Phylogenetic Relationships, Biogeography and Species Delimitation:

A case study on southern African *Silene* (Caryophyllaceae)

Ntwai Moiloa

Department of Biological and Environmental Sciences

Faculty of Science

UNIVERSITY OF GOTHENBURG

Gothenburg 2023

grantwae.moiloa@gmail.com

ISBN: 978-91-8069-433-9 (PRINT)

ISBN: 978-91-8069-434-6 (PDF)

Tillgänglig via http://hdl.handle.net/2077/77979

Printed by Stema Specialtryck AB, Borås, Sweden, 2023

Table of Contents

Abstract

The primary aim of this thesis was to explore the phylogenetic relationships and historical biogeography of African members of the plant genus *Silene* (Caryophyllaceae), with special focus on the eight native southern African taxa. Three loci (the nuclear ribosomal internal transcribed spacer region, and the *rps16* and *matK* regions from the plastid genome) generated using traditional Sanger sequencing were used to infer the phylogenetic positions of African taxa, while sequence data from 28 low copy nuclear loci obtained through the target capture method were used to explore shallow phylogenetic relationships within southern African taxa. Bayesian multispecies coalescent methods (StarBeast2, STACEY) were used for species-tree estimation with historical diffusion models (GEO_SPHERE) used simultaneously to infer the biogeographic origins of the monophyletic groups identified. The results indicate that southern African members of *Silene* belong to two different groups (sect. *Elisanthe* and *Silene*) which are relatively distantly related and have different colonisation histories in southern Africa. For North African species of *Silene*, similar analytical approaches resolved the phylogenetic positions of many hitherto understudied taxa but with lack of resolution in the deeper nodes. By leveraging the large amount of sequence data produced via target capture from a comprehensive sample of southern African section *Elisanthe* the monophyly and phylogenetic relationships of *S. rigens*, *S. saldanhensis*, and *S. ornata*, all local endemics to the South African southwest coast, were resolved and in agreement with a recently established taxonomy. However, the results also indicate that the widespread *S. undulata* may not be monophyletic. Several recovered well supported clades within *S. undulata* are congruent with geographical distribution rather than ecology, indicating a spatial differentiation pattern. The work carried out in this thesis demonstrates that target capture sequencing is a valuable method for generating informative sequence data and useful for resolving phylogenetic relationships at shallow levels. Additionally, the thesis also demonstrates how phylogenetic analyses performed under explicitly parameterized statistical models such as the multispecies coalescent can be expanded by incorporation of other sources of information (e.g., biogeographic) to better understand evolutionary relationships of young lineages and thus inform or test existing taxonomic classifications.

Keywords: *Elisanthe*, Illumina sequencing, Multispecies coalescent model, NGS, phylogenetics, phylogeography, species-trees, STACEY, target-capture.

Sammanfattning

Det primära syftet med denna avhandling var att utforska de fylogenetiska släktskapsförhållandena och den historiska biogeografin hos afrikanska representanter för blomväxtsläktet *Silene* (Caryophyllaceae), med särskilt fokus på de som förekommer i södra Afrika. Tre DNA-sekvensregioner (den nukleära ribosomala interna transkriberade spacer-regionen och *rps16*- och *matK*-regionerna från plastidgenomet) genererade med traditionell Sanger-sekvensering användes för undersöka de fylogenetiska positionerna för afrikanska taxa, medan 28 nukleära multi-lokus sekvenser genererades för att utforska grunda fylogenetiska samband inom sektionen *Elisanthe* från Södra Afrika. Bayesianska koalescentmetoder (StarBeast2, STACEY) användes för art-trädskattning och historiska diffusionsmodeller (GEO_SPHERE) användes för att spåra deras biogeografiska ursprung. Resultaten visar att taxa förekommande i södra Afrika tillhör två olika grupper (sektionerna *Elisanthe* och *Silene*) som är relativt avlägset besläktade med varandra och har olika kolonisationshistorier i södra Afrika. För Nordafrikanska taxa klargör resultaten de fylogenetiska positionerna för många hittills lite, eller inte alls studerade arter i sektionen *Silene*. Genom att utnyttja den stora mängden sekvensdata som producerats från ett omfattande urval av representanter för sektionen *Elisanthe* i Södra Afrika påvisades släktskapsförhållandena mellan *S. rigens*, *S. ornata* och *S. saldanhensis* (alla från Sydafrikas sydvästkust), i överensstämmelse med en nyligen etablerad taxonomi. Resultaten indikerar också att den geografiskt mera utbredda *S. undulata* kan vara monofyletisk. Flera väl understödda grupper inom arten är geografiskt begränsade snarare än ekologiskt, vilket indikerar geografiskt betingad differentiering. Denna avhandling visar att "Target capture sequencing" är en värdefull metod för att generera informativ DNA-sekvensdata och användbar för att lösa fylogenetiska samband på grunda nivåer. Dessutom visar avhandlingen också hur fylogenetiska analyser utförda under explicita statistiska modeller kan utökas genom att införliva andra informationskällor (t.ex. biogeografiska) för att bättre förstå evolutionära samband mellan unga linjer och därmed informera eller testa befintliga taxonomiska klassificeringar.

List of manuscripts

This thesis is based on the following papers and referred to in the text by roman numerals:

I. **Moiloa**, **N.A**., Mesbah, M., Nylinder, S., Manning, J., Forest, F., de Boer, H.J., Bacon, C.D., Oxelman, B., 2021. Biogeographic origins of southern African *Silene* (Caryophyllaceae). *Mol. Phylogenet. Evol*. 162, 107-199. https://doi.org/10.1016/j.ympev.2021.107199

Authorship contribution from **Ntwai Moiloa** (according to CRediT): Conceptualization, Investigation, Formal analysis, Data curation, Methodology, Validation, Writing - original draft, Visualization.

II. Mesbah, M*., **Moiloa N.A***, Sáez, L., Oxelman, B. A phylogenetic study of the genus *Silene* (Caryophyllaceae) in North Africa. *Manuscript*

Authorship contribution from **Ntwai Moiloa** (according to CRediT): Conceptualization (in part–biogeographical analysis), Data Curation, Formal analysis, (in part–biogeographical analysis and preparation of Figures), Investigation, Methodology, Visualization, Writing – Review & Editing.

* Denotes shared first authorship

III. **Moiloa**, **N.A**., Dludlu, M.N., Bello, A., Shaik, Z., Muasya, A.M., Oxelman, B., 2022. Chapter 19. Systematics and evolution. In: de Boer, H., Rydmark, M.O., Verstraete, B., Gravendeel, B. (Eds). *Molecular identification of plants*: *from sequence to species*. Advanced Books. https://doi.org/10.3897/ab.e98875

Authorship contribution from **Ntwai Moiloa** (according to CRediT): Conceptualization, Project administration, Writing - Original Draft.

IV. **Moiloa**, **N.A**., Oxelman, B. A Phylogenetic Study of Southern African Members of *Silene* Section *Elisanthe* (Caryophyllaceae) Inferred from Target Capture Sequence Data. *Manuscript*

Authorship contribution from **Ntwai Moiloa** (according to CRediT): Conceptualization, Data Curation, Formal analysis, Investigation, Methodology, Validation, Project administration, Visualization, Writing - Original Draft

1. Introduction

1.1.1 The genus *Silene* L. (Caryophyllaceae)

Silene is a genus comprising *ca*. 870 species distributed globally, mainly diverse in the Mediterranean Basin and in Central and Western Asia (Oxelman *et al*., 2013; Jafari *et al*., 2020). Plants in the genus are mainly annuals or short-lived perennials with a few species growing as small shrubs or geophytes (Manning & Goldblatt, 2012a). The taxonomic classification has historically been based on a number of morphological characters *viz*. calyx, seed and petal morphology; carpel number, length of carpophore (i.e., the internode between the calyx and corolla), indumentum (glandular or eglandular), number of styles and stamens, number of calyx teeth, inflorescence type (i.e., monochasium or dichasium) (e.g., Naciri *et al*., 2017). However, these morphological characters show high levels of homoplasy, resulting in classifications which are incongruent with most phylogenetic studies on the genus (Oxelman & Lidén, 1995; Oxelman *et al*., 2001; Eggens, 2007; Petri & Oxelman, 2011; Rautenberg *et al*., 2012). The use of morphological characters with high levels of homoplasy has impacted the stability of classifications in the genus to an extent where the circumscription of the genus has expanded since Linnaeus (1753). Most late 20th Century studies on the genus have historically adopted the taxonomic classification by Chowdhuri (1957) which recognized 44 sections and 45 subsections. However, phylogenetic studies (e.g., Oxelman & Lidén 1995; Oxelman *et al*., 1997, 2001; Popp & Oxelman, 2004; Frajman *et al*., 2009a, b; Rautenberg *et al*., 2010; Petri & Oxelman, 2011; Toprak *et al*., 2016; Đurović *et al*., 2017; Du Pasquier *et al*., 2017; Naciri *et al*., 2017) focusing on specific groups in the genus did not show support for Chowdhuri's (1957) infrageneric classification. Moreover, contrary to Chowdhuri (1975), a recent infrageneric circumscription by Jafari *et al*. (2020) identified 33 sections placed in four distinct clades comprising three subgenera *viz*. *S.* subg. *Lychnis* (L.) Greuter, *S.* subg. *Behenantha* (Otth) Torr. & A.Gray, *S.* subg. *Silene* (Fig. 1).

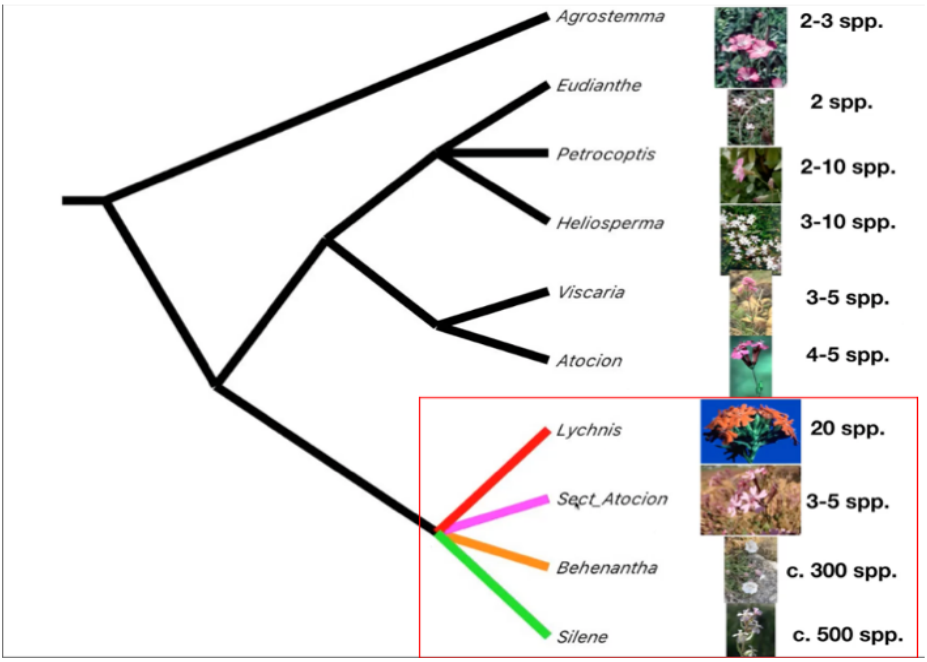

Figure 1: Cladogram summarising generic relationships in Sileneae. The genus *Silene* with infrageneric groups is highlighted by a red box.

In the African context, about 150 native species are currently recognized, with the majority of the species distributed in North Africa and few found in sub-Saharan and southern Africa (Manning & Goldblatt, 2012a; Dobignard & Chatelain, 2011). The native African species are placed in nine sections belonging to the three subgenera recognized by Jafari *et al*. (2020). Subgenus *Silene* in Africa is represented by *S*. sect. S*iphonomorpha* Otth, *S*. sect. *Muscipula* (Tzvelev) Oxelman, F.Jafari & Gholipour and *S*. sect. *Portenses* F.Jafari & Oxelman, and *S*. sect. *Silene*, which is most diverse in North Africa (Maire, 1963). The subgenus *Behenantha* in Africa is represented by *S*. sect. *Behenantha* Otth, *S*. sect. *Melandrium* (Röhl.) Rabeler, *S*. sect. *Sedoides* Oxelman & Greuter, *S*. sect. *Conoimorpha* Otth and *S*. sect. *Elisanthe* (Fenzl) Ledeb., the latter section is diverse and includes species endemic to southern Africa (Manning & Goldblatt, 2012a; Jafari *et al*., 2020). Although the recent phylogenetic circumscription by Jafari *et al*. (2020) included some African taxa, the sampling was not comprehensive and the available classifications are largely based on previous regional treatments (e.g., Sonder, 1860; Turrill, 1956; Wild 1961; Maire, 1963; Wickens, 1976; Gilbert, 2000; Bocquet, 1977; Masson,

1989; Goldblatt & Manning 2000; Manning & Goldblatt, 2012a, b; Snijman, 2013). With the high levels of endemism in the understudied African species, it is essential that studies focusing on the taxonomy and phylogenetic placement of the African species are conducted. The taxonomic and phylogenetic position of the southern African and North African species are addressed in Manuscripts **I** and **II**, respectively.

1.1.2 Southern African *Silene*

Silene in southern Africa is represented by eight species mainly distributed along the coastal regions of southern Africa, with a few species extending further inland into the drier parts of the sub-continent. The centre of diversity of *Silene* in southern Africa is in the Greater Cape Floristic Region of southern Africa where most of the species are endemic (Manning & Goldblatt, 2012a, b; Snijman, 2013). Taxonomically, native southern African *Silene* have previously been misidentified by several biologists, mistakenly identifying the species as European e.g., Thunberg (1794) and Burman (1768). Thus, some taxa remained undescribed for almost a century, with Sonder (1860) being the first to present a comprehensive classification of the southern African taxa. Although there has been a number of studies on the native taxa since Sonder (1860) (e.g., Rohrbach, 1869; Bocquet, 1977; Masson, 1989; and Goldblatt & Manning, 2000), the major taxonomic revision is by Manning & Goldblatt (2012a). In their taxonomic revision, Manning & Goldblatt (2012a) recognized eight species native to southern Africa are mostly diverse in the predominantly winter rainfall region. Subsequent to the revision by Manning & Goldblatt (2012a), a recent phylogenetic study by Jafari *et al*. (2020) have demonstrated that the eight species belong to two distantly related sections (*Elisanthe* and *Silene*). The sections, *Elisanthe* and *Silene*, are phylogenetically placed in subg. *Behenantha* and subg. *Silene*, respectively (Fig. 1) (Jafari *et al*., 2020).

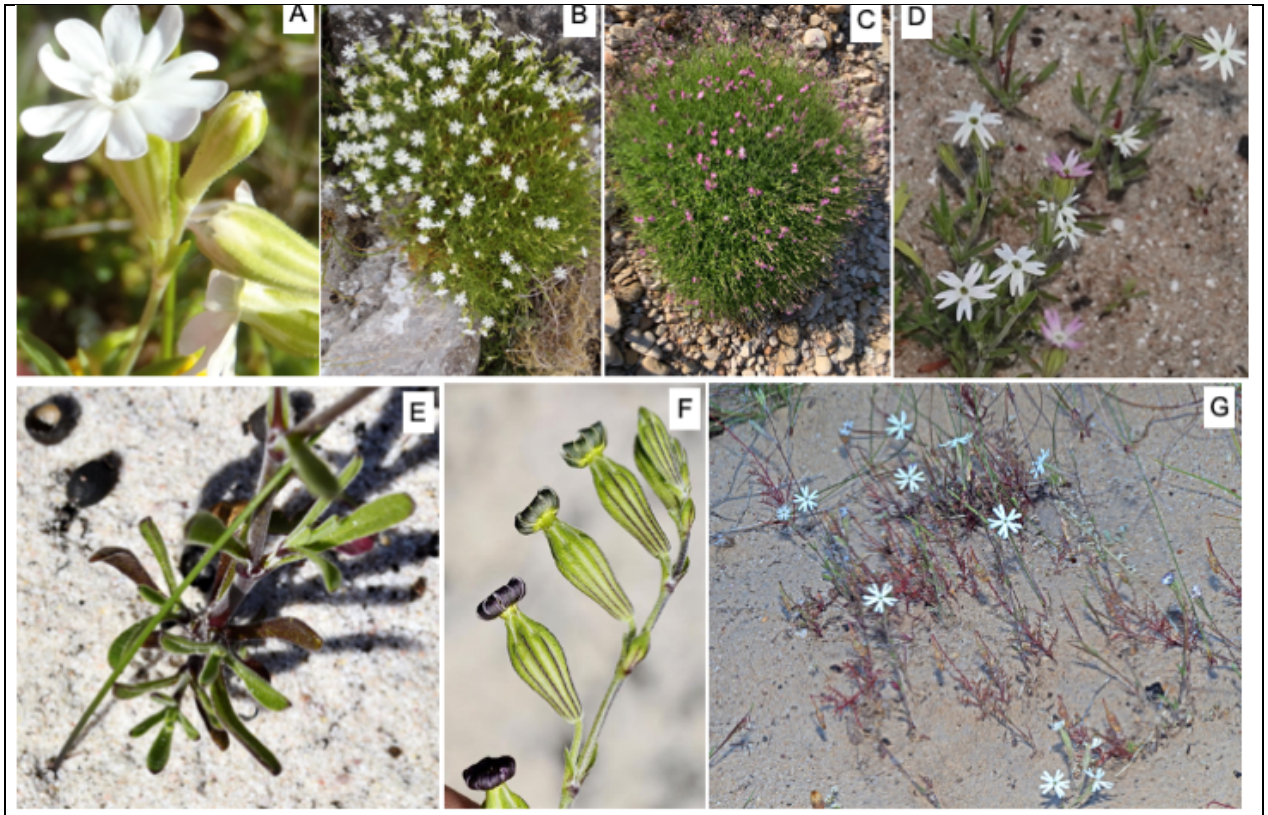

Figure 2.1: Composite plate showing morphological diversity of southern African members of section *Silene*: A–C, *S. mundiana*; D–F, *S. aethiopica* subsp. *aethiopica*; G, *S. aethiopica* subsp. *longiflora*. Images obtained from iNaturalist (https://www.inaturalist.org/) and used under the Creative Commons licence (CC BY-NC). Image credits: A–B, Nicola van Berkel; C, smatt853; D, desertnaturalist; E–F, Richard Adcock; G, Felix Riegel.

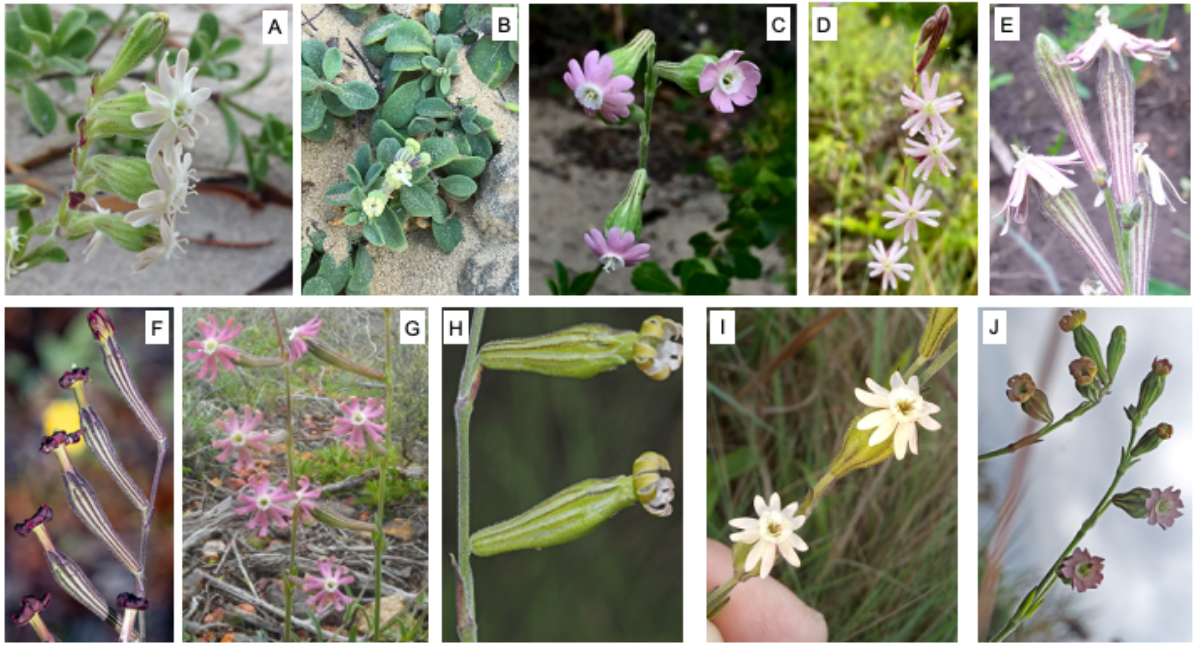

Figure 2.2: Composite plate showing morphological diversity of southern African members of section *Silene*: A–B, *S. crassifolia* subsp. *crassifolia*; C, *S. crassifolia* subsp. *primuliflora*; D–E, *S. burchellii* subsp. *burchellii*; F–G, *S. burchellii* subsp. *pilosellifolia*; H–I, *S. burchellii* subsp. *modesta*; J, *S. burchellii* subsp. *multiflora*. Images obtained from iNaturalist (https://www.inaturalist.org/) and used under the Creative Commons licence (CC BY-NC). Image credits: A, Micky Orrey; B, David Hoare; C, Brendan Cole; D, Wendy Hogarth; E, Wendy Hitchcock; F, Magriet B; G, Petra Broddle; H, Felix Riegel; I, Garth Aiston; J, Felix Riegel.

In southern Africa, section *Silene* is represented by four species which are morphologically and ecologically distinct *viz*. *S. burchellii* Otth, *S. crassifolia* L., *S. aethiopica* Burm., and *S. mundiana* Eckl. & Zeyh. Species in this section are diagnosed by a monochasial inflorescence (albeit with varying numbers of flowers borne on each inflorescence; Fig 2.1 & 2.2), reniform seed testa with undulate peripheral wings and a hilar notch (Fig. 3), indumentum with eglandular hairs, a calyx lacking conspicuous anastomoses between reticulate veins (Sonder, 1860; Rohrbach 1869; Bocquet 1977; Masson 1989; Goldblatt & Manning, 2000; Manning & Goldblatt, 2012a).

Silene mundiana is a woody perennial geophyte diagnosed by a few flowered inflorescence (usually 1–2 flowers per inflorescence), ovate calyx lobes, clavate calyx, and the plants usually grow to form tufted sprawling mats (Fig 2.1 A–C), discoid seeds (Fig. 3 A–B) (Manning & Goldblatt, 2012b). The species is restricted to the limestone outcrops of De Hoop National Park of the Overberg Region (Fig. 4A) (Manning & Goldblatt, 2012 a, b). *Silene aethiopica* is an annual species diagnosed by an ovoid to lanceolate calyx, a carpophore which may either be shorter (*S. aethiopica* subsp. *aethiopica*) or longer (*S. aethiopica* subsp. *longiflora* J.C.Manning & Goldblatt) than the capsule, with an erect or decumbent habit (Fig 2.1. D–G), discoid seeds (Fig. 3 CD) (Manning & Goldblatt, 2012a). The species is endemic to the winter rainfall region extending from the Richtersveld in the West along the South coast towards Laingsburg, where *S. aethiopica* subsp. *aethiopica* is a common form mainly restricted to the Western parts of the distribution and is replaced by *S. aethiopica* subsp. *aethiopica* in the North-Eastern parts further inland (Fig. 4 B). A peculiar maritime form with fleshy leaves and short calyx, *S. dewinteri* Bocquet, was identified by Bocquet (1977), but later rejected by Manning & Goldblatt (2012a). *Silene crassifolia* is a prostrate geophytic perennial species diagnosed by succulent to leathery suborbicular leaves which may either be felted with a shorter calyx (*S. crassifolia* subsp. *crassifolia*) or may be puberulous with a longer calyx (*S. crassifolia* subsp. *primuliflora* (Eckl. & Zeyh.) J.C.Manning & Goldblatt), a subglobose and conspicuously pleated capsule (Fig. 2.2 A–C) (Sonder, 1860; Bocquet & Kiefer 1878; Masson 1989; Manning & Goldblatt, 2012a). The species is distributed along the coast from Saldanha to the margins of Kwazulu Natal, with *S. crassifolia* subsp. *crassifolia* mainly occurring to the West of the Breede River while *S. crassifolia* subsp. *primuliflora* occurs mostly to the Eastern parts of the Breede River (Fig. 4 C) (Manning & Goldblatt, 2012a, b).

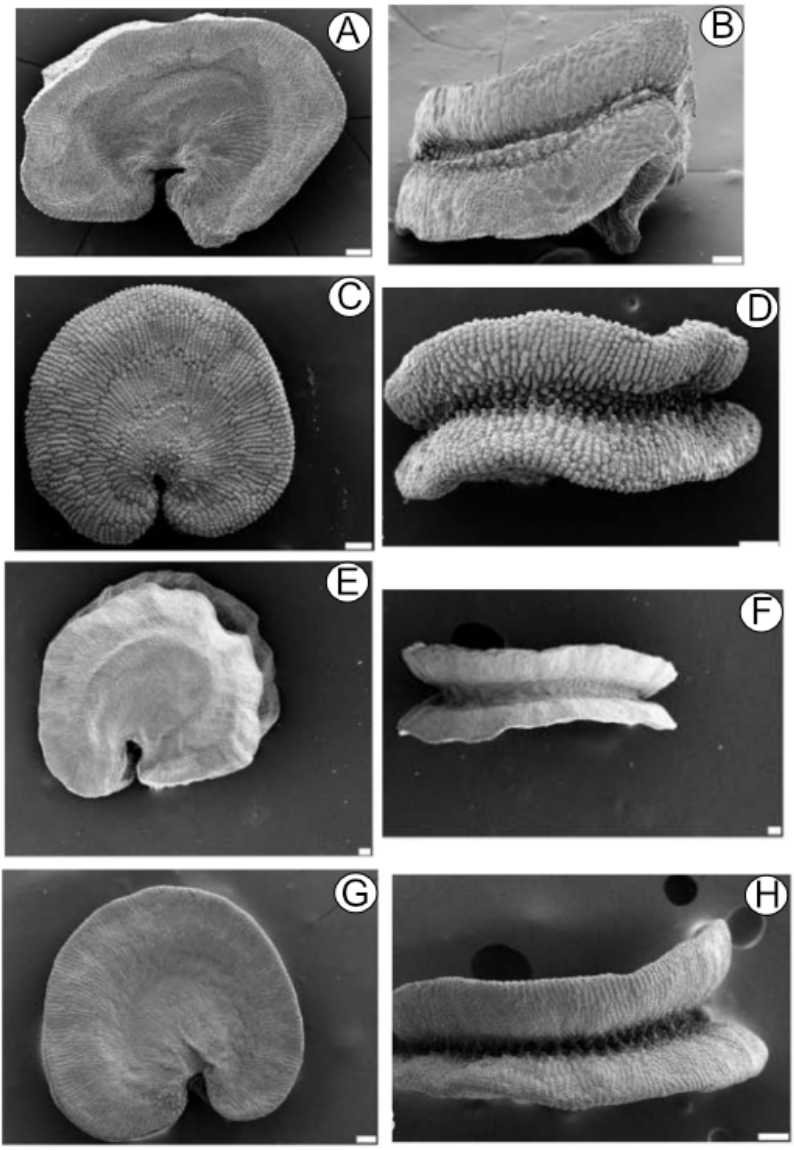

Figure 3: SEM micrographs showing the morphological variability of seed testa in members of southern African section *Silene*: A–B, *S. mundiana*; C–D, S. aethiopica; E–F, *S. crassifolia*; G–H, *S. burchellii*. Scale bar = 100 µm. Images adapted from Manning & Goldblatt (2012a) and used with permission under licence from SANBI GRE.

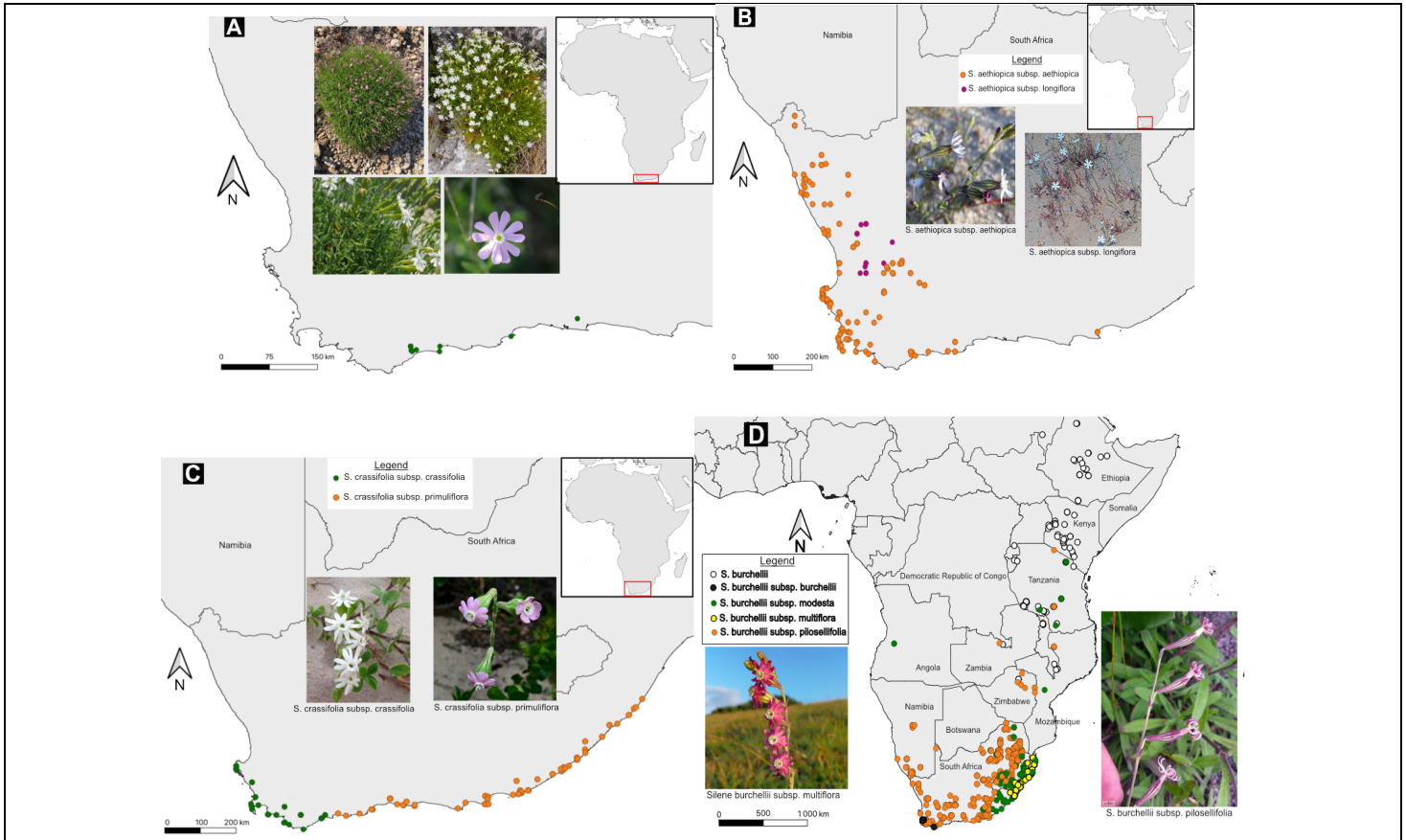

Figure 4: Geographical distribution of southern African members of section *Silene*: A, *S. mundiana*; B, *S. aethiopica*; C, *S. crassifolia*; D, *S. burchellii*. Occurrence records obtained from determined preserved specimens listed in Manning & Goldblatt (2012a). Specimens whose identity could not be verified were not included in the mapping. Images obtained from iNaturalist (https://www.inaturalist.org/) and used under the Creative Commons licence (CC BY-NC). Image credits: *S. mundiana*, smatt853 (top left), Nicola van Berkel (top right), Felix Riegel (bottom left), Aaron Jones (bottom right); *S. aethiopica* subsp. *aethiopica*, Cecile Roux; *S. aethiopica* subsp. *longiflora*, Felix Riegel; *S. crassifolia* subsp. *crassifolia*, Micky Orrey; *S. crassifolia* subsp. *primuliflora*, Brendan Cole; *S. burchellii* subsp. *multiflora*, Craig Peter; *S. burchellii* subsp. *pilosellifolia*, Chris Vynbos.

Silene burchellii is the most diverse and widespread species among the southern African members of section Silene, with a distribution across southern Africa extending along the East African Rift mountains to Ethiopia (Fig. 4 D) (Wickens, 1976; Manning & Goldblatt, 2012b; Snijman, 2013). *Silene burchellii* is a perennial geophyte diagnosed by a globose capsule, mostly curved (i.e., not linear) carpophore, broad oblanceolate to obovate leaves (Fig. 2.2 D–J). Manning & Goldblatt (2012a) recognized four subspecies which are morphologically distinct and geographically restricted, these include: (i.) *S. burchellii* subsp. *burchellii*, which is diagnosed by decumbent and compactly branched habit, a medium sized calyx (Fig. 2.2 D–E) and is distributed from Gansbaai to the Agulhas (Fig. 4 D); (ii) *S. burchellii* subsp. *pilosellifolia* (Cham. & Schltdl.) J.C.Manning & Goldblatt is diagnosed by suberect habit and the longest calyx (mostly curved; Fig. 2.2 F–G), with a distribution across the entire southern Africa (Fig. 4 D); (iii) *S. burchellii* subsp. *modesta* J.C.Manning & Goldblatt is diagnosed by an erect habit, linear subglabrous leaves with a long carpophore (Fig. 2.2 H–I), and distributed in the Eastern parts of southern Africa extending along the East African Rift Mountains to Ethiopia (Fig. 4 D); (iv) *S. burchellii* subsp. *multiflora* J.C.Manning & Goldblatt is diagnosed by a decumbent to tufted habit with the broadest leaves and an inflorescence with the highest number of flowers and a short calyx (Fig. 2.2 J), with a distribution mainly in the KwaZulu Natal area extending Northwards further inland towards the Drakensberg (Fig. 4 D) (Turrill, 1954; Hedberg, 1954, 1957; Manning & Goldblatt, 2012a, b; Snijman, 2013).

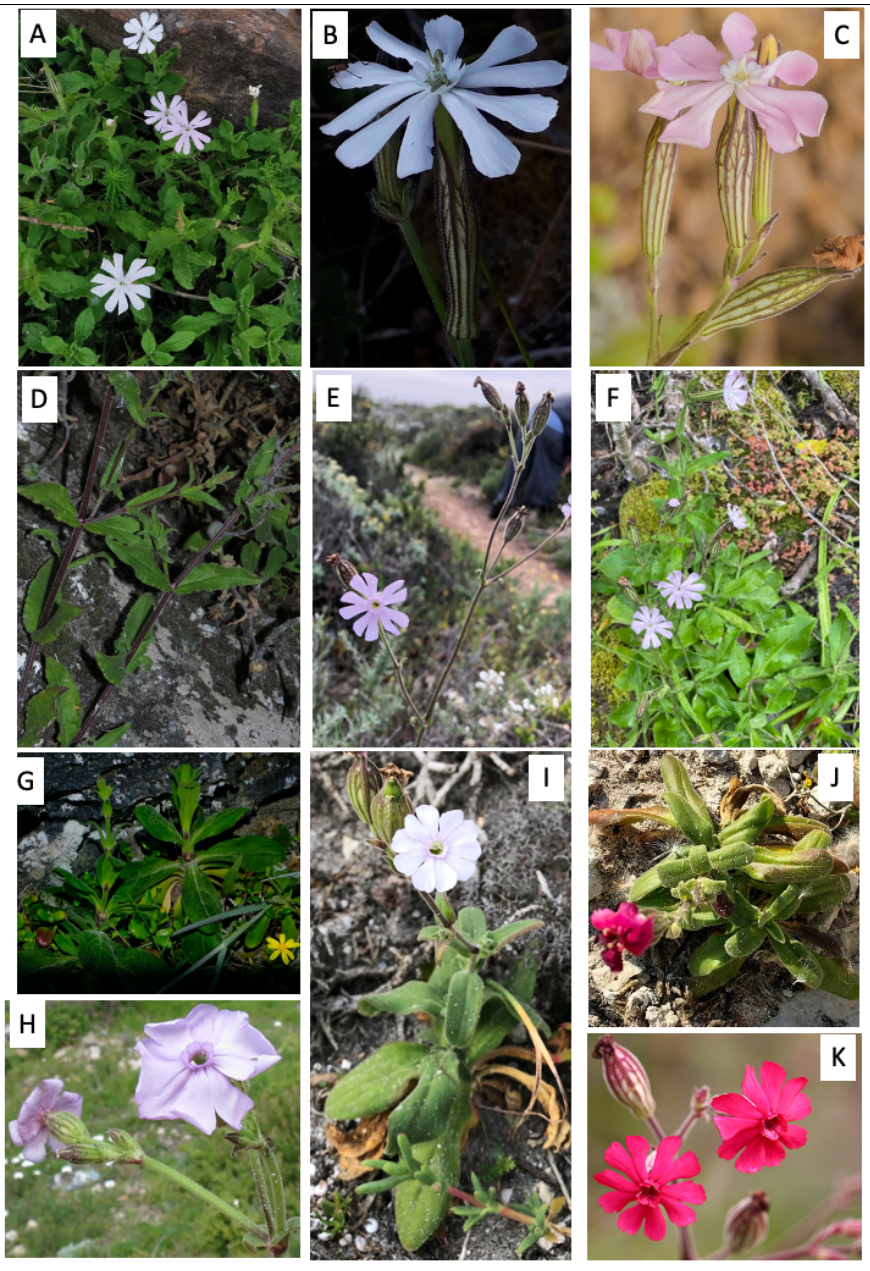

Figure 5: Composite plate showing morphological diversity of southern African members of section *Elisanthe*: A–C, *S. undulata*; D–F, *S. rigens*; G–I, *S. saldanhensis*; J–K, *S. ornata*. Images obtained from iNaturalist (https://www.inaturalist.org/) and used under the Creative Commons licence (CC BY-NC). Image credits: A, Annerie Senekal; B, Sally Adam; C, Brendan Cole; D, Jeremy Gilmore; E, Bengt Oxelman; F, Barbara Laurie; G, Nick Helme; H, kooscl; I–J, Bengt Oxelman; K, Bastiaan Notebaert.

The second group represented in southern Africa is section *Elisanthe*, comprising four species *viz*. *S. undulata* Ait., *S. rigens* J.C.Manning & Goldblatt, *S. ornata* Ait., and *S. saldanhensis* J.C.Manning & Goldblatt. Species in this section are diagnosed by a few-flowered dichasial inflorescence (Fig. 5), globose seed testa lacking undulate peripheral wings (Fig. 6), indumentum with glandular hairs, unilocular ovaries and a calyx with conspicuous anastomosis between reticulate veins (Sonder, 1860; Chater & Walters, 1964; Manning & Goldblatt, 2012a, b; Snijman, 2013). *Silene saldanhensis* is a perennial species diagnosed by a compact tufted branching habit, lanceolate leaves, an urn-shaped calyx with conspicuous ribs, overlapping bifid petal limbs coloured mauve to purple (Fig. 5 G–I), colliculate seed testa (Fig. 6 A–B), flowering from September to October (Manning & Goldblatt, 2012a, b). The species is endemic to the calcareous dunes of Saldanha Bay, with collections only known from a handful of localities around Saldanha Bay, Hopefield and Postberg Reserve (Fig. 7 A) (Manning & Goldblatt, 2012a). The species is currently listed as endangered (EN) in the Red List of South African Plants, with populations sizes decreasing due to habitat loss as a result of infrastructure development as well as agriculture (von Staden & Claassens, 2014). *Silene ornata* is a perennial species diagnosed by a sprawling/straggling growth habit, cauline spreading leaves, calyx lacking conspicuous ribs, non-overlapping petal limbs with a deep carmine colour (Fig. 5 J–K), tuberculate seed testa (Fig. C–D), flowering from September to October (Manning & Goldblatt, 2012a, b). Similar to *S. saldanhensis*, *S. ornata* is endemic to the Saldanha Bay area extending to the West Coast National Park, where the species occupies rocky limestone and granite outcrops (Fig. 7 A) (Raimondo *et al*., 2009; Manning & Goldblatt, 2012b). *Silene ornata* is listed as vulnerable (VU) in the Red List of South African Plants threatened by industrialization and agriculture in unprotected areas, but with thriving populations in protected areas such as the Postberg Reserve in the West Coast National Park (Raimondo *et al*., 2009; von Staden & Helme, 2014). The two species (i.e., *S. saldanhensis* & *S. ornata*) are closely allied and share close morphological similarity in several aspects but differ mainly in flower colour (mauve and carmine, respectively, Fig. 5 G–K) as well as the ecological niches and substrate (i.e., soils derived from different geological strata) they occur on.

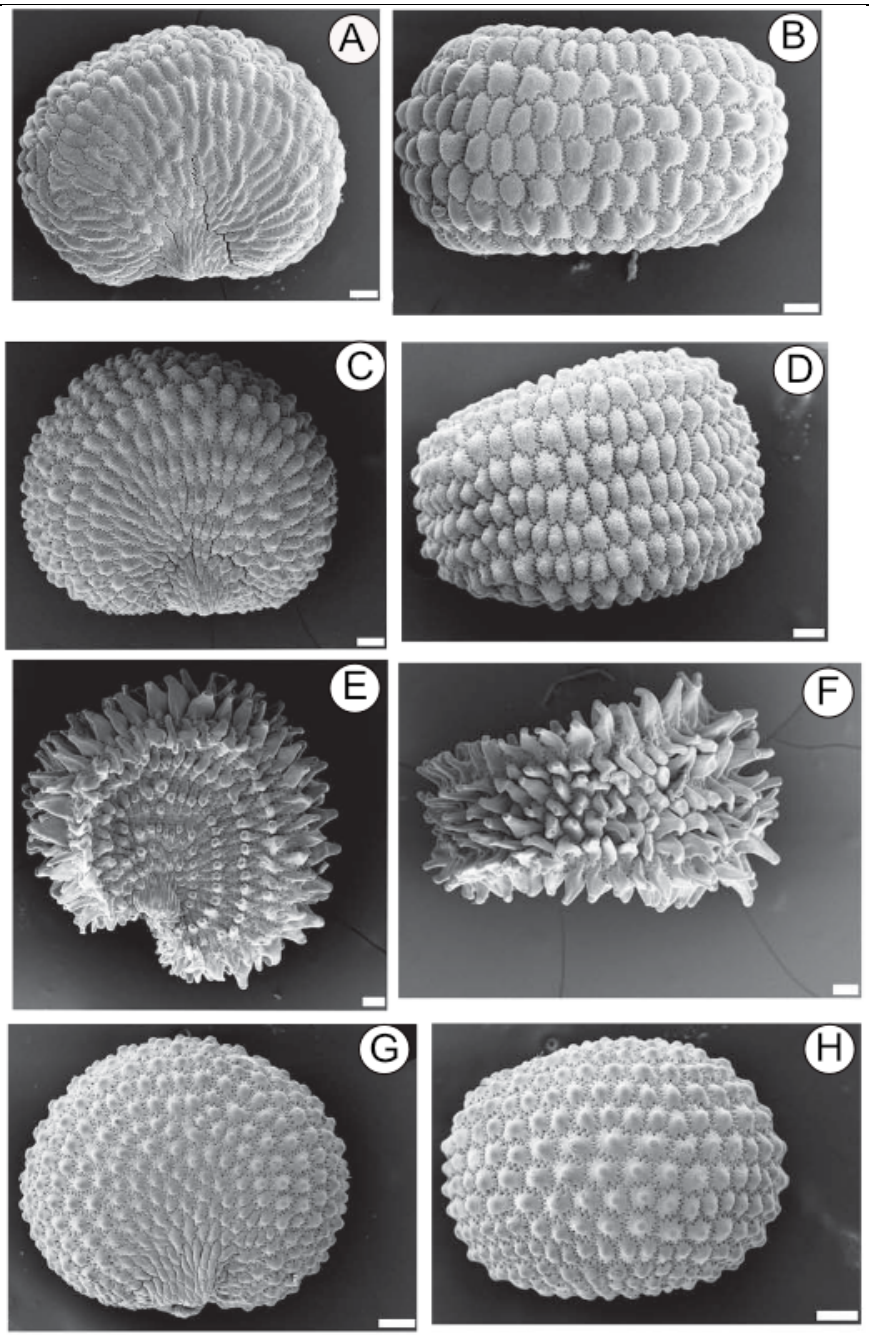

Figure 6: SEM micrographs showing the morphological variability of seed testa in members of southern African section *Elisanthe*: A–B, *S. saldanhensis*; C–D, *S. ornata*; E–F, *S. rigens*; G–H, *S. undulata*. Scale bar = 100 µm. Images adapted from Manning & Goldblatt (2012a) and used with permission under licence from SANBI GRE.

Silene rigens is a tufted perennial species diagnosed by long stiff erect stems, spatulate leaves, non-overlapping cuneate petals with a pale pink to mauve colour (Fig. 5 D–F), echinate seeds (Fig. 6 E–F), flowering from September to October (Manning & Goldblatt, 2012b). The species occurs on deep calcareous sands, distributed from Saldanha Bay extending southwards to the Cape Flats extending eastward towards Hermanus (Fig. 7 A). *Silene rigens* is listed as near threatened (NT) in the Red List of South African Plants, with major population threats being from competition with invasive alien plants, habitat loss due to agriculture and industrialization (von Staden *et al*., 2014). *Silene undulata* is the most diverse and widespread species among the southern African members of section *Elisanthe* with a distribution across most parts of southern Africa reaching the highlands of Zimbabwe (Fig. 7 B, with the exclusion of some more arid desert-like areas e.g. Great Karoo and the Kalahari) (Manning & Goldblatt, 2012a, b; Snijman, 2013). The species is perennial diagnosed by multibranched sprawling stems, oblanceolate to spatulate cauline undulate leaves, subulate calyx lobes, non-overlapping cuneate petals with a white to pink colour (Fig. 5 A–C), and colliculate to tuberculate seed testa (Fig. 6 G–H) (Rohrbach, 1869; Goldblatt & Manning, 2000; Manning & Goldblatt, 2012a, b; Snijman, 2013). The phenology differs depending on where the species occurs; with plants in the winter rainfall region flowering in August to December, while plants in the summer rainfall region flower in November to as late as June in some parts (Manning & Goldblatt, 2012a, b). *Silene undulata* is morphologically diverse across the entire distribution range, where there are distinct maritime forms, winter rainfall forms, summer rainfall to arid inland forms; which all demonstrate variability in growth habit, calyx and carpophore length, vestiture, capsule size and shape (Sonder, 1860; Manning & Goldblatt, 2012a). Variation in these characters have been the basis of the recognition of these forms at species level e.g., the carpophore length has been emphasised in segregating between *S. bellidioides*, *S. undulata* and *S. capensis*, which have a comparably short, intermediate, and long carpophore, respectively (Sonder, 1860). The most recent taxonomic revision however did not recognize the above-mentioned species mainly due to the insufficiency of morphological characters enabling confident diagnosis. Thus, according to the current taxonomic classification by Manning & Goldblatt (2012a), *S. undulata* is divided into two subspecies, *S. undulata* subsp. *undulata* (Fig. 5 A–B) and *S. undulata* subsp. *polyantha* J.C.Manning & Goldblatt (Fig. 5 C). *Silene undulata* subsp. *polyantha* is a taxon restricted to the eastern parts of southern Africa (Fig. 7 B; KwaZulu Natal and Swaziland), diagnosed by more compact branching stems, a comparably short capsule and carpophore (distinctly shorter than most forms of *S. undulata* subsp. *undulata*) (Manning & Goldblatt,

2012a). Both subspecies are currently listed as least concern (LC) in the Red List of South African Plants (Pooley, 2003; Raimondo *et al*., 2009; von Staden, 2014).

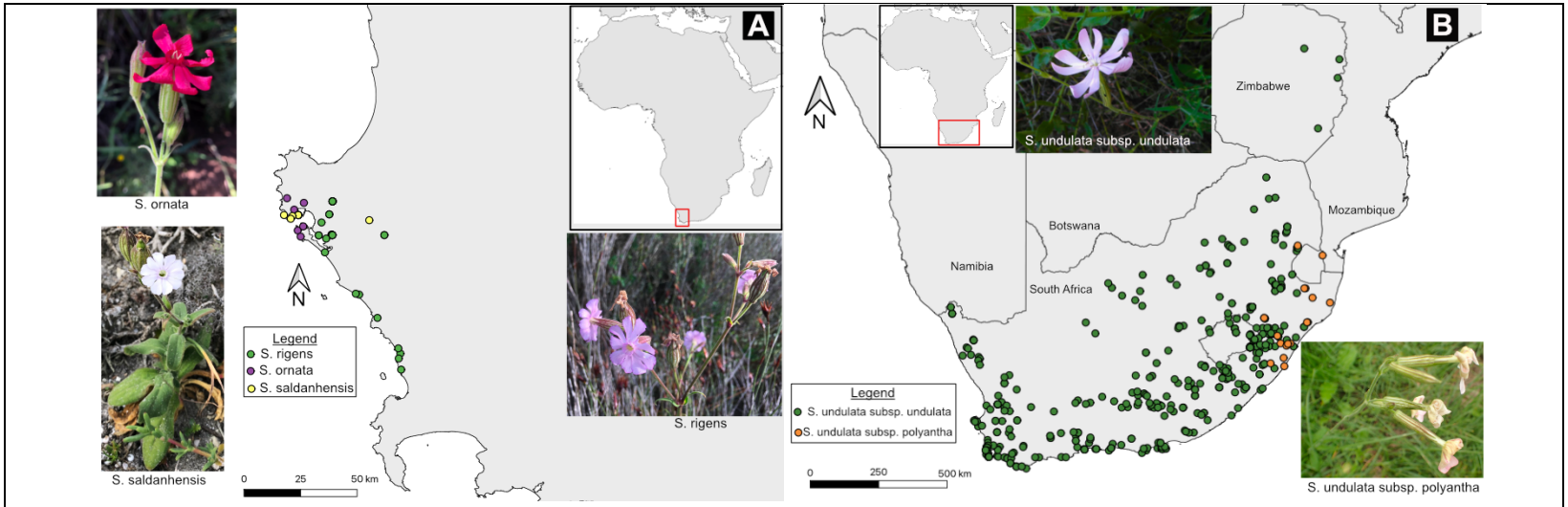

Figure 7: Geographical distribution of southern African members of section *Elisanthe*: A, *S. ornata*, *S. rigens* & *S. saldanhensis*; B, *S. undulata*. Occurrence records obtained from determined preserved specimens listed in Manning & Goldblatt (2012a). Specimens whose identity could not be verified were not included in the mapping. Images obtained from iNaturalist (https://www.inaturalist.org/) and used under the Creative Commons licence (CC BY-NC). Image credits: *S. ornata*, Micky Orrey; *S. saldanhensis*, Bengt Oxelman; *S. rigens*, Peter Zika; *S. undulata* subsp. *undulata*, Jeffrey Michael; *S. undulata* subsp. *polyantha*, Peter Warren.

1.2. The Greater Cape Floristic Region, a center for diversity and endemism

The Greater Cape Floristic Region (GCFR; sensu Born *et al*., 2007) is a megadiverse biodiversity hotspot located at the southwestern most tip of Africa. This biologically rich region which spans across parts of southern Africa (Fig. 8) is characterised by high levels of species richness and endemism. The flora of the region has received much attention over the years, during which several authors attempted to unpack the floristic uniqueness which characterise the region and to define the boundaries of the region (e.g., Drège, 1843–1844; Bolus, 1875; Marloth 1908; Weimarck 1941; Bond & Goldblatt 1984). The history of the region's various circumscriptions is documented elsewhere (e.g., Manning & Goldblatt, 2012b). Here, I give an overview of the region as currently circumscribed by Born *et al*. (2007); a delimitation adopted by Snijman (2013) and Manning & Goldblatt (2012b). Geographically, the GCFR covers the area from Hottentots Bay in South Namibia along the Atlantic coast extending further south along the coast, covering the southernmost tip of Africa in the Western Cape of South Africa and extending eastwards to Port Elizabeth/Gqeberha (Fig. 8; Snijman, 2013). Interestingly there is a decrease in seasonality of the rainfall regime when moving from West to East and extending further inland (Deacon *et al*., 1992; Cowling *et al*., 1997; Manning & Goldblatt, 2000; Linder, 2003). In addition to varying rainfall regimes, the region is also characterised by edaphic heterogeneity which includes a variety of soil types derived from geologically different substrates e.g., sandstone, shale, dolomitic, limestones and calcareous derived soils (Bond & Goldblatt, 1984; Linder, 1985; Mucina & Rutherford, 2006; Cowling *et al*., 2009). Additionally, there are sharp differences in altitudinal gradients which give rise to various floral compositions depending on the substrate (i.e., soil type) and altitude (Goldblatt & Manning, 1996; Cowling *et al*., 2009). The GCFR is broadly divided into two subregions: the Core Cape Subregion (CCR; Manning & Goldblatt, 2012b) and the Extra Cape Subregion (ECR; Snijman, 2007). Following the Biome concept (see Mucina, 2018 and references therein), the two subregions broadly coincide with the Fynbos and Succulent Karoo biomes, respectively, which both are considered part of the most biologically rich and endangered ecoregions (Mittermeier *et al*. 1999). Both biomes are to a large extent sharply distinct, comprising specific climatic, edaphic, ecological and floristic compositions, which are briefly described below. In general, the delineation of the boundaries according to Mucina & Rutherford (2006) and Born *et al*. (2007), there are sharp distinction areas where the CCR can be differentiated from the ECR e.g., the slopes of the Bokkeveld Mountains where there is a sharp change in floristic composition gradient. Whereas in transitional zones it is difficult to segregate between the two

subregions, particularly in areas characterized by mainly azonal floristic elements e.g., the Overberg region (Mucina & Rutherford, 2006; Snijman, 2013).

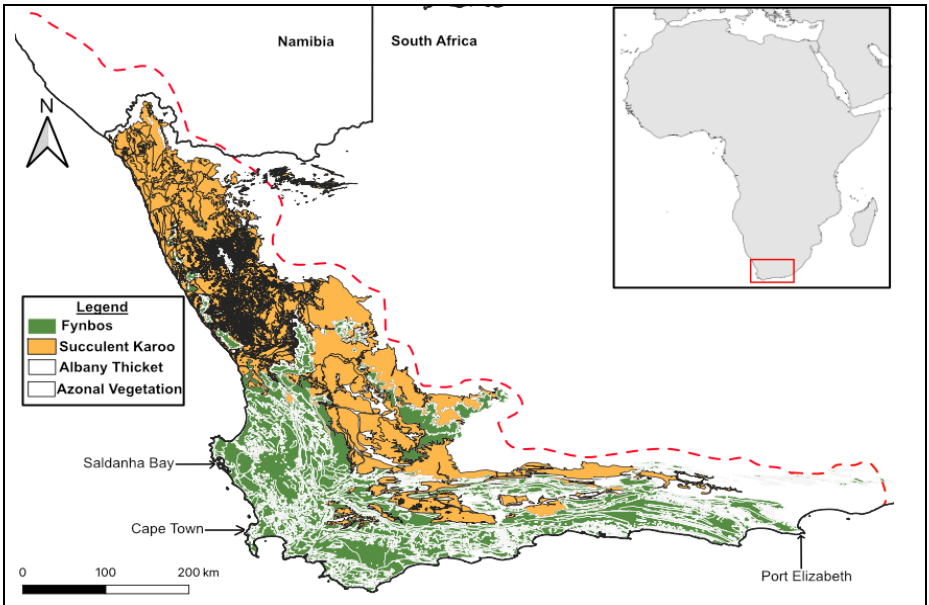

Figure 8: An overview of the geographical extent of the Greater Cape Floristic Region demarcated by red dashed lines. The Core Cape Subregion represented by the Fynbos Biome is shown in green; the Extra Cape Subregion represented by the Succulent Karoo is shown in orange.

The CCR, formerly known as the Cape Floristic Region (CFR; Goldblatt & Manning, 2000) is generally well studied and continues to receive much more attention compared to the ECR. Covering a land area of ca 90 769 km^2, the subregion harbours a diversity of about 9383 angiosperm species (Manning & Goldblatt, 2012b). However, the subregion is rather compositionally biased, with lineage richness in particular clades to the exclusion of others, resulting in an assemblage of lineages giving rise to the defining elements of the Fynbos biome, i.e., *Protea*, E*rica* and *Restio* (lineages belonging to the families Proteaceae, Ericaceae and Restionaceae, respectively). Within the CCR, the Fynbos biome is further characterised by predominantly four vegetation types *viz*. fynbos heathland, renosterveld, Afromontane forest (although currently in decline), Strandveld vegetation, which are mainly defined based not only on their floristic assemblages, but importantly on the substrate on which they occur i.e. they normally occur on soils derived from different substrates and thus have different moisture

retention characteristics (Bergh *et al*., 2014). For example, the fynbos vegetation mainly occurs on nutrient poor sand-derived soils poor in nitrogen and phosphorus and low moisture retention, whereas renosterveld vegetation mainly occurs on nutrient rich shale derived soils comprising a sharply distinct floral composition (Goldblatt & Manning, 2000).

The ECR on the other hand, is an understudied subregion harbouring mainly dwarf shrubby, mostly succulent shrublands (Mucina & Rutherford, 2006). The region spans *ca*. 98 869 km^2 encompassing eight ecogeographic regions (Snijman, 2013), which can be broadly equated to vegetation types at different levels. Briefly, the ECR ecoregions are namely, from North descending Southwards towards the CCR: Southern Namib, Gariep, Namaqualand Hardeveld, Namaqualand Sandveld, Kamiesberg Mountains, Knestervlakte, Western Mountain Karoo, Tanqua southern-succulent Karoo (see Snijman, 2013 for an overview). Similar to the vegetation types in the CCR, the ecoregions found in the ECR are characterised by specific ecology, geological and climatic regimes (Mucina & Rutherford, 2006). The ECR landscape contrasted to the CCR, mainly comprises coastal salt pans, rocky Mountain ranges, pebbly quartz derived patches, calcareous sandy soils; and climatically the region is comparably drier compared to the CCR, and it experiences a variety of rainfall regime (i.e., summer, winter and aseasonal). However, within the EC there are also elements of the Fynbos, Desert, and Azonal elements present (Rutherford & Westfall, 1986; Cowling *et al*. 1999); making the landscape quite variable and distinct when compared to the CCR.

The GCFR has long been an arena for botanists interested in understanding the drivers of the observed patterns of high species richness and endemism, which have generally been attributed to high rates of speciation with low rates of extinction (Linder, 2003). There are several hypotheses which have been emphasised by various authors, with the most popular being related to environmental factors (i.e., adaptive radiation) being a major driver accounting for high speciation rates (Goldblatt & Manning, 2002; Linder, 1985, 2003). Environmental factors suggested to drive speciation include topography, soil types, rainfall availability and seasonality, and pollinator specialisation (Rourke, 1972; Cowling, 1990; Kurzweil *et al*., 1991, Verboom *et al*., 2004; Goldblatt & Manning, 1996). For example, one of the most noteworthy observed patterns related to climate in the GCFR, is that there is a decline in levels of diversity and endemism when moving longitudinally from West to East across the region; a pattern which has been coined Levyns' Law (Cowling *et al*., 2018). Levyns' Law is primarily attributed to seasonality of rainfall in the GCFR, whereby areas experiencing predominantly

winter rainfall are associated with high levels of diversity and endemism, while areas experiencing aseasonal rainfall (i.e., those in the eastern parts of the GCFR towards Port Elizabeth/Gqeberha; Fig. 8) are associated with lower levels of diversity and endemism (Linder & Vlok, 1991; Cowling *et al*., 2018). Thus, the importance of rainfall seasonality has been emphasised in discussions related to the evolution of the GCFR, where earlier authors emphasised rainfall seasonality not just as an important factor which gave rise to the contemporary GCFR (i.e. the mid-Miocene onset of the Mediterranean-type climate in the GCFR, but see van Santen & Linder, 2020) but also as a key driver of speciation in the region (Linder & Vlok, 1991; Linder 2003, 2005; Linder & Hardy, 2004). For example, experimental studies (e.g., Latimer *et al*., 2009; Carlson *et al*., 2011) focussing on the white proteas (*Protea* sect. *Exsertae*) have demonstrated that rainfall seasonality is positively correlated with diversification of the taxa, where most diversity is found in winter rainfall areas compared to less diversity in areas receiving rainfall year-round. This study is but one of the few empirical ones demonstrating the correlation between diversification and rainfall seasonality and it is still unclear how rainfall seasonality can act as a driver to diversification. Apart from rainfall seasonality, other environmental factors such as edaphic heterogeneity, pollinator availability and specialisation, as well as topography have also been studied as potential drivers of diversification. For example, Verboom *et al*. (2015) demonstrated that high-elevation endemic Cape lineages belonging to the Asteraceae (*Stoebe* L., *Syncarpha* DC.), Proteaceae (*Leucadendron* R.Br. and *Protea* L.), Cyperaceae (*Tetraria* Beauv.), and Restionaceae (*Elegia* L. and *Thamnochortus* P.J.Bergius) exhibit phylogenetic niche conservatism positively linked to topography where their ranges are limited by elevation compared lineages occupying low-lying areas. Moreover, van der Niet & Johnson (2009) highlighted that apart from climate and topography, other environmental factors such as pollinator availability and specialisation may be a key driver of speciation in some Cape lineages. Thus, emphasising the important point that different environmental/ecological factors play a different role as drivers of speciation in the Cape lineages, including previously understudied factors such as pollination systems. The GCFR remains a fascinating biodiversity hotspot which allows for tackling unaddressed evolutionary and ecological questions.

1.3 DNA sequencing: from Sanger to Third Generation sequencing

The discipline of molecular systematics, like other areas of biology, has gained immensely from the development of and the advances made in high-throughput next generation sequencing (NGS) methods (Dapprich *et al*., 2016). Next generation sequencing methods

provide the ability to generate sequence data from thousands of loci across the genome, providing more genomic information content necessary for improving the effort to infer better resolved phylogenies. Although NGS methods have the capacity to generate whole genome sequence (WGS) data, there are associated downstream challenges mainly related to the complexity of the genomes of the taxa being studied, sequence data storage and bioinformatic processing strategies when dealing with whole genomes (Andermann *et al*., 2020). Therefore, studies on phylogenetics utilising NGS data have focused on generating sequence data from phylogenetically informative loci across the genome obtained through targeted sequencing approaches (Faircloth *et al*., 2014; Lemmon *et al*., 2012; Jones & Good, 2016). In addition, targeted sequencing is more cost effective (Davey *et al*., 2011) compared to WGS because for example; genomic complexity is much more reduced when only specific loci are targeted; read depth and coverage is substantially improved due to the ability to generate much more sequence reads for the targeted loci; the complexity of the bioinformatic processing challenges for assembling sequence reads from targeted loci are less/reduced compared to those of assembling whole genomes (McKain *et al*., 2018; Johnson *et al*., 2019). Moreover, targeted sequencing allows for including samples with fragmented DNA which would otherwise not be appropriate for WGS e.g., samples obtained from herbarium specimens, which are known to contain highly fragmented DNA over time.

Targeted capture sequencing (also referred to as target enrichment or bait capture) has since emerged a popular method frequently used in plant phylogenetic studies (e.g., de Sousa *et al*., 2014; Nicholls *et al*., 2015; Johnson *et al*., 2019; Esterman *et al*., 2021; Acha & Majure, 2022; Ferreira *et al*., 2022; Thureborn *et al*., 2022; Michel *et al*., 2022), particularly those which include samples obtained from herbarium specimens (Jones & Good, 2016; Johnson *et al*., 2019; Hale *et al*., 2020). The target capture method involves firstly deciding on whether to utilise universal (e.g., Angiosperm353; Johnson *et al*., 2019) or designing custom RNA-bait sets which are more specific to the study group. In general, studies focusing on relatively divergent taxa (e.g., family level) universal bait sets are normally utilised, whereas for studies on closely related species custom bait sets are more appropriate. There are several ways to generate custom bait sets either from existing transcriptome or whole genome sequence data (see Andermann *et al*., 2020 and Woudstra *et al*., 2022 for details). The RNA-baits are then hybridised with the relevant complementary indexed DNA library fragments followed by enrichment of the captured fragments using PCR (Andermann *et al*., 2020). The captured enriched libraries are then sequenced on the sequencing platform of choice depending on

factors such as the e.g., fragment length of the DNA libraries, number of samples for multiplexing and the specific preservation strategies for such samples, computational storage and capacity to handle data output for a particular sequencing technology.

1.4 Biogeography

Phylogeography (Avise *et al*., 1987) is a continuously growing field involving the use of genomic and geographical data to better understand the geographical and evolutionary histories of taxa. There are several models usually employed in phylogenetic biogeography where the landscape may either be considered as discrete unit e.g., hierarchical vicariance models (HVMs), island models, and reticulate models; or as a continuous unit e.g., diffusion models (Lemmon & Lemmon, 2008; Lemey *et al*., 2010; Pybus *et al*., 2012; Bouckaert, 2016). A shared characteristic of discrete models is that transitions of taxa between discontinuous units (i.e., discrete areas) is emphasised, whereas continuous models emphasise tracing the movement of taxa within a continuous area. Reticulate models are primarily based on the premise that landmasses undergo cycles of splitting and fusion; while with HVMs, landmasses undergo fragmentation through the emergence of dispersal barriers (e.g., geological, climatic, or biotic) separating the recently fragmented areas. Island models are quite similar to reticulate models in that landmasses undergo splitting, with the focus being on dispersal events between the different islands (Ronquist & Sanmartín, 2011). Diffusion models emphasise a random walk movement (i.e., dispersal) of taxa across a continuous landscape (Lemmon & Lemmon, 2008; Lemey *et al*., 2010; Ronquist & Sanmartín, 2011; Bouckaert, 2016). In principle, both the discrete and continuous biogeographic models can be implemented following Parsimony, Maximum Likelihood and Bayesian inference. Thus, my primary emphasis is on the differences between the biogeographic models and not on the inference methods.

Common to the discrete models is that distribution areas of taxa are defined prior to analysis. The definition of distribution areas may be according to specific criteria e.g., areas defined based on current distribution records which may be obtained from various databases (e.g., GBIF–https://www.gbif.org/; iNaturalist–https://www.inaturalist.org/; or herbarium specimen records). The use of distribution records to define distribution areas may include an analytical step where the distribution records are summarised to generate an estimated distribution range for taxa (e.g., Edler *et al*., 2017). Alternatively, areas may also be defined according to natural boundaries such as continents and islands. Defining areas for species endemic to islands seems to be straightforward compared to defining discrete areas for widespread taxa with overlapping

distributions on continents for example. Generally, there is no standardised specific criteria for defining species distribution area and in most cases, this is done implicitly based on the study group and the question at hand. The necessity of defining justifiable distribution areas for taxa has been a major criticism of discrete models, particularly for areas without any natural borders. In addition, another criticism to discrete models is that optimization of ancestral areas is based on a fixed tree or a distribution of trees which were inferred before the biogeographic analysis. This means that uncertainty in tree parameter estimation and phylogeographic model estimation are not jointly taken into account during inference (but see Lemey., *et al*., 2009). However, despite these criticisms, discrete models have remained popular and continue to be utilised in many historical biogeography studies. Several implementations of discrete models have been developed e.g., parsimony-based models such as dispersal vicariance analysis (DIVA; Ronquist, 1997) where inference is based on optimising ancestral area estimation by minimization on the overall costs of vicariant range changing events across a phylogenetic tree. Another example is the dispersal extinction cladogenesis model (DEC; Ree *et al*., 2005; Ree & Smith, 2008) where changes in geographical range resulting from occurrence of biological events (dispersal and extinction) are modelled along a branch (anagenesis) or at a node (cladogenesis) on a phylogenetic tree in continuous time. Popular software for historical biogeographical inference includes LAGRANGE (Ree, 2005), BayArea (Landis *et al*., 2013), RASP (Yu *et al*., 2015) and BioGeoBEARS (Matzke, 2013).

Continuous diffusion models have thus far received less attention compared to discrete models, particularly in historical biogeography analyses compared to epidemiological analyses where continuous models remain popular. Although there are ML implementations of diffusion models in general, here I focus on the Bayesian Inference (BI) implementation in a species-tree approach where the model is briefly defined as: given a tree T and root location $R_{location}$, what is the probability of the sampling locations $Sample_{LatDD–LonDD}$? This therefore means that continuous (and discrete) diffusion models are an expansion of the standard BI phylogenetic posterior equation with an addition of two parameters, the sample locations (L) and the precision rate matrix describing the relaxed random walk (RRW) process (Lemey *et al*., 2009, 2010). Thus, continuous diffusion models are based on the premise that migration or geographical movement of taxa can be modelled following a well understood diffusion process (i.e., Brownian Diffusion; BD). However, an implicit assumption of the standard form of BD, is that the model describes a random homogeneous movement/diffusion process which does not change over time. This implicit assumption is of-course not applicable when considering

that the migration of taxa is not homogeneous across space and time, meaning that this general implicit assumption of BD needs to be relaxed. Relaxing the assumptions of the standard BD gives rise to a relaxed random walk (RRW) model which maintains constant diffusion rates along single branch on the phylogeny while allowing diffusion rates to vary between branches (i.e., heterogeneous diffusion) across the tree (Dellicour *et al*., 2019). The theoretical framework behind RRW is similar to uncorrelated relaxed molecular clock models which have been applied to molecular sequence data (Drummond *et al*., 2006). Lemey *et al*. (2010) implemented a RRW model where they defined a precision matrix for diffusion rate which is rescaled according to a specific rate scalar for each branch across a tree. Earlier implementations of continuous diffusion models (e.g., Lemey *et al*., 2009) were primarily based on conditioning over a gene-tree i.e., there was no explicit distinction between gene-trees and species-trees. This presented major fundamental flaws when considering the importance of differentiating between gene-trees and species-trees (see section on Phylogenetics). Given this, recent implementations of diffusion models have been extended to infer ancestral areas following a species-tree approach (e.g., Nylinder *et al*., 2014). Despite the continuing developments however, continuous diffusion models have been less used in historical biogeography studies, with just a handful of examples for plants (e.g., Nylinder *et al*., 2014; Đurović *et al*., 2017; Johansson *et al*., 2018; Min Choo *et al*., 2020) and reptiles (Leaché *et al*., 2017). Ronquist & Sanmartín (2011) highlighted a major shortcoming of diffusion models is the tendency to produce oversaturated local diffusion patterns when analysis is at longer temporal and geographic scales. Additionally, another important consideration is that because diffusion models use sample locations as data, a non-random sampling will produce a biased effect on the results, a caveat that holds true for molecular sequence data also. However, the abovementioned criticisms have not been comprehensively tested with empirical data, and the performance/limitations of diffusion models in general remain largely unknown/untested (but see Kalkauskas *et al*., 2021). The models remain sophisticated in spatial inference as they allow for incorporation of different sources of spatial information (spatial heterogeneity, to be included in the analysis while also allowing accounting for uncertainty in parameter estimation. Additionally, the use of continuous diffusion models where sampling locations are represented as geographical positioning system coordinates (GPS coordinates latitude–longitude decimal degrees), means that the definition of arbitrary discretization of areas is no longer required prior to analysis, thus, reducing ambiguity and providing a more straightforward criteria for defining an occurrence location of a sample. In this thesis, I explore the utility of BI continuous diffusion

models to infer the historical biogeography of *Silene* at the infrageneric level (i.e., sections) (Manuscripts **I** and **II**).

1.5 Phylogenetics and species delimitation

Previously, there was little distinction made between gene-trees and species-trees, with the terms sometimes used interchangeably to refer to species relationships. Avise *et al*. (1987) emphasised the distinction between gene-trees and species-trees, and most importantly highlighted the discordance between gene and species-trees. This discordance between gene and species-trees is at least partly due to incomplete lineage sorting (ILS; Avise *et al*., 1987) which is sometimes referred to as deep coalescence (Maddison, 1997). However, there are other biological processes which are responsible for gene-tree species-tree discordance such as migration (e.g., horizontal or lateral gene transfer, hybridization and recombination) (Degnan & Rosenberg, 2009). In recent decades, there have been efforts focussed on computational development of phylogenetic inference models able to estimate species relationships while accounting for the most common cause of gene-tree species-tree discordance (i.e., ILS). One of the currently popular approaches to phylogenetic inference is the multispecies coalescent model (MSC). The MSC is an extension of the Wright-Fischer genetic drift model (WF) where gene genealogies are modelled for multiple populations constrained within a species tree, while also taking ancestral demographic parameters into account (Rannala *et al*., 2020). With the increase in capacity to produce large datasets from multiple loci, studies utilising MSC models have one of two approaches which they can infer species-trees: (a) Summary methods which involve a two-step approach where gene-trees are inferred first from gene alignments, then the resulting gene-trees are used as input for estimating species-trees (Mirarab & Warnow, 2015; Mirarab *et al*., 2016; Zhang *et al*., 2018); or (b) Co-estimation methods where gene alignments are used as input to simultaneously estimate gene and species-trees (Liu & Pearl, 2007; Heled & Drummond, 2010; Jones *et al*., 2015; Jones, 2017; Ogilvie *et al*., 2017; Douglas *et al*., 2022). Co-estimation methods have gained popularity due to their accuracy compared to summary methods, however for most studies, a balance between dataset size and computational efficiency determines which method to apply.

During the past decade, Bayesian Inference methods have been developed to delimit species under the MSC while using different kinds of sequence data e.g., SNPs (Leaché *et al*., 2014), or gene alignments (Jones *et al*., 2014; Jones, 2017). These Bayesian models are implemented in different ways in order to improve computational and inference robustness e.g., reversible

jump MCMC while searching tree space using nearest-neighbour interchange proposals (as implemented in BPP; Flouri *et al*. 2018), birth-death-collapse model where samples falling below a certain tree height parameter threshold (epsilon) are collapsed into a single species (as implemented in DISSECT/STACEY; Jones *et al*., 2014; Jones, 2017, and SPEEDEMON; Douglas *et al*., 2022). These methods have become popular in the past decade and have been implemented in several species delimitation studies (e.g. Toprak *et al*., 2016). With the ability to generate large amounts of high-throughput sequence data, it has become important that these methods are improved to be able to handle even larger datasets e.g. the SNAPPER for SNP data, and StarBeast3 (Douglas *et al*., 2022), SPEEDEMON (Douglas *et al*., 2022), are recent implementations derived from the above mentioned models, particularly aimed at handling larger datasets and computational speed and efficiency. These models provide an interesting approach to implement on inferring phylogenetic relationships and species limits for previously understudied groups such as the southern African *Silene* using high throughput sequence data.

2. Objectives

The overarching aim of this thesis is to advance the understanding of the evolutionary history and current diversity of southern African *Silene* using high throughput sequence data coupled with explicit Bayesian coalescent models. To achieve this, the thesis is made up of four chapters, each having specific questions they are addressing, which in turn feed into answering the two broader questions of this thesis:

1. Understanding and disentangling the phylogenetic relationships and historical biogeography of *Silene* in North and southern Africa (Focus of Manuscript **I** and **II**) we address the following question:
    (i) What is the phylogenetic position and biogeographic origins of the previously unexplored North and southern African *Silene*?

Manuscript I focusses on determining the phylogenetic position of the southern African *Silene* using three DNA loci (ITS, *rps16* and *matK*) to generate a genus-level phylogeny. Additionally, utilizing location/collection data to jointly co-estimate the phylogeny as well as the historical biogeography of the southern African *Silene*. Similarly, **Manuscript II** follows the same approach of determining the phylogenetic position of the North African members of section Silene using two DNA loci (ITS and *rps16*). Additionally, the study addresses previously ambiguous taxonomic nomenclature. Moreover, historical biogeographic inference to determine the biogeographic origins of North Africa section *Silene* is explored. Particularly, interest is in elucidating the geographical history between the North and southern African members of section *Silene*. Therefore, the two manuscripts share the same goals of addressing similar unaddressed questions on phylogeny and biogeography of *Silene* in Africa.

2. Exploring the utility of target capture sequence data in addressing species relationships and limits using coalescent-based methods (Focus of Manuscripts **III** and **IV**).

(i) Statistical advances in phylogenetic tree reconstruction models now allow us the ability to infer robust phylogenies and test different species delimitation hypotheses. Do currently available phylogenetic coalescent models together with availability/capacity to generate of high-throughput sequence data enable us to better infer robust phylogenies and estimate species limits?

(ii) What are the species relationships within southern African members of section *Elisanthe*?

(iii) Is there a consensus between phylogeny and taxonomic species classifications in section *Elisanthe*?

Manuscript III provides a brief overview of the advances made in the development of statistical models used in molecular phylogenetics. In particular, the manuscript highlights the importance and relevance of the multispecies coalescent model in modern phylogenetics, and how the model provides an opportunity to delimit species in a statistically defined and testable framework using empirical data. **Manuscript IV** then explores the utility of how the multispecies coalescent model can be used in a phylogenetic framework to infer species delimitation hypotheses and test the robustness existing taxonomic classifications on species limits. To achieve this, we focus on members of southern African section *Elisanthe*.

3. Summary of thesis chapters

3.1 Manuscript I

The most recent taxonomic revision of native members of southern African *Silene* species was done by Manning & Goldblatt (2012a), who recognized eight species. Since the revision by Manning & Goldblatt (2012a) there hasn't been a phylogenetic study on *Silene* which included all native southern African species, and their phylogenetic position has thus remained largely unknown. For example, the recent infrageneric circumscription study by Jafari *et al*. (2020), only included two out of eight species of the native southern African taxa. In this manuscript, the phylogenetic placement of the southern African members of *Silene* was investigated. We sampled all eight currently recognized southern African *Silene* species from across their known distribution. We generated sequence data for three genetic loci (ITS, *rps16*, *matK*), which were then combined with existing complementary sequence data for the different *Silene* species. To ensure that our sampling included representation of all currently recognized taxa in the genus, we added our newly generated sequence data for the southern African taxa to the complementary multiple sequence alignments (MSA) of Jafari *et al*. (2020) which included a comprehensive sampling of all the currently recognized infrageneric groups in the genus (i.e., subgenera and subsections). To infer the phylogenetic relationships within the genus, we generated a species-tree using StarBEAST2 (Ogilvie *et al*., 2017) as implemented in BEAST2 (Bouckaert *et al*., 2019).

Results from our phylogenetic analyses placed the southern African taxa in two distantly related clades (Fig. 1 in **Manuscript I**). Our results supported the monophyly of section *Elisanthe* where all four (*S. undulata*, *S. rigens*, *S. ornata* and *S. saldanhensis*) southern African taxa, including two Eurasian relatives formed part of a strongly supported clade (PP=1; Fig. 1 in **Manuscript I**). These results were congruent with those of Jafari *et al*. (2020), who despite their limited sampling recovered a sister relationship between *S. undulata*, *S. noctiflora* and *S. turkestanica*. The second clade we identified corresponding to the broadly circumscribed section *Silene* by Jafari *et al*. (2020), which included a sub-Saharan clade. The identified sub-Saharan clade comprised the remaining four southern African taxa (i.e., *S. burchellii*, *S. mundiana*, *S. aethiopica* and *S. crassifolia*) as well as the west-central African endemic *S. biafrae*, including other accessions of *S. burchellii* collected from several parts of East Africa (Kenya, Tanzania and Ethiopia). Within the sub-Saharan clade, we identified a sister relationship between a clade comprising southern African members (PP=1) and the west-

central African endemic *S. biafrae* together with accessions of *S. burchellii* collected outside southern Africa (PP=1). Although the sequence data showed little variation within the sub-Saharan clade, the sequence variation was still sufficient to highlight a moderately supported geographical structure where the southern African taxa are distinct from the Central and East African taxa.

In addition to inferring the phylogenetic position of southern African *Silene*, we also investigated the biogeographic origins of the southern African taxa. To estimate the ancestral areas of the southern African Silene, we utilized the GEO_SPHERE package (Bouckaert, 2016) as implemented in BEAST2 (Bouckaert *et al*., 2019). The model implemented in the GEO_SPHERE package is a continuous diffusion model where ancestral areas are estimated according to a diffusion process following a relaxed random walk (RRW) along the branches of the phylogeny (Lemey *et al*., 2010; Bouckaert, 2016). Diffusion models have an advantage over other popular models used in historical biogeography in that they firstly allow for the use of point locations to represent collection sites for each sample, thus negating the need for a priori defining distribution areas for taxa. Secondly, while it is common with other popular methods to infer a species-tree separately followed by ancestral area estimation conditioning on a single tree obtained from the species-tree analysis; diffusion models allow for joint species-tree and ancestral area estimation during inference, which means that uncertainty is taken into account during estimation. Our analysis entailed coding geographical information for each sample as point coordinates i.e., in the form of GPS coordinates represented in decimal degrees. Species-tree inference was carried concurrently with the ancestral area estimation where three data partitions were included in the analysis: Two partitions comprised sequence data from nuclear (ITS) and plastid (*rps16* and *matK*), while geographical partition comprised point locality coordinates. This joint-inference setup meant that geography was modelled as a trait on the species-tree, and is not implemented to influence the species-tree reconstruction (i.e., topology, branch lengths etc.) but is expected to evolve/diffuse along the tree following a RRW.

Results from our ancestral area estimation demonstrate that the two southern African groups colonized the region in different ways. Southern African members of section *Silene* colonized the region from North Africa via the East-Central Africa Rift Valley onto the southwestern parts of southern Africa during the Pleistocene (1.28 Ma; 95% HPD: 0.8–1.79 Ma; Fig. 2b in **Manuscript I**). While members of section *Elisanthe* colonized southern Africa via long

distance dispersal from Eurasia during the Pleistocene (1.37 Ma; 95% HPD: 0.86–1.83; Fig 2c in **Manuscript I**).

3.2 Manuscript II

The aim of this study was to determine the phylogenetic position and redress the nomenclatural status of the previously understudied North African *Silene*. Despite the extensive global and regional taxonomic generic treatments, the North African taxa remain relatively poorly explored, with the two most comprehensive taxonomic classifications done by Rohrbach (1867) followed by Maire (1963). There are *ca*. 144 species recognised in North Africa, with the bulk (*ca*. 56) being narrow endemics to Morocco, Algeria, Tunisia, or Libya (Rahou & Amssa, 2003; Dobignard & Chatelain, 2011). Section *Silene* is most diverse and well-represented in North Africa, with the majority of the understudied species tentatively placed in the section based on earlier taxonomic classification by Maire (1963). However, most of the native North African species have not been included in phylogenetic studies on the genus (e.g., Oxelman & Lidén, 1995; Desfeux & Lejeune, 1996; Oxelman *et al*., 1997; Oxelman *et al*., 2001; Popp & Oxelman, 2001, Popp *et al*., 2004; Popp *et al*., 2008; Erixon & Oxelman, 2008; Frajman *et al*., 2009b; Greenberg & Donoghue, 2011; Ghahremaninejad *et al*., 2014; Jafari *et al*., 2020). Elucidating the phylogenetic position and infrageneric relationships of North African section *Silene* is therefore most important, particularly because the region is the centre of diversity for the section. We collected samples from 124 native North African taxa and generated sequence data for ITS (124 accessions) and *rps16* (105 accessions). Our newly generated sequences were combined with complementary *Silene* sequences for the same loci obtained from GenBank. We generated gene-trees for both loci, followed by species-tree inference under the multispecies coalescent model performed on a subset of the dataset initially used for gene-trees.

Our results demonstrate that North African taxa belong to a strongly supported and widely circumscribed *S.* sect. *Silene* clade. In addition, we have demonstrated that other North African taxa belong to the sections *Siphonomorpha* Otth., *Muscipula* Oxelman, Jafari & Gholipour, *Behenantha* (Otth) Torr. & Gray, *Sedoides* Oxelman & Greuter, *Coinomorpha* Otth. Moreover, our results corroborated the recently suggested expanded phylogenetic circumscription of sect. *Silene* (Jafari *et al*., 2020) which includes North African representatives previously assigned to sections *Atocion*, *Dipterospermae*, *Fruticulosae*, *Nicaeenses*, *Scorpioides*, and *Succulentae,* thus rejecting the previous taxonomic assignment by Maire (1963). Our results demonstrate that none of the abovementioned groups previously recognised by Maire (1963) are supported as monophyletic at the sub-sectional level and can therefore not be recognised. For example, the previously recognised section *Fruticulosae*, whose type is *S. ciliata*, is recovered as not

monophyletic forming a clade several annual species *viz. S. micropetala* Lag. (= *S. cisplatensis* Cambess, see Jafari *et al.* 2020), *S. scabriflora* Brot., *and S. tuberculata* (Ball) Talavera (clade II; Fig. 1; Manuscript II). Congruent with previous studies (e.g., Kyrkou *et al*., 2015), *S. ciliata* comprises two clades which are geographically structured with the Iberian clade including *S. legionensis* Leg. nested within; given that the type of *S. ciliata* is from the Pyrenees, our results support the recognition of *S. ciliata s.s*. as confined to the Iberian Peninsula. The eastern Balkan-Italian clade with sequences labelled *S. ciliata* is clearly distinct from the Iberian clade and should thus, in agreement with Küpfer (1974), be recognized as *S. graefferi* Guss. Our results therefore highlight that the perennial habit is a poor diagnostic trait for the section *Fruticulosae*, particularly because the results demonstrate that the trait undergone multiple switches between annual and perennial. Furthermore, all North African species previously assigned to section *Fruticulosae* have now been shown by our results to phylogenetically belong elsewhere (Fig. 1; Manuscript II).

Another example is that of *S. nocturna*, which forms two strongly supported (PP=0.95–0.98) separate clades which are morphologically difficult to distinguish (clades XX and XIII; Fig. 1; Manuscript II), requiring further study. Among the *S. nocturna* clades, we recovered accessions labelled *S. nocturna* which were strongly supported (PP=1.0) recovered in clade XX, which is diagnosed by fewer flowers in the inflorescence, shorter calyces, and seeds with a narrow dorsal furrow. Additionally, compared to the *S. nocturna s*.s.in clade XIV, the accessions in clade XX have petiolate lower stem leaves. We consider the plants in clade XIV to conform well to the concept of *S. nocturna* of Sáez *et al*. (2022) and highlight that further sampling is necessary in order to better diagnose clade XX which we here determine as requiring further investigation. Similar is the case observed in *S. pomeli* (clades XV and XVII), where several infraspecific names have been applied to the taxa invariably. Although the infrageneric relationships are not fully resolved, the specimen labelled *S. pomeli* in clade XIV matches the Moroccan *S. pomeli* subsp. *adusta*, whereas the other two in clade XVII are consistent with the nominal subspecies which has a more eastern distribution. Our results thus highlight that the "*adusta*" taxon probably deserves recognition at the species level. Other identified clades include the strongly supported (PP=1) clade XXI which taxonomically corresponding to subsect. *Rubellae* (Batt.) Chowdhuri treated by Oxelman (1991) as the *S. diversifolia* Otth., group. Within the *S. diversifolia* comprising specimens from a large part of the Mediterranean Basin, we identified a second clade (PP=0.99) grouping the Western Iberian species *S. bergiana* Lindm., together with the Northern Moroccan taxon *S. volubilitana* Braun-Blanq., &

Maire. Furthermore, withing clade XXI, we identified a third clade (PP=0.96) corresponding to *S. turbinata* Guss., known from Southern Italy and the Numidian part of the Maghreb. Our results highlight that the previously named *S. rubella* subsp. *segetalis* (Dufour) from the North African plants (Dobignard & Chatelain, 2011) should be rejected and with priority given to *S. diversifolia* and *S. turbinata.*

3.3 Manuscript III

Plant systematics is a field of study concerned with understanding the evolutionary history of biodiversity through utilisation of taxonomic and phylogenetic framework. Taxonomy is focussed on identifying, describing, classifying, and naming of organisms. On the other hand, phylogenetics is more focussed on studying the evolutionary history and relationships of organisms. Both taxonomy and phylogenetics are an integral part of systematics. The focus of this manuscript is to provide an overview of the current state of the field of systematics. Particularly the emphasis is on (i) how the ability to generate large amounts of high-throughput sequence data has benefited phylogenetics, and (ii) how advances in computational capacity and development of statistical models are revolutionising modern phylogenetics. Additionally, the discussion focussed on how taxonomy and modern phylogenetics can both be mutually beneficial amid such advancements in sequencing capacity and development of explicit statistical models. In phylogenetics there are several kinds of data sources which have historically been utilised e.g., anatomical, chemical, cytology, embryology, palynology, geography, and genetic data sources. Molecular (genetic) data has seen considerable popularity in phylogenetic studies because of the ongoing developments in high-throughput next generation sequencing technologies, which allow the generation of sequence data at larger scales. There are different types of genomic data that can be produced using various NGS methods such as whole genome sequencing (WGS), transcriptomics, proteomics, genotype by sequencing, and target capture. In principle, genomic data produced by any of the abovementioned NGS methods (i.e., whole genome sequencing or sequencing a subset of loci from the genome) can readily be coupled with explicit statistical models to infer phylogenies. In this manuscript we focus on methods which target a subset of loci (a summary is provided in section 1.3) from the genome because they allow effective studies on sequence variability even at low levels (e.g., species or populations), which sets the context for this thesis.

In parallel with advances in sequencing technologies, there have been developments in computational capacity and statistical methods available to study the Tree of Life. There are several phylogenetic approaches to studying the Tree of Life using sequence data, which can be grouped into three categories *viz*. distance, optimality and Bayesian approaches. Algorithmic phylogenetic approaches e.g., Neighbour-joining (NJ; Saitou & Nei, 1987), unweighted pair group method with arithmetic mean (UPGMA; Sokal & Michener, 1958) employ a stepwise clustering algorithm given a dataset to build a phylogenetic tree. Optimality phylogenetic approaches e.g., minimum evolution (Kidd & Sgaramella-Zonta, 1971; Rzhetsky

& Nei, 1993), maximum parsimony (MP; Farris, 1970; Fitch, 1971), and maximum likelihood (ML; Felsenstein, 2004), on the other hand use predefined criteria to estimate the posterior probability distribution of possible trees given the data through identifying trees with the highest posterior probability from a landscape of possible trees given model parameters (Wiley & Lieberman, 2011). In Bayesian approaches, it is not necessary to compute posterior probabilities for all possible trees from the full landscape, rather it is possible to sample parts of the full tree space using the Markov Chain Monte Carlo (MCMC) tree-searching process.

Developments in Bayesian phylogenetic approaches have advanced to allow for estimating phylogenetic trees from multiple loci while taking into consideration that the gene genealogies do not necessarily share the same histories. This is termed gene-tree discordance, which is a result of processes such as incomplete lineage sorting (ILS) and migration (Degnan & Rosenberg, 2009; Edwards, 2009; Heinrich *et al*., 2009). To account for gene-tree discordance caused by ILS, Bayesian phylogenetic approaches have been developed under the multispecies coalescent model which is an extension of the Kingman coalescent model (Kingman, 1982). The MSC is an explicit statistical model which defines "species" as the branches on the species-tree and models gene genealogies for multiple populations constrained within a species-tree under a set of Wright-Fisher assumptions (Fisher, 1930; Wright, 1931). The Wright-Fisher assumptions include random mating, no population structure and instantaneous speciation with no gene flow between species; however, these assumptions do not always hold in empirical cases and thus resulting in violations to the model. Nonetheless, development of the MSC model has advanced modern phylogenetics by allowing authors to generate phylogenies using a statistically explicit model while accounting for the ubiquitous cause of gene-tree discordance. There are two classes of approaches to inferring species trees under the MSC model *viz*. summary and co-estimation approaches. Summary approaches take unrooted gene trees generated using other methods as input for species tree estimation (Liu *et al*., 2010; Mirarab *et al*., 2014, 2015). Co-estimation methods take sequence alignments from different loci as input and jointly estimates the gene-trees and species-tree (Liu & Pearl 2007; Heled & Drummond 2010; Ogilvie *et al*., 2017). Although the MSC model is able to handle ILS, the model is only able to handle low levels of migration (i.e., the other cause of gene -tree discordance); where migration can either be modelled as continuous (i.e., isolation with migration; Hey & Nielsen, 2004) or discrete (i.e., multispecies network coalescent models; Wen *et al*., 2018; Flouri *et al*., 2020; Rannala *et al*., 2020).

The MSC model is a robust statistical framework which has benefited modern phylogenetics with the ability to estimate species trees and objectively test various hypotheses on species limits. The MSC model can be used in combination with criteria such as monophyly (Hennig, 1950) to discover natural phylogenetic groups (i.e., monophyletic groups) which are supported by other aspects such as morphological diagnosability, taxonomic conservatism, ecology, geography, and physiology in an integrative taxonomy approach (Dayrat, 2005). Such an approach integrates phylogenetics and taxonomy in a sense that information from other sources is included during the estimation of species trees and the testing of species boundaries. It is important to note that when following such an approach, the recognition of monophyletic groups should always precede the formulation of taxonomic rankings. Although, the robustness and clarity of the MSC is that it offers an explicit definition that the species in such a model are the branches in the tree, it is important to note that violation of the assumptions will always lead to inconsistencies in the recognition of these "branches/species". When there are violations to the MSC assumptions, the addition of more data will inflate the number of species estimated (Leaché *et al*., 2019). On the other hand, for recently diverged groups with lack of sufficient genetic variation, the number of species may be underestimated even if the assumptions are met. Nonetheless, the MSC model coupled with the current state of DNA sequencing will continue to revolutionise modern phylogenetics and enable it to focus on previously unexplored avenues of previously unexplored hypotheses about species relationships and boundaries.

3.4. Manuscript IV:

The focus of this study is to infer the species-level phylogeny of southern African members of section *Elisanthe* (Caryophyllaceae) and to test the phylogenetic support for the current taxonomic classification. Although the genus *Silene* (Caryophyllaceae) is mostly represented and diverse around the Mediterranean Basin, there are several species distributed across the northern (further inland) and southern parts of the African continent. In Africa, *S.* section *Elisanthe* is represented in southern Africa by five perennial herbaceous taxa *viz*. *S. saldanhensis*, *S. ornata*, *S. rigens*, *S. undulata* subsp. *undulata* and *S. undulata* subsp. *polyantha* (Manning & Goldblatt, 2012a). *Silene saldanhensis*, *S. ornata* and *S. rigens* are endemic to the Core Cape Subregion in the Cape, while *S. undulata* subsp. *polyantha* is restricted to KwaZulu Natal–Swaziland region and *S. undulata* subsp. *undulata* is widely distributed across southern Africa. Although the southern African members of section *Elisanthe* have taxonomically been revised (Manning & Goldblatt, 2012a) and their phylogenetic position recently determined (Jafari *et al*., 2020; Moiloa *et al*., 2021), the species relationships within the section remain unclear.

We collected tissue from 96 accessions from the field and herbarium specimens representing the currently recognised species across southern Africa. From the 96 accessions, 77, 12, 3, 4 accessions were determined (according to the Manning & Goldblatt, (2012a) classification) as representative of *S. undulata*, *S. rigens*, *S. ornata* and *S. saldanhensis*, respectively. Deoxiribose nucleic acid (DNA) was extracted from plant tissue using the Qiagen Nucleospin® Plant II extraction kit (Macherey-Nagel, Düren, Germany) or a modified cetyltrimethylammonium bromide (CTAB) for field collected and herbarium sampled specimens, respectively. We prepared for Illumina sequencing using the NEBNext® UltraTM II FS DNA Library Prep Kit for Illumina® (New England Biolabs, Ipswich, MA, USA), which were indexed with NEBNext® Multiplex Oligos dual-index sets 1–4. Libraries were divided into three sequencing batches based on the preservation method (field collected versus herbarium specimens) and collection year. Our subdivision of libraries meant that field collected samples (with expectedly less fragmented DNA) were sequenced on the Illumina MiSeq instrument with a read length of 300 bp paired-end; while libraries of samples from collected from herbarium specimens (with expectedly more fragmented DNA) were sequenced using the Illumina NovaSeq instrument with a read length of 150 bp and 250 bp paired-end for herbarium material. We multiplexed the libraries in 12-plex, followed by target capture

hybridization performed custom taxon-specific bait sets targeting 48 low copy nuclear genes following the MYBaits protocol v5.0 (Arbor Biosciences, Ann Arbor, MI, United States).

The resulting raw FASTQ sequence data were assembled using custom scripts described in Cangren *et al*. (in prep). The assembly process involved performing two rounds of trimming and quality filtering using Trimmomatic v0.39 (Bolger *et al*., 2014). This was followed by generation of contigs which were BLASTED against probe sequences where contigs matching probes were used as internal references for the *de novo* assembly. The resulting *de novo* assembled references were then subjected to mapping of reads for an iterative-mapping based reference assembly in order to extend the initial *de novo* derived reference using SPAdes v3.15.5 (Bankevich *et al*., 2012). Reads were then mapped onto the resulting reference. In addition, we downloaded and included reference sequences for ITS (MT036578) and *rps16* (MT062329) in order to capture off target reads corresponding to those loci. Reads were then mapped against the generated reference, followed by allele phasing using WhatsHap v1.8 (Patterson *et al*., 2015; Martin *et al*., 2023). Multiple sequence alignment (MSA) for the phased sequence data was generated using MAFFT v7.467 (Katoh & Standley, 2013). The alignments were visualised in AliView v1.28 (Larsson, 2014). Alignments which had less than 50% of the taxa represented in the matrices were excluded from all downstream analyses. This resulted in 28 alignments for the phylogenetic analyses.

We carried out phylogenetic reconstruction in a Bayesian species-tree inference framework under the multispecies coalescent model using co-estimation (STACEY v1.3.0.1; Jones, 2017) and summary methods (ASTRAL III v5.7.8; Zhang *et al*., 2018). Input gene-trees for ASTRAL III were generated using IQ-TREE v2.0.3 (Minh *et al*., 2020), with the GTR +G +I specified as the substitution model and the analysis run for 1,000 ultrafast bootstrap replicates (Hoang *et al*., 2018). The resulting gene-trees were used as input to infer the species-tree using ASTRAL III (Zhang *et al*., 2018) under default settings with node support evaluated using local posterior probabilities (LPP; Zhang *et al*., 2018). For the STACEY analysis, two independent analyses were run for one billion generation sampling every 10,000 generations. We generated a maximum clade credibility (MCC) tree containing posterior probability support values using TreeAnnotator v2.7.3 (Bouckaert *et al*., 2019).

Our analysis strongly supports the hypothesis that *S. ornata* and *S. saldanhensis*, the Saldanha Bay endemics are monophyletic and sister species. The two species share morphological similarities in characters such as lanceolate leaves, stem vestiture (i.e., presence of long

glandular hairs mixed with sparse short eglandular hairs). However, the species occupy different ecological niches where S. *ornata* occupies rocky limestone outcrops while *S. saldanhensis* occupies deep sandy calcareous dunes. Additionally, there are distinct morphological traits which set the species apart e.g., *S. ornata* is diagnosed by a sprawling habit, carmine coloured fully bifid flower petals, and tuberculate seed testa; while *S. saldanhensis* is diagnosed by a more compact habit, mauve coloured partially bifid (approximately half) flower petals and colliculate seed testa (Manning & Goldblatt, 2012a). Our results therefore support the taxonomic recognition (Manning & Goldblattt, 2012a) of *S. ornata* and *S. saldanhensis* as different species. In addition, the distinct differences in floral colour and occupation of different ecological niches suggest that pollinator specialization may also be an important ecological factor driving divergence between the two species. Ecological speciation (*sensu* Schluter, 2001) is a pattern that has been reported by several studies on Cape flora with different emphasis on e.g., pollinator specialization (van der Niet & Johnson, 2009) and edaphic heterogeneity (Goldblatt, 1979; Ellis *et al*., 2006; Verboom *et al*., 2009).

Another West Coast endemic, *S. rigens*, was strongly-supported as monophyletic and sister to the *S. ornata–S. saldanhensis* alliance. *S. rigens* is morphologically different from the sister species in being characterised by tall erect stems which are covered exclusively by glandular hairs (while sister species have a mix of both glandular and eglandular), small flowers with a calyx having a distinctively short and carpophore, and echinate seeds. With a strong relationship between the West Coast taxa, our results strongly suggest that phylogenetic structure may be segregated by geographic identity. This pattern has also been observed in the *S. saxifraga* group distributed in the Balkan Peninsula (Đurović *et al*., 2017). *Silene rigens* occurs on sandy calcareous substrata, similar to *S. saldanhensis*, but with a wider distribution along the West Coast. Our results further highlight the importance of calcareous substrata and the role played on diversification of sect. *Elisanthe* in the Cape, particularly given that the sister pair (i.e., *S. rigens* and the *S. ornata–S. saldanhensis* alliance) are restricted to such substrata.

Results from the two inference methods both recovered *S. undulata* as paraphyletic. We observe that the widespread *S. undulata* is a highly diverse taxon with some geographical signal observed in the strongly-supported internal clades in our phylogeny, but with low levels of support for most clades. Our analyses do not support the intraspecific (i.e., recognition of *S. undulata* susbsp. *undulata* and *S. undulata* subsp. *polyantha*) classification of *S. undulata* but rather highlight that the taxon is diverse with some level of differentiation highlighted by the

geographically coherent strongly supported clades recovered. Interestingly, although the clade was not supported, accessions which formed a clade with *S. undulata* subsp. *polyantha* were all collected from the Natal and Lesotho region. Other geographically coherent clades include the strongly-supported Cape Point clade and the Agulhas–Hermanus clade, both of which resemble morphologically distinct coastal forms which are different from the common inland forms. In their taxonomic revision, Manning & Goldblatt (2012a) indicated the morphological diversity exhibited by coastal forms, and our results where the Cape coastal forms are strongly-supported as distinct further highlight this. Further investigation focussing on *S. undulata* is required to better understand the intraspecific patterns observed from our phylogenetic results as well as the morphological patterns highlighted by Manning & Goldblatt (2012a).

Our results demonstrate the utility of target capture for studying and resolving phylogenetic relationships even at lower (i.e., species) levels. Additionally, the results presented in our study provide an example of how combining taxonomy and phylogeny can contribute to identifying systematics questions requiring deeper scrutiny in the pursuit of understanding the evolutionary history and diversification of section *Elisathe* in southern Africa.

4. Discussion and Conclusion

This thesis addressed four questions which are discussed in relation to the results below:

(i) What is the phylogenetic position and biogeographic origins of the previously unexplored North and southern African *Silene*?

The results from manuscripts I and II provide new insights and understanding into the diversity and phylogenetic positions of North and southern African taxa. Making use of genomic loci (ITS, *rps16*, and *matK*) to complement the existing sequences for *Silene* at genus level, the phylogenetic placement of African taxa was elucidated. For the southern African taxa, manuscript I provided clarity to the proposed taxonomic classification by Manning & Goldblatt (2012a), who proposed that Silene in southern Africa was represented by three groups. The results however, demonstrated Silene in southern Africa was represented by two distantly related groups (sections *Elisanthe* and *Silene*). To a broader extent, the results highlighted that most commonly used morphological traits in infrageneric classification is impacted by the homoplasious nature of the traits which has resulted in somewhat flawed groupings. This is clearly demonstrated in the case of southern African taxa where life cycle strategy (annual vs perennial) has been emphasised in segregating between groups where for example *S. aethiopica* was previously placed in *S.* sect *Dipterospermae* based on the taxon being an annual. The results from manuscript I however, indicate that the use of life cycle strategy as a character to segregate between infrageneric classification is not reliable as there has been multiple switches between annual and perennial. Additionally, results for the ancestral area estimation for southern African taxa reflected the phylogenetic results. For example, the radiation of *S.* sect. *Silene* (which has a somewhat connected distribution in Africa) depicted a biogeographical history pattern which explains the current distribution pattern in a sense that the inferred ancestral area of section *Silene* was estimated to be around the Mediterranean Basin; where a radiation directed from the Mediterranean Basin towards the South to colonise Central-East Africa and finally southern Africa. This pattern is reflective of the current distribution of *Silene* in Africa, where the section is most represented and diverse in North Africa, with a few species occurring in Central-East Africa and several species occurring in southern Africa. Another striking pattern is the disjunct distribution of section *Elisanthe* where two species occur in Eurasia and the rest of the species are distributed in southern Africa. This pattern was also observed in the results where the inferred ancestral area for section *Elisanthe* was found to be in the eastern parts of Eurasia where long distance dispersal southwards to colonise southern

Africa was observed. An intriguing aspect which could not be addressed in this thesis is determining the key drivers of the current distribution patters of *Silene* in southern Africa where most species are narrowly distributed to a single Biome (or in most cases vegetation type), while a few species have a wide distribution extending across multiple Biomes. Biogeographic results for the North African taxa did not provide a clear picture of the ancestral areas and radiation of section *Silen*e in North Africa. The reason for the unclear biogeographic results is thought to be due to the sampling strategy used for the North African taxa, where most widespread taxa are not unique to North Africa and are represented elsewhere in Eurasia, compounded by narrowly distributed North African endemics (some known from single localities). These reasons affect the performance of the diffusion model in that biased sampling will always produce distorted (and sometimes also oversaturated) diffusion estimates.

(ii) Do currently available phylogenetic coalescent models together with availability/capacity to generate of high-throughput sequence data enable us to better infer robust phylogenies and estimate species limits?

(iii) What are the species relationships within southern African members of section *Elisanthe*?

(iv) Is there a consensus between phylogeny and taxonomic species classifications in section *Elisanthe*?

Questions (ii), (iii), and (iv) were all addressed in manuscript IV, with some theoretical and philosophical aspects discussed in manuscript III. The results demonstrate clearly that target capture is a robust method having ample utility in generating sequence data sufficiently informative to disentangle relationships at lower taxonomic levels. The design of custom baits is an important step which can be quite challenging and sometimes expensive, depending on the available sequence information for the group being studied. However, the target capture method provides good utility for phylogenetics studies as it is more efficient than whole genome sequencing because in most cases it is not necessary to sequence the entire genome in order to get a resolved phylogenetic tree. Phylogenetic models such as the MSC provide a robust framework where the produced multilocus sequence data which are expected to have different genealogical histories can be analysed in a statistically rigorous manner. Using target capture sequence data (28 loci) and coalescent-based Bayesian inference (STACEY and ASTRAL), the species relationships in section *Elisanthe* were elucidated. In addition, the recovered species relationships demonstrated that the diversification of section *Elisanthe* in

southern Africa is closely linked with geographical structure. This is consistent with general hypotheses (e.g., ecological speciation; Verboom *et al*., 2009; van der Niet & Johnson, 2009) aimed at explaining the drivers of speciation in the Cape flora. Perhaps the most important findings from manuscript IV is that phylogenetic results supported taxonomic classification and also indicated other geographical and genetic structure which were not captured in the taxonomy of the section. For example, *S. undulata* is widespread in southern Africa occurring across the entire subcontinent. And from our phylogenetic results, the monophyly of this taxon could not be determined or rejected due to moderate to low posterior probabilities, particularly for clades which were recovered outside the core *S. undulata* clade (e.g., samples collected from Vredendal which were recovered at the root position in the STACEY tree). When considering species in the MSC sense, this result may be interpreted as an example of how lack of variability in genomic loci used coupled with violations of the assumptions of the MSC model may produce inaccurate phylogenetic estimates. Although, not performed in this study, it would be interesting to be able to test how empirical data fit the MSC model (see Reid *et al*., 2013). Such a test would allow for better interpretation of results such as those obtained for the *S. undulata* clade (Manuscript IV Fig 5A, B). On the other hand, it was interesting to recover strongly–supported sister relationships for the monophyletic West Coast endemics, which, given the strong phylogenetic and geographical signal is perhaps an indication minimal genetic migration among the West Coast endemics even though they represent a recent radiation.

5. Future prospects

There are still a lot of open questions which we have only just scratched the surface on. I will discuss a few of the areas where I foresee myself focusing on contributing the most.

First and foremost, my intention is to implement the skills I have learnt to other understudied groups in the Greater Cape Floristic Region. There is currently little known about the phylogenetic relationships, drivers of distribution patterns and diversification of the genus *Aspalathus* in the Greater Cape Floristic Region. Apart from the much known about the phytochemical properties of rooibos (*Aspalathus linearis*), not much attention has been given to understanding the systematics and evolutionary history of the genus. To date, the only existing phylogenetic studies on the genus are that of Boatwright *et al*. (2008) and Edwards & Hawkins (2007), and the sampling from these studies were not at all close to being comprehensive where about half of the taxonomic species were sampled and sequencing was performed for a handful of loci. The results from these studies produced unresolved trees because the sequenced markers were highly conserved. Therefore, my goal is to undertake a postdoctoral fellowship in the Cape where I will study *Aspalathus* using the techniques I learnt during my PhD. In that sense, I will carry out a phylogenomic study on *Aspalathus* where my first project will be to generate target capture sequence data which will be used to infer the phylogeny of *Aspalathus*. With a resolved phylogeny at hand, we will have the opportunity to revise the genus altogether because the most recent revision was by Dahlgren (1988). In addition, the sequence data produced will allow us to test Dahlgren's (1988) infrageneric classifications. Given that *Aspalathus* shows great potential to being an economically utilized genus, it is important that studies on the species limits are conducted to form a solid starting point for further studies such as the phytochemical exploration of suite of compounds present across the genus. With such unstable taxonomic classifications regarding species boundaries at the moment, most phytochemical studies are hindered by the inability to formulate studies testing the utility or suitability of other *Aspalathus* species potentially also having the similar phytochemical properties which may also be used commercially. Therefore, providing a new circumscription using HTS data coupled with explicit models like the MSC to delimit species will provide a substantial line of evidence to re-circumscribe the genus and provide end users of taxonomy with robust classifications.

The Greater Cape Floristic Region is a fairly well studied biodiversity hot-spot with high levels of species richness and endemism (Cowling & Pressey, 2001; Goldblatt & Manning, 2002;

Linder, 2003). As a botanist it is perhaps one of the greatest areas to study, however, most local researchers studying the flora in the region are constrained (whether due to funding or lack of expertise in new techniques) to stick to methods such as Sanger sequencing for most phylogenetic studies. This limits the ability to the kinds of questions that can be addressed given the Sanger sequence data they would have at hand. Therefore, as a researcher with focus in phylogenomics, I see myself as a bridge to the skills gap when it comes to elucidating the evolutionary histories of the Cape legumes e.g., *Lebeckia*, *Rafnia*, *Wiborgiella*, *Psoralea*, *Otholobium* (maybe extended to other non-legume groups as well, depending on the level of cross-collaborations I will have the opportunity to carry out). The lack of funding may still be a hindrance to conducting phylogenomic studies at the level, which is in tune with the international arena, but I believe that given the networks I have developed through my training as a PhD student, I am better equipped to foster fruitful collaborations to bridge the gap. It is however noteworthy to highlight that considerations regarding computational capacity (storage and bioinformatics) are of course important bottlenecks which may affect projects; however, the computational problems are a higher-level challenge to all conducting genomics studies in general.

Species delimitation remains a topic intensely discussed in literature, particularly, the development of software which can handle large datasets. For example, the current popular species delimitation programs (e.g., STACEY, BP&P, SNAPP, SNAPPER) are only able to handle a certain amount of data before being rendered computationally intensive and sometimes unfeasible to run in reasonable time. Therefore, my other interest as I develop my bioinformatics skills further post my PhD, I intend to actively participate in the development of such software. My interest is to delve into testing and comparing the consistency of these different programs using the datasets I generated during my PhD. Given that native southern African *Silene* comprise 8 recognized species, I have had the opportunity to test the robustness of the different species delimitation programs. The programs implement different models and can sometimes not be directly compared due to the different assumptions behind the models. However, I believe that it is precisely why my approach would be important because it would assist the developers of such models with the much-needed feedback regarding the utility of their programs. At a theoretical level, I believe that the conversation and quest for finding an answer to the species problem must be continued, and it is a topic which I find fundamental and a driver of most systematic studies. Therefore, I intend to continue being part of this aspect of research, but not necessarily restricting my career to doing testing of species limits only on

a single group, because every group present its own set of challenges, and that is exciting for me.

We are in an era where we are able to generate large amounts of HTS data. During my time as a PhD student assembling my own data, it has become clear that HTS data assembly is not as straightforward and simple as one might consider. There are a large number of uncertainties regarding the accuracy of the assembly pipelines which exist. I believe that this issue is not given the attention with as much gravity it deserves. This is important because if we assemble data erroneously, then the downstream analyses are hugely affected. In order to determine the severity of inaccuracy during assembly, a lot of testing is required, and also a lot of background knowledge regarding the approaches used by each of the assemblers. Also, during each stage of the assembly, it is important to be able to examine the intermediate results for each step-in order to pinpoint where errors in assembly are most likely to arise. For example, during my assembly stages, I identified that generation of references (de novo and iterative reference creation) and the mapping stages for shorts reads, there was a tendency for the assemblers to include divergent sequences at both trailing ends of the reads. There are a number of ways to solve this e.g., masking divergent ends of the reads, however, it is as much important to find the cause of this behaviour by the assemblers. Given that I have generated large amounts of sequence data, my intention is to look further into how the assemblers work and try to pinpoint the cause of overall erroneous or divergent sequences in the final alignments. To achieve this, I will work with colleagues I met during my time at BioEnv who are actively developing assembly pipelines for HTS data.

Lastly, I believe that my permanent place is in the sequencing facilities and the biotechnology sector where I see myself being more a genomics consultant and project coordinator for sequencing projects. Even though we are able to produce large amounts of HTS data, I believe that my home continent Africa is lagging behind in being able to also use these modern technologies for their research projects. Therefore, my focus in the long term is to actively work on ensuring that we obtain funding to have sequencing facilities in major parts of Africa where researchers can not only utilize the facilities but also ensure genomics capacity and skills development in the continent, in that sense I would be considered as defecting from actively working in academia. But I believe that it is a necessary ambition if Africa is to at all catch up with the advances in genomics.

6. References

Acha, S., Majure, L.C., 2022. A new approach using targeted sequence capture for phylogenomic studies across Cactaceae. *Genes*. 13,350.

Andermann, T., Torres-Jiménez, M.F., Matos-Maraví, P., Batista, R., Blanco-Pastor, J.L., Gustafsson, A.L.S., Kistler, L., Liberal, I.M., Oxelman, B., Bacon, C.D., Antonelli, A.A., 2020. A Guide to Carrying Out a Phylogenomic Target Sequence Capture Project. *Front. Genet*. 10,1407.

Avise, J.C., Arnold, J., Ball, R.M., Bermingham, E., Lamb, T., Neigel, J.E., Reeb, C.A., Saunders, N.C., 1987. Intraspecific phylogeography. *Annu. Rev. Ecol. Evol. Syst*. 18, 489-522.

Bengtson, A., Nylinder, S., Karis, P.O., Anderberg A.A., 2015. Evolution and diversification related to rainfall regimes: diversification patterns in the South African genus *Metalasia* (Asteraceae–Gnaphalieae). *J. Biogeogr*. 42, 121–131.

Bergh, N., Verboom, G., Rouget, M., Cowling, R., 2014. *Vegetation types of the Greater Cape Floristic Region*. In: Allsopp, N., Colville, J.F., Verboom, G.A. (Eds.), *Fynbos: Ecology, Evolution and Conservation in a Megadiverse Region*. Oxford University Press. 1–25.

Bettisworth, B., Smith, S.A., Stamatakis, A., 2023. Lagrange-NG: The next generation of Lagrange. *Syst. Biol*. 72, 242–248.

Boatwright, J.S., Le Roux, M.M., Wink, M., Morozova, T., van Wyk, B.-E., 2008. Phylogenetic relationships of tribe Crotalarieae (Fabaceae) inferred from DNA sequences and morphology. *Syst. Bot*. 33, 752–61.

Bocquet, G., 1977. *Silene dewinteri*, a new species of the Caryophyllaceae from the south-Western Cape. *Bothalia*. 12, 309–311.

Bocquet, G., Kiefer, H., 1978. The Ecklon & Zeyher collection at the Compton Herbarium with special reference to the species of *Silene* (Caryophyllaceae). *Bericht der Schweizerischen Botanischen Gesellschaft*. 88, 7–19.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 30(15), 2114–2120.

Bolus, H., 1875. Extract from a letter of Harry Bolus, Esq., F.L.S., to J.D. Hooker, Pres. R.S. *Journal of the Linnean Society*. 14, 482–484.

Bond, P., Goldblatt, P., 1984. Plants of the Cape flora: a descriptive catalogue. *Journal of South African Botany*. Suppl. Vol. 13.

Born, J., Linder, H.P., Desmet, P., 2006. The Greater Cape Floristic Region. *J. Biogeogr*. 34, 147–162.

Bouckaert, R., 2016. Phylogeography by diffusion on a sphere: whole world phylogeography. *PeerJ*. 4, e2406.

Bouckaert, R., Vaughan, T.G., Barido-Sottani, J., Duchêne, S., Fourment M., Gavryushkina A., Heled, J., Jones, G., Kühnert, D., De Maio, N., Matschiner, M., Mendes, F.K., Müller, N.F., Ogilvie, H.A., Plessis, L., Popinga, A., Rambaut, A., Rasmussen, D., Siveroni, I., Suchard, M.A., Wu, C-H., Xie, D., Zhang, C., Stadler, T., Drummond, A.J., 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS comput. Biol*. 15, e1006650.

Brundin, L., 1972. Phylogenetics and biogeography. *Syst. Zool.* 21, 69-79.

Burman, N.L., 1768. *Flora indica: cui accedit series zoophytorum indicorum, nec non Prodromus florae capensis*. Cornelius Haak, Leiden.

Carlson, J.E., Holsinger, K.E. Prunier, R., 2011. Plant responses to climate in the Cape Floristic Region of South Africa: evidence for adaptive differentiation in the Proteaceae. *Evolution*. 65, 108–124.

Chater, A.O., Walters, S.M., 1964. *Silene* L. In Tutin, T.G.T., Heywood, V.H., Walters, S.M., Webb, D.A. *Flora europaea* 1: Lycopodiaceae to Platanaceae. Cambridge University Press.

Chowdhuri, P.K., 1957. Studies in the genus *Silene*. *Notes Roy*. *Bot*. *Gard*. *Edinburgh*. 22, 221–278.

Cowling, R.M., 1990. Diversity components in a species-rich area of the Cape Floristic Region. *J. Veg. Sci*. 1, 699–710.

Cowling, R.M., Esler, K.J., Rundel, P.W., 1999. Namaqualand, South Africa—an overview of a unique winter-rainfall desert ecosystem. *Plant Ecol*. 142, 3–21.

Cowling, R.M., Pressey, R.L., 2001. Rapid plant diversification: planning for an evolutionary future. *PNAS*. 98, 5452-5457.

Cowling, R.M., Procheş, Ş., Partridge, T.C., 2009. Explaining the uniqueness of the Cape flora: incorporating geomorphic evolution as a factor for explaining its diversification. *Mol. Phylogenet. Evol*. 51(1), 64–74.

Cowling, R.M., Richardson, D.M., Pierce, S.M., 1997. *Vegetation of Southern Africa*. Cambridge University Press, UK.

Dahlgren, R., 1988. Crotalarieae (*Aspalathus*). pp. 1-430 in: Leistner, O.A., (Eds.), *Flora of southern Africa*, vol. 7, part 3, fase. 6. Pretoria: CTP Book Printers.

Dapprich, J., Ferriola, D., Mackiewicz, K., Clark, P.M., Rappaport, E., D'Arcy, M., Sasson, A., Gai, X., Schug, J., Kaestner, K.H., Monos, D., 2016. The next generation of target capture technologies - large DNA fragment enrichment and sequencing determines regional genomic variation of high complexity. *BMC Genomics*. 17, 486.

Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M., Blaxter, M.L., 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat*. *Rev*. *Genet*. 12, 499–510.

Dayrat, B., 2005. Towards integrative taxonomy. *Biological Journal of the Linnean Society*. 85, 407–415.

de Sousa, F., Bertrand, Y.J.K., Nylinder, S., Oxelman, B., Eriksson, J.S., Pfeil, B.E., 2014. Phylogenetic Properties of 50 Nuclear Loci in Medicago (Leguminosae) Generated Using Multiplexed Sequence Capture and Next-Generation Sequencing. *PLoS ONE*. 9(10), e109704.

Deacon, H.J., Jury, M.R., Ellis, F., 1992. *Selective regime and time*. In: Cowling, R.M., (Eds.), *The Ecology of Fynbos: Nutrients, Fire and Diversity*. Oxford University Press, Cape Town, 6–22.

Degnan, J.H., Rosenberg, N.A., 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol*. *Evol*. 24, 332–340.

Dellicour, S., Lemey, P., Artois, J., Lam, T.T., Fusaro, A., Monne, I., Cattoli, G., Kuznetsov, D., Xenarios, I., Dauphin, G., Kalpravidh, W., Von Dobschuetz, S., Claes, F., Newman, S.H., Suchard, M.A., Baele, G., Gilbert, G., 2019. Incorporating heterogeneous sampling probabilities in continuous phylogeographic inference—Application to H5N1 spread in the Mekong region. *Bioinformatics*. 2019, 1–7.

Desfeux, C., Lejeune, B., 1996. Systematics of Euro-Mediterranean *Silene* (Caryophyllaceae): Evidence from a phylogenetic analysis using ITS sequences. *Comptes Rendus de l'Académie des Sciences. Série III, Sciences de la Vie*. 319, 351–358.

Dobignard, A., Chatelain, C., 2011. Index synonymique de la flore d'Afrique du Nord, volume 3, Dicotyledonae Balsaminaceae à Euphorbiaceae. *Conservatoire et Jardin botaniques de la Ville de Genève, hors-série* 11b.

Douglas, J., Jiménez-Silva, C., Remco Bouckaert. R., 2022. StarBeast3: Adaptive Parallelized Bayesian Inference under the Multispecies Coalescent. *Syst. Biol*. 71(1), 901–916.

Drège, J.F., 1843–1844. *Zwei pflanzengeographischen documenten*. Besondere Beigabe zur Flora 2 (1843), Leipzig.

Drummond, A., Ho, S., Phillips, M., Rambaut, A., 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biol*. 4, e88.

Du Pasquier, P.-E., Jeanmonod, D., Naciri, Y., 2017. Morphological convergence in the recently diversified *Silene gigantea* complex (Caryophyllaceae) in the Balkan Peninsula and south-western Turkey, with the description of a new subspecies. *Bot. J. Linn. Soc*. 183, 474–493.

Đurović, S., Schönswetter, P., Niketić, M., Tomović, G., Frajman, B., 2017. Disentangling relationships among the members of the *Silene saxifraga* alliance (Caryophyllaceae): Phylogenetic structure is geographically rather than taxonomically segregated. *Taxon*. 66, 343–364.

Edler, D., Guedes, T., Zizka, A., Rosvall, M., Antonelli, A., 2017. Infomap Bioregions: Interactive mapping of biogeographical regions from species distributions. *Syst. Biol*. 66(2), 197–204.

Edwards, D., Hawkins, J.A., 2007. Are Cape floral clades the same age? Contemporaneous origins of two lineages in the genistoids s.l. (Fabaceae). *Mol. Phylogenet. Evol*. 45, 952-970.

Edwards, S.V., 2009. Is a new and general theory of molecular systematics emerging? *Evolution*. 63, 1–19.

Eggens, F., Popp, M., Nepokroeff, M., Wagner, W.L., Oxelman, B., 2007. The origin and number of introductions of the Hawaiian endemic *Silene* species (Caryophyllaceae). *Am. J. Bot*. 92(2), 210–218.

Ellis, A.G., Weis, A.E., Gaut, B.S., 2006. Evolutionary radiation of "stone plants" in the genus Argyroderma (Aizoaceae): unravelling the effects of landscape, habitat and flowering time. *Evolution*. 61, 39-55.

Erixon, P., Oxelman, B., 2008. Reticulate or treelike chloroplast DNA evolution in *Sileneae* (Caryophyllaceae)? *Mol. Phylogenet. Evol*. 48, 313–325.

Eserman, L.A., Thomas, S.K., Coffey, E.E.D., Leebens-Mack, J.H., 2021. Target sequence capture in orchids: developing a kit to sequence hundreds of single- copy loci. *Appl. Plant Sci*. 9, e11416.

Faircloth, B.C., Branstetter, M.G., White, N.D., Brady, S.G., 2015. Target enrichment of ultraconserved elements from arthropods provides a genomic perspective on relationships among hymenoptera. *Mol. Ecol. Resour*. 15, 489–501.

Farris, J.S., 1970. Methods for computing wagner trees. *Syst. Biol*. 19, 83–92.

Felsenstein, J., 2004. *Inferring phylogenies*, 2nd ed. Sinauer Associates, Sunderland, Massachusetts.

Ferreira, P., Batista, R., Andermann, T., Groppo, M., Bacon, C.D., Antonelli, A., 2022. Target sequence capture of Barnadesioideae (Compositae) demonstrates the utility of low coverage loci in phylogenomic analyses. *Mol. Phylogenet. Evol*. 169, 107432.

Fisher, R.A., 1930. *The genetical theory of natural selection*. Clarendon Press, Oxford.

Fitch, W.M., 1971. Toward defining the course of evolution: minimum change for a specific tree topology. *Syst. Biol*. 20, 406–416.

Flouri, T., Jiao, X., Rannala, B., Yang, Z., 2020. A Bayesian implementation of the multispecies coalescent model with introgression for phylogenomic analysis. *Mol*. *Biol*. *Evol*. 37, 1211–1223.

Frajman, B., Eggens, F., Oxelman, B., 2009a. Hybrid origins and homoploid reticulate evolution within *Heliosperma* (Sileneae, Caryophyllaceae)–A multigene phylogenetic approach with relative dating. *Syst*. *Biol*. 58, 328–345.

Frajman, B., Heidari, N., Oxelman B., 2009b. Phylogenetic relationships of *Atocion* and *Viscaria* (Sileneae, Caryophyllaceae) inferred from chloroplast, nuclear ribosomal, and low-copy gene DNA sequences. *Taxon*. 58, 811–824.

Ghahremaninejad, F., Angaji, A., Etemad, M., Vahidynia, F., Attar, F., 2014. Molecular taxonomy and phylogeny of *Silene* species (Caryophyllaceae) using DNA-based markers. *J. Bio. & Env. Sci.* 4(4), 125-132.

Gilbert, M.G., 2000. Caryophyllaceae. In Edwards, S., Tadesse, M., Demissew, S., Hedberg, I., *Flora of Ethiopia and Eritrea*. 2(1), 196–228. Addis Ababa university, Addis Ababa and Uppsala university, Uppsala.

Goldblatt, P. Manning, J.C., 2002. Plant diversity of the Cape region of Southern Africa. *Ann. Mo. Bot. Gard*. 89, 281–302.

Goldblatt, P., 1979. Biology and systematics of *Galaxia* (Iridaceae). *S. Afr*. J. *Bot*. 45, 385-423.

Goldblatt, P., Manning, J.C., 1996. Phylogeny and speciation in *Lapeirousia* subgenus *Lapeirousia* (Iridaceae: Ixioideae). *Ann. Mo. Bot. Gard.* 83, 346–361.

Goldblatt, P., Manning, J.C., 2000. Cape plants. A conspectus of the Cape flora of South Africa. *Strelitzia* 9. National Botanical Institute, Cape Town & Missouri Botanical Garden, St. Louis.

Goldblatt, P., Manning, J.C., 2002. Plant diversity of the Cape region of Southern Africa. *Ann*. *Mo*. *Bot*. *Gard*. 89, 281-302.

Greenberg, A. K., Donoghue, M. J., 2011. Molecular systematics and character evolution in Caryophyllaceae. *Taxon*. 60(6), 1637–1652

Greuter, W., 1995. *Silene* (Caryophyllaceae) in Greece: a subgeneric and sectional classification. *Taxon*. 44, 543–581.

Hale, H., Gardner, E.M., Viruel, J., Pokorny, L., Johnson, M.G., 2020. Strategies for reducing per-sample costs in target capture sequencing for phylogenomics and population genomics in plants. *Appl. Plant Sci.* 8, e11337.

Hedberg, O., 1954. Caryophyllaceae. *Svensk Botanisk Tidskrift Utgifven af Svenska Botaniska Foreningen*. 48, 199–210.

Hedberg, O., 1957. Afroalpine vascular plants: a taxonomic revision. *Symbolae botanicae Upsalienses*. 15, 1–411.

Heinrich, M., Edwards, S., Moerman, D.E., Leonti, M., 2009. Ethnopharmacological field studies: a critical assessment of their conceptual basis and methods. *J. Ethnopharmacol*. 124, 1–17.

Heled, J., Drummond, A.J., 2010. Bayesian inference of species trees from multilocus data. *Mol. Biol. Evol*. 27, 570–580.

Hennig, W., 1950. *Grundzüge einer Theorie der phylogenetischen Systematik*. Deutscher Zentralverlag, Berlin.

Hennig, W., 1966. *Principles for a phylogenetic systematics*. Univ. Illinois Press.

Hey, J., Nielsen, R., 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics*. 167, 747–760.

Jafari, F., Zarre, S., Gholipour, A., Eggens, F., Rabeler, R. K., Oxelman, B., 2020. A new taxonomic backbone for the infrageneric classification of the species-rich genus *Silene* (Caryophyllaceae). *Taxon*. 69(2), 337-368.

Johansson, U.S., Nylinder, S., Ohlson, J.I., Tietze, D.T., 2018. Reconstruction of the late Miocene biogeographical history of tits and chickadees (Aves: Passeriformes: Paridae): A comparison between discrete area analyses and probabilistic diffusion approach. *J. Biogeogr*. 45, 14–25.

Johnson, M. G., Pokorny, L., Dodsworth, S., Botigué, L. R., Cowan, R. S., Devault, A., Eiserhardt, W.L., Epitawalage, N., Forest, F., Kim, J.T., Leebens-Mack, J.H., Leitch, I.J., Maurin, O., Soltis, D.E., Soltis, P.S., Wong, G.K., Baker, W.J., Wickett., N.J., 2019. A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-medoids clustering. *Syst. Biol.* 68, 594–606.

Jones, G., 2017. Algorithmic improvements to species delimitation and phylogeny estimation under the multispecies coalescent. *J. Math. Biol*. 74, 447–467.

Jones, G., Aydin, Z., Oxelman, B., 2015. DISSECT: an assignment-free Bayesian discovery method for species delimitation under the multispecies coalescent. *Bioinformatics*. 31, 991–998.

Jones, M.R., Good, J.M., 2016. Targeted capture in evolutionary and ecological genomics. *Mol. Ecol.* 25, 185–202.

Kalkauskas, A., Perron, U., Sun, Y., Goldman, N., Baele, G., Guindon, S., De Maio, N., 2021. Sampling bias and model choice in continuous phylogeography: Getting lost on a random walk. *PLoS Comput. Biol*. 17(1), e1008561.

Katoh, K., Standley, D. M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol*. 30(4), 772–780.

Kidd, K.K., Sgaramella-Zonta, L.A., 1971. Phylogenetic analysis: concepts and methods. Am. J. Hum. *Genet*. 23, 235–252.

Kingman, J.F.C., 1982. The coalescent. Stoch. *Process. Their Appl*. 13, 235–248.

Küpfer, P., 1974. Recherches sur les liens de parenté entre la flore orophile des Alpes et celle des Pyrénées. *Boissiera*. 23, 113-131.

Kurzweil, H., Linder, H.P., Chesselet, P., 1991. The phylogeny and evolution of the *Pterygodium–Corycium* complex (Coryciinae, Orchidaceae). *Plant Syst. Evol*. 175, 161–223.

Kyrkou, I., Iriondo, J. M., García-Fernández, A., 2015. A glacial survivor of the alpine Mediterranean region: phylogenetic and phylogeographic insights into *Silene ciliata* Pourr.(Caryophyllaceae). *PeerJ*. 3, e1193.

Landis, M.J., Matzke, N.J., Moore, B.R., Huelsenbeck, J.P., 2013. Bayesian analysis of biogeography when the number of areas is large. *Systematic Biology*. 62, 789–804.

Larsson, A., 2014. AliView: a fast and lightweight alignment viewer and editor for large data sets. *Bioinformatics*. 30(22), 3276-3278.

Latimer, A.M., Silander, J.A., Rebelo, A.G. Midgley, G., 2009. Experimental biogeography: the role of environmental gradients in high geographic diversity in Cape Proteaceae. *Oecologia*. 160, 151–162.

Leaché, A.D., Grummer, J.A., Miller, M., Krishnan, S., Fujita, M.K., Böhme, W., Schmitz, A., Lebreton, M., Ineich, I., Chirio, L., Ofori-boateng, C., Eniang, E.A., Greenbaum, E., Rödel, M.O., Wagner, P., 2017. Bayesian inference of species diffusion in the West African Agama agama species group (Reptilia, Agamidae). *System. Biodivers*.15(3), 192–203.

Leaché, A.D., Zhu, T., Rannala, B., Yang, Z., 2019. The spectre of too many species. *Syst. Biol*. 68, 168–181.

Lemey, P., Rambaut, A., Drummond, A.J., Suchard, M.A., 2009. Bayesian phylogeography finds its roots. *PLoS Comput. Biol*. 5(9), e1000520.

Lemey, P., Rambaut, A., Welch J.J, Suchard, M.A., 2010. Phylogeography takes a relaxed random walk in continuous space and time. *Mol. Biol. Evol*. 27(8), 1877–1885.

Lemmon, A.R., Emme, S.A., Lemmon, E.M., 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. *Syst. Biol*. 61, 727–744.

Lemmon, A.R., Lemmon, E.M., 2008. A likelihood framework for estimating phylogeographic history on a continuous landscape. *Syst. Biol*. 57, 544–561.

Levyns, M.R., 1964. Migrations and origins of the Cape Flora. *Trans. R. Soc. S. Afr*. 37, 85–107.

Linder, H.P., 1985. *Gene flow, speciation, and species diversity patterns in a species-rich area: the Cape Flora*. In Vrba, E.S. (Eds). *Species and speciation*. Pretoria, Transvaal Museum Monograph 4, 53–57.

Linder, H.P., 2003. The radiation of the Cape Flora. *Biol. Rev*. 78, 597–638.

Linder, H.P., 2005. Evolution of diversity: the Cape Flora. *Trends Plant Sci*. 10, 536–541.

Linder, H.P., Hardy, C.R., 2004. Evolution of the species-rich Cape flora. *Philos*. *Trans. Royal Soc*. *B*. 359, 1623–1632.

Linder, H.P., Vlok J.H., 1991. The morphology, taxonomy and evolution of *Rhodocoma* (Restionaceae). *Plant Syst. Evol*. 175, 139–160.

Linnaeus, C., 1753. *Species plantarum*. Salvius, Stockholm.

Liu, H., Feng, C-L., Luo, Y-B., Chen, B-S., Wang, Z-S., Gu, H-Y., 2010. Potential challenges of climate change to orchid conservation in a wild orchid hotspot in southwestern China. *Bot*. *Rev*. 76, 174–192.

Liu, L., Pearl, D.K., 2007. Species trees from gene trees: reconstructing Bayesian posterior distributions of a species phylogeny using estimated gene tree distributions. *Syst*. *Biol*. 56, 504–514.

Maddison, W.P. 1997. Gene trees in species trees. *Syst*. *Biol*. 46, 523– 536.

Maire, R., 1963. *Flore de l'Afrique du nord (Maroc, Algérie, Tunisie, Tripolitaine, Cyrénaïque et Sahara)*. Edition Paul Lechevalier, Paris. 10: 229.

Manning, J.C., Goldblatt P., 2012b. *Plants of the Greater Cape Floristic Region 1: The Core Cape. Strelitzia*. 29. South African National Biodiversity Institute, Pretoria.

Manning, J.C., Goldblatt, P., 2012a. A taxonomic revision of the southern African native and naturalized species of *Silene* L. (Caryophyllaceae). *Bothalia*. 42, 147–186.

Marloth, R., 1908. Das Kapland, insonderheit das Reich der Kapflora, das Waldgebiet und die Karroo, pflanzengeographisch dargestelt. *Wissenschaffliche Ergebnisse der deutschen TiefseeExpedition auf dem Dampfer 'Valdivia'* 1898–1899 (ed. by C. Chun). 1–427. Gustav Fischer, Jena.

Martin, M., Ebert, P., Marschall, T., 2023. Read-Based Phasing and Analysis of Phased Variants with WhatsHap. In: Peters, B.A., Drmanac, R., (Eds) *Haplotyping. Methods in Molecular Biology*. vol 2590. Humana, New York, NY.

Masson, D., 1989. *Silene vlokii* D.Masson *sp. nov*., new species of Caryophyllaceae for South Africa. *Candollea*. 44, 485–491.

Matzke, N. J., 2013. *BioGeoBEARS: BioGeography with Bayesian (and Likelihood) Evolutionary Analysis in R Scripts*. University of California, Berkeley, CA.

McKain, M.R., Johnson, M.G., Uribe-Convers, S., Eaton, D., Yang, Y., 2018. Practical considerations for plant phylogenomics. *Appl. Plant Sci.* 6, e1038.

Michel, T., Tseng, Y.-H., Wilson, H.P., Chung, K.-F., Thomas, D.C., Kidner, C.A., 2022. a hybrid capture bait set for *Begonia*. *Edinb*. *J*. *Bot*. 79, 1–33.

Min Choo, L., Forest, F., Wieringa, J.J., Bruneau, A., de la Estrella, M., 2020. Phylogeny and biogeography of the *Daniellia* clade (Leguminosae: Detarioideae), a tropical tree lineage largely threatened in Africa and Madagascar. *Mol. Phylogenet. Evol*. 146, 106752.

Mirarab, S., Nguyen, N., Guo, S., Wang, L-S., Kim, J., Warnow, T., 2015. PASTA: ultra-large multiple sequence alignment for nucleotide and amino-acid sequences. *J*. *Comput*. *Biol*. 22, 377–386.

Mirarab, S., Reaz, R., Bayzid, M.S., Zimmermann, T., Swenson, M.S., Warnow, T., 2014. ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics*. 30, i541-8.

Mittermeier, R.A., Myers, N., Mittermeier, G.C., 1999. *Hotspots: Earth's biologically richest and most endangered terrestrial ecoregions*. Agrupación Sierra Madre, S.C., Mexico City.

Moiloa, N. A., Mesbah, M., Nylinder, S., Manning, J., Forest, F., de Boer, H. J., Bacon, C.D., Oxelman, B. 2021. Biogeographic origins of southern African *Silene* (Caryophyllaceae). *Mol*. *Phylogenet*. *Evol*. 162, 107-199.

Mucina, L., 2018. Biome: evolution of a crucial ecological and biogeographical concept. *New Phytol*. 222, 97-114.

Mucina, L., Rutherford, M.C., 2006. *The vegetation of South Africa, Lesotho and Swaziland*. South African National Biodiversity Institute, Pretoria.

Naciri, Y., Du Pasquier, P.-E., Lundberg, M., Jeanmonod, D., Oxelman, B., 2017. A phylogenetic circumscription of *Silene* sect. *Siphonomorpha* (Caryophyllaceae) in the Mediterranean Basin. *Taxon*. 66, 91–108.

Nicholls, J.A., Pennington, R.T., Koenen, E.J.M., Hughes, C.E., Hearn, J., Bunnefeld, L., Dexter, K.G., Stone, G.N., Kidner, C.A., 2015. Using targeted enrichment of nuclear genes to increase phylogenetic resolution in the neotropical rain forest genus *Inga* (Leguminosae: Mimosoideae). *Frontiers in Plant Science*. 6, 710.

Nylinder, S., Lemey, P., De Bruyn, M., Suchard, M.A., Pfeil, B.E., Walsh, N., Anderberg, A.A., 2014. On the biogeography of *Centipeda*: a species-tree diffusion approach. *Sys. Biol*. 63, 178–191.

Ogilvie, H.A., Bouckaert, R.R., Drummond, A.J., 2017. Starbeast2 brings faster species tree inference and accurate estimates of substitution rates. *Mol*. *Biol*. *Evol*. 34, 2101–2114.

Oxelman, B., 1991. *Silene diversifolia* Otth and related species in Europe. *Botanical journal of the Linnean Society*. 115-117.

Oxelman, B., Lidén, M., Berglund, D.,1997. Chloroplast rps16 intron phylogeny of the tribe Sileneae (Caryophyllaceae). *Plant Syst. Evol*. 206 (1), 393-410.

Oxelman, B., Lidén, M., 1995. Generic boundaries in the tribe *Sileneae* (*Caryophyllaceae*) as inferred from nuclear rDNA sequences. *Taxon*. 44, 525-542.

Oxelman, B., Lidén, M., Rabeler, R.K., Popp, M., 2001. A revised generic classification of the tribe Sileneae (Caryophyllaceae). *Nordic*. *J*. *Bot*. 20, 743–748.

Oxelman, B., Rautenberg, A., Thollesson, M., Larsson, A., Frajman, B., Eggens, F., Petri, A., Aydin, Z., Töpel, M., Brandtberg-Falkman, A., 2013. Sileneae taxonomy and systematics. http://www.sileneae.info. Accessed on 28 July 2023.

Patterson, M., Marschall, T., Pisanti, N., van Lersel, L., Stougie, L., Klau, G.W., Schönhuth, A., 2015. WhatsHap: Weighted Haplotype Assembly for Future-Generation Sequencing Reads. *J*. *Comput*. *Biol*. 22(6), 498-509.

Petri, A., Oxelman, B., 2011. Phylogenetic relationships within *Silene* (Caryophyllaceae) section *Physolychnis*. *Taxon*. 60, 953–968.

Pooley, E., 2003. *Mountain flowers: a field guide to the flora of the Drakensberg and Lesotho*. Natal Flora Publications Trust, Durban.

Popp, M., Erixon, P., Eggens, F., Oxelman, B., Ranker, T.A., 2005. Origin and evolution of a circumpolar polyploid species complex in *Silene* (Caryophyllaceae) inferred from low copy nuclear RNA polymerase introns, rDNA, and chloroplast DNA. *Syst. Bot.* 30, 302-313.

Popp, M., Gizaw, A., Nemomissa, S., Suda, J., Brochmann, C., 2008. Colonization and diversification in the African 'sky islands' by Eurasian *Lychnis* L. (*Caryophyllaceae*). *Journal of Biogeography.*, 35(6), 1016-1029.

Popp, M., Gizaw, A., Nemomissa, S., Suda, J., Brochmann, C., 2008. Colonization and diversification in the African 'sky islands' by Eurasian *Lychnis* L. (*Caryophyllaceae*). *J. Biogeogr*. 35(6), 1016-1029.

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

Popp, M., Oxelman, B., 2004. Evolution of a RNA polymerase gene family in *Silene* (Caryophyllaceae)– incomplete concerted evolution and topological congruence among paralogues. *Syst. Biol*. 53, 914–932.

Pybus, O.G., Suchard, M.A., Lemey, P., Bernardin, F.J., Rambaut, A., Crawford, F.W., Gray, R.R., Arinaminpathy, N., Stramer, S.L., Busch, M.P., Delwart, E., 2012. Unifying the spatial epidemiology and molecular evolution of emerging epidemics. *PNAS*. 109(37), 15066–15071.

Rahou, A., Amssa, M., 2003. Essai d'une synthèse d'affinité des espèces du genre *Silene* L. au Maroc. *Bulletin de l'Institut Scientifique*. (25), 43-51.

Raimondo, D., von Staden, L., Foden, W., Victor, J.E., Helme, N.A., Turner, R.C., Kamundi, D.A., Manyama, P.A., 2009. *Red List of South African Plants*. Strelitzia 25. South African National Biodiversity Institute, Pretoria.

Rannala, B., Edwards, S.V., Leaché, A., Yang, Z., 2020. The multi-species coalescent model and species tree inference, In Scornavacca, C., Delsuc, F., Galtier, N., (Eds.), *Phylogenetics in the Genomic Era*. p. 3.3:1-3.3:21.

Rautenberg, A., Hathaway, L., Oxelman, B., Prentice, H.C., 2010. Geographic and phylogenetic patterns in *Silene* section *Melandrium* (Caryophyllaceae) as inferred from chloroplast and nuclear DNA sequences. *Mol*. *Phylogenet*. *Evol*. 57: 978–991.

Rautenberg, A., Sloan, D.B., Aldén, V., Oxelman, B., 2012. Phylogenetic Relationships of *Silene multinervia* and *Silene* Section *Conoimorpha* (Caryophyllaceae). *Syst*. *Bot*. 37 (1), 226–237.

Ree, R.H., Moore B.R., Webb, C.O., Donoghue, M.J., 2005. A Likelihood Framework for Inferring the Evolution of Geographic Range on Phylogenetic Trees. *Evolution*. 59, 2299–2311.

Ree, R.H., Smith, S.A., 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Syst. Biol*. 57, 4–14.

Rohrbach, P., 1867. Conspectus systematicus specierum generis *Silenes* (Heliospermatisatque Elisanthes generibus exclusis). *Ann. Sci. Nat. Bot*. 5(8), 369-382.

Ronquist, F., 1997. Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. *Syst. Biol*. 46, 195–203.

Ronquist, F., Sanmartín, I., 2011. Phylogenetic Methods in Biogeography. *Annu. Rev. Ecol. Evol. Syst*. 42, 441–64.

Rourke, J.P., 1972. Taxonomic studies on *Leucospermum* R. Br. Journal of *S. Afr. J. Bot*. Supplementary Volume. 8, 1–194.

Rutherford, M.C., Westfall, R.H., 1986. Biomes of southern Africa—an objective categorization. *Mem. Bot. Surv. S. Afr*. 54, 1–98.

Rzhetsky, A., Nei, M., 1993. Theoretical foundation of the minimum-evolution method of phylogenetic inference. *Mol*. *Biol*. *Evol*. 10, 1073–1095.

Sáez, L., Mesbah, M., López-Alvarado, J., Bacchetta, G., El Mokni, R., Peruzzi, L., Oxelman, B., 2022. Re-establishment of *Silene neglecta* Ten. (Caryophyllaceae) with taxonomic notes on some related taxa. *PhytoKeys*. 195, 143-160.

Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425.

Schluter, D., 2001. Ecology and the origin of species. *Trends Ecol. Evol.* 16, 372–380.

Snijman, D.A., 2013. *Plants of the Greater Cape Floristic Region: The Extra Cape flora*. Strelitzia 30. South African National Biodiversity Institute, Pretoria.

Sokal, R.R., Michener, C.D., 1958. *A statistical method for evaluating systematic relationships*. University of Kansas Science Bulletin 28, 1409–1438.

Sonder, O.W., 1860. *Caryophylleae*. In W. Harvey, W., Sonder, O.W., *Flora capensis*. 1, 120–151. L. Reeve, London.

Thunberg, C.P., 1794. *Prodromus plantarum capensium*, vol. 1. Edman, Uppsala.

Thureborn, O., Razafimandimbison S.G., Wikström N., Rydin C., 2022. Target capture data resolve recalcitrant relationships in the coffee family (Rubioideae, Rubiaceae). *Front. Plant Sci.* 13, 967456.

Toprak, Z., Pfeil, B.E., Jones, G., Marcussen, T., Ertekin, A.S., Oxelman, B., 2016. Species delimitation without prior knowledge: DISSECT reveals extensive cryptic speciation in the *Silene aegyptiaca* complex (Caryophyllaceae). *Mol. Phylogenet. Evol.* 102, 1–8.

Turrill, W.B., 1954. New varieties of *Silene burchellii. Kew Bulletin* 1954, 57–58.

Turrill, W.B., 1956. *Silene*. In Turrill, W.B., Milne-Redhead, E., *Flora of tropical East Africa: Caryophyllaceae*: 31–34. Crown Agents for Oversea governments and Administrations, London.

Van der Niet, T., Johnson, S.D., 2009. Patterns of plant speciation in the Cape Floristic Region. *Mol. Phylogenet. Evol.* 51, 85–93.

van Santen, M., Linder, H.P., 2020. The assembly of the Cape flora is consistent with an edaphic rather than a climatic filter. *Mol. Phylogenet. Evol.* 142, 1–11.

Verboom, G.A., Archibald, J.K., Bakker, F.T., Bellstedt, D.U., Conrad, F., Dreyer, L.L., Forest, F., Galley, C., Goldblatt, P., Henning, J.F., Mummenhoff, K., Linder, H.P., Muasya,

A.M., Oberlander, K.C., Savolainen, V., Snijman, D.A., van der Niet, T., Nowell, T.L., 2009. Origin and diversification of the Greater Cape flora: Ancient species repository, hot-bed of recent radiation, or both? *Mol. Phylogenet. Evol*. 51, 44-53.

Verboom, G.A., Bergh, N.G., Haiden, S.A., Hoffmann, V., Britton, M.N., 2015. Topography as a driver of diversification in the Cape Floristic Region of South Africa. *New Phytol*. 207, 368–376.

Verboom, G.A., Linder, H.P., Stock, W.D., 2003. Phylogenetics of the grass genus *Ehrharta*: evidence for radiation in the summer-arid zone of the South African Cape. *Evolution*. 57, 1008–1021.

von Staden, L., 2014. *Silene undulata* Aiton subsp. *polyantha* J.C.Manning & Goldblatt. National Assessment: Red List of South African Plants version 2020.1. Accessed on 14 July 2023.

von Staden, L., 2014. *Silene undulata* Aiton subsp. *undulata*. National Assessment: Red List of South African Plants version 2020.1. Accessed on 14 July 2023.

von Staden, L., Claassens, J.G., 2014. *Silene saldanhensis* J.C.Manning & Goldblatt. National Assessment: Red List of South African Plants version 2020.1. Accessed on 14 July 2023.

von Staden, L., Helme, N.A., 2014. *Silene ornata* Aiton. National Assessment: Red List of South African Plants version 2020.1. Accessed on 14 July 2023.

von Staden, L., Manning, J.C., Goldblatt, P., 2014. *Silene rigens* J.C.Manning & Goldblatt. National Assessment: Red List of South African Plants version 2020.1. Accessed on 14 July 2023.

Weimarck, H., 1941. Phytogeographical groups, centres and intervals within the Cape flora. *Acta Universitatis lundensis, Nova Series, Sectio 2, medica, mathematica, scientiae rerum naturalium*. 37, 3–143.

Wen, D., Yu, Y., Zhu, J., Nakhleh, L., 2018. Inferring phylogenetic networks using phylonet. *Syst. Biol*. 67, 735–740.

Wickens, G.E., 1976. *Flora of Jebel Marra (Sudan Republic) and its geographical affinities*. Her Majesty's Stationery Office, London.

Wild, H., 1961. *Silene*. In Exell, A.W., Wild, H., *Flora zambesiaca* 1, 350–355. Crown Agents for Oversea governments and Administrations, London.

Wiley, E.O., Lieberman, B.S., 2011., *Phylogenetics: theory and practice of phylogenetic systematics*, 2nd ed. John Wiley & Sons, Inc., Hoboken, New Jersey.

Woudstra, Y., Quatela, A-S., Kidner, C., Viruel, J., Zuntini, A., Martin, M.D., Michel, T., Grace, O.M., 2022. Chapter 14. Target capture. In: de Boer, H., Rydmark, M.O., Verstraete, B., Gravendeel, B., (Eds) *Molecular identification of plants: from sequence to species*. Advanced Books.

Wright, S., 1931. Evolution in Mendelian populations. *Genetics*. 16, 97–159.

Yu, Y., Blair, C., He, X., 2020. RASP 4: Ancestral State Reconstruction Tool for Multiple Genes and Characters. *Mol Biol Evol*. 37(2), 604-606.

Yu, Y., Harris, A.J., Blair, C., He, X., 2015. RASP (Reconstruct Ancestral State in Phylogenies): a tool for historical biogeography. *Mol Phylogenet Evol*. 87, 46-9.

Zhang, Z., Rabiee, M., Sayyari, E., Mirarab, S., 2018. ASTRAL-III: Polynomial Time Species Tree Reconstruction from Partially Resolved Gene Trees. *BMC Bioinform*. 19 (S6), 153.

7. Acknowledgements

I would like to extend my gratitude to:

Funding organizations who provided generous financial support during my PhD. These include: The Plant.ID Innovative Training Network of the European Research Council Horizon 2020 programme under the Marie Skłodowska-Curie grant agreement No 765000. Further funding is also acknowledged from the Nilsson-Ehle Endowments (2020–11–11), Adlerbertska Stipendier (2019/20), Lars Hiertas Minne (2020–8–12).

My supervisor Bengt Oxelman, for giving me the opportunity to carry out my PhD in such an exciting environment. Your enthusiasm, willingness to teach, patience and support throughout the journey have been exceptional. You are a wealth of knowledge and I have learnt a lot from you during our random chats at the office and out in the field. Your mentoring has transformed me into a better scientist, filled with curiosity and a questioning mind never accepting the status quo. I appreciate your availability and dedication to always provide motivation particularly during the covid times when it all seemed doom and gloom with shortage of reagents and kits for the laboratory work. Lastly, I would like to thank you and Maria Jordemån Oxelman for assisting during field trips, and also for always welcoming me to your home for a scrumptious meals filled with lots of insight into Swedish culture, I will treasure those moments.

Christine Bacon and Hugo de Boer, for your support and always being there to offer guidance, particularly during the challenging times of my journey. I appreciate your being there to always provide alternative ideas on how to ensure progress on my projects, particularly during the planning and execution stages of the laboratory work and data analyses. Thank you.

My former examiner Mari Källersjö, and current examiner Mats Olson, for always showing up during my time of need, your support and always being available to have a chat and making me see the bigger picture to keeping track that my PhD journey progresses smoothly. During the times of frustration and stress you were always there to give an objective ear followed by very good advice delivered in probably the humblest of ways, I couldn't have asked for better examiners. Thank you.

Colleagues and friends Anne-Sophie Quatela, Yannick Woudstra, Isabelle Sindiswa Giddy, Zaynab Shaik, Patrik Cangren, Patrik Svedberg, Pedro Inostroza, Heléne Aronsson, Tobias Andermann, Olivier Manzi, Kelebileone Kaisara, Keitumetse Tsubane, Patience Mushonga,

and Audrey Madakiwa. I appreciate and treasure the good conversations, exchange of ideas and fun times we had during my time in Gothenburg.

Sven Toresson, Lovisa Stenbäcken and Anna Ansebo, for your continued help guidance to how to always get things to work at Botan, even during the difficult covid times when everything seemed not to work. Despite how busy you were, on any given day you always had a solution to any challenge I encountered, without ever complaining or giving a cold shoulder. For your assistance I say thank you and dedicate this thesis to you.

My collaborators Stephan Nylinder, Melilia Mesbah, John Manning, Félix Forest and Paola de Lima Ferreira, for all your invaluable inputs on the different manuscripts, and your patience throughout the developmental stages of the manuscripts. I enjoyed working with each of you and am looking forward to continued collaboration.

The curators and directors of the following herbaria: GB, GRA, PRE, KEI, NH, BOL, NBG, KMG, BNRH, BLFU, PUC, GOFS, S and UPS. I appreciate your permission to come visit the herbaria to study the *Silene* specimens and most importantly for assisting with sampling tissue from the herbarium specimens. I enjoyed our discussions on the current knowledge gaps concerning southern African flora.

The computational analyses, data handling and storage were enabled by resources (project numbers SNIC2022/22–748; SNIC2022/5–399; NAISS2023/23–298) by the National Academic Infrastructure for Supercomputing in Sweden (NAISS) and the Swedish National Infrastructure for Computing (SNIC) at the Tetralith cluster partially funded by the Swedish Research Council through grant agreements no. 2022-06725 and no. 2018-05973. Further computational analyses, data handling and storage were also provided by Sigma2 - the National Infrastructure for High Performance Computing (Saga project no. NN9623K) and Data Storage (NIRD project no. NS9623K) in Norway. The authors acknowledge support from the National Genomics Infrastructure in Stockholm funded by Science for Life Laboratory, the Knut and Alice Wallenberg Foundation and the Swedish Research Council, and SNIC/Uppsala Multidisciplinary Centre for Advanced Computational Science for assistance with massively parallel sequencing and access to the UPPMAX computational infrastructure. The authors acknowledge support from the National Genomics Infrastructure in Stockholm funded by Science for Life Laboratory, the Knut and Alice Wallenberg Foundation and the Swedish Research Council, and SNIC/Uppsala Multidisciplinary Centre for Advanced Computational

Science for assistance with massively parallel sequencing and access to the UPPMAX computational infrastructure.

My fallen soldiers Keitumetse “Sesi” Molefe, Ketumile “Mecro” Molefe, Victor “Master” Mogorosi, Thato “Chuku” Nawane, Kgosi Tebogo Charles Moiloa III, loosing you during my PhD journey was the most devastating thing to experience. In my heart you will continue to reside. I dedicate this thesis to you and will always treasure the moments we spent together, the social support and random discussions we had about importance of purpose and goals. Gofitlha re kopana gape letla dula lele menaganong yaka.

Lastly, to my family, Ketshabile Moiloa, Tsholofelo Moiloa, Sandra Moiloa, Keorapetse Moiloa, Olorato Moiloa, Goitseone Pule, Nyiko Moiloa, Ntwanano Moiloa, Tintswalo Moiloa and my partner Oarabile Gagoopane. You guys have been a great source of strength and support during my entire PhD journey. Even when you did not understand much about my work, you still showed up whenever I needed you and always gave encouragement. Ke le leboga go menagane Bahurutshe, I dedicate this thesis to you all.

Dipako le dipoko di tshwanetse Ramasedi wa Magodimo oo kogodimodimo gonne ka ene re kgona tsotlhe.