. Ziquala, Mt. Menagesha, Honku and Abela Peaks) and Tanzania (Mbeya Peak) and Madagascar (Tsiavajavona Mountains).

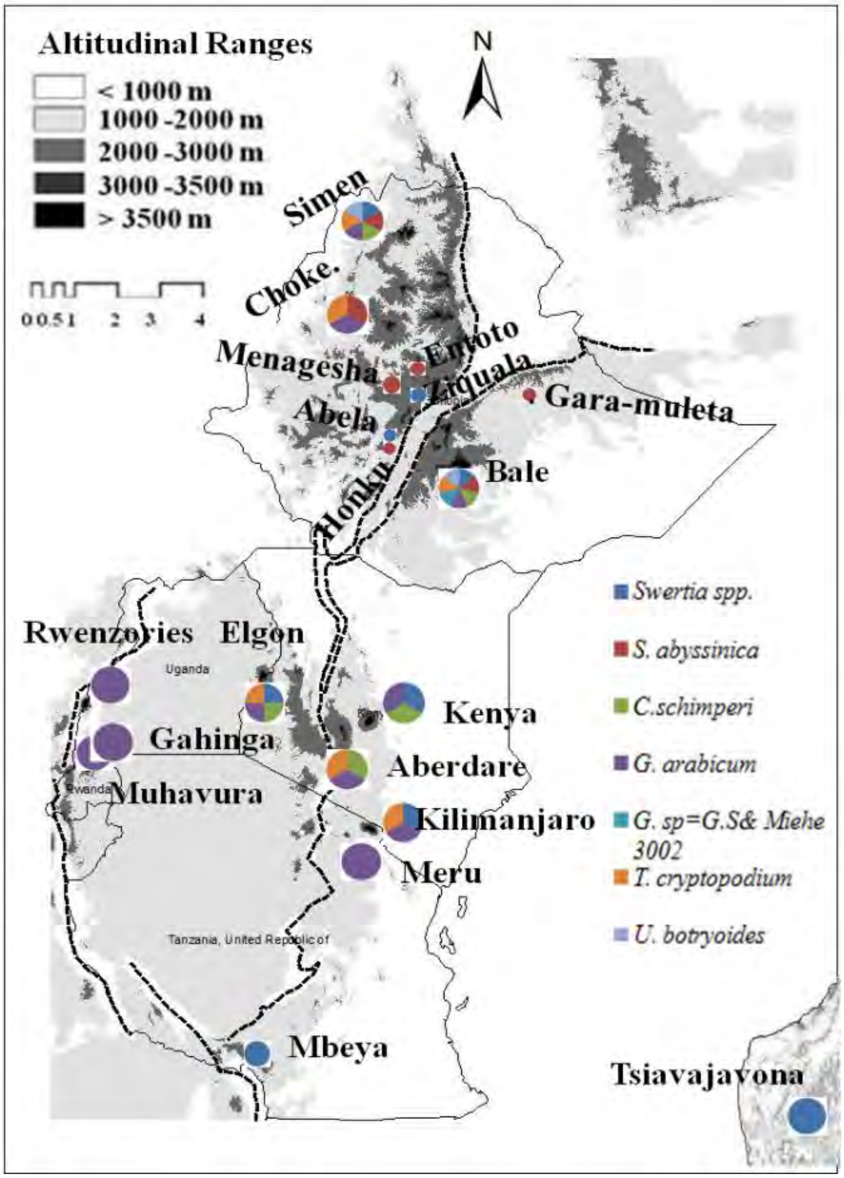


Fig. 9 Sampling map of all the six taxa investigated. Colours on the pie-charts indicate the sampled taxa in the corresponding mountain or mountain system.

Table 2 High mountains and mountain systems from which samples were collected

Mountain/Mountain system	**Country**	**Taxa collected**
Simen Mountains	Ethiopia	*S. abyssinica, S.pumila, S. engleri, S. scimperi, G. arabicum, T. cryptopodium, C. schimperi, U. botryoides*
Mt. Choke	Ethiopia	*S. abyssinica, G. arabicum, T. cryptopodium*
Mt. Ziquala	Ethiopia	*S. tetrandra*
Mt. Entoto	Ethiopia	*S. abyssinica*
Mt. Menagesha	Ethiopia	*S. abyssinica*
Mt. Gara Muleta	Ethiopia	*S. abyssinica*
Bale Mountains	Ethiopia	*S. crassiuscula* subsp. *robust, S. volkensii* var *baleensis, S. macrosepala* subsp. *microsperma, S. abyssinica, S.pumila, G. arabicum, Geranium sp. nov.= G. & S. Miehe 3002, T. cryptopodium, C. schimperi, U. botryoides*
Abela peaks	Ethiopia	*S. quartiniana*
Honku peaks	Ethiopia	*S. abyssinica*
Mt. Elgon	Kenya	*S. subnivalis, S. uniflora, S. kilimandscarica, G. arabicum, T. cryptopodium, C. schimperi*
Mt. Aberdare	Kenya	*G. arabicum, T. cryptopodium, C. schimperi*
Mt. Kenya	Kenya	*S. volkensii* var *volkensii*, *G. arabicum, C. schimperi*
Mt. Kilimanjaro	Tanzania	*S. crassiuscula* subsp. *crassiuscula*, *G. arabicum, T. cryptopodium*
Mt. Meru	Tanzania	*G. arabicum*
Mbeya peak	Tanzania	*S. usambarensis*
Mt. Ruwenzori	Uganda	*G. arabicum*
Mt. Muhavura	Uganda	*G. arabicum*
Tsiavajavona Mts.	Madagascar	*S. rosulata*

4.2 Study plants

4.2.1 *Swertia* L.

Swertia is a large genus in the *Gentianella* lineage (subtribe Swertiinae, family Gentianaceae) and includes more than 130 species (Ho, Xue & Wang, 1994; Albert & Struwe, 2002; Chassot & von Hagen, 2008). The genus is known for its diversity in morphology and ploidy level. It is distributed in temperate regions of the northern hemisphere and in tropical Africa (Meusel *et al.*, 1978; von Hagen & Kadereit, 2001; Chassot *et al.*, 2001; Struwe *et al.*, 2002; Joshi & Joshi, 2008). The highest species diversity is found in the Himalayas, where the genus probably originated (Ho *et al.*, 1994).

All *Swertia* species have one or two fimbriate or naked nectarines at the base of the corolla lobe (Nemomissa, 1997). They are either annual or perennial herbs with erect, procumbent, climbing or cushion life forms. The stems vary from an acaulescent form to a length of 1.5 m, and are usually four-angled. Leaves are opposite or alternate (rarely whorled) with no prominent petioles (either sessile or with a narrow modification at the base). The flowers are either solitary, or occur in an inflorescence which mostly is a corymb having several to many flowers (Nemomissa, 2002; Nemomissa, 2006). The corolla and calyx lobes are highly variable and are used as characters to delimit taxa (Nemomissa, 1997).

A total of 25 species are known from Africa, distributed mostly in the high mountains of North East and East Africa in the altitudinal range from 950 to 4500 m a. s. l. (Nemomissa, 1997). They normally grow on moist black soil, in grasslands, along streams, in boggy areas, under shades of big rocks in rock outcrop habitats and as understory plants in forests.

4.2.2 Carduus schimperi

Carduus schimperi belongs to the family Asteraceae, tribe Cynareae (commonly known as „thistle"). The genus is known for its very spiny leaves and free stamen filaments. *C. schimperi* is distinguished from other members of the genus by its sessile capitula at the center of a leaf rosette (Fig. 10). The species is subdivided into three subspecies: subsp. *platyphyllus*, subsp. *nanus* and subsp. *schimperi*. The size of corolla and shape of leaves are used as key characters for the subspecies. The corolla length in subsp. *platyphyllus* is the shortest (15-17 mm) while that of subsp. *schimperi* is the longest (26-40 mm). The corolla lobe of subsp. *nanus* is intermediate in size (19-24 mm). The leaves of subsp. *nanus* and subsp. *schimperi* are similar in shape (narrowly oblanceolate to oblong or elliptic in outline) while the leaves of subsp. *platyphylus* are obovate to broadly ovate in outline (Jeffery and Beentje, 2000).

Carduus schimperi subsp. *platyphyllus* is endemic to Mt. Kenya growing in moorlands, the heath zone or open rocky sites between 3000 and 4300 m a.s.l, whereas the other two subspecies are more widely distributed. *C. schimperi* subsp. *nanus* occurs on Mt. Elgon (both on the Ugandan and Kenyan sides), Mt. Aberdare and in mountains of the Congo (Kinshasa). It grows in grassland or moorland with an altitudinal range between 2550 and 4050 m a. s.l. The typical subspecies (subsp. *schimperi*) is distributed in the Massai district of Kenya and Tanzanya, the Nyeri District of Kenya, Mbulu District of Tanzania, the mountains of Sudan, and the Bale and Simen Mountains of Ethiopia. This subspecies mainly grows in afro-montane and afro-alpine grasslands, ericaceous bushlands, in open *Juniperus* forest between altitudes of 1800 and 4600 m a. s.l. (Jeffery & Beentje, 2000; Tadesse, 2004).

Fig. 10. *C. schimperi* from Simen Mountains (Photo taken by Linda de Volder)

4.2.3 Trifolium cryptopodium

Trifolium cryptopodium is a member of the family Fabaceae. Compound leaves that are digitately 3-foliolate are a distinguishing feature of the genus. *T. cryptopodium* has a perennial life form with a thick and woody root stock. The stems are usually creeping and root at the nodes, forming large mats. For distinguishing characters of this species see Thulin (1989). The species is distributed in grasslands and rocky habitats of high mountains in Ethiopia and East Africa.

4.2.4 Geranium arabicum and Geranium sp. nov.= G. & S. Miehe 3002

Geranium arabicum and *Geranium sp. nov.* belong to the family Geraniaceae. The genus is distinguished from other members of the family by having a small schizocarp (1-3.5 cm long), fewer stamens and a rostrum of mericarps that rolls upward at dehiscence. Distinguishing features for *G. arabicum* are the leaf lamina, which is pentagonal, and the flowers that usually appear in pairs (Fig. 11). The color and size of the corolla is highly variable in the species; however, the pattern of leaf dissection distinguishes the three subspecies, subsp. *latistipulatum*, subsp. *arabicum* and subsp. = *Ash* (the third subspecies not yet formally described, see Flora of Ethiopia and Eritrea). The leaf-blades of subsp. *latistipulatum* are deeply to shallowly pinnatisect with narrow oblong segments separated by wide sinuses, usually more than halfway to the midrib. Leaf-blades of the typical subspecies are clearly pentagoneous with the main lobe rhombic in outline. *G. arabicum* subsp. = *Ash* has more or less kidney-shaped leaf-blades. Some of the materials on which the identification of this subspecies was made are very similar to *G. kilimandscharicum* from the higher mountains of East Africa (Gilbert and Vorster, 2000). According to Kokowaro (1971), *G. kilimandscaricum* occurs at higher altitude than *G. arabicum*, which occurs below 3200m a. s. l. in the East African mountains. However, their similarity and the occurrence of intermediate forms might be due to introgression and hybridization between the two species.

All of the subspecies of *G. arabicum* are very widely distributed except for subp. = *Ash*, which only has been reported from the Bale Mountains. *Geranium arabicum* subsp. *latistipulatum* occurs in *Erica* forest of the Ethiopian and East African mountains with an altitudinal range from 1000 to 3150 m a. s. l. *G. arabicum* subsp. *arabicum* is more widely distributed with a wider altitudinal range of 1100 to 3950 m. It occurs in herb layer of forests, along streams, and damp grasslands of the afromontane and afro-alpine environments. Besides the Ethiopian and East African mountains, it occurs in Egypt, Zambia, Zimbabwe and Cameroon, Nigeria and Madagascar and Yemen (Gilbert and Vorster, 2000).

In the process of determination of our collected samples, it was found that most of the collections we originally referred to as *G. arabicum* from the Tanzanian and Kenyan Mountains, except one from each country (TZ0316 and KN1037), exhibited a close resemblance to the high altitude species, *G. kilimandscharicum*. Kokowora (1971) also remarked the absence of sufficient distinguishing characters between these two species.

Absence of a stem distinguishes *Geranium sp. nov. = G. & S. Miehe* 3002 (informally accepted as a new species in Flora of Ethiopia and Eritrea) from *G. arabicum*. The exceptionally overlapping leaves and the sessile flowers are the key characters that made it recognized as a separate species. Leaf-blades are deeply 3-lobed and kidney-shaped. The latter characters are highly similar to those of *G. arabicum* subsp. = Ash, and this potentially new species may be a specialized ecotype (Gilbert and Vorster, 2000).

Fig. 11 *G. arabicum* (source: Madagascar image Gallery)

4.2.5 Swertia abyssinica

S. abyssinica is widely distributed in the high mountains of Ethiopia, Eritrea, Sudan, Kenya, Tanzania, Uganda, Malawi, Mozambique, Zambia, Cameroun and Equatorial Guinea (Nemomissa, 1997). The species occurs at an altitudinal range between 1500 and 4150 m. It occupies a highly variable ecological habitat and grows on moist black soil, in volcanic rocks, in montane open grassland, on very steep slopes, in humid and shady areas, wet areas among mosses at forest edges, in seasonally flooded meadows, and in ericaceous scrub in association with *Hagenia abyssinica, Juniperus procera,* and *Erica arborea.*

Like many *Swertia* species, *S. abyssinca* shows large morphological variation. It exhibits a great range of variation in the shapes and apices of the leaves, calyx and corolla lobes, and the length of the pedicel (Nemomissa, 1994; 1997). Nemomissa (1997) indicated that *S. abyssinica* is a diploid species with a chromosome number of 2n=20. Kebede (2000) reported 2n = 20 and 40 after including more populations of *S. abyssinica*. A previous molecular data based study by Bekele (2007) of the species complex of *S. abyssinica* found that the variation between *S. abyssinica* and *S. lugardae* reflects continuous variation along an altitudinal gradient, and that similar high-altitude forms (*'S. lugardae'*) has originated independently from low-altitude forms *(S. abyssinica* s.str.) in different Ethiopian mountains. Therefore, in the present study we treated all of these populations as a single species under the name *S. abyssinica*.

4.2.6 Umbilicus botryoides

The genus *Umbilicus* is a member of the family Crassulaceae. The genus is characterized by its peltate, petiolate leaves, racemose inflorescence and joined petals. *U. botryoides* is a glabrous perennial herb with a tuberous or rhizomatous rootstock. The inflorescence type is typically terminal, racemose and five-merous. This species is distributed in well-drained, moist, shady sites on rocks, soil banks or as epiphytes in forests. The species is distributed across at altitudinal range from 1350 to 4150 m, and the rock outcrops of the high mountains in Ethiopia and East Africa are favorable sites (Gilbert, 1989).

4.3 Sampling

Fresh young leaf materials of six taxa (*Swertia* spp., *G. arabicum, Geranium sp. nov. = G. & S. Miehe* 3002, *C. schimperi*, *T. cryptopodium* and *U. botryoides*) were collected from 17 mountain systems in Ethiopia, East Africa and Madagascar between 2007 and 2009 (see Table 2 & Fig. 9 above). Whenever possible, four geographically separated populations of each species were sampled per mountain system, with each population represented by five individuals. The five individuals were collected from each corner and mid point of a 100m^2 plot. In places where it is not possible to make such a plot (e.g. extended rock outcrops), individuals were collected in every 100m. Besides, efforts were made to separate populations by at least 1km distance and with good geographical barriers. Leaf samples were preserved in silica gel and voucher specimens of all sampled individuals were pressed. The five vouchers from each population were deposited in the following herbaria: Natural History Museum, University of Oslo, Norway (O; 1 voucher); Addis Ababa University, National Herbarium of Ethiopia (ETH; 3 vouchers);

and the fifth voucher was deposited according to country of collection: East African Herbarium (EA), Kenya; Sokoine University of Agriculture (SUA), Tanzania and Makarare University (MU), Uganda.

Twenty taxa of the African *Swertia* were collected from 10 mountain systems of East Africa, Ethiopia and Madagascar (Table 3). Previously collected silica-dried leaf tissue from ETH and EA Herbaria were included in the phylogenetic analyses of *Swertia*, when collection of fresh materials was not successful (Table 3).

Whenever possible, 20 individuals from four populations were collected for each species from a mountain or a mountain system. The three subspecies of *C. schimperi* were collected from five mountains in Ethiopia (Simen & Bale Mts.) and Kenya (Mt. Elgon, Aberdare and Kenya). Sufficient sampling of *G. arabicum* and *Geranium sp. nov. = G. & S. Miehe* 3002 was obtained from 10 mountains of Ethiopia and East Africa. Samples of *T. cryptopodium* were collected from Simen, Choke and Bale mountains in Ethiopia, from Mt. Elgon and Aberdare in Kenya, and from Mt. Kilimanjaro in Tanzania. Although sampling from the East African Mountains was not successful for *S. abyssinica*, sufficient sampling was performed from the Ethiopian mountains (Simen, Choke, Menagesha, Entoto and Bale Mts.; Gara Muleta, Mt. Ziquala, and Honku Peak). Samples of *Umbilicus botryoides* were collected from the Simen and Bale Mts. (see Table 2 and Fig. 9).

Table 3 Material of *Swertia* used in this study. Population ID follows the DNAbank of the National Herbarium of Ethiopia and NCB, Natural History Museum, University of Oslo.

No	Taxon	Population ID	Country	Locality	Altitude (m)	Latitude	Longitude
1	*S. abyssinica* Hochst.	ET-0047-2	Ethiopia	Simen Mountains	3570	13.15.99N	38.06.46E
2	*S. adolfi-friderici* Mildbr. & Gilg.	TZ-0710-2	Tanzania	Mbeya peak	2730	08.50.18S	33.22.25E
3	*S. brownii* J. Shah	UG-2005-2	Uganda	Mt. Gahinga	2600	01.21.85S	29.38.67E
4	*S. crassiuscula* Gilg. subsp. *crassiuscula* Gilg. var. *leucantha* Sileshi	TZ-0334-2	Tanzania	Mt. Kilimanjaro	2770	03.10.37S	37.31.10E
5	*S. crassiuscula* Gilg. subsp. *robusta* Sileshi	ET-0678-2	Ethiopia	Bale Mountains	3700	06.46N	39.45E
6	*S. engleri* Gilg. var. *engleri*.	ET-0136-2	Ethiopia	Simen Mountains	3720	13.17.11N	38.07.10E
7	*S.* aff. *abyssinica*	ET-1454-2	Ethiopia	Bale Mountains	3480	07.00.44N	39.42.59E
8	*S. intermixta* A. Rich.*	Igoma-08	Tanzania	Mbeya	2700	08.50S	33.22E
9	*S. kilimandscharica* Engl.	KN-0365-2	Kenya	Mt. Elgon	3580	01.06.13N	34.37.29E
10	*S. macrosepala* Gilg. subsp. *macrosepala*	UG-2061-2	Uganda	Mt. Gahinga	4130	01.22.96S	29.40.67

11	*S. macrosepala* Gilg. subsp. *microsperma* Sileshi	ET-0681-1	Ethiopia	Bale Mountains	3450	06.50N	39.43E
12	*S. pumila* Hochst. ex Hook f.	ET-0222-2	Ethiopia	Simen Mountains	3700	13.12.04N	38.16.01E
13	*S. quartinina* A. Rich.	ET-1304-2	Ethiopia	Abela Peak	1990	06.54.94N	38.30.01E
14	*S. rosulata* (Baker) Klack	MD-0006-2	Madagascar	Tsiafajavona Mountains	2240	19.21.48S	47.15.52E
15	*S. scandens* T.C.E.Fr.[+]	9707254-3/1	Kenya	Mt. Kenya			
16	*S. schimperi* Griseb.	ET-0187-2	Ethiopia	Simen Mountains	3550	13.16N	38.06E
17	*S. squamigera* Sileshi[*]	S-1994	Tanzania				
18	*S. subnivalis* T.C.E.Fr.	KN-0004-2	Kenya	Mt. Elgon	4220	01.07.43N	34.36.11E
19	*S. tetrandra* Hochst.	ET-1294-2	Ethiopia	Mt. Ziquala	2300		
20	*S. uniflora* Mildbr.	KN-0193-1	Kenya	Mt. Elgon	4040	01.07.08N	34.35.20E
21	*S. usambarensis* Engl.	TZ-0711-2	Tanzania	Mbeya Peak	2620	08.50.21S	33.22.35E
22	*S. volkensii* Gilg. var. *baleensis* Sileshi	ET-0680-2	Ethiopia	Bale Mountains	4050	06.46N	39.45E
23	*S. volkensii* Gilg.var. *volkensii*	KN-0922-2	Kenya	Mt. Kenya	4190	00.08.01S	37.16.59E
24	*S. welwitschii* Engl.[*]	R-1994	Tanzania	Sumbawanga			

[+] Silica-dried leaves from Addis Ababa University National Herbarium

[*] Herbarium material from East African Herbarium.

4.4 DNA extraction

Total genomic DNA was extracted from silica-dried leaf tissue and herbarium materials using a GeneMole® robot and the Mole StripTM Plant DNA Kit (QIAGEN, Nordic, Oslo, Norway) or DNeasyTM Plant Mini Kit (QIAGEN, Valencia, CA). Prior to loading the plant material to the GeneMole®, the following modifications were performed: leaf tissue was mechanically ground in 2.0 µL tubes with two tungsten carbide beads for 2 min at 15 Hz in a mixer mill (MM301, Retsch GmbH & Co., Haan, Germany), 300-350 µL of lysis buffer was added to the crushed material, vortexed, spun briefly, incubated on a heat block for 10 min at 65°C, and centrifuged at 14000 Hz for 2 min. 200 µL of the lysate was transferred into new tubes and loaded to the GeneMole® robot, which was set to produce 100 µL of elution.

4.5 Methods of the phylogenetic and biogeographic study

4.5.1 DNA amplification and sequencing

Two non-coding plastid regions (the *trn*L-F region [both the intron and the intergenic spacer] and the *trn*S-fM intergenic spacer), one coding plastid region (*mat*K) and the nuclear ribosomal internal transcribed spacer (ITS) were amplified. These regions were selected because they were used by Chassot *et al*. (2001) and Gielly and Taberlet (1996), from which we used data for comparison. In addition, the *mat*K and the *trn*L intron were used to reconstruct the phylogeny of tribes in the family Gentianaceae (Struwe *et al*., 2001). The latter regions were used in the analysis of divergence times in African *Swertia*. PCR was performed in a volume of 10 µL containing 2.12 µL ddH2O, 1 µL 10x PCR buffer, 1 µL MgCl2 (25 mM), 1 µL dNTP (10 mM), 1 µL BSA (1%), 0.4 µL of each primer, 0.08 µL Taq DNA polymerase (Applied Biosystems, Foster City, CA, USA), and 3 µL DNA template of unknown concentration. We used the c and f primers (Taberlet et al., 1991) to amplify the *trn*L-F region, the *trn*S and *trn*fM primers (Shaw *et al.*, 2005; Demesure, Sodzi and Petit, 1995) for the *trn*S-fM region, and the ITS4 and ITS5 primers (White *et al.*, 1990) for the ITS region. Primers used for the amplification of two neighboring fast-evolving regions of *mat*K were *mat*K 1198F (ctgtgttagatatacgaatacc), *mat*K 1581R (cttgatacctaacataatgcat), *mat*K 1729F (aagggtctatataaagcaatt) and *mat*K 2053R (ttagcrcaagayagtcgaagta) following Thiv *et al.* (1999a). The PCR program for *trn*L-F was 80^{o}C 5 min, (94oC 1 min, 50^{o}C 1 min, 72^{o}C 2 min) x 35, 72^{o}C 5 min. The program for *trn*S-fM was 80^{o}C 5 min, (94oC 30 s, 55^{o}C 30 s, 72^{o}C 2 min) x 30, 72^{o}C 5 min (Shaw *et al.*, 2005). The program used to amplify the ITS and the *mat*K regions was 95^{o}C 5 min, (94^{o}C 30 s, 50^{o}C 30 s, 72^{o}C 2 min) x 30, 72^{o}C 5 min. The PCR products were checked on 1% agarose gel and cleaned with 10x dilution of ExoSAP-IT (USB Corp., Cleveland, OH, USA).

The purified PCR products were diluted 15-30 times, depending on the intensity of the bands on the electrophoresis gel. Sequencing was carried out with the same primers as for PCR using the BigDye Terminator v 1.1 Cycle Sequencing kit (Applied BiosystemsTM, Foster City, CA, USA). The volume of the sequencing reaction was 10 µL, composed of 6.3 µL ddH2O, 1.8 µL sequencing buffer, 0.4 µL BigDye, 0.5 µL of each primer and 1 µL PCR product. The thermal cycling was 96oC 1 min, (96oC 10 s, 50oC 5 s, 60oC 4 min) x 35, final extension 60oC 10 min. The sequencing products were cleaned using Sephadex and run on an ABI Prism 3130 genetic analyzer. Sequences were edited using CodonCode aligner v3.5 (CodonCode Corporation, Dedham, MA, USA). Plastid sequences from 36 species that belong to subtribe Swertiinae and three species of subtribe Gentianinae (outgroups) were retrieved from GenBank for the reconstruction of phylogeny and inference of biogeography of the African *Swertia* (see Chassot *et al.*, 2001 for detail information of these sequences). Six additional sequences, which belong to the lineage where plants with fossil evidence are included, were also retrived from GenBank for the analysis of molecular divergence time (Appendix 1).

4.5.2 Phylogenetic data analyses

The sequences of the two plastid DNA regions (*trn*L-F and *trn*S-fM) were concatenated for all included species except *S. welwitschii*, for which amplification of the *trn*S-fM region failed. The ITS matrix contained all included species except *S. squamigera*, for which amplification failed. Each data set was analysed separately and in combination (the two plastid DNA regions) following the same procedure. Multiple sequence alignment was done using the ClustalW option in SeaView v4.0 (Galtier, Gouy and Gautier, 1996) and adjusted manually. MEGA v5.05 (Tamura *et al.*, 2011) was used to analyze the sequences‟ characteristics, such as length of sequence, number of variable characters and parsimony-informatives. MrModeltest v2.3 (Nylander, 2004) was used to select the best fit substitution models for each of three partitions (corresponding to a continuous DNA region). The Input data for MrModeltest was executed with PAUP* v. 4.0b10 (Swofford, 2003). PhyML was used as implemented in SeaView v4.0 (Galtier *et al.*, 1996) for maximum likelihood (ML) bootstrapping with 500 replicates. Bayesian analyses were run with MrBayes v3.1.2 (Ronquist & Huelsenbeck, 2003) as two parallel MCMC (Marcov Chain Monte Carlo) analyses, each run for one million (for the plastid DNA regions) and ten million generations (for the nuclear ribosomal DNA region), with one cold and three heated chains and sampled every 100th (plastid DNA regions) and 1000th generation. Convergence was diagnosed using the standard deviation of split frequencies (SDSF). Twenty five percent of the generations were discarded as burn-in and the remaining were summarized in a 50%-majority rule consensus tree with posterior probabilities (PP) as a measure of clade support. Trees were visualized using v 3.1 (Rambout, 2009).

4.5.3 Molecular dating analysis

A Bayesian approach as implemented in BEAST 1.6.1 (Drummond and Rambaut, 2007) was used to estimate the divergence time of the African *Swertia* from species of other continents and within the African *Swertia*. The molecular data were analysed with a GTR substitution model with gamma (γ) distributed rate variation, an uncorrelated relaxed lognormal clock, and a Yule tree prior with a constant speciation rate per lineage. MrModeltest was also used to select the best fit substitution models for each of the three DNA regions used in this analysis.

Three calibration points, two fossil and one calibration from a previous molecular dating analysis (Yuan *et al*., 2005; Favre, 2010) were applied in the analysis. Three fossil records are known in the family Gentianaceae: seed of *Voyrioseminites magnus* (Trivedi and Chaturvedi, 1972), pollen and flower of *Pistillipollenites mcgregorii* (Pocknall, 1987; Crepet and Daghlian, 1981), and pollen of *Lisianthius* (Graham, 1984). The validity of the fossil of *Pistillipollenites mcgregorii* was uncertain (Struwe *et* al., 2001) until Martínez-Millán (2010) confirmed its reliability. *Pistillipollenites mcgregorii* was not assigned to any extant family until a thorough morphological examination was performed. Among all the investigated extant species, the pollen of *Pistillipollenites* *mcgregorii* has shown the highest similarity to pollen of the four species sampled from *Macrocarpae* (Crepet and Daghlian, 1981). Recent studies on phylogeny and biogeography of taxa in the family Gentianaceae used the fossil evidence to estimate divergence dates (e.g. Yuan *et al*., 2005; and Favre *et al*., 2010). The fossil evidence from *Pistillipollenites mcgregorii* (*Macrocarpae*) was dated to the Lower Eocene while that of *Lisianthius* was from Late Eocene (Pocknall, 1987; Graham, 1984). Two different dating approaches were considered to further reconstruct two independent chronograms using the *mat*K (the codon region) sequences. In the first approach, two calibration points were considered: 55.5 Mya (the earliest age in Lower Eocene) for the root, referring to the age of the *Macrocarpae* and 37.2 Mya (the earliest age in the Late Eocene), referring to the age of *Lisianthius* pollen. In the second approach a third calibration point was put at the crown of the Swertiinae, which was assumed to be 15 Mya (Yuan *et al*., 2003). The ages of the Epochs were set according to the Geologic time scale by Gradstein *et al*. (2004) and Ogg *et al*. (2008). The calibration points in both appraches were used to estimate the divergence time of the African *Swertia*. The estimated divergence time was latter used to further estimate the ancestral ages of the African *Swertia* using concatenated regions of DNA (*mat*K, *trn*L and ITS).

The Markov Chain Monte Carlo (MCMC) was run for 10 million generations, sampling hypotheses every 1000 generations. The output was visualized using Tracer (Rambaut & Drummond, 2003), making sure that parameter values were fluctuating at stable levels. Based on these results, the first 1 million generations were discarded as burn-in, and the remaining samples were summarized as a maximum sum of clade credibility tree with mean divergence times and HPD intervals of age estimated in TreeAnnotator and visualized in FigTree v 3.1.

The analyses were repeated by excluding all genera but *Swertia* for the concatenated data set. The estimated divergence date for the African *Swertia,* obtained from the previous analysis was used to further reconstruct the divergence date of the two African *Swertia* clades, i.e. the divergence date of Asian *Swertia* was put at the root. In this case, the two Asian *Swertia* species were considered as outgroups.

4.6 Methods for AFLP phylogeographic and genetic diversity analyses

4.6.1 AFLP Primer test, primer selection and PCR

Two primers were tested on eight samples followed by an extended test of twelve primers on a larger sample size for each species. After a thorough examination of the resulting bands, two or three promising primer combinations were selected for the final analysis (the primer combinations used to each species are given in Table 4).

Table 4 Primer combinations used to generate AFLP data for each of the five species

Taxa	Primer combinations
	*EcoR*I/*Mse*I
C. schimpei	ACA/CAC (FAM), ACT/CAG (FAM), ACG/CTC (VIC)
T. cryptopodium	AGA/CTC (FAM), ACA/CAC (VIC)
G. arabicum	ATG/CGA (FAM), ACA/CAC (VIC), AGC/CTG (NED)
S. abyssinica	ACC/CAG (FAM), ATG/CGG (FAM)
U. botryoides	AGA/CTG (FAM), AGG/CAA (VIC), AAC/CAT (NED)

AFLP data generation for each of the five species was undertaken following the standard AFLP procedure. To carry out the restriction-ligation, a master mix composed of 1.1 µL of 10x T4 ligase buffer, 1.1 µL of 0.5M NaCl, 0.55 µL of 1mg/mL BSA (bovine serum albumin), 0.02 µL of 50 U/µL *Mse*I 0.125 µL of 40U/µL *EcoR*I, 0.2 µL of 5 U/µL, T4-DNA ligase, 1 µL 10 µM *Mse*I-adapter, 1 µL of 10 µM *EcoR*I adapter and 0.41 µL of ddH_2O was prepared. Newly started adapters were always denatured at 95°C for 5 minutes and kept on a flat surface at room temperature for about 10 minutes before use. A 9 µL of the master mix was added to 2 µL of sample and subjected to a short spin. The mixture was incubated at 37°C for three hours on a PCR machine.

A master mix containing 1.25 µL of 10x Taq buffer, 0.75 µL of 25 mM $MgCl_2$, 1 µL of 10 mM dNTP, 0.25 µL of 10 µM *EcoR*I primer (Eco-A), 0.25 µL of 10 µM *Mse*I primer (Mse-C), 0.075 µL of 5 U/ µL Ampli*Taq* DNA Polymerase, and 7.45 µL of ddH_2O was prepared to undertake the pre-selective PCR. The restriction-ligation product was diluted 10x and a 1.5 µL of the product was mixed with 11 µL of the pre-selective master mix and subjected to short spin. The program used for the pre-selective PCR was: 72°C 2 min, (94°C 30 s, 56°C 30 s, 72°C 2 min) x 30, 72°C 10 min.

The pre-selective PCR was diluted 10x and 2.5 µL of each sample was added to a 10 µL selective master mix. The master mix comprised 1.25 µL of 10x Gold Taq buffer, 1.25 µL of 25 mM $MgCl_2$, 0.1 µL of 1 mg/mL BSA, 1 µL of 10 mM dNTP, 0.1 µL of 10 µM fluorescent *EcoR*I primer (E.AXX), 0.25 µL of 10 µM *Mse*I primer (M.CXX), 0.1 µL of 5 U/ µL Ampli*Taq* Gold DNA Polymerase and 5.95 µL of ddH_2O. After a short spin, the mixture was subjected to a selective PCR run using a program of 95°C 10 min, (94°C 30 s, 65°C 1 min, 72°C 1 min) x 13; (94°C 30 s, 56°C 1 min, 72°C 1 min) x 23, 72°C 10 min.

For each individual, 2.0 µL 6-FAM, 2.0µL VIC and 3.0µL NED labeled selective PCR products were mixed and add to a master mix of 11.7 µL formamide and 0.3µL GENESCAN ROX 500 size standard before tunning on an ABI3100 sequencer (Applied Biosystems).

4.6.2 AFLP data analysis

AFLP markers were scored as present (1) or absent (0) using GeneMapper version 4.0 with the size range of 50–500 base pairs (bp). While scoring, careful examination of the bands was carried out in such a way that peaks of low intensity were included in the analyses when they were unambiguous, and removed from analyses when ambiguous. An error rate test was performed to ensure reproducibility and reliability of the results (Bonin *et al*., 2004). Ten percent of the total sample size was selected at the beginning from each species to test for reproducibility (Appendix 2). Error rate calculations and data cleaning were done for each primer combination separately. All duplicates were removed and matrices for all primers were combined in the final analyses.

Principal Coordinate Analyses (PCoAs) were performed using NTSYSpc 2.1 (Rohlf, 2000) on matrices of pair-wise DICE similarities, which is equivalent to Nei & Li (1979), and assumes similarities based on the sharing of present bands. Gene diversity (h) according to Nei and Li (1979) and rarity (DW) were calculated as implemented in AFLPdat (Ehrich, 2006). Neighbor-joining (NJ) trees based on genetic distance were constructed for all species with bootstrap support using TreeCon v 1.3b (Peer and Wachter, 1994). The no admixture model and uncorrelated allele frequencies were selected in analyses that employed STRUCTURE v 2.3.3 (Pritchard *et al.* 2000) to investigate the population structure. The analysis was performed with K ranging from 1-10 and for each K, 10 replicate runs with burn-in period of 2000000 and 1000000 iterations were used. A web-based tool, STRUCTURE HARVESTER (Earl and Holdt, 2011), was used to visualize the outputs from STRUCTURE analysis.

AMOVAs (Analyses of Molecular Variance) were performed to investigate the partitioning of genetic variation (Excoffier *et al*., 1992) at three different hierarchical levels (among groups, among populations within groups and within populations) using ARLEQUIN V. 3.5 (Scheneider *et al*. 2000). AFLPOP v 1.1 (Duchesne and Bernatchez, 2002) was used to examine the presence of individuals with mixed genetic makeup from different groups. AFLPOP assigns individuals to potential source populations based on the likelihood of observing their AFLP genotype in the source populations, given the population frequencies of AFLP bands. The program is also used for assigning potential hybrid populations of different genealogy (F1, F2, and first backcrosses to each parental population), based on estimation of allele frequencies in the parental populations by assuming Hardy–Weinberg (H-W) equilibrium. The source populations that were considered as parental populations were defined based on the group assignment obtained from the STRUCTURE analysis. Individuals assigned to a different group from that expected were assumed as unknown samples. The procedure of allocation was performed with the formula 1/sample size + 1 as Zero frequency correction values; 1-1/sample size + 1 as Unit frequency correction values; with probability of population assignment calculated from 1000 re-sampled individuals. Two criteria were used to detect hybrids: they should assign to a hybrid class with likelihood 100 times higher than to the source population from which they were sampled, and their probability should be statistically significant to reject the probability of a member of the source population while their probability of being a hybrid should not be rejected. i.e. P was set to be 0.05 among the 1000 re-sampled individuals of the source populations, 5 % score likelihoods less than or equal to the likelihood that the individuals being members to the source population indicates potential hybrids. The procedure was repeated with likelihood 10 times and 1 time higher than the source population from which they were sampled. The same procedure was repeated by employing the option that allows allocation of unknown individuals as member of the source populations, F1, F2 or a backcross population, for individuals that were not allocated to any group following the first option.

5 Results

5.1 Results from phylogenetic and biogeographic study

Phylogenetic trees inferred from the Bayesian analysis (SDSF = 0.007 for plastid DNA data set, and 0.012 for the nuclear ribosomal DNA data set) and maximum likelihood analyses were very similar in topology. Thus, only the Bayesian trees with posterior probability values are presented here.

5.1.1 Plastid DNA phylogeny of the African Swertia

The lengths of sequences were 696-798 bp for the *trn*L-F region and 935-1071 bp for the *trn*S-fM region. We were not able to amplify the complete *trn*S-fM region for *S. scandens* and *S. intermixta*. The number of variable characters was 232 (21.5%) of which 112 (10.4%) were parsimony-informative in the aligned *trn*L-F sequences, and 161 (12.9%) of which 78 (6.2%) were parsimony-informative in the aligned *trn*S-fM sequences. A GTR+G substitution model was selected for both partitions.

In all analyses, the African species of *Swertia* were grouped into two distantly related clades with high support (PP=1, ML Bootstrap support=91-97; bootstrap support not shown) but poor internal resolution (Fig. 12). Two Himalayan species of *Swertia* (*S. cuneata* and *S. cordata*) were inferred as members of these otherwise African clades (indicated with arrows).

Three Asian species of *Swertia* and two species of *Gentianella* formed a polytomy with the African Clade 1 with high support (PP=0.98). The African Clade 1 contained a basal polytomy formed by the Himalayan *S. cuneata*, three of the African species, and three

African subclades: 1) a subclade containing *S. usambarensis*, *S. intermixta*, *S. welwitschi, S.* aff. *abyssinica, S. crassiuscula* subsp. *robusta*, *S. S. brownii*; PP=0.74, 2) a subclade containing two accessions of *S. crassiuscula* subsp. *crassiuscula* and *S. engleri*; PP=0.76, and 3) a subclade containing *S. pumila*, *S. engleri* and *S. abyssinica*; PP=0.97). Subclades 1 and 2 included different accessions of *S. crassiuscula* while subclades 2 and 3 contained different accessions of *S. engleri*. *Swertia crassiuscula* subsp. *robusta* and *S.* aff. *abyssinica*, both collected from the Bale Mountains, grouped together with high support (PP=1) in the first subclade. Two other subspecies of *S. crassiuscula*, collected from Mt. Kilimanjaro and Mt. Kenya, grouped together in the second subclade. One accession of *S. engleri* var. *engleri*, collected from Simen Mountains, grouped together with *S. crassiuscula* subsp. *crassiuscula* in the second subclade while the second accession, collected from the same mountain, grouped in the third subclade (Fig. 12)

The Himalayan *S. angustifolia* was inferred as sister to the African Clade 2, which included the Himalayan *S. cordata* (PP=1). A clade including the type species of the genus *Swertia*, *S. perennis*, was resolved as sister to the clade consisting of *S. angustifolia* and the African Clade 2 (PP=0.96). The Himalayan *S. cordata* and three African subclades formed a basal polytomy in the African Clade 2. The three African subclades were 1) a subclade including *S. volkensii* var. baleensis, *S. macrosepala* subsp. *microsperma*, 2) a subclade including *S. subnivalis,* two accessions of *S. volkensii* var. *volkensii, S. uniflora* and *S. macrosepala* subsp. *macrosepala*, and 3) a subclade including *S. scandens, S. adolfi-friderici, S. squamigera,* two accessions of *S. kilimandscarica* and *S. schimperi*. Different accessions of two species, *S. macrosepala* and *S. volkensii*, grouped into different subclades. In both cases the accessions were assigned to different subspecies (Fig. 12).

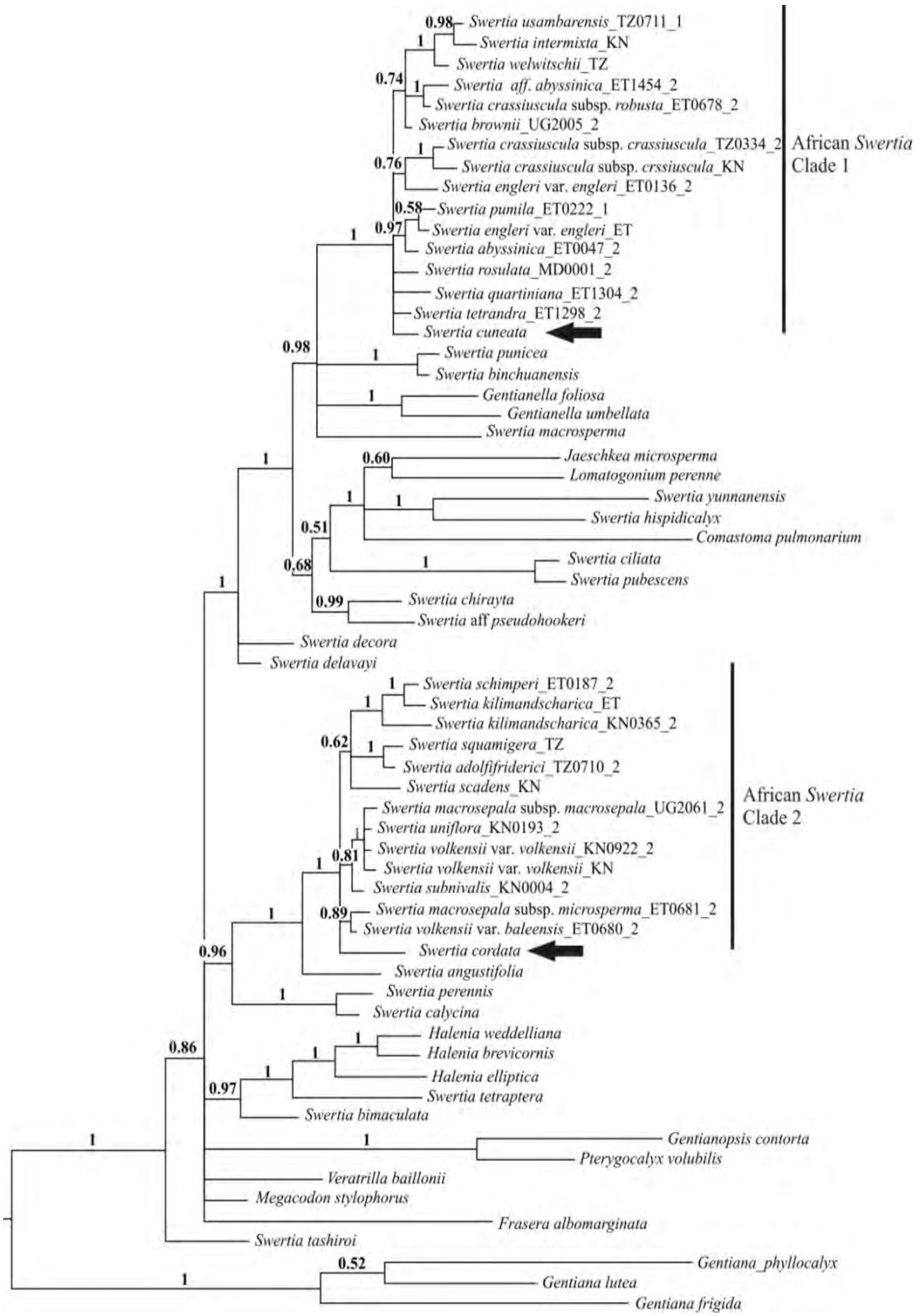

0.98
Swertia usambarensis_TZ0711_1
1
Swertia intermixta_KN
Swertia welwitschii_TZ
0.74
1
Swertia aff. abyssinica_ET1454_2
Swertia crassiuscula subsp. robusta_ET0678_2
Swertia brownii_UG2005_2
1
Swertia crassiuscula subsp. crassiuscula_TZ0334_2
0.76
Swertia crassiuscula subsp. crssiuscula_KN
Swertia engleri var. engleri_ET0136_2
0.58
Swertia pumila_ET0222_1
0.97
Swertia engleri var. engleri_ET
Swertia abyssinica_ET0047_2
Swertia rosulata_MD0001_2
Swertia quartiniana_ET1304_2
Swertia tetrandra_ET1298_2
Swertia cuneata
African Swertia Clade 1
1
0.98
1
Swertia punicea
Swertia binchuanensis
1
Gentianella foliosa
Gentianella umbellata
Swertia macrosperma
1
0.60
Jaeschkea microsperma
Lomatogonium perenne
1
1
Swertia yunnanensis
Swertia hispidicalyx
0.51
Comastoma pulmonarium
0.68
1
Swertia ciliata
Swertia pubescens
1
0.99
Swertia chirayta
Swertia aff pseudohookeri
Swertia decora
Swertia delavayi
1
Swertia schimperi_ET0187_2
1
Swertia kilimandscharica_ET
Swertia kilimandscharica_KN0365_2
0.62
1
Swertia squamigera_TZ
Swertia adolfifriderici_TZ0710_2
Swertia scadens_KN
African Swertia Clade 2
Swertia macrosepala subsp. macrosepala_UG2061_2
Swertia uniflora_KN0193_2
1
Swertia volkensii var. volkensii_KN0922_2
0.81
Swertia volkensii var. volkensii_KN
1
Swertia subnivalis_KN0004_2
0.89
Swertia macrosepala subsp. microsperma_ET0681_2
Swertia volkensii var. baleensis_ET0680_2
1
Swertia cordata
0.96
Swertia angustifolia
1
Swertia perennis
Swertia calycina
1
Halenia weddelliana
1
Halenia brevicornis
0.86
Halenia elliptica
1
0.97
Swertia tetraptera
Swertia bimaculata
1
Gentianopsis contorta
1
Pterygocalyx volubilis
Veratrilla baillonii
Megacodon stylophorus
Frasera albomarginata
Swertia tashiroi
0.52
Gentiana_phyllocalyx
1
Gentiana lutea
Gentiana frigida

Fig. 12 Phylogenetic tree inferred from Bayesian analysis of the plastid DNA (*trn*L-F and *trn*S-fM) data set. Posterior probability values are indicated above the branches. The arrows indicate the Himalayan *Swertia* species that are placed within the otherwise African clades (ET= Ethiopia, KN= Kenya, TZ= Tanzania, UG= Uganda).

5.1.2 Nuclear DNA (ITS) phylogeny of the African Swertia

The length of the sequences for the ITS region varied between 406 and 625 bp. The number of variable characters was 281 (38.7%), of which 177 (24.4%) were parsimony-informative. The best fit model of substitution selected was GTR+G. The trees inferred from both the Bayesian and maximum likelihood analyses were poorly resolved among the genera in subtribe Swertiinae. However, the African *Swertia* were placed in two distinct groups (PP = 0.9-1.0 and ML Bootstrap support = 72-75; Bootstrap support not indicated; Fig. 13).

Unlike in the plastid DNA analysis, no Asian species or clade appeared to be sister to the African Clade 1. *Swertia engleri* var. *engleri* (ET0136-2) and the two accessions of *S. crassiuscula* subsp. *crassiuscula* resolved from the rest of the species in the clade with support value of PP=1. The second accession of *S. engleri* var. *engleri* (ET) was resolved as sister to a clade containing all remaining accessions (PP=1). In this clade, *S.* aff. *abyssinica* was unresolved relative to three subclades (all with PP=1), of which one contained *S. abyssinica*. The three subclades were 1) *S. brownii*-*S. quartiniana* (PP = 1), 2) *S. pumila*-*S. tetrandra* (PP = 1); and 3) *S. usambarensis*- *S. crassiuscula* subsp. *robusta* (PP = 1; Fig. 13).

The trend shown by the African species in Clade 2 was quite similar to that of the plastid DNA analysis. The Himalayan species (*S. cordata*) was resolved as sister to this clade. *S.*

kilimandscharica ET was resolved from the rest of the species in this clade with a support value of PP = 0.86. The two varieties of *S. volkensii*, *S. uniflora*, *S. subnivalis*, and the two subspecies of *S. macrosepala* were grouped together with a support value of PP=0.96, leaving the rest of the species in the clade unresolved. One of the subspecies of *S. macrosepala* grouped with *S. subnivalis* with high support (PP=0.99) within the subclade.

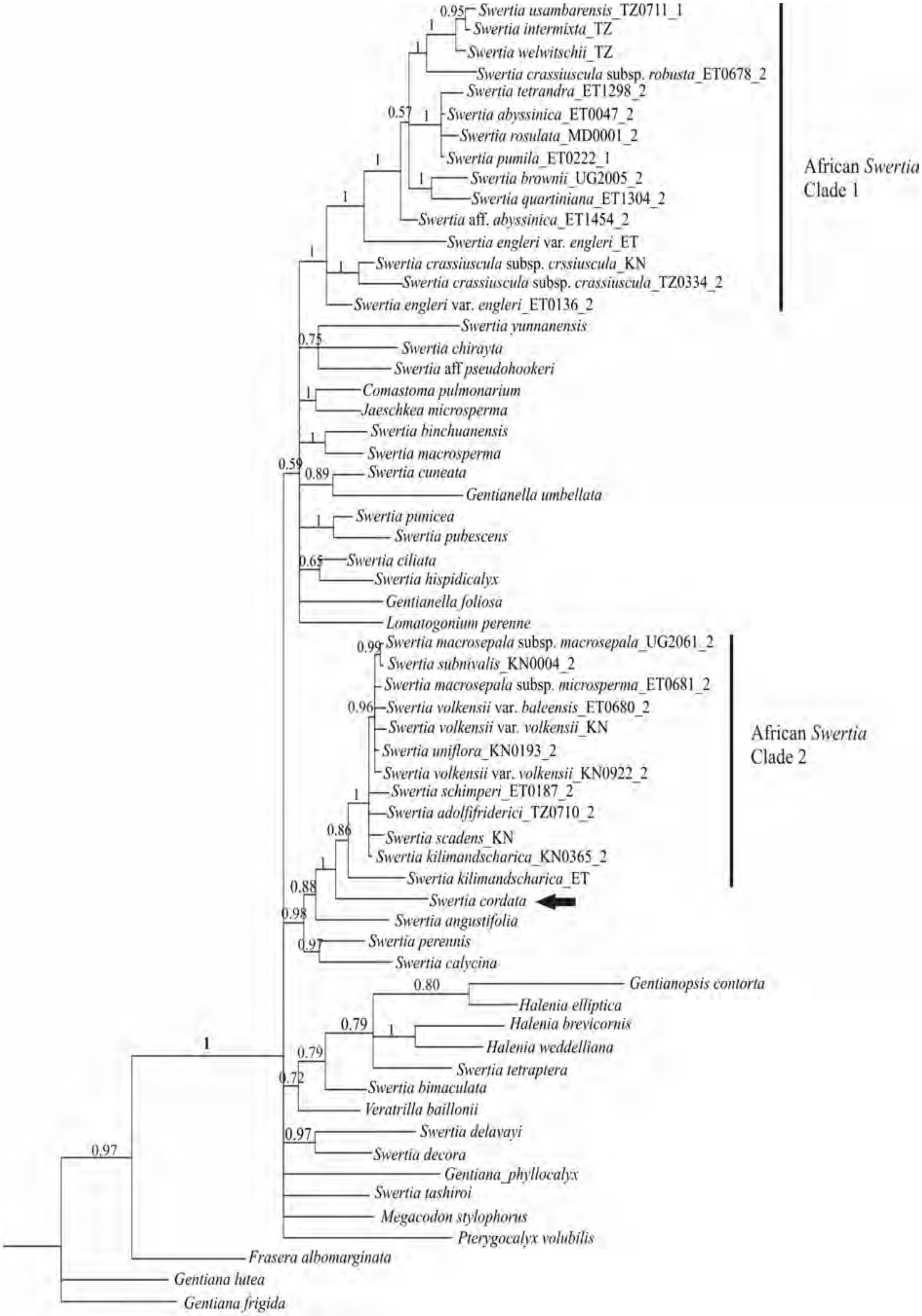
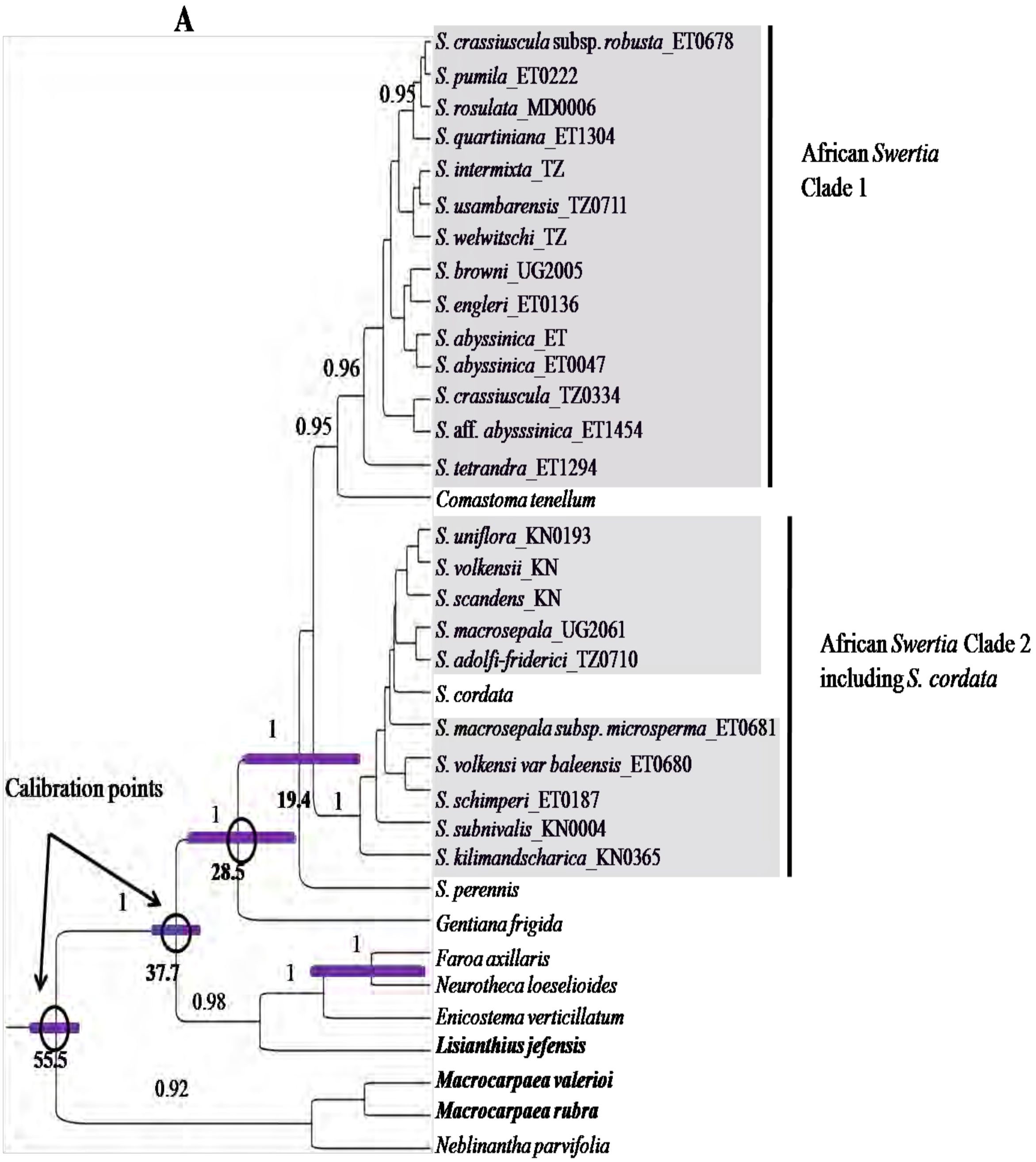

0.95
Swertia usambarensis_TZ0711_1
1
Swertia intermixta_TZ
1
Swertia welwitschii_TZ
Swertia crassiuscula subsp. robusta_ET0678_2
Swertia tetrandra_ET1298_2
0.57
1
Swertia abyssinica_ET0047_2
Swertia rosulata_MD0001_2
1
Swertia pumila_ET0222_1
1
Swertia brownii_UG2005_2
1
Swertia quartiniana_ET1304_2
Swertia aff. abyssinica_ET1454_2
Swertia engleri var. engleri_ET
1
1
Swertia crassiuscula subsp. crssiuscula_KN
Swertia crassiuscula subsp. crassiuscula_TZ0334_2
Swertia engleri var. engleri_ET0136_2
African Swertia
Clade 1
Swertia yunnanensis
0.75
Swertia chirayta
Swertia aff pseudohookeri
1
Comastoma pulmonarium
Jaeschkea microsperma
1
Swertia binchuanensis
Swertia macrosperma
0.59
0.89
Swertia cuneata
Gentianella umbellata
1
Swertia punicea
Swertia pubescens
0.65
Swertia ciliata
Swertia hispidicalyx
Gentianella foliosa
Lomatogonium perenne
0.99
Swertia macrosepala subsp. macrosepala_UG2061_2
Swertia subnivalis_KN0004_2
Swertia macrosepala subsp. microsperma_ET0681_2
0.96
Swertia volkensii var. baleensis_ET0680_2
Swertia volkensii var. volkensii_KN
Swertia uniflora_KN0193_2
Swertia volkensii var. volkensii_KN0922_2
1
Swertia schimperi_ET0187_2
Swertia adolfifriderici_TZ0710_2
0.86
Swertia scadens_KN
Swertia kilimandscharica_KN0365_2
1
Swertia kilimandscharica_ET
African Swertia
Clade 2
0.88
Swertia cordata
0.98
Swertia angustifolia
0.97
Swertia perennis
Swertia calycina
0.80
Gentianopsis contorta
Halenia elliptica
0.79
1
Halenia brevicornis
Halenia weddelliana
1
0.79
Swertia tetraptera
0.72
Swertia bimaculata
Veratrilla baillonii
0.97
Swertia delavayi
Swertia decora
0.97
Gentiana_phyllocalyx
Swertia tashiroi
Megacodon stylophorus
Pterygocalyx volubilis
Frasera albomarginata
Gentiana lutea
Gentiana frigida

Fig. 13 Phylogenetic tree inferred from Bayesian analysis of the nuclear ribosomal DNA (ITS) data set. The arrow indicates the Himalayan *Swertia*, which was resolved as sister to clade 2 of the African *Swertia* (ET= Ethiopia, KN= Kenya, TZ= Tanzania, UG= Uganda).

5.1.3 Molecular dating and divergence time

The best fit models of substitution selected were GTR+G for the *mat*K region, GTR for the *trn*L region and GTR+I for the ITS region. The maximum clade credibility dated phylogenetic trees are presented in Figs 12 and 13 summarizing the major node ages. The trees were inferred based on the use of two different calibration points, which resulted in two dating hypotheses. Approach 1 (Figs. 14A, 15A) resulted in older ages at different nodes of the topology than was the case for Approach 2 (Fig. 14B, 15B). The age of the split between the subtribes Gentianinae and Swertiinae ranged from 28.5 Mya (Approach 1; Fig. 14A) to 15.1 (Approach 2; Fig. 14B) whereas, the divergence date of the African *Swertia* ranged between 19.4 Mya (Approach 1; Fig. 14A) to 11.6 Mya (Approach 2; Fig. 14B).

The estimated divergence time for the two African *Swertia* clades (the age of their most recent common ancestor), based on the phylogeny constructed using *mat*K, *trn*L and ITS sequences, was found to be between 14.1 Mya (Middle Miocene) in Approach 1 (Fig. 15A) and 9.07 (Late Miocene) in Approach 2 (Fig. 15B). Clade 1 (*S. crassiuscula-abyssinica* Group) appeared to be older than Clade 2 (*S. kilimandscharia-macrosepala* Group) in both setups. The outcome of Approach 1 (Fig. 15A) indicated that the ancestral ages of both clades were in the Late Miocene, i.e., 8.0 Mya (Clade 1) and 6.9 Mya (Clade 2). The result of divergence date analysis based on three calibration points (Approach 2) showed that the estimated ancestral age of Clade 1 to be 4.8 Mya while that of Clade 2 was 3.9 Mya; both in the Pliocene epoch (Fig. 15B).



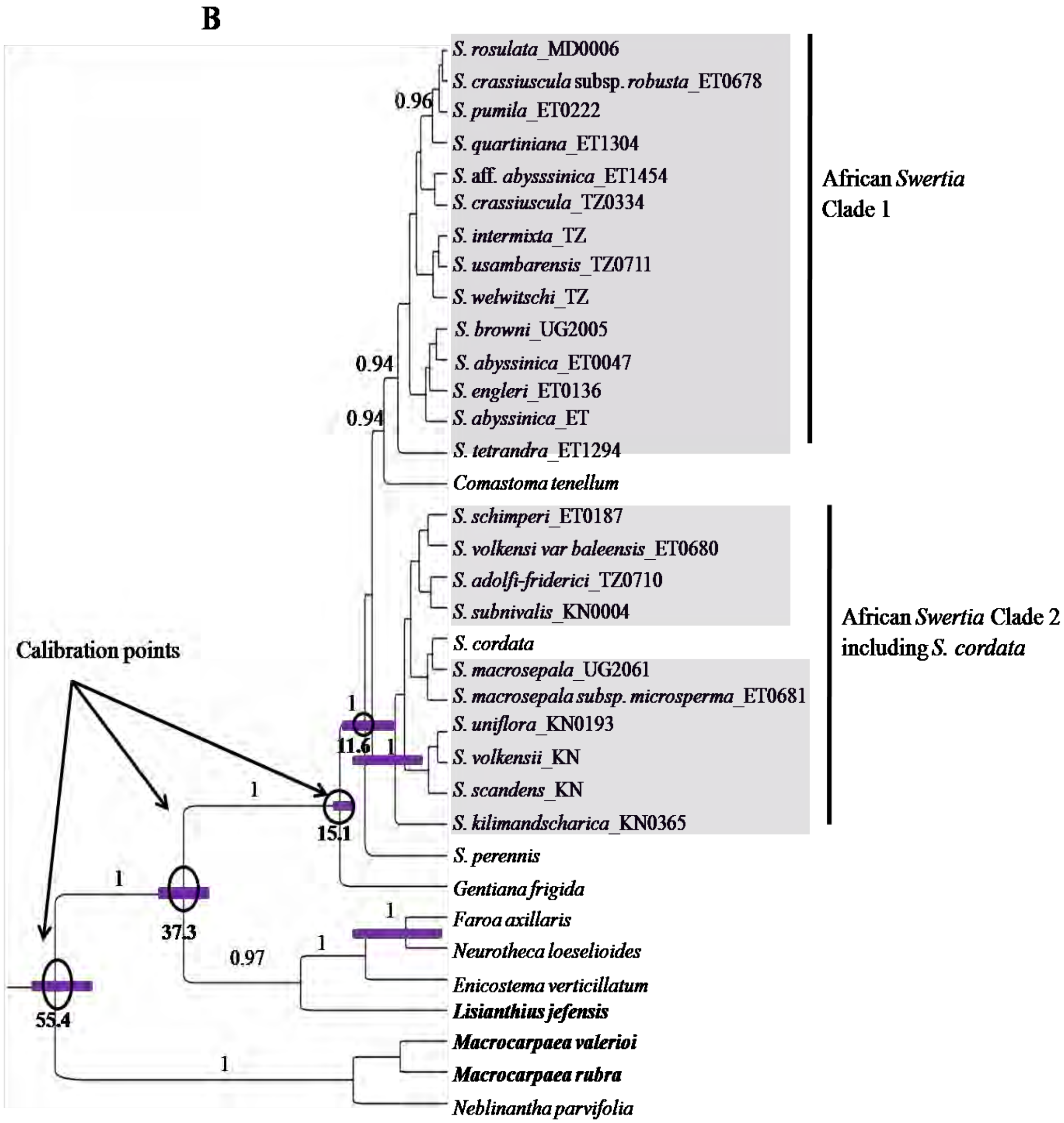
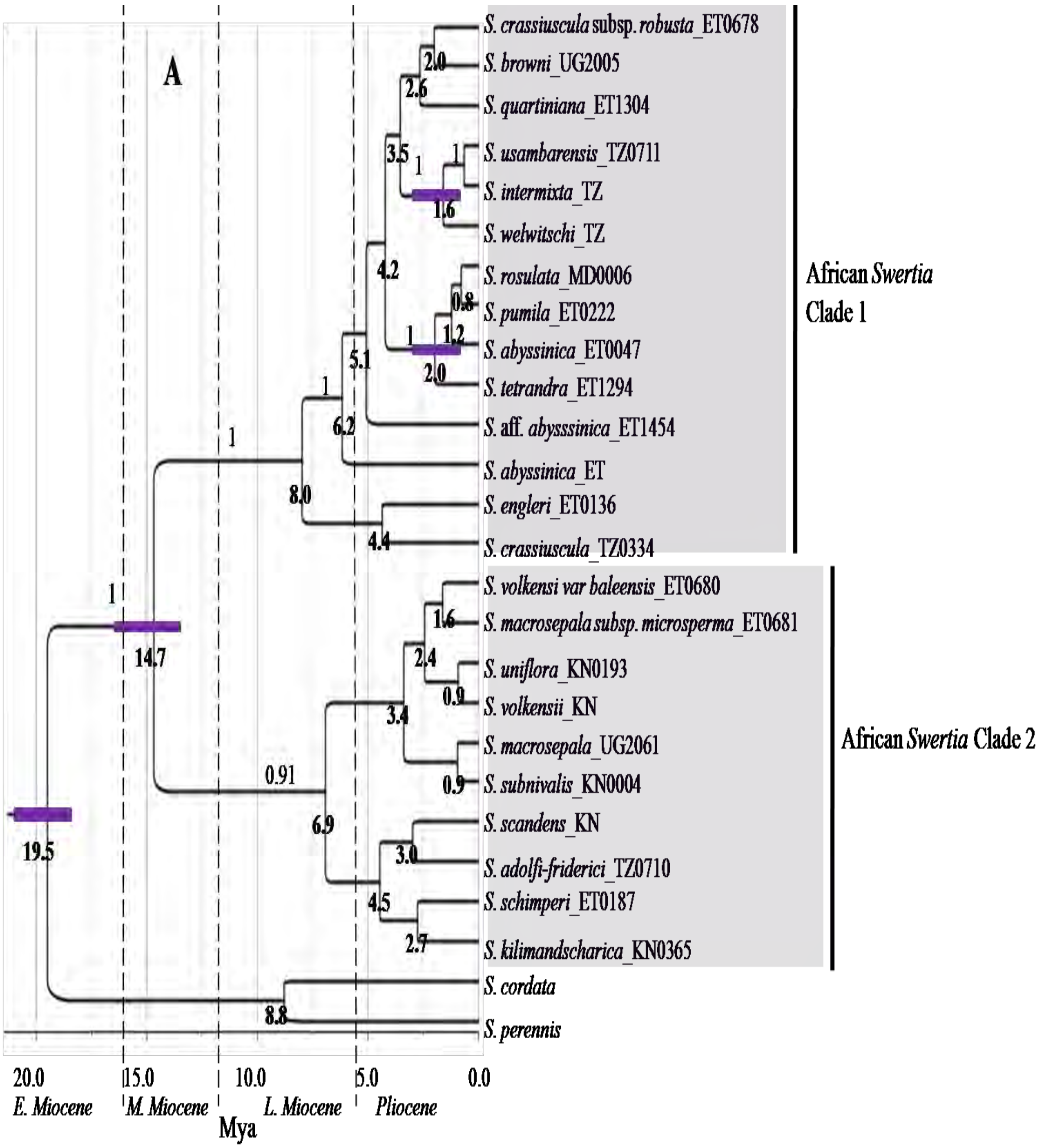


Fig. 14 Maximum credibility dated phylogenetic trees inferred from Bayesian analysis of *mat*K sequences to estimate the divergence time of the African *Swertia*. A) Approach 1: use of two calibration points to estimate the divergence time of the African Swertia. B) Approach 2: use of three calibration points to estimate the divergence time of the African *Swertia*. Posterior probabilities (only PP > 0.9) are shown above branches and estimated ages of the respective node are shown below branches „Node bars" indicate the confidence intervals on each node.


A
S. crassiuscula subsp. robusta_ET0678
S. browni_UG2005
S. quartiniana_ET1304
S. usambarensis_TZ0711
S. intermixta_TZ
S. welwitschi_TZ
S. rosulata_MD0006
S. pumila_ET0222
S. abyssinica_ET0047
S. tetrandra_ET1294
S. aff. abysssinica_ET1454
S. abyssinica_ET
S. engleri_ET0136
S. crassiuscula_TZ0334
S. volkensi var baleensis_ET0680
S. macrosepala subsp. microsperma_ET0681
S. uniflora_KN0193
S. volkensii_KN
S. macrosepala_UG2061
S. subnivalis_KN0004
S. scandens_KN
S. adolfi-friderici_TZ0710
S. schimperi_ET0187
S. kilimandscharica_KN0365
S. cordata
S. perennis
African Swertia Clade 1
African Swertia Clade 2
2.0
2.6
3.5
1
1
1.6
4.2
0.8
1
1.2
2.0
5.1
1
6.2
1
8.0
4.4
1
14.7
1.6
2.4
0.9
3.4
0.9
0.91
6.9
3.0
4.5
2.7
19.5
8.8
20.0
15.0
10.0
5.0
0.0
E. Miocene
M. Miocene
L. Miocene
Pliocene
Mya

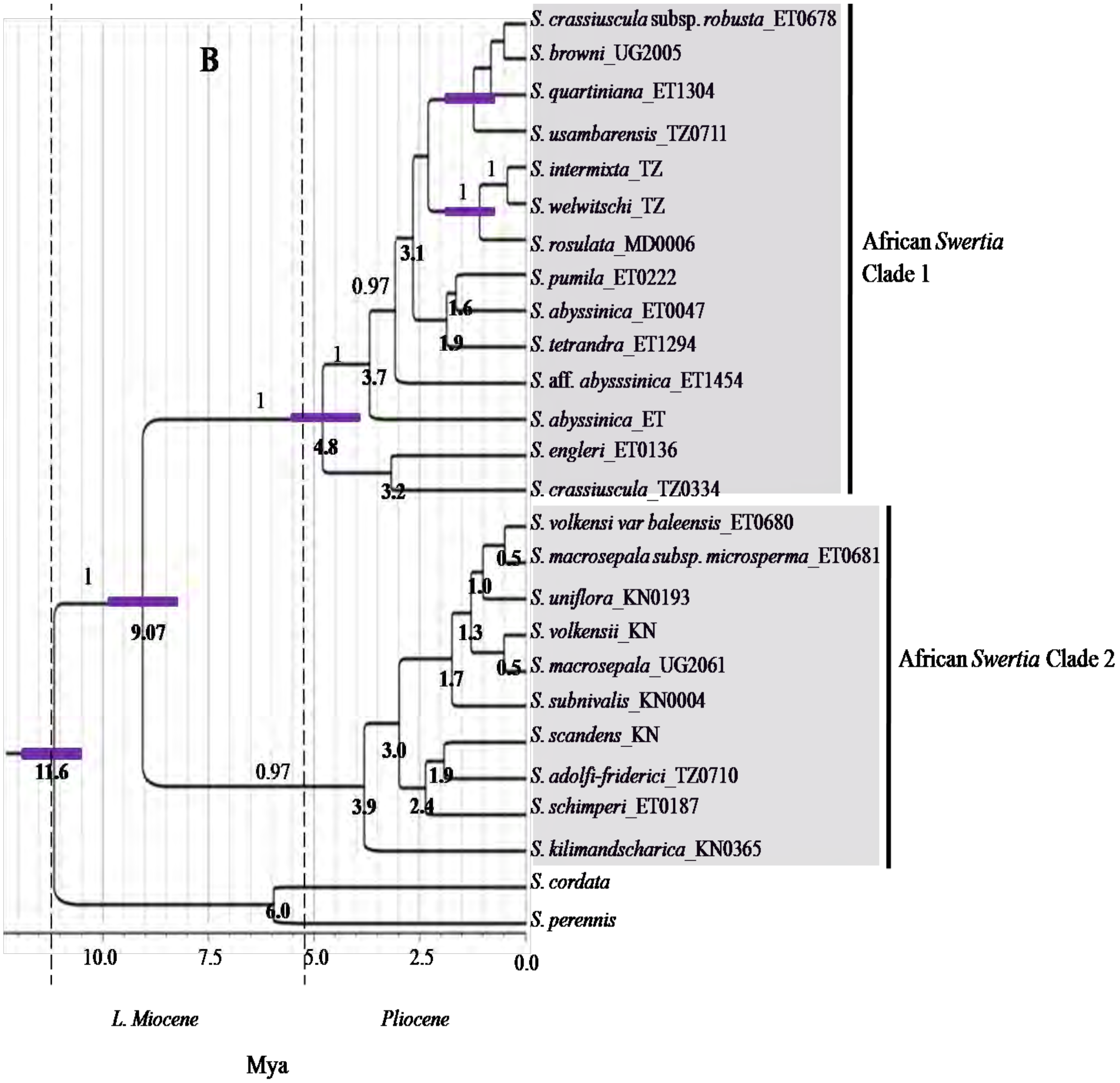


Fig. 15 Maximum credibility dated phylogenetic trees inferred from Bayesian analysis of the three concatenated regions of DNA (*mat*K, trnL and ITS) to estimate the ancestral ages of the African *Swertia*. A) Based on the estimated divergence time from Approach 1. B) Based on the estimated divergence time from Approch 2. Posterior probabilities (only PP > 0.9) are shown above branches and estimated ages of the respective node are shown below branches. „Node bars‟ indicate the confidence intervals on each node.

5.2 Results from comparative phylogeography and genetic diversity study

Carduus schimperi

A total of 119 individuals from 24 populations of *C. schimperi* were used for the final AFLP analysis. A total of 541 markers were produced by employing the final selections of primer combinations, of which 241 (44.55%) were polymorphic. The markers were 98.3% reproducible. The PcoA extracted 19% of the variation along the first axis and 11 % along the second axis, producing three groups (Fig. 16). Group 1 (G1) comprised populations from Mt. Kenya, Group 2 (G2) comprised populations from Mt. Elgon, Mt. Aberdare and the Bale Mts., and Group 3 (G3) comprised populations from the Simen Mts. (Fig. 16). G1 and G3 represented populations referred to subsp. *platyphylus* and subsp. *schimperi*, respectively. G2 represented populations from two subspecies, subsp. *nanus* (from Mt. Elgon and Aberdare) and subsp. *schimperi* (from the Bale Mts.). In G2, the populations from the Bale Mts. tended to cluster together as did those from Mt. Aberdare and Mt. Elgon.

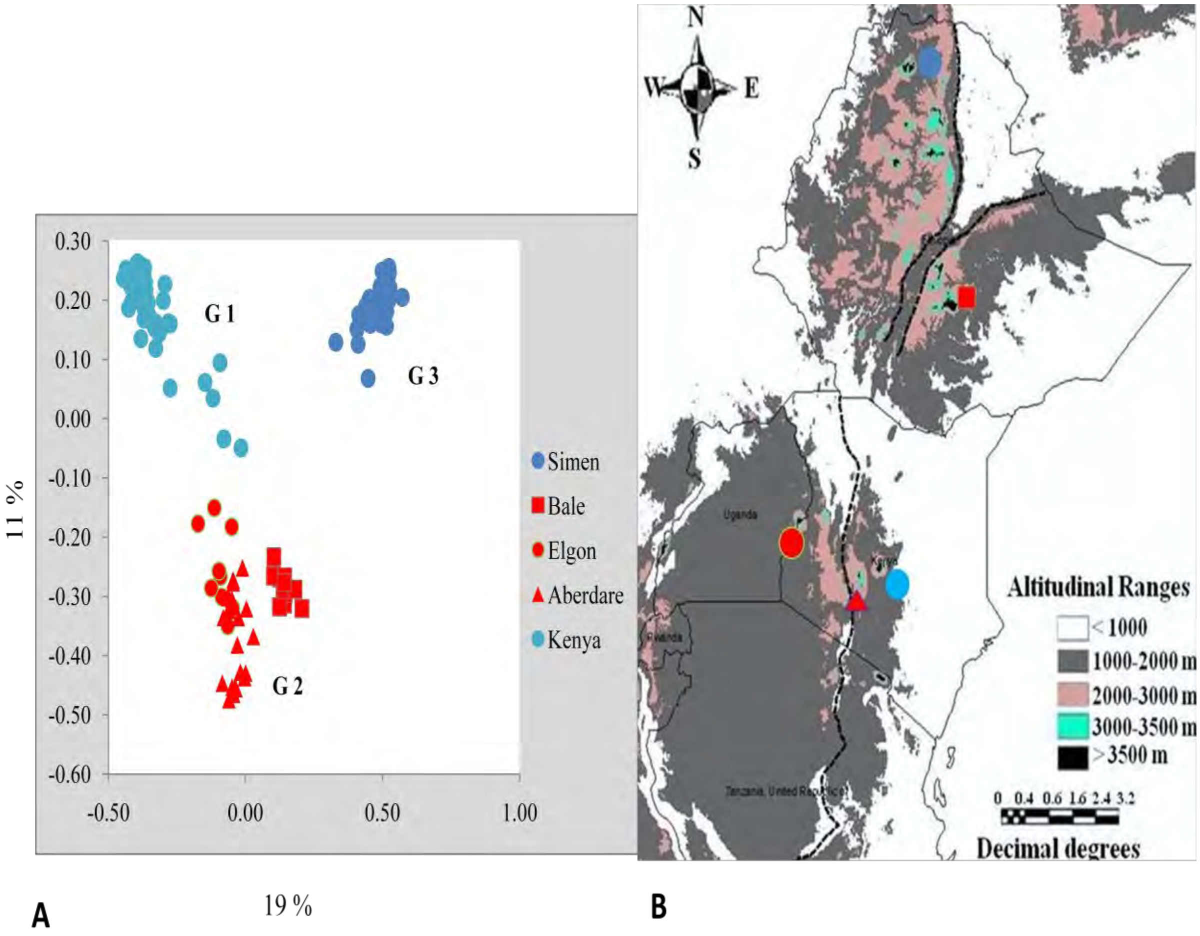


Fig. 16 (A) PcoA showing three groups of *C. schimperi*. (B) Geographical origin of populations with colours corresponding to the groups in the PcoA and shapes indicating mountains

The same grouping was also retrieved in the neighbor-joining tree (Fig. 17). G2 and G3 were supported with bootstrap values of 98 (Fig. 17B, C).

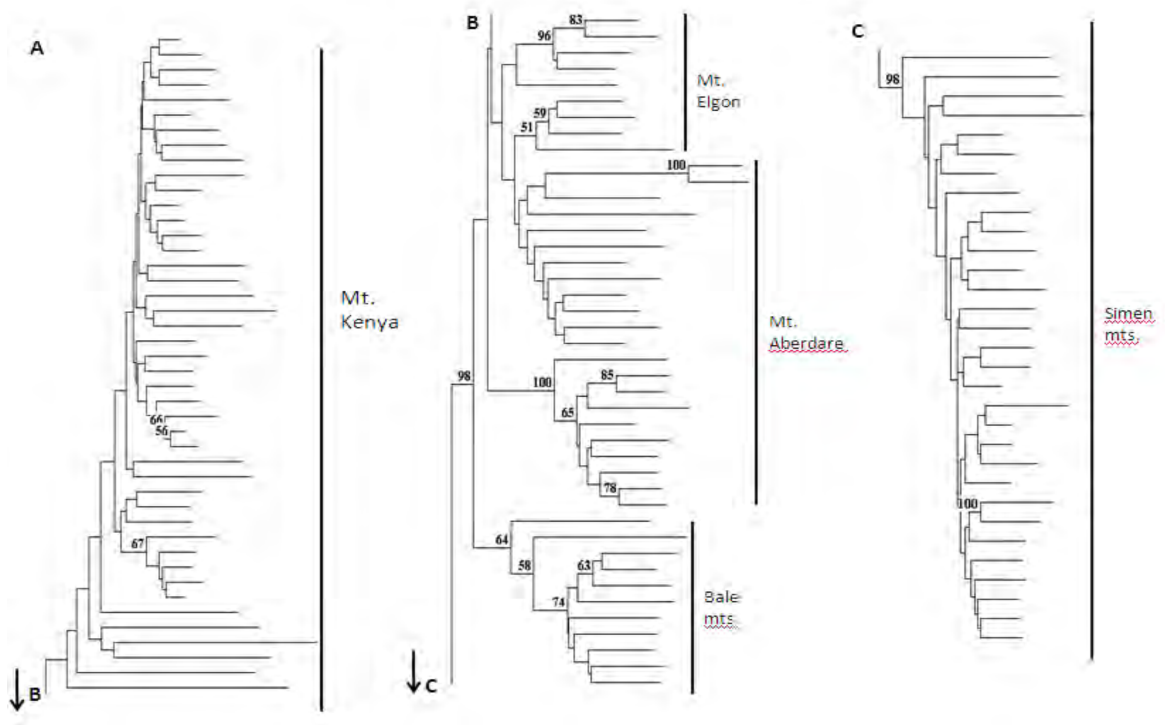


Fig. 17A-C: NJ tree showing the relationship among *C. schimperi* populations collected from four mountains in Ethiopia and Kenya

The structure analysis yielded three groups (K = 3; Fig. 18A, B) identical to those inferred from the PcoA with the single exception of one individual from Mt. Kenya. This individual (KN1009-2) was assigned to a different group (Group 2) together with samples from Mt. Aberdare, Mt. Elgon and Bale Mts. In the AFLPOP analysis, this individual was recognized as the hybrid G 1 X G 2 with a likelihood 100 times higher than to any one of the source populations (G 1 & G 2) and a P value of less than 0.05.

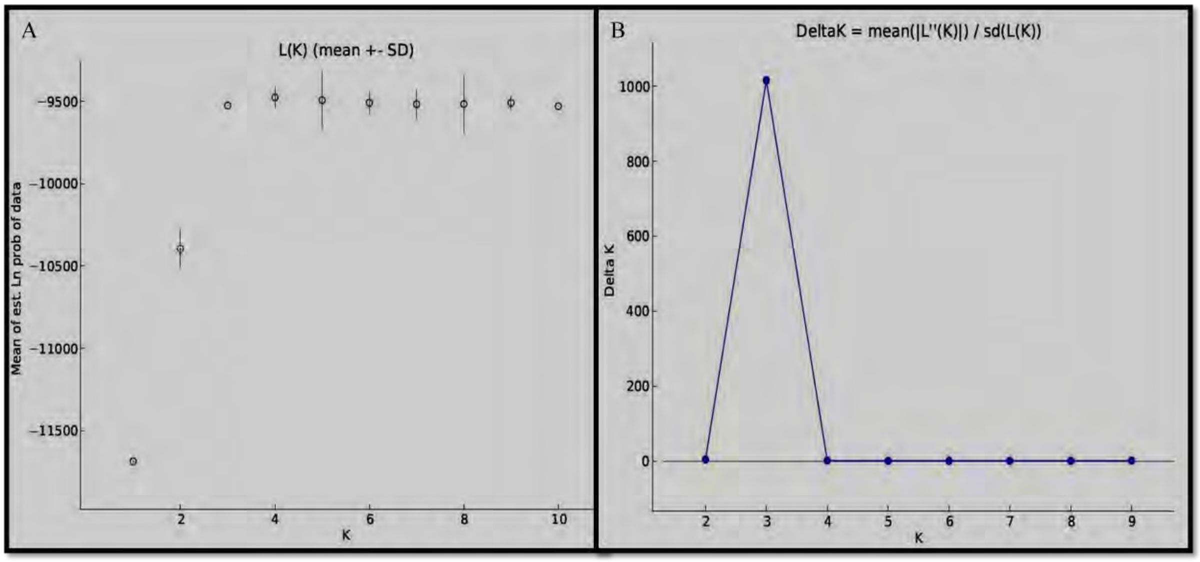


Fig. 18 (A) Log probability of data L(K) of *C. schimperi*, as a function of K for 10 structure runs; (B) Rate of change in the probability between successive runs, delta K, as a function of K.

Table 5 presents the average estimates of gene diversity, Shannon‟s diversity index and rarity in each population. The average gene diversity (h) and Shannon‟s diversity index (I) among individuals of a population were 0.57 and 0.55, respectively. A population from Mt. Elgon (KN250) showed the highest gene diversity followed by two populations from Mt. Kenya (KN807 & KN979) and a population from the Bale Mts. (ET1426). The lowest gene diversity was recorded in three populations from Mt. Kenya (KN0904, KN0870 and KN1035). A low level of gene diversity was observed in the populations from the Simen Mts, while intermediate levels of gene diversity were recorded in populations sampled from Mt. Aberdare. The population with highest genetic distinctiveness (DW) was recorded from Mt. Elgon (KN0250) while the population with the lowest DW value was from the Simen Mts. (ET0024).

Table 5 Gene diversity, rarity and number of individuals in populations of *C. schimperi*

Population name	**Locality**	**No. of individuals**	**h (gene diversity)**	**I (Shannon's index)**	**DW (Rarity of AFLP markers)**
ET0024	Simen Mts.	5	0.036229205	0.0377	2.007343
ET0117	Simen Mts.	5	0.039926063	0.0429	2.382602
ET0270	Simen Mts.	4	0.042821935	0.0414	2.874949
ET0392	Simen Mts.	5	0.05767098	0.0572	3.608225
ET0487	Simen Mts.	5	0.037338262	0.0388	2.00266
ET0618	Simen Mts.	3	0.043130006	0.0361	3.724105
ET1426	Bale Mts.	6	0.071719039	0.0748	5.27982
ET1480	Bale Mts.	4	0.057917437	0.0543	5.532643
KN0250	Mt. Elgon	4	0.12138016	0.1127	10.1416
KN0299	Mt. Elgon	5	0.060998152	0.0582	5.56407
KN0606	Mt. Aberdare	5	0.06987061	0.0644	5.158187
KN0683	Mt. Aberdare	5	0.052495379	0.0547	2.927306
KN0689	Mt. Aberdare	5	0.05767098	0.055	5.466223
KN0707	Mt. Aberdare	5	0.063216266	0.0608	5.820315
KN0807	Mt. Kenya	5	0.072828096	0.0722	4.461941
KN0828	Mt. Kenya	5	0.063216266	0.0646	3.758772
KN0870	Mt. Kenya	4	0.033271719	0.0288	2.621927
KN0904	Mt. Kenya	4	0.032347505	0.0286	2.104462
KN0928	Mt. Kenya	5	0.055083179	0.0596	2.579319
KN0960	Mt. Kenya	4	0.070240296	0.0642	4.992738
KN0979	Mt. Kenya	5	0.074676525	0.0709	4.838098
KN1009	Mt. Kenya	4	0.070856439	0.0674	2.863058
KN1032	Mt. Kenya	4	0.049907579	0.0479	3.281866
KN1035	Mt. Kenya	3	0.035736291	0.0282	2.873505

In the hierarchical AMOVA, 34.42% of the variation was obtained among the three groups, 14.31% among populations within groups and 51.27% within populations (Table 6).

Table 6 Analysis of molecular variance (AMOVA) for AFLP diversity in *C. schimperi*

Source of variation	d.f.	Sum of squares	Variance of components	Percentage of variation
Among three groups	2	817.99	10.45	34.42
Among populations within groups	21	739.93	4.35	14.31
Within populations	85	1323.42	15.57	51.27
Total	108	2881.34	30.37	

Trifolium cryptopodium

A total of 89 individuals from 19 populations of *T. cryptopodium* was used in the final AFLP analysis. The three selected primer combinations produced a total of 297 markers, of which 180 (68.7%) were polymorphic. The markers were 98.3% reproducible. In the PcoA, 16% of the total variation was accounted for by the first axis and 6% by the second axis, resulting in two groups. Group 1 (G1) was composed of the populations from the Simen Mts. and Mt. Choke, and Group 2 (G2) contained the populations from Mt. Elgon, Mt. Aberdare and the Bale Mts. (Fig. 19). Similar groupings were obtained in the neighbor-joining tree although the bootstrap support was low, BP=50 (Fig. 20).

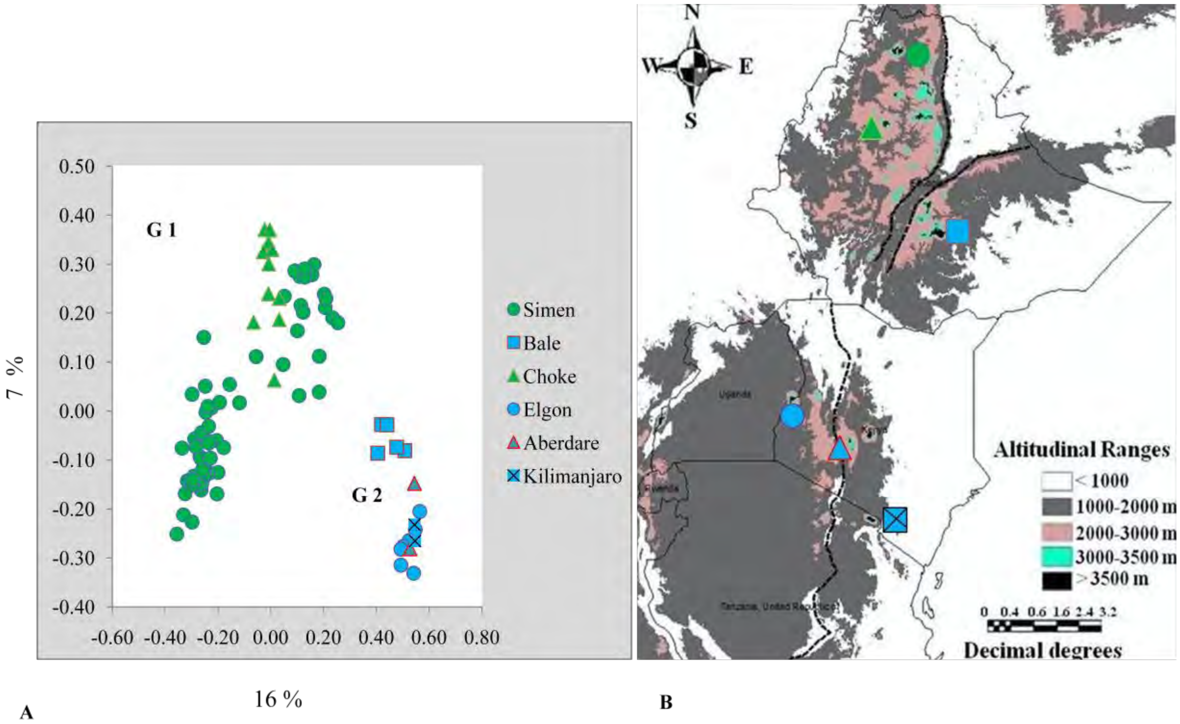


Fig. 19 (A) PcoA showing two groups in *T. cryptopodium*. (B) Geographical origin of the populations with colours corresponding to the PcoA groups and shapes representing mountains

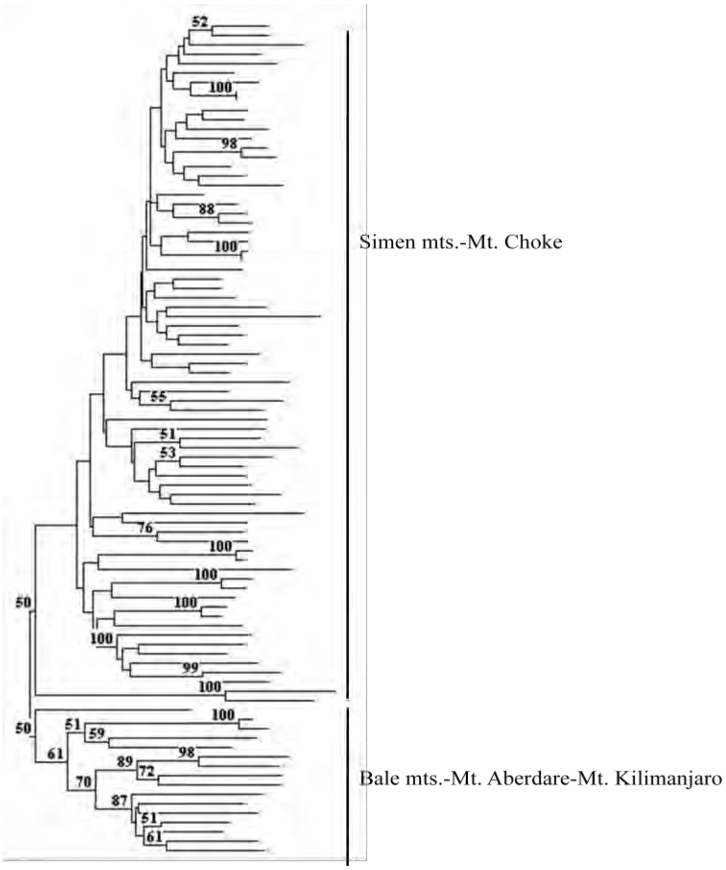


Fig. 20 NJ tree showing the relationships among the *T. cryptopodium* populations collected from five mountains of Ethiopia and East Africa

The structure analysis identified two groups as in the PcoA (K = 2; Fig. 21). Exceptions were two individuals from Simen Mts. Following this analysis, these individuals were subjected to a hybrid test. They were identified as F1 hybrids between the populations of the two groups (G 1 X G 2) with likelihood 100 times higher than to any one of the source populations (G 1 & G 2) and a P value of less than 0.05.

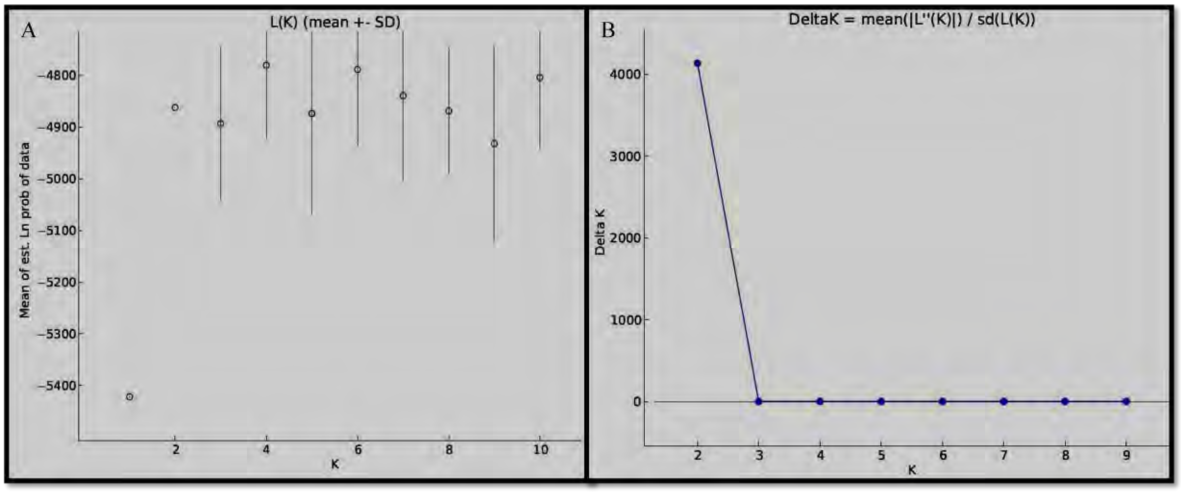


Fig. 21 (A) Log probability of data L(K) of *T. cryptopodium* as a function of K for 10 structure runs; (B) Rate of change in the probability between successive runs, delta K, as a function of K.

The average estimated gene diversity (h) among individuals of a population was 0.79, while the Shannon‟s diversity index was 0.68. A population from the Bale Mts. (ET0829) exhibited the highest genetic and Shannon‟s diversity values, followed by one from Mt. Aberdare (KN0488) and one population from Mt. Elgon (KN359). The lowest values were recorded for a population from the Simen Mts., ET0297 (Table 7). The largest DW value was recorded in a population from Mt Aberdare (KN0488), and the lowest DW value was recorded in the same population from the Simen Mts. (ET0297).

Table 7 Gene diversity, rarity and no of individuals analyzed from the populations of *T. cryptopodium*

Population name	**Locality**	**No. of individuals**	**h (gene diversity)**	**I (Shannon's index)**	**DW (Rarity)**
ET0059	Simen Mts.	9	0.094269	0.0997	2.933798
ET0115	Simen Mts.	4	0.085302	0.0753	2.589314
ET0127	Simen Mts.	4	0.062992	0.0541	2.293546
ET0232	Simen Mts.	3	0.112861	0.0928	5.332124
ET0297	Simen Mts.	3	0.036745	0.0303	0.840096
ET0374	Simen Mts.	5	0.056693	0.0564	1.289203
ET0409	Simen Mts.	5	0.064567	0.0611	1.750361
ET0466	Simen Mts.	5	0.052756	0.0515	1.742797
ET0567	Simen Mts.	5	0.041732	0.0385	1.238817
ET0829	Bale Mts.	4	0.129921	0.1249	5.988421
ET1362	Mt. Choke	5	0.075591	0.0688	3.189426
ET1374	Mt. Choke	2	0.047244	0.0286	2.314695
ET1394	Mt. Choke	4	0.086614	0.0769	2.631397
KN0322A	Mt. Elgon	3	0.094488	0.0772	5.319332
KN0359	Mt. Elgon	3	0.115486	0.0955	5.832665
KN0488	Mt. Aberdare	2	0.122047	0.0738	7.401181
TZ0340	Mt. Kilimanjaro	2	0.070866	0.0429	7.063846

An AMOVA revealed that more than 50 % of the variation was due to within population variation, slightly over 30 % among groups, and 15 % among populations within groups (Table 8).

Table 8 Analysis of molecular variance (AMOVA) for AFLP diversity in *T. cryptopodium*

Source of variation	**d.f.**	**Sum of squares**	**Variance of components**	**Percentage of variation**
Between the two groups	1	166. 20	6.52	30.79
Among populations within each group	17	389.57	3.20	15.11
Within populations	49	561.200	11.45	54.10
Total	67	1116.97	21.17	

Geranium arabicum* and *Geranium sp. nov.

After data cleaning, a total of 243 samples were used in the final analysis. The three selected primers produced a total of 541 markers, of which 306 (56.56%) were polymorphic. The markers were 98.1% reproducible. Two genetic groups were distinguished in the PcoA in which 13% of the total variation was explained by the first axis and 11% by the second axis. Group 1 included most of the East African samples of *G. arabicum* (Mt. Kenya, Mt. Aberdare, Mt. Elgon, Mt. Meru and Mt. Kilimanjaro) while G2 comprised all populations of *G. arabicum* from the Simen Mts., Mt. Choke, and the Bale Mts. in Ethiopia, and Mt. Muhavura and Mt. Ruwenzori in Uganda (Fig. 22). Furthermore, one population from Mt. Kilimanjaro grouped with G2.

Four populations of *Geranium* sp. *nov.* = *G. S. & Miehe*. 3002 were analysed. Whereas one complete population and four individuals from the remaining three populations were placed within Group 1, the remaining fourteen individuals from the three populations were placed within Group 2 (Fig. 22).

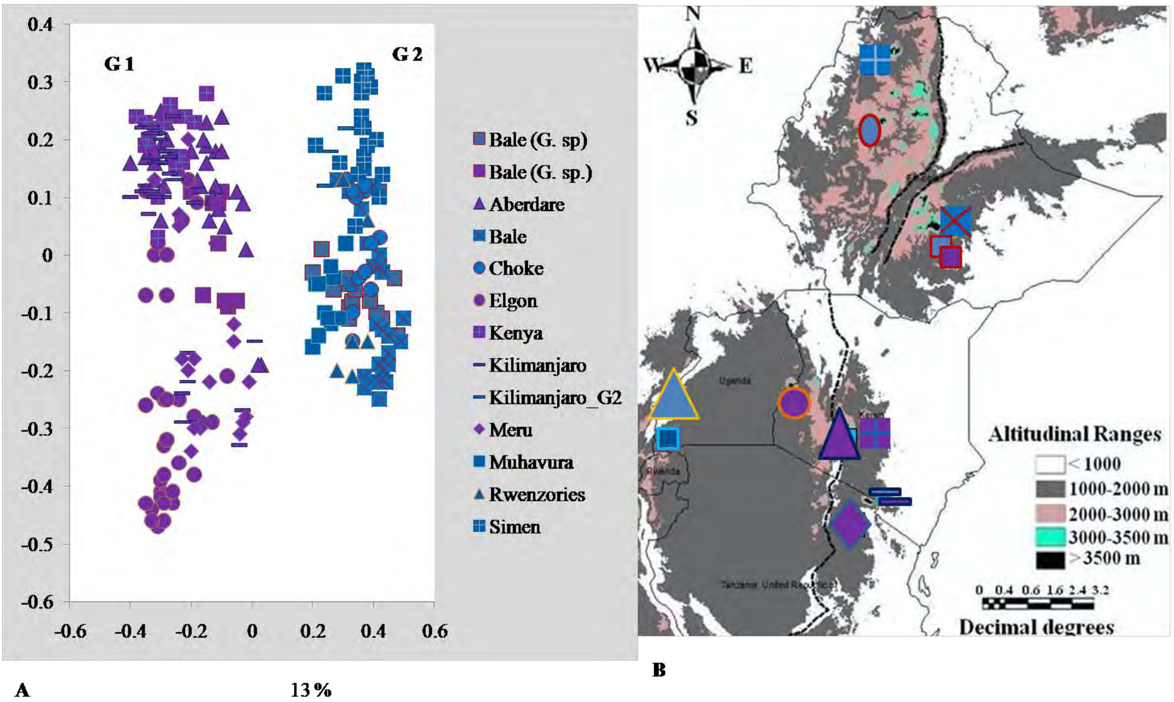


Fig. 22 (A) PcoA showing two groups of *G. arabicum* and *Geranium sp. nov.* (B) Geographical origin of the populations with colours corresponding to the PcoA groups and shapes indicating mountains/*Geranium* sp.

The NJ tree demonstrated the same two groups, but with poor bootstrap support (BP values less than 50, not shown). The populations of *Geranium* sp. nov. were placed similarly in the NJ tree (Fig. 23A-F).

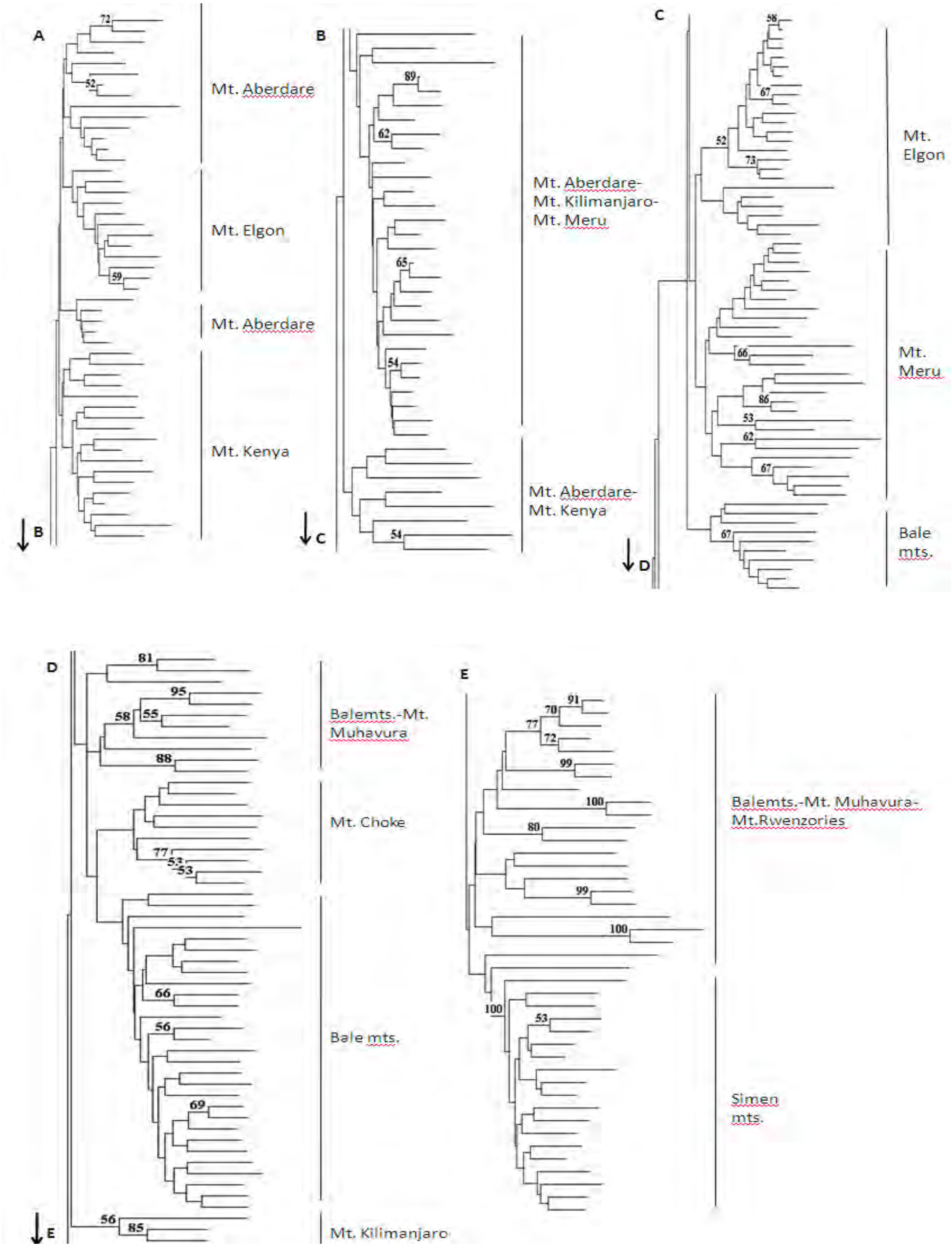


Fig. 23 A-F NJ tree showing the relationship among the *G. arabicum* and *Geranium* sp. nov. populations collected from four mountains of Ethiopia and East Africa

The STRUCTURE analysis resolved the same two groups (K = 2; see Fig. 24) as seen in the PcoA. The three populations of *Geranium* sp. nov. in which individuals were assigned to two different groups, were subjected to a hybridization test. None of the these individuals were grouped with either G1 or G2 with a likelihood 100 times higher than the source populations (G1 & G2) with a P value ≤ 0.05. No individuals were grouped with either source population even if the likelihood value was lowered to 10 times higher than the source populations with a P value ≤ 0.05.

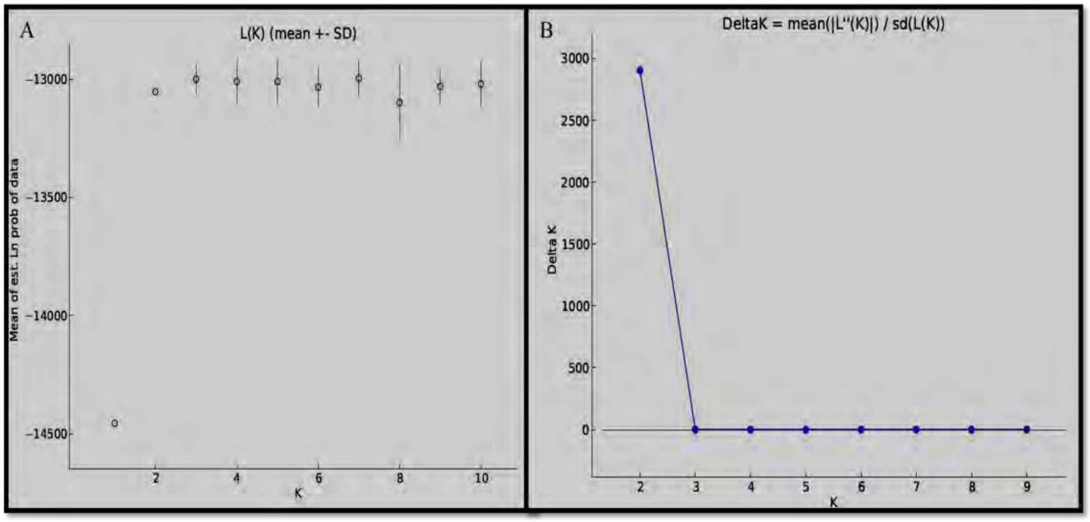


Fig. 24 (A) Log probability of data L(K) of *G. arabicum* and *Geranium* sp. nov. as a function of K for 10 structure runs, B) Rate of change in the probability between successive runs, delta K, as a function of K

The average values of gene diversity were h = 0.036 and I = 0.033. The most diverse population was recorded on Mt. Muhavura (UG2251; h = 0.097; I = 0.087) and the diversity decreased from Mt. Meru (TZ0503), the Bale Mts. (ET0974), to Mt. Aberdare (KN0464). A population from Mt. Elgon (KN0028) exhibited the lowest values of diversity (h = 0.0095; I = 0.0096). The population from Mt. Muhavura that exhibited highest gene diversity also had the highest DW value while a population from Mt. Kilimanjaro (TZ209) exhibited the lowest index of rarity (Table 9).

Table 9 Gene diversity, rarity and no of individuals analysed in the populations of *G. arabicum* and *Geranium* sp. nov.

Population name	**Locality**	**No. of individuals**	**h (gene diversity)**	**I (Shannon's index)**	**DW (Rarity)**
ET0036	Simen Mts.	3	0.060367	0.0448	1.875911
ET0109	Simen Mts.	5	0.040157	0.0416	1.028938
ET0139	Simen Mts.	5	0.049606	0.0466	1.062208
ET0250	Simen Mts.	5	0.025984	0.0264	0.788038
ET0677*	Bale Mts.	5	0.029134	0.0259	1.147298
ET0827	Bale Mts.	5	0.030709	0.0308	0.735396
ET0954*	Bale Mts.	5	0.035433	0.033	0.680015
ET0974*	Bale Mts.	5	0.065354	0.0575	1.896974
ET1030*	Bale Mts.	5	0.044094	0.0406	1.104513
ET1373	Mt. Choke	5	0.029921	0.0308	0.674145
ET1403	Mt. Choke	5	0.058268	0.0544	1.515572
ET1420	Bale Mts.	5	0.06063	0.0567	1.542758
ET1447	Bale Mts.	5	0.038583	0.0333	1.246848
ET1476	Bale Mts.	5	0.033858	0.0312	1.535352
ET1485	Bale Mts.	4	0.019029	0.0168	0.632182
KN0028	Mt. Elgon	5	0.009449	0.0096	0.19211
KN0097	Mt. Elgon	3	0.013123	0.0102	0.337691
KN0180	Mt. Elgon	4	0.019029	0.0156	0.44037
KN0225	Mt. Elgon	5	0.020472	0.0221	0.290392
KN0256	Mt. Elgon	5	0.017323	0.0176	0.549008

KN0279	Mt. Elgon	4	0.019029	0.0168	0.419702
KN0320	Mt. Elgon	4	0.026247	0.0233	0.696716
KN0464	Mt. Aberdare	4	0.062336	0.0577	1.874277
KN0480	Mt. Aberdare	5	0.044094	0.0401	0.870395
KN0484	Mt. Aberdare	3	0.041995	0.0324	1.07391
KN0542	Mt. Aberdare	5	0.02126	0.0199	0.940045
KN0643	Mt. Aberdare	3	0.020997	0.0174	0.286637
KN0647	Mt. Aberdare	5	0.028346	0.0243	0.811817
KN0651	Mt. Aberdare	5	0.016535	0.0145	0.345627
KN1037	Mt. Kenya	4	0.020341	0.0175	1.102137
KN1089	Mt. Kenya	11	0.017609	0.0234	0.435252
TZ0010	Mt. Kilimanjaro	5	0.043307	0.0394	0.958584
TZ0041	Mt. Kilimanjaro	5	0.01811	0.0168	0.630973
TZ0209	Mt. Kilimanjaro	3	0.010499	0.0083	0.172849
TZ0305	Mt. Kilimanjaro	5	0.033858	0.0312	0.848899
TZ0316	Mt. Kilimanjaro	3	0.044619	0.0351	0.770384
TZ0327	Mt. Kilimanjaro	5	0.011024	0.0093	0.388584
TZ0384	Mt. Meru	5	0.041732	0.0415	0.78384
TZ0461	Mt. Meru	5	0.04252	0.0393	0.844285
TZ0501	Mt. Meru	4	0.059711	0.0497	1.825089
TZ0503	Mt. Meru	5	0.073228	0.065	1.827502
TZ0809	Mt. Kilimanjaro	4	0.035433	0.0314	0.485742
UG2215	Mt. Gahinga	5	0.051181	0.0488	1.118576
UG2251	Mt. Muhavura	5	0.09685	0.0871	2.659463
UG2388	Ruwenzori Mts.	5	0.027559	0.0277	0.732888

***Geranium sp. nov. = G. S. Miehe* 3002**

The AMOVA showed that most of the variation (53%) was contained within populations, while equivalent amounts were present among populations within groups (25%) and between groups (22%; Table 10).

Table 10 Analysis of molecular variance (AMOVA) for AFLP diversity in *G. arabicum* and *Geranium* sp. nov.

Source of variation	**d.f.**	**Sum of squares**	**Variance of components**	**Percentage of variation**
Between two groups	1	346.37	3.53	21.87
Among populations within groups	40	1092.32	4.02	24.86
Within populations	154	1325.90	8.61	53.27
Total	195	2764.58	16.16	

Swertia abyssinica

The final AFLP dataset for *S. abyssinica* included a total of 152 samples from 28 populations. The two primer combinations produced a total of 363 markers of which 134 (36.95%) were polymorphic. The reproducibility of the markers was 98.3%. In the PcoA, only 5% of the total variation was accounted for by the first axis and 4% by the second axis. Samples were not partitioned into distinct groups (Fig. 25). Similarly, in the NJ tree (Fig. 26) and STRUCTURE analyses (Fig. 27), no genetic groups were identified.

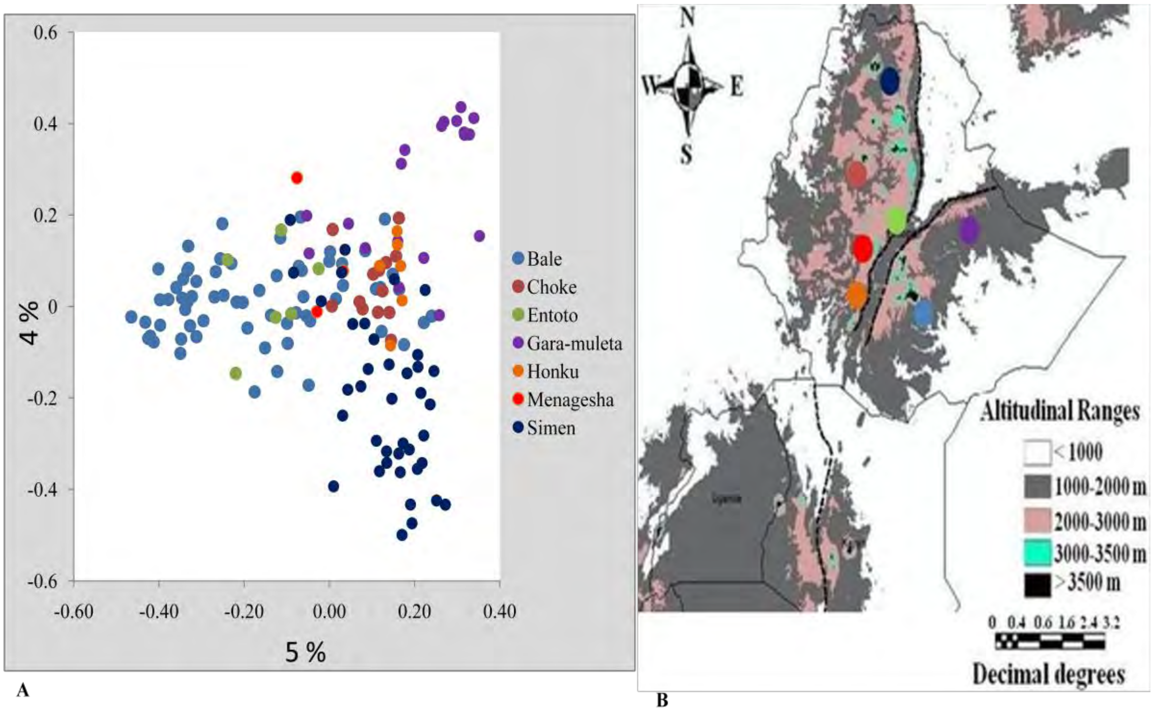


Fig. 25 (A) PcoA of *S. abyssinica*, (B) Geographical origin of the populations with colours representing mountains

The average diversity within populations was 0.094 (h) and 0.084 (I). The population with the highest gene diversity was from Mt. Choke (ET1393) and that with the lowest from the Bale Mountains, ET1414 (Table 11). The AMOVA (Table 12) showed that 80% of the total variation was contained within populations.

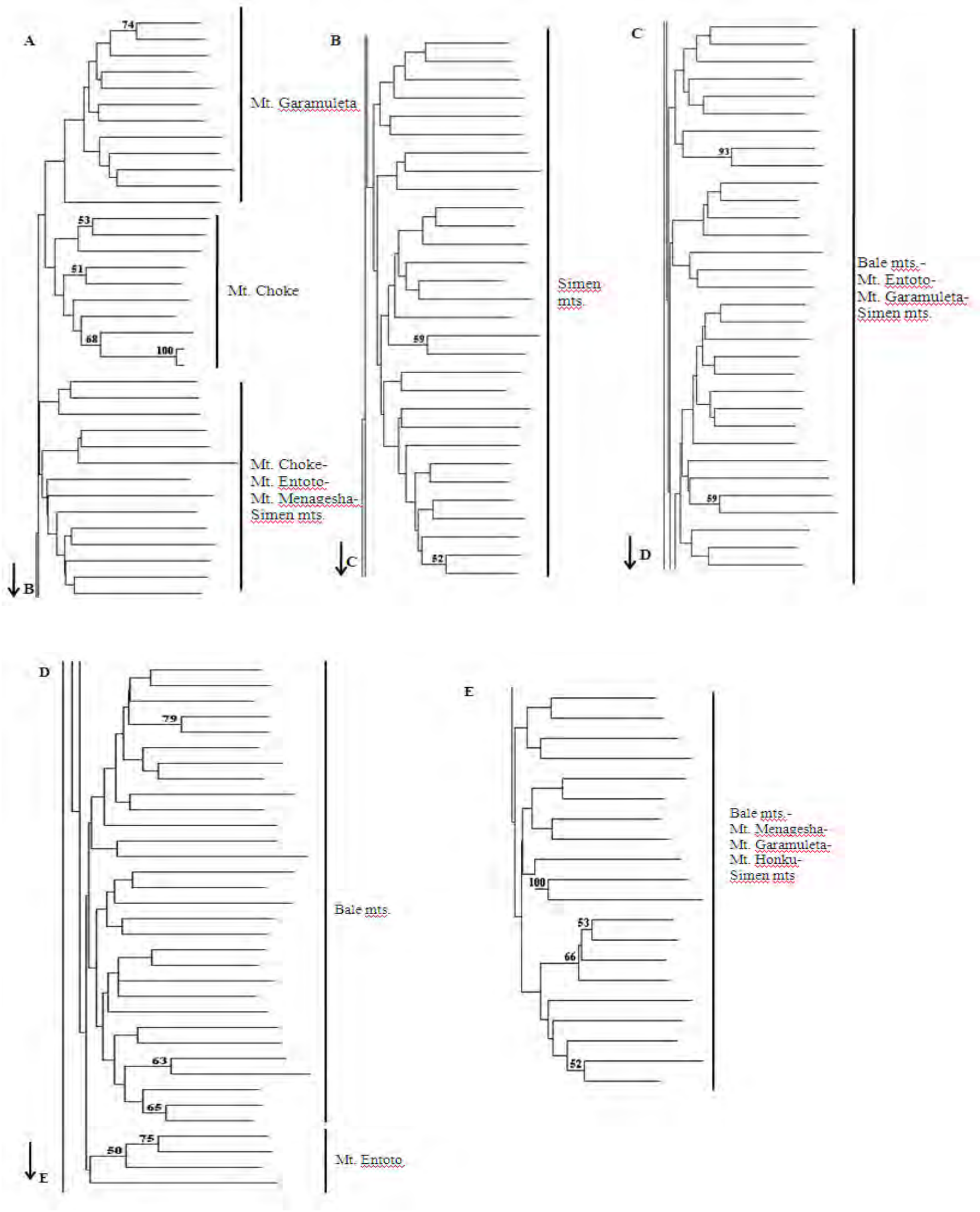


Fig. 26 NJ tree showing the relationship among the *S. abyssinica* populations collected from seven mountains of Ethiopia

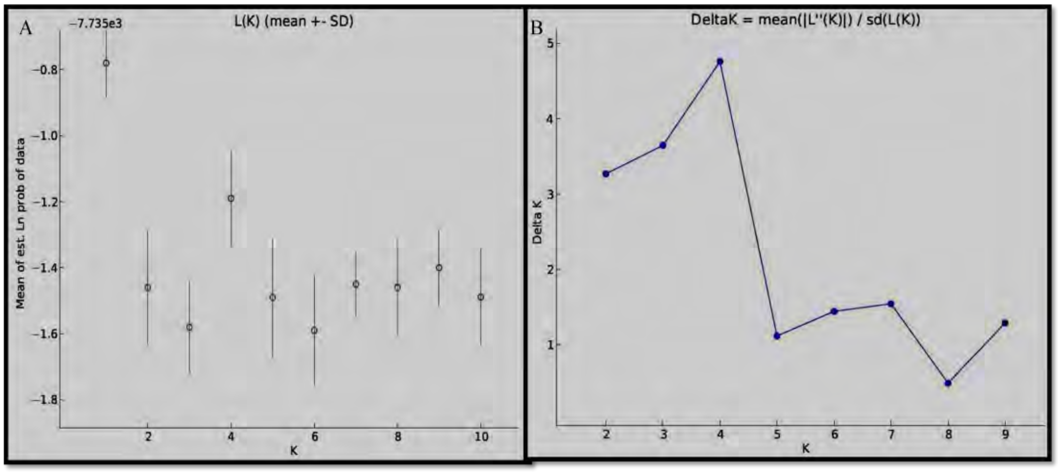


Fig. 27 (A) Log probability of data L(K) for *S. abyssinica*, as a function of K for 10 structure runs, (B) Rate of change in the probability between successive runs, delta K, as a function of K.

Table 11 Gene diversity, rarity and no of individuals analysed of *S. abyssinica*

Population name	**Locality**	**No. of individuals**	**h (gene diversity)**	**I (Shannon's index)**	**DW (Rarity)**
ET0047	Simen Mts.	5	0.122047	0.1095	2.588176
ET0065	Simen Mts.	5	0.095276	0.0862	1.473417
ET0121	Simen Mts.	5	0.069291	0.0628	1.549741
ET0176	Simen Mts.	4	0.108268	0.0928	2.295734
ET0268	Simen Mts.	3	0.062992	0.0491	0.916539
ET0314	Simen Mts.	4	0.077428	0.0688	1.723409
ET0336	Simen Mts.	4	0.072178	0.0629	1.142034
ET0393b	Simen Mts.	3	0.097113	0.0751	1.392777
ET0529	Bale Mts.	5	0.11811	0.1052	1.77031
ET0653	Bale Mts.	5	0.076378	0.0712	0.900139
ET0821	Bale Mts.	4	0.073491	0.0648	1.252117
ET0938	Bale Mts.	4	0.113517	0.0978	1.780032
ET0991	Bale Mts.	4	0.097113	0.082	1.883253
ET1031	Bale Mts.	5	0.092126	0.0852	1.364701
ET1297	Menagesha Mts.	2	0.11811	0.0714	2.882264

ET1299	Entoto Mts.	5	0.114173	0.1047	2.48255
ET1309	Mt. Choke	5	0.124409	0.1244	3.630246
ET1357	Mt. Choke	4	0.089895	0.079	1.83834
ET1376	Mt. Choke	4	0.105643	0.0991	2.221883
ET1393	Honku peak.	5	0.134646	0.1266	3.34979
ET1414	Bale Mts.	4	0.061024	0.0544	1.18884
ET1436	Bale Mts.	5	0.092913	0.0846	1.601673
ET1468	Bale Mts.	4	0.064304	0.0548	0.799979
ET1502	Bale Mts.	5	0.107087	0.1011	2.516576
ET1506	Bale Mts.	5	0.08189	0.0754	1.764683
ET1508	Bale Mts.	4	0.091207	0.0783	1.942879
ET1516	Bale Mts.	5	0.092913	0.0914	2.23832
ET1527	Gara Muleta	19	0.081181	0.0908	1.467469

Table 12 Analysis of molecular variance (AMOVA) for AFLP diversity in *S. abyssinica*

Source of variation	**d.f.**	**Sum of squares**	**Variance of components**	**Percentage of variation**
Among mountains	6	369.12	2.10	10.91
Among populations within mountains	21	493.65	1.88	9.75
Within populations	108	1651.569	15.29	79.33
Total	135	2514.34	19.27	

Umbilicus botrioydes

Umbilicus botrioydes was collected from only two mountains in Ethiopia, the Bale and Simen Mts. The total number of samples was 38 from seven populations. The three primer combinations produced 246 markers, of which 102 (41.46%) were polymorphic. The AFLP profiles were 98.1% reproducible. In the PCoA, some populations from the Simen Mts. tended to cluster together whereas those from Bale Mts. fell into two different groups (Fig. 28). The first axis explained 12 % and the second axis 11 % of the total variation. Neither the NJ tree (Fig. 29) nor the STRUCTURE analysis (Fig. 30) identified a clear genetic structure.

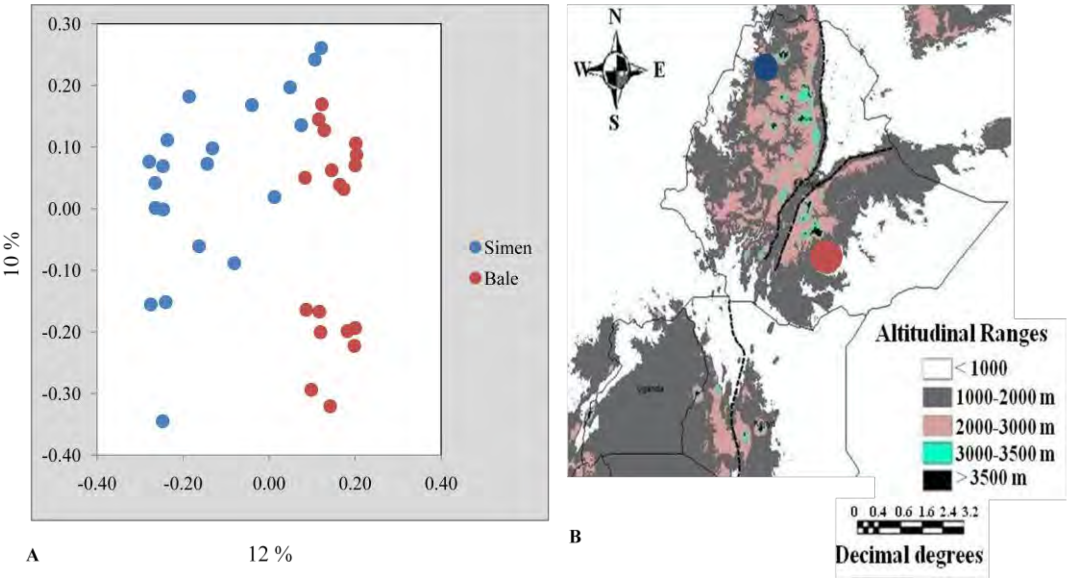


Fig. 28. (A) PcoA of *U. botryoides.* (B) Geographical origin of the populations with colours representing mountains

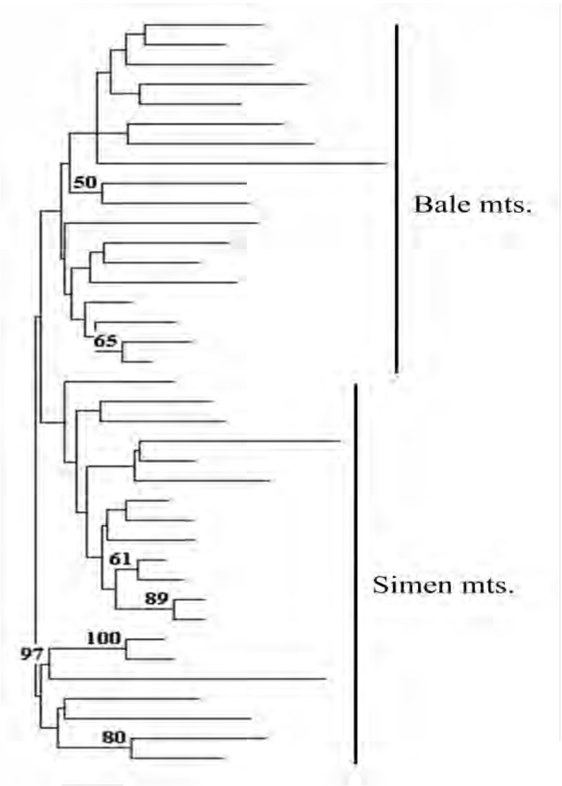


Fig. 29 NJ tree showing the relationships among the *U. botryoides* populations collected from two mountains in Ethiopia (Simen and Bale Mts.)

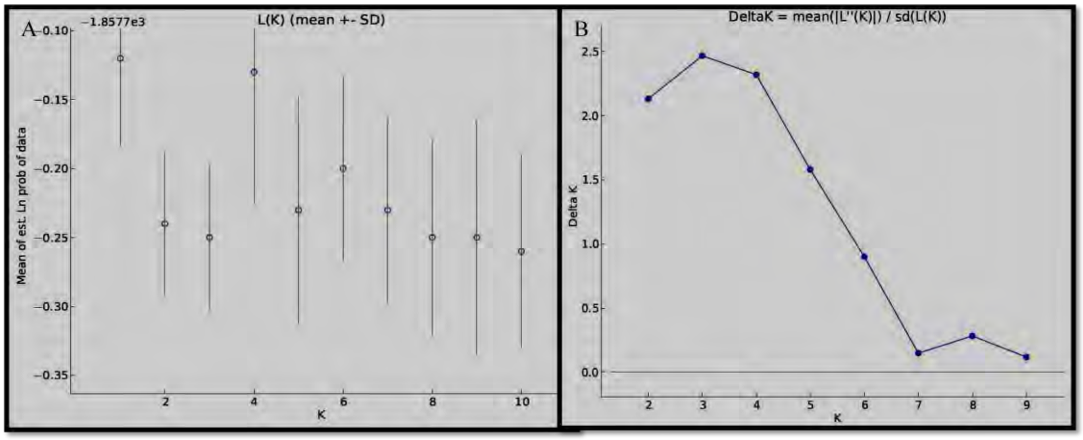


Fig. 30 (A) Log probability of data L(K) of *U. botryoides*, as a function of K for 10 structure runs. (B) Rate of change in the probability between successive runs, delta K, as a function of K.

Table 13 presents gene diversity (h) and Shannon‟s diversity index (I) values. The highest gene diversity occurred in a population from the Bale Mts. (ET0945) while the lowest level was recorded in a population from the Simen Mts. (ET0295).

Table 13 Gene diversity, rarity and no of individuals analysed of *U. botryoides*

Population name	Locality	No. of individuals	h (gene diversity)	I (Shannon's index)	DW (Rarity)
ET0178	Simen Mts.	5	0.102439	0.0997	4.26106
ET0230	Simen Mts.	2	0.101626	0.0615	5.199292
ET0295	Simen Mts.	4	0.04065	0.034	2.603916
ET0408	Simen Mts.	4	0.091463	0.0847	4.750198
ET0663	Bale Mts.	5	0.101626	0.0975	5.719634
ET0866	Bale Mts.	5	0.104878	0.0996	5.572407
ET0945	Bale Mts.	5	0.130081	0.1242	8.28389

The AMOVA revealed that within-population variation accounted for almost 80% of the total variation (Table 14).

Table 14 Analysis of molecular variance (AMOVA) for AFLP diversity in *U. botryoides*

Source of variation	d.f.	Sum of squares	Variance of components	Percentage of variation
Between mountains	1	45.90	1.93	12.78
Among populations within mountains	5	83.15	1.09	7.25
Within populations	23	277.250	12.05	79.97
Total	29	406.300	15.07	

6 Discussion

6.1 Phylogeny, biogeography and molecular dating of African *Swertia*

6.1.1 Phylogeny and biogeography

Our results add to a growing body of evidence suggesting that a major part of the afro-alpine flora was formed after local diversification initiated by repeated long-distance colonizations from the Northern Hemisphere (see overview in Gehrke & Linder, 2009). In the arctic-alpine and afro-alpine *Arabis alpina*, diversification proceeded at the intraspecific level within the two lineages that immigrated independently into Africa from the Middle East (Koch *et al.*, 2006; Assefa *et al.*, 2007; Ehrich *et al.* 2007). Repeated colonizations triggered repeated species radiations in other genera such as *Carex* and *Ranunculus* (Gehrke & Linder, 2009) and, as shown in this study, *Swertia*. The placement of the African *Swertia* into two distantly related and strongly supported clades within the *Gentianella* lineage in our analyses, which included most of the African species (Fig. 12 & 13), provides robust evidence for two introductions of this genus into Africa. This result is contradictory to the suggestion of monophyly of the African *Swertia* by Ho *et al*. (1994) but corroborates earlier suggestions of an „at least twice into Africa" scenario based on morphology (Nemomissa, 1994) and on molecular analysis including four of the African species (Chassot *et al.,* 2001).

The most remarkable finding for the African *Swertia* is the extremely long distance from the inferred source area, the Himalaya Mountains, which are located some 4000-5000 km away from the African mountains. The genus has previously been hypothesized to have originated in southwestern China based on high species diversity and inferred trends in the evolution of morphological characters (Ho *et al.,* 1994), and different African taxa have been suggested to have derived from different Asian lineages based on morphological similarity (Nemomissa, 1994). In our plastid DNA phylogeny, different Himalayan species (*S. cuneata* and *S. cordata*) were inferred as sisters to the two African clades, and the nuclear ribosomal DNA phylogeny also inferred *S. cordata* as sister to clade 2. No species of *Swertia* are known today from the area between the Himalayas and Africa, except the mainly Ethiopian *S. engleri* that extends into small parts of Yemen and Saudi-Arabia. The type species *Swertia perennis*, which is mainly distributed in the European Alps, does not belong to the African clades in our analysis. A stepwise expansion from the Himalayas towards Africa with subsequent extinction in the intermediate area is therefore not likely based on our results. It is thus reasonable to propose that two events of extremely long-distance dispersal directly from the Himalayas to afro-alpine Africa have occurred. Such long-distance dispersal was inferred, based on fossil-calibrated molecular clocks, for the disjunct distribution of *Exacum* in the Indian Ocean basin (Yuan *et al*., 2005). Moreover, the low resolution we found within each African clade indicates rapid *in-situ* diversification after arrival in Africa.

The molecular results presented in this study also have implications for the systematics of *Swertia*. Our phylogenetic reconstruction supported that *Swertia* is strongly paraphyletic relative to other genera in subtribe Swertiinae, as also recognized by Chassot *et al*. (2001), highlighting the need for comprehensive revision of the delimitation of genera. The two African clades inferred here are consistent with the two informal morphological groups suggested by Nemomissa (1997), signifying that the four characters he identified (number of nectaries per corolla lobe, shape and surface of the seed, and surface of the exine) may be phylogenetically informative, especially for the African species. It can also be noted that the Himalayan species which are sister to the African clades, show similar character states as the African members of their respective clade (Chassot *et al.*, 2001, 2008). The taxa in Clade 1 and the respective Himalayan sister species (in the plastid phylogeny) have two nectaries per corolla lobe, ovoid seeds with smooth surface and spinulose exines, whereas the taxa in Clade 2 and the respective Himalayan sister species have a single nectary per corolla lobe, ellipsoid seeds with ridge-like outgrowths and striato-reticulate exines. However, the variation across the entire subtribe suggests that the number of nectaries per corolla lobe have changed more than once and may not provide phylogenetic information at this level. For example, *S. calycina* and *S. perennis*, which have two nectaries per corolla lobe, were included in a clade that also consists of *Swertia* species having one nectary per corolla lobe. Similarly, *S. ciliata* and *S. pubescens*, which have one nectary per corolla lobe, appeared within a clade that also contains *Swertia* species having two nectaries per corolla lobe (Figs. 10 and 11). In the same way, exine sculptures are highly diverse and not suitable to characterize groups across the subtribe Swertiinae (Chassot *et al.*, 2001; Chassot and von Hagen, 2008). On the other hand, the *S. calycina*- *S. macrosepala* subsp. *macrosepala* clade (Figs. 10 and 11) has distinctive seed shape and testa, leaving all the other species of *Swertia* to have seeds more or less rounded with a smooth to warted testa (Chassot *et al.*, 2001). Thus, seed shape and surface appear to be more reliable taxonomic characters at a higher taxonomic level.

Some internal relationships could be inferred within the two African clades, of which some are inconsistent with the current morphological taxonomy. In Clade 1, two accessions of *S. crassiuscula* subsp. *crassiuscula*from Mt. Kilimanjaro and Mt. Kenya grouped together while a third accession (*S. crassiuscula* subsp. *robusta*) of this species from the Bale Mountains grouped in a different subclade. Two accessions of *S. engleri* var. *engleri*, both from Simen Mountains, also grouped separately. *Swertia* aff. *abyssinica*, which is quite distinct morphologically, grouped with *S. crassiuscula* subsp. *crassiuscula* in our plastid DNA analyses, and not with *S. abyssinica* in any of the analyses. We examined the floral characters of five individuals of *S*. aff. *abyssinica*, and found that they were intermediate between those of *S. abyssinica* and *S. crassiuscula* subsp. *robusta*, calling for more detailed morphological as well as molecular analyses to resolve its origin and relationships. We also found other potential inconsistencies such as the relationships between the two subspecies of *S. macrosepala* and the two varieties of *S. volkensii*. More detailed studies are necessary to address whether the inconsistencies in African *Swertia* are caused by incomplete lineage sorting or introgression/chloroplast capture, or both (Rieseberg, 1991; Soltis & Kuzoff, 1995; Sang *et al*., 1997).

6.1.2 Molecular divergence time

The two different calibration point procedures produced different estimates of divergence time. However, earlier studies (Yuan *et al*., 1996) have suggested that the origin of the subtribe Swertiinae took place 15 Mya (the third calibration point in this study). This was substantiated by the evolution of temperate vegetation in Western Himalayas that took place at the end of Oligocene and the beginning of Miocene, and the rise of the mountains providing the alpine climate. These hypotheses have been considered in subsequent studies of molecular divergence time for some taxa in the family Gentianaceae (Yuan *et al*., 2003; 2005; Favre *et al*., 2010). Therefore, the second procedure (the analysis with three calibration points) is considered as likely to produce the most probable inference in the present study. Nonetheless, assuming that the divergence time between the subtribe Gentianinae and Swertiinae is about 15 Mya may slightly underestimate the age of the two subtribes (Favre *et al*., 2010), suggesting the need for further detailed studies including the whole geographical range of Gentaniaceae.

There are still not any absolute dating studies published for afro-alpine plant lineages based on fossil-calibrated molecular clocks. Both immigrations of *Arabis alpina* were suggested to have occurred in the Pleistocene based on inferred mutation rates (Koch *et al*., 2006), whereas an ongoing study based on a fossil-calibrated molecular clock suggests that immigration of the ancestral Eurasian lineage of *Lychnis* took place earlier, some 4 Myr ago (Gizaw *et al*. unpublished data; see also Popp *et al*., 2008). It is likely that long-distant colonizations of the afro-alpine "sky islands" occurred at different times for different lineages, as was suggested by Donoghue (2011) and Popp *et al*. (2011) for the extremely disjunct bipolar plants. The result from the fossil-calibrated molecular dating of the African *Swertia* corroborates this hypothesis rather than the hypothesis of Late Cretaceous to the Oligocene Epoch colonization as suggested by Ho *et al*. (1994). On the other hand, the molecular dating analysis adds to the parellel findings that the afro-alpine environment was colonized by different species at different temporal scales and some colonization happened well before to the Pleistocene Epoch (Gizaw, *et al*. unpublished data; Popp *et al*., 2008). Therefore, the findings in this study may support the hypothesis of Pleistocene refugia in Africa as detailed by Nichol (1999).

Linder (2008) correlates rapid and recent radiations with the availability of new habitats or space. He further explained rapid and recent radiation as the early stage of speciation, in which diversification dominates over extinction. Taking the estimated divergence time of the African *Swertia* into account and considering the overall low levels of resolution within both of the African clades, the speciation in the African *Swertia* could be explained as a result of rapid *in-situ* diversification.

6.2 Comparative phylogeography and genetic diversity of some afro-alpine plant taxa

6.2.1 Phylogeography of the studied afro-alpine taxa

We found strong phylogeographic structure in some of the studied afro-alpine taxa. Unexpectedly, in two of the studied taxa (*C. schimperi* and *T. cryptopodium*), we observed that the populations from the mountains of Northern Ethiopia were divergent from those of Southern Ethiopia (the Bale Mountains), the latter being genetically more similar to the more distant East African mountains (Kenya, Tanzania and Uganda). We also found evidence of occassional gene flow across the Great Rift Valley, as previously observed in *Arabis alpina* (Ehrich *et al*., 2007). Our findings add to a growing body of evidence suggesting that initial patterns of genetic divergence among isolated afro-alpine 'sky islands' are confoundingly overlaid by occasional long-distance dispersals, which not always occur among the geographically most closely situated mountains or mountain groups (AFROALP-II, unpublished data).

In *Carduus schimperi*, we found three quite distinct genetic groups corresponding to the three reported subspecies, except that the Bale populations grouped with East African populations of ssp. *nanus* rather than with the Simen populations of ssp. *schimperi*. Interestingly, the Aberdare populations did not group with those of the geographically most closely situated mountain, Mt. Kenya, but rather with the distant mountains of Elgon and Bale. The occurrence of three distinct genetic groups in *C. schimperi* suggests long-term isolation in three different areas. Each of two of the groups is confined to a single mountain based on the material we had available for analysis. One is confined to Mt Kenya, in accordance with previous reports of ssp. *platyphyllus* being endemic to this mountain. The other is confined to the Simen Mts and corresponds to ssp. *schimperi*, but this subspecies has also been reported from other areas in East Africa and Sudan from where we did not have material available. The third genetic group, corresponding to ssp. *nanus*, consisted of the populations from Mt. Elgon, the Aberdare Mts, and the Bale Mts (this subspecies has also been reported from one mountain in Congo). The high degree of genetic similarity observed within this group suggests that it initially diverged in isolation in one mountain area and more recently colonized the other mountains. Most diversity and highest genetic distinctiveness (DW) were found in Mt. Elgon, suggesting that this mountain may have served as the source for more recent colonizations of the Aberdare Mts and the Bale Mts.

We found however low levels of diversity and distinctivity in the Simen Mountains (ssp. *schimperi*), suggesting that this subspecies may have originated in the southern parts of its distribution area, from where we did not have samples available, and more recently colonized the Simen Mts. Interestingly, a single plant from Mt Kenya was identified as a hybrid between the Mt. Kenya-endemic subspecies (ssp. *platyphyllus*) and ssp. *nanus*, whose nearest occurrence is in the neighbouring Aberdare Mts. This finding suggests recent dispersal of ssp. *nanus* from Aberdare to Mt. Kenya, resulting in hybridization with the local ssp. *platyphyllus*. Since our allocation test assigned this plant to an F1 hybrid, it is not yet known whether complete genetic barriers have arisen between the two subspecies or whether introgression may occur between them.

In *Trifolium cryptopodium*, we recognized two quite distinct genetic groups, one confined to northern Ethiopia (the Simen Mts and Choke), the other distributed in the East African mountains but also in southern Ethiopia (the Bale Mountains). However, in the PCoA, the Bale populations appeared to be somewhat intermediate between the two groups. This pattern suggests long-term isolation in two areas, possibly followed by dispersal across the Great Rift Valley in Ethiopia. This hypothesis is supported by the identification of two plants from the Simen Mts as F1 hybrids, suggesting that they originated after dispersal of the other lineage into the Simen Mts, across the Rift from the Bale Mts. The high diversity and genetic distinctiveness we observed in some of the East African populations of the East African/Bale group may suggest that it originated in the south and later dispersed to the Bale Mts, but the fairly high diversity and distinctivity in the Bale Mts suggest that this happened a long time ago.

In *Geranium* spp., we recognized two genetic groups, but these were not very distinct and in several cases, individual populations/mountains were split between the two groups. We found that much of the genetic variation was contained within populations, suggesting that the species is outcrossing and that much of the morphological polymorphism observed in afro-alpine *Geranium* is maintained as intrapopulational polymorphism. The lack of clear geographic structuring of the genetic diversity also suggests a quite recent history of intermountain colonization in *Geranium*. As some populations are known to occur at quite low altitudes (down to 1000 m), it is likely that gradual migration between some mountains has been facilitated by interglacial forest bridges (Hedberg, 1964; Kebede *et al*., 2007), but the considerable distance across lowlands between Ethiopia and East Africa also suggests that direct long-distance dispersals have taken place (cf.

Hedberg, 1970). The importance of long-distance dispersal in shaping the afro-alpine flora is now increasingly recognized (Hedberg, 1964; 1970; Koch *et al*., 2006; Ehrich *et al.*, 2007; Gehrke & Linder, 2009; and the current study on *Swertia*).

We did thus not find any genetic distinction between low-altitude populations and high-altitude populations corresponding to a taxonomic distinction between *G. arabicum* and *G. kilimandscharicum*, supporting the conclusion of Kokowora (1971) that only a single species, *G. arabicum*, can be recognized. We neither found evidence for recognizing any subspecies within *G. arabicum*, nor that the putative new species ('*Geranium* sp. nov.) is genetically distinct from *G. arabicum*. In our analysis, the populations of '*Geranium* sp. nov.' from the Bale Mountains, initially distinguished from other Bale populations based on the rosette-like arrangement of the leaves caused by extremely reduced internodes, were very similar genetically to the other Bale populations. This suggests that their deviating morphology is caused by environment-induced phenotypic plasticity or possibly by high-altitude adaptation in a few genes, and is without taxonomic significance. Notably, the deviating populations were only found above 4000 m a.s.l and only occupy exposed habitats disturbed by the giant mole rat, whereas typical *G. arabicum* occupies well sheltered habitats and grows among grasses. The reduced growth form of '*Geranium* sp. nov.' may therefore have developed to withstand the effects of the strong high-altitude winds. On the other hand, the *Geranium* sp. nov. populations were split between the two genetic groups, which might indicate that they rather represent a mixture between two slightly divergent genetic lineages which met in the Bale Mountains.

Our AFLP analysis of the widespread *Swertia abyssinica* only included the Ethiopian part of its distribution area, and we did not recognize divergent genetic groups within this species. It shows however an interesting gradual genetic structuring corresponding to geography, extending continuously across the Great Rift Valley in Ethiopia. The populations from Gara Muleta, the Simen Mts., and the Bale Mts. occupied the extremes of the PCoA plot while mixtures of populations from other mountains appear in the middle. Geographically, the latter mountains are located more or less centrally in Ethiopia, at about mid-distance between the three extremes. Thus, although indicating some degree of genetic isolation with geographic distance, our data suggest that the Great Rift Valley has not acted as a major barrier against gene flow across the Ethiopian Highlands. This might be expected in a species such as *Swertia abyssinica*. Although is occurs in the highest alpine areas in a densely tufted growth form (previously recognized as a distinct species, *S. lugardae*), it also extends down to 1500 m and therefore may have been able to gradually cross the Rift Valley during cold periods.

We did not find any distinct phylogeographic structure in the widespread *Umbilicus botryoides*, of which we only had material available from the Simen and Bale Mts. in Ethiopia. Also in this species, we found only gradual genetic differentiation between the two mountains, suggesting that the Rift Valley does not represent a major barrier against gene flow.

7.2. *Taxonomic implications*

Our phylogeographic study has some implications concerning the taxonomy of afro-alpine plants.

Geranium in the afro-alpine environment

It is clear from the above discussion that our genetic data support recognition of only a single polymorphic species of *Geranium* in the afro-alpine environment, *G. arabicum* s. lat., without recognition of any infraspecific taxa. Delimitation of the high-altitude species *G. kilimandscharicum* was mainly based on the shape of the leaves. Reniform 2-4(5)-lobed leaves versus pentagonal 5-lobed leaves have been given as the distinguishing features between *G. kilimandscharicum* and *G. arabicum*, respectively. Our genetic analysis rather suggests that this variation reflects polymorphism within a single species, and therefore, *G. kilimandscharicum* should be reduced to a synonym of *G. arabicum*. This is in agreement with Kokowaro (1971), who reported absence of clear taxonomic characters to distinguish between *G. arabicum* and *G. kilimandscharicum*. Our AFLP results neither supported the recognition of '*Geranium* sp. nov. = G. & S. Miehe 3002' as a species distinct from *G. arabicum*. Its reduced growth form may simply reflect phenotypic plasticity or adaptation to wind-exposed high-altitude conditions based on a few genes only, not warranting taxonomic recognition, or perhaps it represents a mixture between two slightly divergent lineages within *G. arabicum*. Notably, '*Geranium* sp. nov.' and *G. kilimandscharicum* were described having very similar leaves, being reniform with 3-5 lobes (Kokowaro, 1971; Gilbert and Vorster, 2000). Gilbert and Vorster (2000) also suggested that *'Geranium* sp. nov. = G. & S. Miehe 3002' could be an ecotype of *G. arabicum.*

Carduus schimperi

The genetic data presented in this study also have implications for the infraspecific taxonomy of *C. schimperi*. The results support the recognition of three subspecies based on morphology, but may suggest a revision of their circumscription. Currently, only one subspecies, ssp. *schimperi*, has been recognized in Ethiopia, whereas our genetic data suggest that the Bale populations rather belong to the otherwise East African ssp. *nanus*. However, since our collection primarily was carried out in the afro-alpine zone, and we therefore did not collect ssp. *schimperi* in its lower-lying areas in East Africa, we cannot conclude on the circumscription of the subspecies based on our material. An extended study with more material of ssp. *schimperi* and a detailed morphological analysis is required.

7 Conclusions and Recomendations

The findings in this study adds to a growing body of evidence suggesting that long-distance dispersals into the afro-alpine region from remote sources, as well as intermountain dispersals within the afro-alpine region, have occurred more frequently than previously envisioned. This study also suggests that initial patterns of isolation and divergence among different populations have been overlaid by more recent episodes of colonization, leading to confounded patterns of genetic structuring and taxonomic distinctivity. The current study used African *Swertia* spp., *Trifolium cryptopodium*, *Geranium arabicum*, a putative new species of *Geranium*, *Carduus schimperi* and *Umbilicus botryoides* to assess the history of plant colonization and speciation in the high mountains of Ethiopia and East Africa (Kenya, Tanzania and Uganda). The African species of *Swertia* were used in phylogenetic and biogeographic analyses while the remaining species were used to explore phylogeographic history and the level of their genetic diversity across the afro-alpine environment. In the *Swertia* study, we used sequences of two non-coding plastid regions, one coding plastid region (*mat*K) and the nuclear ribosomal internal transcribed spacer (ITS) to investigate the phylogenetic relationships, biogeography and divergence time. We showed that the African species formed two distantly related clades, each closely related to different Himalayan species of the genus. Thus, this study revealed two episodes of colonization of Africa by different lineages of *Swertia* from an extremely distant source, the Himalayas, and a molecular divergence data analysis suggested that the earliest migration into Africa took place in the Miocene. In the AFLP-based phylogeographic studies of the other taxa, several long-distance migrations among the high mountains in Ethiopia and East Africa were indicated. In two species, *Carduus schimperi* and *Trifolium cryptopodium*, the populations from Bale Mountains in Ethiopia were genetically more similar to populations from the East African Mountains than to those from other Ethiopian mountains. The putative new species of *Geranium* (listed as *Geranium sp. nov. = G. S. Miehe 3002* in Flora of Ethiopia and Eritrea as endemic to the Bale Mountains) was not found to be genetically distinct from the widespread and genetically variable *G. arabicum*, suggesting that the characters used to distinguish it merely represent phenotypical plasticity or adaptation to extremely wind-exposed conditions based on a few genes only. We did neither find any evidence to separate *G. kilimandscharicum* as a distinct species nor any subspecies within *G. arabicum*, and thus suggest to recognize the afro-alpine populations of *Geranium* as a single, polymorphic species. In contrast, the previous division of *Carduus schimperi* into three subspecies based on morphology was supported by the genetic data, except that the material from the Bale Mountains which previously had been referred subsp. *schimperi* grouped with the East African subsp. *nanus*.

References

Abbott, R. J., Smith, L. C., Milne, R. I., Crawford, R. M. M, Wolff, K., Balfour, J. 2000. Molecular analysis of plant migration and refugia in the Arctic. *Science* **289**: 1343–1346.

Adams, J. 2008. DNA sequencing technologies. *Nature Education* **1**

Adams, M. D., Fields, C. and Venter, J. C. 1996. Automatic DNA Sequencing and Analysis. San Diego: Academic Press.

Adams, R. P., Morris, J. A., Pandey, R. N., Schwarzbach, A. E. 2005. Cryptic speciation between *Juniperus deltoides* and *Juniperus oxycedrus* (Cupressaceae) in the Mediterranean. *Biochemical Systematics and Ecology* **33**: 771-787.

Albert V. A., Struwe L. 2002. Gentianaceae in context. In: Struwe L & Albert VA (eds.), Gentianaceae: systematic and natural history. Cambridge, Cambridge University Press.

Alsos, I. G., Engelskjøn, T., Gielly, L., Taberlet, P. and Brochmann, C. 2005. Impact on ice ages on circumpolar molecular diversity: insights from an ecological key species. *Molecular Ecology* **14**: 2739–2753.

Alsos, I.G., Eidesen, P.B., Ehrich, D., Skrede, I., Westergaard, K., Jacobsen, G. H., Landvik, J. Y., Taberlet, P. and Brochmann, C. 2007. Frequent Long-Distance plant colonization in the changing Arctic. *Science* **316**: 1606-1609.

APG III. (2009). An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society* **161**: 105-121.

Assefa, A., Ehrich, D., Taberlet, P., Nemomissa, S. and Brochmann, C. 2007. Pleistocene colonization of afro-alpine 'sky islands' by the arctic-alpine *Arabis alpina*. *Heredity* **99**: 133-142.

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

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

Baldwin, B. G. 1992. Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: An example from the Compositaogy. *Molecular Phylogenetics and Evolution* **1** : 3–16.

Brochmann, C., Gabrielsen, T. M., Nordal, I., Landvik. J. Y., Elven, R. 2003. Glacial survival or tabula rasa? The history of North Atlantic biota revisited. *Taxon* **52**: 417–450.

Bronken, P., Taberlet, P., Gielly, L., Brochmann, C. 2001. Chloroplast and nuclear DNA variation on a circumpolar scale: migration history of the clonal Saxifraga cernua. *Bauhinia*: **15**: 69–90.

Bekele, B., Ehrich, D., Kebede, M., Brochmann, C., and Nemomissa, S. 2007. The Swertia abyssinica complex in the afro-alpine mountains: parallel ecoclinal evolution or secondary contact between differentiated glacial gene pools? *Book of Abstract*, *XVIIIth AETFAT Congress*, Yaoundé, Cameroon.

Bonin, A., Bellemain, E. Eidesen, P. B. Pompanon, F., Brochmann, C. and Taberlet, P. 2004. How to track and assess genotyping errors in populations genetics studies. *Molecular Ecology* **13**: 3261-3273.

Brout, C., McKey, D. and Douzery. J. 2004. Differentiation in geographical mosaic of plants coevolving with ants: phylogeny of the *Leonardoxa africana* complex (Fabaceae: Caesalpinoideae) using AFLP markers. *Molecular Ecology*, **13**: 1157-1171.

Campbell, D., Duchesne, P. and Bernatchez, L. 2003. AFLP utility for population assignment studies analytical investigation and empirical comparison with microsatellites. *Molecular Ecology*, **12**: 1993-1998.

Chassot, P., Nemomissa, S., Yuan, Y. M. and Küpfer, P. 2001. High paraphyly of *Swertia* L. (Gentianaceae) in the *Gentianella*-lineage as revealed by nuclear and chloroplast DNA sequence variation. *Plant Systematics and Evolution* **229**: 1-21.

Chassot, P. and von Hagen, K. B. 2008. Pollen morphology of the Swertiinae (Gentianaceae): phylogenetic implications. *Botanical Journal of the Linnean Society* **157**: 323–341.

Cheng, S., Fockler, C., Barnes, W. M., Higuchi, R. 1994. Effective Amplification of Long Targets from Cloned Inserts and Human Genomic DNA. *Proceedings of the National Academy of Sciences of the United States of America* **91** (12): 5695–5699.

Comes, H. P. and Kadereit, J. W. 2003. Spatial and temporal patterns in the evolution of the flora of the European Alpine system. *Taxon* **52**: 451–462.

Crepet, W. L. and Daghlian, C. P. 1981. Lower Eocene and Paleocene Gentianaceae: Floral and Palynological Evidence. *Science, new series* **214**: 75–77.

Crisci, J. V. 2001. The voice of historical biogeography. *Jounrnal of Biogeography* **28**: 157-168.

Demesure, B., Sodzi, N. and Petit, R. J. 1995. A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. *Molecular Ecology* **4**: 129-139.

Despres, L., Gielly, L., Redoutet, B. and Taberlet, P. 2003. Using AFLP to resolve phylogenetic relationships in a morphologically diversified plant species complex when nuclear and chloroplast sequences fail to reveal variability. *Molecular Phylogenetics and Evolution*, **27**: 185-196.

Donghue, M. J. 2011. Bipolar biogeography. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 6341-6342.

Donghue, M. J. and Moore, B. R. 2003. Towards an integrative historical biogeography. *Integrative and Comparative Biology* **43**: 261-270.

Drummond, A.J. and Rambaut, A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology **7**: 214.

Du, F. K., Petit, R. J. and Liu, J. Q. 2009. More introgression with less gene flow chloroplast vs mitochondrial DNA in the Picea asperata complex in China and comparison with other conifers. *Molecular ecology* **18**: 1396-1407.

Duchesne, P. and Bernatchez, L. 2002. AFLPop: A computer program for simulated and real population allocation based on AFLP data. *Molecular Ecology Notes* **3**: 380-383.

Earl, D. A. and von Holdt, B. M. 2011. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* doi: 10.1007/s12686-011-9548-7

Ehrich, D. 2006. AFLPdat: a collection of R functions for convenient handling of AFLPdata. Molecular Ecology. Notes 6: 603-604.

Ehrich, D., Gaudeul, M., Assefa, A., Koch, M., Mummenhoff, K., Nemomissa, S., Intrabiodiv, C., Brochmann, C. 2007. Genetic consequences of Pleistocene range shifts: contrasts between the Arctic, the Alps, and the East African mountains. *Molecular Ecology* **16**: 2542–2559.

Excoffier, L., Smouse, P.E. and Quattro, J.M. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**: 479-491.

Favre, A., Yuan, Y-M., Küpfer, P. and Alvarez, N. 2010. Phylogeny of subtribe Gentianinae (Gentianaceae): biogeographic inferences despite limitations in temporal calibration points. *Taxon*, **59**: 1701-1711.

Fries, T. C. 1923. Die *Swertia*-Arten der Afrikanischen Hochgebirge. *Notizblatt des Botanischen Gartens und Museums zu Berlin-Dahlem. Berlin-Dahlem* **8**: 389-423**.**

Galtier, N., Gouy, M. and Gautier, C. 1996. SEAview and Phylo_Win: two graphic tools for sequence alignment and molecular phylogeny. *Computer Applications in the Biosciences* **12**: 543-548.

Gatrell, A. 1983. Distance and space. A geographical perspective. *Clarendon Press*, Oxford.

Gaudeul, M., Taberlet, P. and Till-Bopttraud, I. 2000. Genetic diversity in an endangered alpine plant *Eryngium alpinum* L. (Apiaceae), inferred from amplified length polymorphism markers. *Molecular Ecology* **9**: 1625-1637.

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

Gilbert, M. G. 1989. Papilionoideae. In: Flora of Ethiopia Vol. 3, Hedberg, I. and Edwards, S. (eds.). Addis Ababa, Uppsala.

Gilbert, M. G. and Vorster, P. 2000. Geraniaceae. In: Flora of Ethiopia & Eritrea, Vol. 2, Part 1, Edwards, S., Tadesse, M. and Demissew, S. and Hedberg, I. (eds.). Addis Ababa, Uppsala.

Gehrke, B. and Linder, H. P. 2009. The scramble for Africa: pan-temperate elements on the African high mountains. *Proceedings of Biological Sciences of the Royal Society* **276**: 2657-2665.

Gillett, J. B., Poihill, R. M. And Verdcourt, B. 1971. Leguminosae (Part 4), subfamily Papilionoideae. In: Flora of Tropical East Africa. Milne-Redhead, E. & Polhill, R. M. (eds.). ISBN, White friars press ltd, London

Goloboff, P. A., Carpenter, J. M., Arias, J. S., Esquivel, D. R. M. 2008. Weighting against homoplasy improves phylogenetic analysis of morphological data sets. *Cladistics* **24**: 758.

Gradstein, F., Ogg, J. G. and Smith., A. 2004. A Geologic time scale. Cambridge.

Graham, A. 1984. Lisianthius pollen from Eocene of Panama. *Annals of the Missouri Botanical Garden* **71**: 987-993.

Graham, S. W. and Olmstead, R. G. 2000. Evolutionary significance of an unusual chloroplast DNA inversion found in two basal angiosperm lineages. *Current Genetics* **37**: 183-188.

Griffin, H.G. and Griffin, A. M. 1993. DNA sequencing: recent innovations and future trends. *Applied Biochemistry and Biotechnology* **38**, 147-159.

Hamby, R. K. and Zimmer, E. A. 1992. Ribosomal RNA sequences as a phylogenetic tool in plant systematics. In: Soltis, P. S., Soltis, D. E. & Doyle, J. J. (eds.), Molecular Systematics of plants. New York, Chapman & Hall, pp: 50-91.

Harison, N., Kinder, K. and Anne, C. 2011. Next generation sequencing and systematics: what can a billion base pairs of DNA sequence data do for you? *Taxon*, **60**: 1552-1566.

Hedberg, O. 1951. Vegetation belts of the East African mountains. *Svensk Botanisk Tidskrift* **45**: 140-202.

________. 1957. Afro-alpine vascular plants: a taxonomic revision. *Symbolae Botanicae Upsalienses* **15**: 1-411.

________. 1961. The phytogeographical position of the afro-alpine flora. *Recent Advances in Botany* **1**: 914-919.

________. 1964: Features of afroalpine plant ecology. *Acta Phytogeographica Suecica* **49**: 1.

________. 1965. Afro-alpine flora elements. *Webbia* **19**: 519-529.

________. 1969. Evolution and speciation in a tropical high mountain flora. *Biological Journal of the Linnean Society* **1**: 135-148.

________. 1970. Evolution of the afro-alpine flora. *Biotropica* **2**: 16-23.

________. 1995. Features of afro-alpine plant ecology. *Ekblad, Västervik*, Sweden.

Hedberg, I. and Hedberg, O. 1979. Tropical-alpine life-forms of vascular plants. *Oikos* **33**: 297-307.

Hedderson, T. A., Chapman, R. L. and Rootes, W. L. 1996. Phylogenetic relationships of bryophytes inferred from nuclear encoded rRNA gene sequences. *Plant Systematics and Evolution* **200**: 213-224.

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

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

Hewitt, G. M. 2004. Genetic consequences of climatic oscillations in the Quaternary. *Proceedings of the Royal Society of London. Biological Sciences* **359**: 183–195.

Ho, T. N., Xue, C.Y. and Wang, W. 1994. The origin, dispersal and formation of the distribution pattern of *Swertia* L. (Gentianaceae). *Acta Phytotaxonomica Sinica* **32**: 525-537.

Hyman, E. D. 1988. A new method of sequencing DNA. *Analytical Biochemistry* **174**: 423–436.

Janssen, P., Coopman, R., Huys, G., Swings, J., Bleeker, M., Vos, P., Zabeau, M. and Kersters, K. 1996. Evaluation of the DNA fingerprinting method AFLP as a new tool in bacterial taxonomy. *Microbiology*, **142**: 1881-1893.

Jeffery, C. and Beentje, H. J. 2000. Compositae (Part I). In: Flora of Tropical East Africa, Beentje, H. J. and Smith, S. A. L. (eds.). Balkema, A. A, ISBN, Royal Botanic Gardens, Kew.

Joshi, K. and Joshi, A. 2008. *Swertia* L. (Gentianaceae) in Nepal Himalaya: checklist, phytogeography, ethnobotany and conservation status. *Ethnobotanical Leaflets* **12**: 361-372.

Judd, W. S., Campbell, C. S., Kellogg, E. A., Stevens, P. F. and Donoghue, M. J. 2007. Plant systematics: a phylogenetic approach. Canada, Sinauer Associates, Inc.

Kebede, M. 2000. Cytogenetic study of species of Swertia L. (Gentianaceae) from Ethiopia. M. Sc. Thesis. Addis Ababa University.

Kebede, M., Ehrich, D., Taberlet, P., Nemomissa, S. and Brochmann, C. 2007. Phylogeography and conservation genetics of a giant lobelia (*Lobelia giberroa*) in Ethiopian and Tropical East African mountains. *Molecular Ecology* **16**: 1233-1243.

Kitching, I., Forey, P., Humphries, C. and Williams, D. 1998. Cladistics: the theory and practice of parsimony analysis. New York, Oxford University Press.

Koch, M. 2002. Genetic differentiation and speciation in prealpine Cochlearia (Brassicaceae): allohexaploid *Cochlearia bavarica* (Brassicaceae) compared to its diploid ancestor Cochlearia *pyrenaica*. Germany and Austria. *Plant Systematics and Evolution* **232** 35–49.

Koch, M. A., Kiefer, C., Ehrich, D., Vogel, J., Brochmann, C., Mummenhoff, K. 2006. Three times out of Asia Minor: the phylogeography of *Arabis alpina* L. (Brassicaceae). *Molecular Ecology* **15** 825–839.

Kokowaro, J. O. 1971. Geraniaceae. In: Milne-Redhead, E. & Polhill, R. M (eds.). Flora of Tropical East Africa. ISBN.

Kolodner, R. And Tewari, K. K. 1979. Inverted repeats in chloroplast DNA from higher plants. *Proceedings of the National Academy of Science* **76**: 41-45.

Kranz, H. D., and Huss, V. A. R. 1996. Molecular evolution of pteridophytes and their relationships to seed plants: evidence from complete 18S rRNA gene sequences. *Plant Systematics and Evolution* **202**: 1-11.

Linder, H. P. 2008. Pant species radiations: where, when , why? *Philosophical Transactions of the Royal Society of Biological Sciences* **363**: 3097-3105.

Metzker, M. L.2005. Emerging technologies in DNA sequencing. *Genome Research* **15**: 1767–1776.

Martínez-Millán, M. 2010. Fossil records and age of the Asteridae. *The Botanical Review* **76**: 83-135.

Mayr, E. and Bock, W. J. 2002. Classification and other ordering systems. *Journal of Zoological Systematics and Evolutionary Research* **40**: 169-194.

Meusel, H., Jager, E., Rauschert, S. and Weinert, E. 1978. Vergleichende Chorologie der Zentraleuropäischen *Flora II (Karten). VEB Gustav Fischer Verlag*. Jena, Germany.

Misra, A., Shasany, K., Shukla, A., Darokar, M., Singh, S., Sundaresan, V., Singh, J., Bagchi, G., Jain, S., Saikia, D. and Khanuja, S. 2010. AFLP markers for identification of *Swertia* species (Gentianaceae). *Genetics and Molecular Research*, **9**: 1535-1544

Morris, A. B, Gitzendanner, M. A, Soltis, D. E, Soltis, P. S. 2005. Commarative phylogeography of Eastern North American trees. *XVII International Botanical Congress Abstracts*. Vienna, Australia.

Morrone, J.J. and Crisci, J.V. 1995. Historical biogeography: introduction to methods. *Annual Review of Ecology and Systematics* **26**: 373-401.

Mueller, U. and Wolfenbarger, L. 1999. AFLP genotyping and fingerprinting. *Tree* **14**: 389-394.

Myburg, A. A., Remington, D. M, O‟Malley, Sederoff, R. R. and Whetten, R. W. 2001. High throughput AFLP analysis using infrared dye labeled and an automated DNA sequencer. *BioTechniques* **30**: 348-357.

Myers, A. A. and Giller, P.S. 1988. Analytical biogeography: An integrated approach to the study of animal and plant distributions. Chapman & Hall, London & New York.

Nei M. and Li, W. 1979. Mathematical model for studying genetic variance in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences*, USA 76, 5269–5273.

Nemomissa, S. 1994. *Swertia* L. (Gentianaceae) in North East Africa. Unpublished D. Phil. Thesis. *University of Vienna*, Austria.

Nemomissa, S. 1997. Floral character-states of the Northeast and Tropical East African *Swertia* species (Gentianaceae). *Nordic Journal of Botany* **17**: 145-156.

Nemomissa, S. 2002. Gentianaceae. In: Flora of Tropical East Africa, Beentje, H. J., & Smith, S. A. L. (eds.). Royal Botanic Gardens, Kew.

Nemomissa, S. 2006. Gentianaceae. In: Flora of Ethiopia and Eritrea Vol. 5, Hedberg, I., Kelbessa, E., Edwards, S., Demissew, S. & Persson, E. (eds.). The National Herbarium , Addis Ababa.

Nichol, J. E. 1999. Geomorphological evidence and Pleistocene refugia in Africa. *The Geographical Journal* **165**: 79-89.

Nylander, J. A. 2004. MrModeltest v2. Program distributed by the author. Evolutionary Biology Centre, Uppsala University.

Ogg, J. G., Ogg, G. and Gradstein, F. 2008. The concise Geologic time scale. Cambridge.

Page, R. D. M. and Charleston, M. A. 1998. Trees within trees: phylogeny and historical associations. *Tree* **13**: 356-359.

Petit, R. J., Aguinagalde, I., de Beaulieu, J-L., Bittkau, C., Brewer, S., Cheddadi, R., Ennos, R., Fineschi, S., Grivet, D., Lascoux, M., Mohanty, A., Müller-Starck, M., Demesure-Musch, B., Palmé, A., Martín, J. P., Rendell, S. and Vendramin, G. G. 2003. Glacial Refugia: Hotspots but not melting pots of genetic diversity. *Science* **300**: 1563-1565.

Piñeiro, R., Aguilar, J., Munt, D., and Feliner, G. 2007. Ecology matters: Atlantic-Mediterranean disjunction in the sand-dune shrub *Armeria pungens* (Plumbaginaceae). *Molecular Ecology*, **16**: 2155-2171.

Pocknall, D. T. 1987. Paleoenvironments and age of the Wasatch formation (Eocene), Powder River Basin, Wyoming. *Palaios* **2**: 368–376.

Popp, M., Gizaw, A., Nemomissa, S., Suda, J. and Brochmann, C. 2008. Colonization and diversification in the afro-alpine 'sky islands' by Eurasian *Lychnis* L. (Caryophyllaceae). *Journal of Biogeography* **35**: 1016-1029.

Popp, M., Mirré, V. Brochmann, C. 2011. A single Mid-Pleistocene long-distance dispersal by bird can explain the extreme bipolar disjunction in crow-berries (Empetrum). *Proceedings of the National Academy of Sciences of the United States of America* **108**: 6520-6525.

Powell, W., Morgante, M., McDevitt, R., Vendramin, G. G. and Rafalski, J. A. 1995. Polymorphic simple sequence repeat regions in chloroplast genomes: applications to the population genetics of pines *Proceeding of the National Acadadamy of Science*. **92**: 7759–7763.

Pritchard, J. K., Stephens, M. and Donnelly, P. J. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155**: 945-959.

Rambout, A. 2009. FigTree v.3.1. http://tree.bio.ed.ac.uk/software/figtree/

Rambaut, A. and Drummond, A. J. 2003. Tracer v 1.5, MCMC Trace analyses package http://beast.bio.ed.ac.uk

Ronquist, F., and Huelsenbeck, J. P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**:1572–1574.

Rohlf, F. 1990. NTSYS-PC. Numerical taxonomy and multivariate analysis system, Version 2.02. Exeter Software, Setauket, New York.

Riddle, B. R., Hafner, D. J. and Jaeger, J. R. 2000. Cryptic vicariance in the historical assembly of a Baja California Peninsular Desert biota. *Proceedings of the National Academy of Sciences of the United States of America* **97**: 14438-14443.

Reisch, C., Poschlod, P. and Wingender, R. 2003. Genetic variation of Saxifraga paniculata Mill. (Saxifragaceae): molecular evidence for glacial relict endemism in central Europe. *Biological Journal of the Linnean Society* **80**: 11–21.

Rieseberg, L. H. and Soltis, D. E. 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary Trends in Plants* **5**: 65-84.

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

Sanger, F., Nicklen, S. and Coulson, A. R. 1977. DNA sequencing with chain-terminating inhibitors.*Proceedings of the National Academy of Sciences of the United States of America* **74**: 5463-5467.

Schneider, S., Kueffer, J., Roessli, D., Excoffier, L. 1997. Arlequin 2. 000: a software for population genetic analysis. *Genetics and Biometry Laboratory*, University of Geneva, Switzerland.

Schendure, J. and Hanlee, J. 2008. Next generation DNA sequencing. *Nature Biotechnology* **26**: 1135-1145.

Schönswetter, P., Tribsch, A., Barfuss, M. and Niklfeld, H. 2002. Several Pleistocene refugia detected in the high alpine plant *Phyteuma globulariifolium* Sternb. and Hoppe (Campanulaceae) in the Europian Alps. *Molecular Ecology*, **11**: 2637-2647.

_______________, Paun, O., Tribsch, A., and Niklfeld, H. 2003. Out of the Alps: colonization of Northern Europe by East Alpine populations of the glacier buttercup Ranunculus glacialis L. (Ranunculaceae). *Molecular Ecology* **12**: 3371–3381.

Shaw, A. J. 2000. Molecular phylogeny and cryptic speciation in the Mosses, *Mielichhoferia elongate* and *M. meilichhoferiana*. *Molecular ecology* **9**: 595-608.

Shaw, A. J. 2001. Biogeographic patterns and cryptic speciation in Bryophytes *Journal of Biogeography* **28**: 256-261.

Shaw, J., Lickey, E. B., Beck, J. T., Farmer, S, B., Liu, W., Miller, J., Siripun, K.C., Winder, C. T., Schilling, E. E. and Small, R. L. 2005. The tortoise and the hare II: relative utility of 21 non-coding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany* **92**: 142-166.

Simpson, M. 2006. Plant systematics. Canada, Elsevier Acadamic Press.

Skrede, I., Eidesen, P., Piñeiro, R. and Brochmann, C. 2006. Refugia, differentiation and postglacial migration in arcticalpine Eurasia, exemplified by the mountain avens (Dryas octopetala L.). *Molecular Ecology* **15**: 1827-1840.

Smith, L.M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C.R., Heiner, C., Kent, S. B. H. & Hood, L. E. 1986. Fluorescence detection in automated DNA sequence analysis. *Nature* **321**, 674-679.G16

Smith, A. P. and Young, T. P. 1987. Tropical alpine plant Ecology. *Annual Review of Ecology and Systematics* **18**: 137-138.

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

Stefanović, S., Jager, M., Deutsch, J., Broutin, J. and Masselot, M. 1998. Phylogenetic relationships of Conifers inferred from partial 28S rRNA gene sequences. *American Journal of Botany* **85**: 688-697.

Stehlik, I. 2003. Resistance or emigration? Response of alpine plants to the ice ages. *Taxon* **52**: 499–510.

Stehlik, I., Schneller, J. J. and Bachmann, K. 2002. Immigration and in situ glacial survival of the low-alpine Erinus alpinus (Scrophulariaceae). *Biological Journal of the Linnean Society* **77**: 87–103.

Struwe L, Kadereit JW, Klackenberg J, Nilsson S, Thiv M, Hagen KB, Albert VA. 2002. Systematics, character evolution, and biogeography of Gentianaceae, including a new tribal and subtribal classification. In: Struwe L, Albert VA, (eds.), Gentianaceae: systematic and natural history. Cambridge, Cambridge University Press, pp: 21–309.

Stuessy, T. F. 2009. Plant taxonomy, 2nd ed.. New York, Columbia University Press.

Swofford, D. L. 2003. PAUP*: Phylogenetic analysis using parsimony (*and other methods) Version 4 Sunderland, Massachusetts.

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

Tadesse, M. 2004. Asteraceae (Compositae). In: Flora of Ethiopia and Eritrea, Vol 4, Part 2. Hedberg, I., Friis, I. and Edwards, S. (eds.). Addis Ababa, Uppsala.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. 2011. MEGA5: Molecular evolutionary genetics analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution* **28**: 2731-2739.

Thiv, M., Struwe, L. & Kadereit, J. W. (1999a): The phylogenetic relationships and evolution of the Canarian laurel forest endemic Ixanthus viscosus (Aiton) Griseb. (Gentianaceae): evidence from matK and ITS sequences, and floral morphology and anatomy. *Plant Systematics and Evolution* **218**: 299–317.

Thulin, M. 1989. Papilionoideae. In: Flora of Ethiopia Vol. III, Hedberg, I. and Edwards, S.(eds.). Uppsala.

Tremetsberger, K., Stuessy, T., Kadlec, G., Urtubey, E., Baeza, C.Beck, S., Valdebenito, H., Ruas, C., and Matzenbacher, N. 2006. AFLP phylogeny of South American species of Hypochaeris (Asteraceae, Lactuceae). *Systematic Botany*, **31**: 610-626.

Tribsch, A., Nswetter, P. and Stuessy, T. 2002. Saponaria pumila (Caryophyllaceae) and the ice age in the European alps. *American Journal of Botany* **89**: 2024-2033.

_______ and Schönswetter, P. 2003. Patterns of endemism and comparative phylogeography confirm palaeoenvironmental evidence for Pleistocene refugia in the Eastern Alps. *Taxon* **52** 477–497.

Trivedi, B. S. and Chaturvedi, S. K. 1972. *Voyrioseminites magnus* genet. sp. nov. a fossil seed from Tertiary coal of Malaya. *Geophytology* **1**: 161–164

Van de Peer, Y., De Wachter, Y. (1994). TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. *Computer Application in the Biosciences* **10**: 569-70.

Von Hagen, K. B. and Kadereit, J. W. 2001. The phylogeny of Gentianella (Gentianaceae) and its colonization of the southern hemisphere as revealed by nuclear and chloroplast DNA sequence variation. *Organisms Diversity and Evolution* **1**: 61-79.

Vos, P., Hogers, R., Bleeker, M. Reijans, M., Lee, T. Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M. and Zabeau, M. 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acid Research*, **23**: 4407-4414

Wang, Z., Baker, A., Hill, G. and Edwards, S. 2003. Reconciling actual and inferred population histories in the house finche (*Carpodacus mexicanus*) by AFLP analysis. *Evolution*, **57**: 2852-2864.

Weising, K, Nybom, H., Wolff, K. Kahal, G. 2005. DNA Fingerprinting in Plants: principles, methods and applications. Taylor and Francis, New York, Singapore.

Wesche, K., Miehe, G. and Kaeppeh, M. 2000. The significance of fire for afro-alpine vegetation. *Mountain Research and Development* **20**: 340-347.

White, F. 1983. The vegetation of Africa: a descriptive memoir to accompany the UNESCO/AETFAT/UNSO vegetation map of Africa. *UNESCO*, Paris.

White, T. J., Bruns, T. D., Lee, S. B, Taylor, J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis MA, Gelfand DH, Sninsky JJ. eds. PCR protocols: a guide to methods and applications. *New York Academic Press*, New York.

Wilcox, L. W., Fuerst, P. A. and Floyd, G. L. 1993. Phylogenetic relationships of four charophycean green algae inferred from complete nuclear encoded small subunit rRNA gene sequences. *American Journal of Botany* **80**: 1028-1033.

Wiley, E. O., Siegel-Causey, D., Brooks, D. R., Funk, V. A. 1991. The Compleat Cladist: A Primer of Phylogenetic Procedures. Lawrence, the University of Kansas Printing Service.

Yuan, Y-M., Küpfer, P. and Doyle, J. J. 1996. Infrageneric phylogeny of the genus Gentiana (Gentianaceae) inferred from nucleotide sequences of the internal transcribed spacers (ITS) of nuclear ribosomal DNA. *American Journal of Botany* **83**: 641–652.

Yuan, Y-M., Wohlhauser, S., Möller, M, Chassot, P., Mansion, G., Grant, J., Küpfer, P. and Klackenberg, J. 2003. Monophyly and relationships of the tribe Exaceae (Gentianaceae) inferred from nuclear ribosomal and chloroplast DNA sequences. *Molecular Phylogenetics and Evolution* **28**:500–517.

Yuan, Y-M., Wohlhauser, S., Möller, M., Klackenberg, J., Callmander, M. W. and Küpfer, P. 2005. Phylogeny and Biogeography of *Exacum* (Gentianaceae): A Disjunctive Distribution in the Indian Ocean Basin Resulted from Long Distance Dispersal and Extensive Radiation. *Systematics Biology* **54**: 21-34.

Zabeau, M. and Vos, P. 1993. Selective restriction fragment amplification: a general method for DNA fingerprints.

Appendix 1 Additional Sequences retrieved from GenBank submitted by Struwe *et al*. 2002, if not otherwise stated.

Taxon	GenBank accession number		
	*mat*K	*trn*L (UAA) intron	ITS
Enicostema verticillatum Engl.	AJ388155, AJ388224		
Faroa axillaris Baker	AJ388159, AJ388228		
Gentiana frigida Haenke	AJ388166, AJ388236		
Lisianthius jefensis A. Robyns & T. S. Elias.	AJ0110522, AJ011451		
Macrocarpaea rubra Malme	AJ388175, AJ388245		
Macrocarpaea valerii Standley	AJ388176, AJ388246		
Neurotheca loeselioides (Spruce ex Progel) Baill.	AJ010524, AJ011453		
Neblinantha parviflolia Maguire	AJ388179, AJ388249		
Swertia abyssinica	AJ388191, AJ388261	AJ4080122	AJ306095, AJ306122*
Swertia cordata Wall.	AJ408034, AJ408027	AJ315209*	AJ318557, AJ410336*
Swertia perennis L.	AJ010528, AJ011457	AJ315218*	AJ580550*

* Sequences submitted by Chassot et al., 2001

Appendix 2 Summary of the total no of individuals and populations successfully analyzed in each species

Species name	**Total no of individuals successfully analyzed**	**Total no of populations successfully analyzed**	**Number of replicates**	**Error rate**
Carduus schimperi	119	24	9	1.75
Swertia abyssinica	152	29	14	1.65
Geranium arabicum	219	41	27	1.86
Geranium sp. nov.= G. S. Miehe 3002	24	4	4	
Trifolium cryptopodium	89	17	21	1.73
Umbilicus botryoides	38	7	8	2.03