DFG Abschlussbericht

[Project Report]

PI 1169/1-2

What drives evolutionary diversification in a biodiversity hotspot? A `next generation´ phylogenetic approach comparing three clades of the Cape Floristic Region´s mega-genus *Erica*

***Erica praecox*, Klein Wellington Sneeukop** (Photo by Seth Musker)

Nicolai M. Nürk*

Lehrstuhl für Pflanzensystematik

Universität Bayreuth

*Report submitted 27.07.2023 on behalf of both applicants: N.M. Nürk and M.D. Pirie; accepted 06.11.2024. Subsequent amendments to the bibliographic data and project-related output are highlighted in red text.

1. General Information

GZ:	PI 1169/1-2
Project number:	251415172
Project title:	What drives evolutionary diversification in a biodiversity hotspot? A \`next generation´ phylogenetic approach comparing three clades of the Cape Floristic Region´s mega-genus *Erica*.
Names of applicants:	(1) Dr. habil. Michael D. Pirie (until 12. 2018) (2) PD Dr. Nicolai M. Nürk (since 12.2018)
Official addresses:	(1) University of Bergen, University Museum of Bergen, The Natural History Collections, Postboks 7800, 5020 Bergen, Norway (2) University of Bayreuth, BayCEER, Dept. of Plant Systematics, Universitätsstraße 30, 95447 Bayreuth, Germany
Names of cooperation partners:	Prof. D. U. Bellstedt (Dept. of Biochemistry, University of Stellenbosch, Stellenbosch, South Africa). Dr. E.G.H. Oliver (Dept. of Botany and Zoology, University of Stellenbosch, Stellenbosch, South Africa). Dr. Nicholas Le Maitre (Dept. of Biochemistry, University of Stellenbosch, South Africa). Dr. Timotheus van der Niet (University of KwaZulu-Natal, South Africa). Malvina Kadlec (Johannes Gutenberg-Universität Mainz, Germany).
Reporting period:	01.12.2018—31.03.2023 (52 month)

2. Summary

The question of what generates biodiversity remains a longstanding classic and is yet still not adequately answered. Since Humboldt and Bonpland (1805; 1846) we know that biodiversity is unequally distributed around the globe with often high diversity in the tropics decreasing towards the poles. Mountain regions with steep ecological gradients and pronounced habitat heterogeneity are unusually diverse, especially in the tropics. Centres of plant diversity outside the tropics, such as the Mediterranean ecosystems, contribute to the uneven distribution in space. Similar to the geographic patterns, diversity is unequally distributed among groups, with radiating lineages often attended by species-poor successive sister clades. In this project, we studied the evolution of biodiversity using its simplest quantity: species richness. We focused on a flowering plant clade, *Erica* (heathers) that shows tremendous differences in species richness among its lineages and that is an iconic member of the fynbos flora of the Cape Floristic Region (CFR) of South Africa, one of the Mediterranean biodiversity hotspots. We tested hypotheses that might explain the incredible diversity of heathers in the CFR. We conducted weeks of fieldwork observing and documenting over 1260 plants, collected more than 700 herbarium specimens and leaf tissue samples, and curated a database of over 65 000 geo-referenced observations. We DNA sequenced three complete genomes, 295 samples by Hyb-Seq, 66 by GBS, and 525 Sanger sequenced, together covering about 425 species. During the project, we also brought together various experts to join forces for *Erica*-specific conservation actions and founded the Global Conservation Consortium for *Erica* (GCC *Erica*). To address our project-specific questions, we reconstructed phylogenetic trees at different evolutionary scales: genus, species, and below species level. Genomic markers and bioinformatic tools were newly developed to suit our purposes. This allowed us to investigate the age and the biogeographic origin of Cape *Erica*. We tested for likely factors proposed as potential causes of the high plant diversity of the CFR, principally those relating to shifts in geographic range, ecological niche, and pollinators. Taken together, a highly dynamic picture of Cape *Erica's* evolution is emerging that points towards geographic features – dissected 'sky island' topography in the CFR, spatial distance, and altitude differences – setting the stage for intensified initial speciation (in isolation). Ecological adaptations dynamically interrelate possibly at finer scales, for example, in situ adaptation to edaphic conditions or more rocky and arid habitats. In parts potentially neutrally driven morphological innovations seem to keep populations separated on secondary contact by influencing pollinator specificity. We inferred these dynamics to be underpinned by genomic processes, including hybridization and horizontal gene transfer. Taken together, neutral and adaptive processes with scale-dependent effects contribute to lineage diversification, and thus, to the megadiverse species richness in *Erica* in the CFR.

2. Zusammenfassung

Die klassisch-zentrale Frage wie Biodiversität entsteht ist vieluntersucht und hinreichend nicht beantwortet. Bereits Humboldt und Bonpland (1805; 1846) zeigten, dass globale biologische Diversität ungleichmäßig verteilt ist, mit einer oft hohen Vielfalt in den Tropen, die zu den Polen hin abnimmt. Gebirge mit allgemein starken ökologischen Gradienten und ausgeprägter Habitat-Heterogenität sind besonders in den Tropen ungewöhnlich artenreich. Pflanzenvielfalt außerhalb der Tropen, wie in den mediterranen Ökosystemen, trägt zur

ungleichen globalen Diversitätsverteilung bei. Ähnlich zu geografischen Mustern ist die Diversität in einzelnen Taxa ungleich verteilt, mit artenreichen Radiationen oft sukzessiv von artenarmen Schwestergruppen begleitet. In diesem Projekt untersuchten wir die Evolution von biologischer Vielfalt anhand ihrer einfachsten Messgröße: Arten. Die megadiversen *Erica* (Heidekräuter) bilden die artenreichste Gattung der Fynbos-Flora im der Kap Region von Südafrika, einem der extratropischen mediterranen Biodiversitätszentren. Wir untersuchten Hypothesen die eventuell die enorme Diversität erklären können. Dafür dokumentierten wir in mehrwöchiger Feldarbeit über 1260 Pflanzen, sammelten mehr als 700 Herbarbelege, und erstellten eine Datenbank mit über 65000 georeferenzierten *Erica* Vorkommen. Drei Genome wurden *de-novo* sequenziert, 295 Proben mittels Hyb-Seq, 66 mittels GBS und 525 mittels Sanger-Sequenzierung; insgesamt ca. 425 Arten. Im Laufe des Projekts brachten wir Wissenschaftler und Experten zusammen und gründeten das ‚Global Conservation Consortium for *Erica*' (GCC Erica). Investigativ rekonstruierten wir phylogenetische Bäume auf verschiedenen evolutionären Ebenen; auf Gattungsniveau, Art- und unterhalb des Artniveaus. Molekularer-genomische Marker und bioinformatische Analysen wurden passgenau für unsere Anforderungen entwickelt. Wir modellierten Alter und Ursprung der Heidekräuter der Kap Region und untersuchten potenziell verantwortliche Faktoren für die hohe Pflanzenvielfalt in der Region, vor allem in Bezug auf Ausbreitung, ökologischen Konservatismus und Bestäuber-Spezifizität. Die gewonnenen Resultate ergeben ein hochdynamisches Szenario, indem räumliche Isolation – differenzierte "sky island" Topografie der Kap Region, räumliche Distanz und Höhenunterschiede – die Grundlage ist für erhöhte Artbildung. Ökologische Anpassungen scheinen eher auf kleineren räumlichen Skalen in dynamischer Wechselbeziehung zu stehen, wie Adaptation an lokaledaphische Bedingungen oder felsig-trockene Lebensräume. Potentiell in Isolation evolvierte, morphologische Innovationen halten Populationen durch Beeinflussung der Bestäubungsspezifität unter sekundärem Kontakt getrennt. Diese Dynamiken werden häufig von genomischen Prozessen überlagert, einschließlich Hybridisierung und horizontalem Gentransfer. Zusammengefasst tragen sowohl neutrale und als auch adaptive Prozesse mit niveauabhängigen Effekten zur evolutionären Diversifizierung bei; somit zur enormen Artenvielfalt in *Erica* in der Kap Region.

3. Progress Report

Background and Objectives of the Project

The Cape Floristic Region (CFR) is a botanical hotspot and represents an ideal system to investigate biological diversity. It is one of the world's biodiversity hotspots (Myers *et al.*, 2000) that is, given its modest geographical extent (Fig. 1; the roughly the size of Portugal), home to a disproportionately high number of plant species, ca. 9 000, most of which are found nowhere else (70% species endemism; Linder, 2003). The megadiverse genus *Erica* is easily the most species-rich (Manning & Goldblatt, 2012), the most prominent and probably most rapidly radiating plant lineage in the CFR (Pirie *et al.*, 2016); with over 690 described 'Cape' species that account for 7% of the flora, found in typically shifting compositions of sympatric species at any given location. *Erica* is a key taxon suited to understanding the dynamics underlying the origin of species in the CFR. By understanding the processes underlying the species richness of the CFR we can better understand the processes underlying the origins of biodiversity in general.

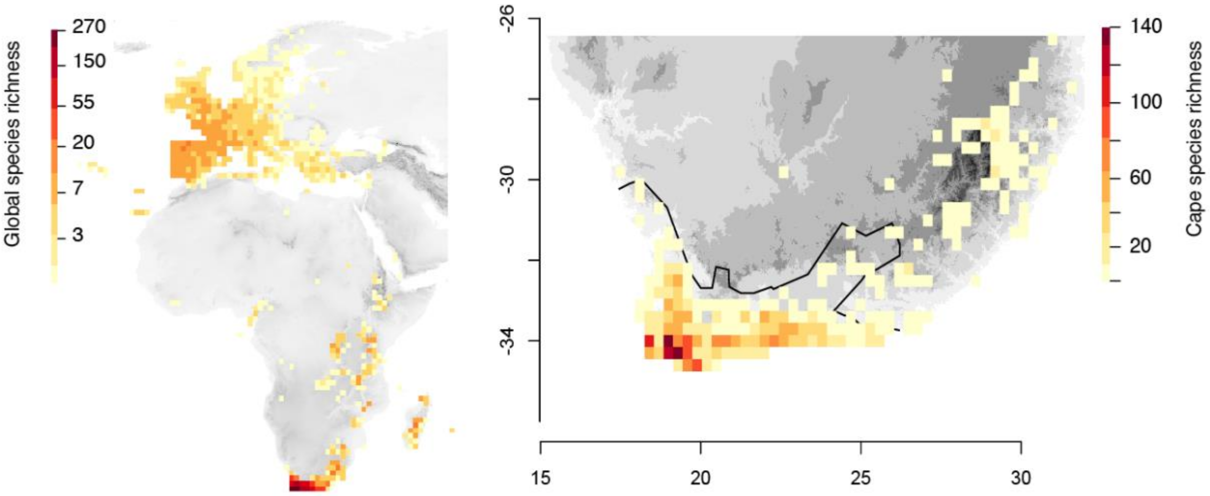

Figure 1 | *Erica* species richness. Note the natural logarithmic scale of global species richness (left panel). The extent of the CFR is indicated by black lines in the Cape species richness map (right panel).

In the first part of the project (2014–2017) we could show that the Cape *Erica* species descend from a common ancestor and that the species radiation in the CFR occurred within the last (6–) 10 (–15) million years. The rather young age makes *Erica* not only the largest but also the most evolutionarily dynamic lineage in the Cape flora (Pirie *et al.*, 2016). Extending from these results, we now focused (**1**) on groups of more closely related Cape *Erica* species: the *imbricata*/*coccinea*-clade (data collected in the first part), the *abietina*/*viscaria*-clade, *E. plukenetii*, and the *E. abietina* species complex. The clades comprise a total of c. 34 species and several intraspecific taxa with often highly localised geographic distributions. They are each characterised by variations in floral characters that are generally related to pollination by different vectors. The recent divergence of the species and populations in these clades means that phylogenetic evidence for evolutionary events, such as dispersals, may remain unmasked by subsequent events. However, reticulate processes, such as hybridisation, and coalescent stochasticity can result in considerable differences between gene trees when comparing closely related individuals and clades, as is apparent in our *imbricata*/*coccinea*-clade dataset. Accurate phylogenetic inference is only possible under these circumstances if we sample numerous, meaningfully resolved independent gene trees and model the processes underlying their differences appropriately. Therefore, we extended approaches from the first part of the project (cf. Kadlec *et al.*, 2017) and focused (**2**) on an *Erica*-optimised hybridisation capture and high-throughput sequencing pipeline (*i.e.* target capture sequencing experiments, Hyb-Seq for short) primarily designing a set of genomic targets selected to provide phylogenetic information at different evolutionary scales. This involved the use of ‘target probes’ designed from known sequences – such as from a list of orthologous gene sequences of flowering plants, or genomes/transcriptomes ideally of more closely related species – to bind to targeted regions of template DNA, which are thereby selectively enriched for high throughput sequencing.

Extending from the data and protocols obtained in the first part of the project, we aimed at testing whether species diversification in the spectacularly diverse CFR is driven by (*i*) geographic range shifts (allopatric speciation), (*ii*) differential adaptations to pollinators with differing distributions (Grant-Stebbins model), (*iii*) differential adaptations to soil types or factors associated with elevation, and (*iv*) geographic area-/pollinator-specific rates of speciation and/or extinction. To test our hypotheses using phylogenetic inference approaches we used both megaphylogenies covering as much as possible diversity known in *Erica* and coalescence-based species tree estimation based on NGS-derived genomic data.

Work programme, data produced, and scientific output

With the start of the project in December 2018, we initially aimed at improving the genomic resources required to design a genetic **target set** ('target probes') that serve our needs for phylogenetic information at different evolutionary scales, that is, at genus, (sub-) clade, species, and possibly population level. Whole-genome shotgun sequencing was done for three *Erica* species (*E. trimera* (Engl.) Beentje from the East African highlands; *E. cinerea* L., widespread across western Europe; and *E. cerinthoides* L. widespread in the CFR and further east in South Africa; the latter two cultivated at Bergen Botanical Garden) on an Illumina NovaSeq 6000 SP flow cell (Musker, 2023) (Musker et al., 2024a).

High-quality assemblies of the three new ***Erica* genomes**, together with three meanwhile publicly available, well-annotated *Rhododendron* transcriptomes (Soza *et al.*, 2019; Yang *et al.*, 2020; Zhang, L-G *et al.*, 2022) were used to refine the target set of Kadlec *et al.* (2017); for future compatibility also considering 'universal' target sets such as the Angiosperm353 probes (Mega353 reference set; McLay *et al.*, 2021). By using the combined information from the *Erica* genomes (genus-specific) and the annotated transcriptomes (more distant relationships) in target design, we were able to specifically select for (1) putative orthologous gene targets and (2) longer intergenic regions, such as introns and other non-coding sequences by theory diverging more at shallower evolutionary scales. Our DFG-funded doctoral candidate, Seth Musker, developed new bioinformatic pipelines (TARGETVET; resource information see below, point 4.2) (Musker et al., 2024a) greatly improving existing approaches to ortholog detection in both computational speed and accuracy (MSA free approach; quantitative detection based on WGS read depth and assembly-target overlap ratios). This resulted in a highly informative target set comprising 285 targets – the 'Erica285' set, with a combined length of 1.16 Mbp, and mean target length of 3.95 kb – and which had very high recovery rates (>99%) across analysed *Erica* species. Hyb-Seq experiments included 295 samples representing 152 species (1—14 samples per species, 75% of species represented by >3 samples). Recovered supercontigs had a mean length of 4.61 kb across all samples signifying the new target set is highly promising for phylogenetic inference in *Erica*.

Collaborative field **collections** were done in March 2019 and again in October 2022 collecting herbarium specimens and silica samples for molecular analyses particularly but not only hunting for species not yet included in published phylogenies. Due to travel restrictions during the COVID-19 pandemic Ph.D. student Musker was forced to remain for an extended period in South Africa throughout the years 2020 and 2021. This situation changed the balance of access to facilities versus field sites, leading us to the pragmatic decision to expand the project's scope to include a population genetic study on a recalcitrant species complex, the *E. abietina* aggregate endemic to the Cape Peninsula. Altogether 710 specimens were collected under research permits by SANParks (South African National Parks reference number to Nürk: NUEN/AGR/004—2019/2019-2023/V1, to Musker: CRC/2019- 2020/004–2019/V1) and by CapeNature (Permit number to Pirie: AAA006-

00001-0028, to Musker: CN35-31-8281) and permanently stored at Compton Herbarium (NBG). In the periods in which it was possible to take advantage of access to lab facilities at the University of Cape Town, Musker also managed to extensively test and develop wet lab DNA extraction protocols that deliver genomic DNA of *Erica* species good enough in quality and quantity for target capture sequencing experiments (DNA extraction in *Erica* is particularly challenging, e.g., Bellstedt *et al.*, 2010), and that is publicly available (Musker 2023) (Musker et al., 2024a).

In the meantime, we continued working on a **mega-phylogeny** for *Erica* aiming at high species coverage. Based on 21 concatenated Sanger-sequenced genetic markers (>12.5 kb, cpDNA and nrDNA; i.e. two gene trees with minimal supported topological conflict) the mega-phylogeny contains 730 accessions today representing 6 outgroups (4 species) and 551 *Erica* species (587 specific and subspecific taxa), which represents 67% of the 827 species accepted following the taxonomic concepts of E.G.H. Oliver (Oliver & Forshaw, 2012) [65% of 851 currently recognised (non-hybrid) species (Elliot et al. 2024) following the taxonomic concepts of E.G.H. Oliver (Oliver et al. 2024)]. We engaged in further collaborative work aimed at bringing research results to bear on conservation of threatened species that will be presented in a themed collection of papers under the title “Systematics, natural history, and conservation of *Erica* (Ericaceae)” (https://phytokeys.pensoft.net/browse_topical_collection_documents.php?collection_id=216&journal_name=phytokeys). This special issue, guest-edited by affiliates of the GCC *Erica* consortium including Musker and Pirie represents the latest insights and state of knowledge in interdependent topics including nomenclature, taxonomy, phylogeny, ecology, and *ex-situ* conservation of *Erica* species. Showcased in the open access, peer-reviewed journal *PhytoKeys*, these papers will present a multi-disciplinary resource aiming to inform in the multifaceted challenge of conservation of *Erica* species in the face of the dramatic ongoing biodiversity disaster (Pirie *et al.*, in prep.). Our further contributions will include nomenclatural novelties (Nelson et al., 2023) and new species descriptions (Hoekstra et al., submitted); a comprehensive database of *Erica* taxonomic names (species, subspecies, and varieties) under the World Flora Online (WFO; Elliott et al., 2024; current and previous versions of which are openly available: https://wfoplantlist.org/plant-list/taxon/wfo-4000013772-2023-06?page=1); an updated version of the *Erica* Identification Aid (Oliver et al., 2024); review of the taxonomic work on Erica by E.G.H. and I.M. Oliver (Nelson et al., 2024); a review of *Erica* in Madagascar (Hackel et al., 2024); phylogenomic resources (Musker et al., 2024a); an analysis of phylogenetic diversity and threat status of *Erica* taxa for conservation prioritisation (Pirie et al., 2024); and a gap analysis of ex situ collections of threatened *Erica* species (in prep., as a separate report).

Scale-dependent results:

Genus level: Pirie *et al.* (2016) showed that Cape *Erica* is monophyletic, and that the genus most likely originated somewhere in the Palearctic. We used nonparametric biogeographic model testing to evidence its **biogeographic history** and to investigate how *Erica* got into the Cape. We tested for different classical dispersal scenarios, for example, a ‘Cape to Cairo’ model that includes long-distance dispersal from Europe to South Africa *versus* a ‘southwards stepping-stone’ model (Cairo to Cape), considering also models that reflect climatic similarity and/or geographic distance. Climatic envelopes of species of five geographic regions of global *Erica* occurrence (based on n = 6818 occurrences) and geographic distances across regions were used to parameterize the ecologic/geographic models. Produced evidence points towards a historical biogeographic scenario close to the 'southwards stepping-stone' model that seem, however, primarily guided by a combination of climatic niche conservatism and dispersal distance constraints (Fig. 2). A notable estimate is the low rate of long-distance dispersal and establishment with a mean rate of 0.05 events $*$ Ma^{-1} across the genus, one every 20 Ma. These low rates also highlight idiosyncrasy in the biodiversity assembly of the CFR, in which one successful dispersal event led to the tremendous species radiation that generated the prime flowering plant group of the Cape flora of today (Pirie *et al.*, 2019).

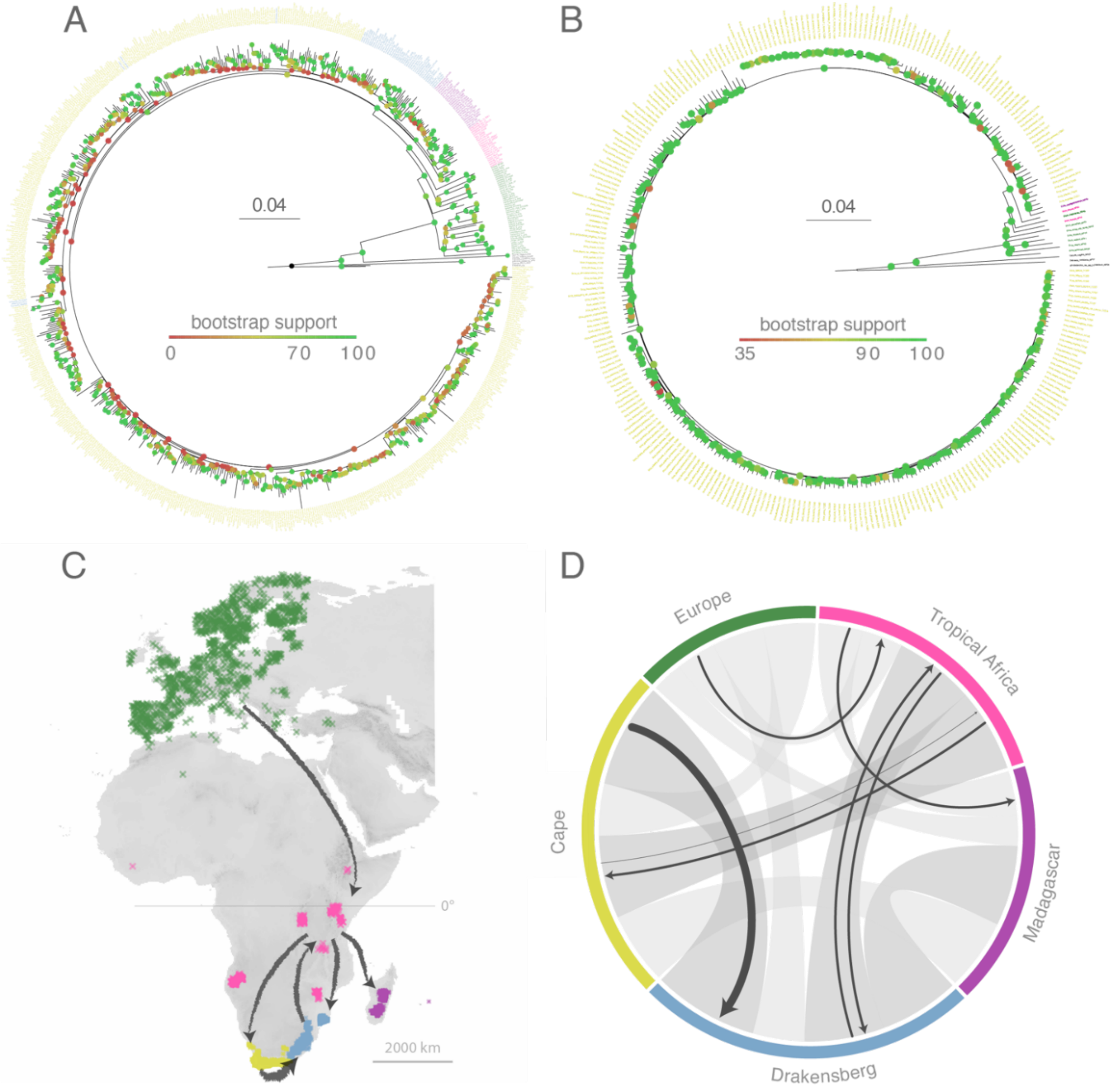

Figure 2 | *Erica* phylogenetics and historical biogeography. (**A**) Mega-phylogeny of *Erica* based on 21 Sanger-sequenced regions (>12.5 kb, ntax = 730); tip names follow color code of C and D; node support is RAxML fast bootstrap; *Empetrum nigrum* L. is used for rooting. (**B**) NGS phylogeny of (Cape) *Erica* based on 285 genomic Hyb-Seq targets (1.16 Mbp, ntax = 295); node support is IQ-TREE ultrafast bootstrap; *Rhododendron rex* H.Lév. is used for rooting. (**C**) Dispersal scenario inferred from the best-fit biogeographical model (parameterized using geographic and climatic distances among the five regions of global *Erica* occurrences) and (**D**) rates of dispersal [events $*$ Ma^{-1}; scaled arrow size] inferred by 1000 stochastic character mappings under the same model overlayed on a chord diagram of climatic niche similarity (Schoener's *D*) among geographic regions of *Erica* occurrences (Pirie *et al.*, 2019). Note the different scales of bootstrap support in A and B.

Using the **genomic sequences** of the 'Erica285' target set generated in the project we analysed 32 accessions included both in the genomic data set and in the mega-phylogeny, including two outgroup samples. We used quartet-based coalescence methods that account for gene tree uncertainty (estimation error) when summarising sets of gene trees under the MSC for species tree estimation (in wASTRAL-h; Zhang, C *et al.*, 2022). The produced genus-wide species tree is identical to the concatenated mega-phylogeny (after pruning non-identical samples) besides the position of *E. australis* sister to the 'African' clade in the ASTRAL species tree but nested within the mega-phylogeny (Fig. 2; Musker, 2023, Fig. 2.13). This result confirms that, firstly, the 'Erica285' target set is suited to phylogenetic inference at the genus level and, secondly, the concatenated mega-phylogeny, although generally not supported in its backbone, provides a stable hypothesis about deeper relationships in *Erica* (Musker *et al.*, in prep.-b) (Musker 2023; Musker et al., 2024a).

***Erica abietina/viscaria*-clade level**: A central part of the resources available in the project went into a comprehensively sampled phylogenomic tree estimation of the species in the *E. abietina/viscaria* clade. DNA extraction and Hyb-Seq were done for 133 samples representing 19 species including 37 subspecific taxa with multiple samples per taxon, plus one putative hybrid (based on morphological character assessment) and two successive sister species used as outgroups in phylogenomic inference. Concatenated trees (IQ-TREE) and wASTRAL-h trees were visually compared for congruence by tanglegrams and node support indices (bootstrap support values, etc.). Besides a few differences often among samples within species, two major discordances may be noteworthy: the putative hybrid sample *E. abietina* subsp *atrorosea* x *E. viscaria* subsp *viscaria* SM403, and one *Erica grandiflora* subsp *grandiflora* SM505 that occurs in sympatry with its IQ-TREE sister tip but groups with other *grandiflora* samples in the ATRAL tree. It is known that concatenation-based tree inference is less sensitive than coalescence-based methods to conflicting topological signals between different regions of the genome, such as might be caused by ILS or introgression, especially when those signals are relatively weak (Giarla & Esselstyn, 2015; Jiang *et al.*, 2020). We tested for the effect of hybrid inclusion on phylogenetic inference comparing node support of ASTRAL trees inferred with and without the putative hybrid samples included. Since nodes shared in both trees (114 out of 133) did not have notably different local posterior probability values on average (paired t-test: $mean\ diff. = 0.0072, df = 111, p = 0.313$) this result confirms that inclusion of hybrids has a rather negligible effect on phylogeny inference under the MSC in both node support and topology (Musker *et al.*, in prep.-a) (Musker, 2023).

***Erica abietina* population level**: The *E. abietina* aggregate is endemic to the Cape Peninsula separated from the rest of the fynbos vegetation of the CFR by the plains of Cape Town. Four subspecific taxa are currently recognized: *abietina*, *diabolis*, *atrorosea*, and *constantiana*. We tested for population structure, isolation-by-distance (IBD), admixed individuals, and phylogenetic relationships of lineages using genotyping-by-sequencing (GBS; RAD-type sequencing approach; Elshire *et al.*, 2011) on 65 samples with ≥6 samples per taxon, and several of unclear affiliation to subspecies, plus four outgroups. After careful data curation, the final alignments had a length of 4.9 Mbp and almost 300 kb SNPs. Exploratory analyses based on principal components and STRUCTURE-type analyses (PCA and Admixture model) identified admixed individuals and population graphs were used to visualize network structure at different scales (using NetView P; Steinig *et al.*, 2016). Maximum likelihood phylogenetic inference of the total data set and a reduced data set with admixed individuals removed (n = 48) amended findings on population structure. Results indicate the presence of six independently evolving lineages, at least one of which represents cryptic diversity (*atrorosea* North vs. *atrorosea* South). Tests on hybrids and their specific parents (using NewHybrids; Anderson, 2008) confirmed cases of recent hybridization and backcrossing suggesting formation of stable hybrid populations. Species tree estimation under the polymorphism-aware PoMo model (implemented in IQ-TREE v2; Schrempf *et al.*, 2019) on the n = 48 'non-admixed' dataset confirmed the existence of six independent lineages, while pairwise F_{ST}-*based* species delimitation (using ADMIXTOOLS2; Bhatia *et al.*, 2013) did not justify new species recognition within the *E. abietina* complex. Remarkably, flower tube length associated with pollination vectors (birds to bees) shifted twice within the species: once in *E. a. constantiana* and independently in *E. a. diabolis.* The second especially is remarkable, as IBD in the *E. a. diabolis* lineage and its sister, *E. a. abietina*, shows significant spatial structure of genotypes (Mantel test, $r = 0.7,\ p = 0.001$ in *abietina* and $r = 0.6, p = 0.021$ in *diabolis*) suggesting effective lineage-specific pollen set in sympatry. These results show that floral differentiation acts in concert with geographic isolation to maintain reproductive isolation (Grant-Stebbins model). Additionally, species network analysis based on $f3$-statistics (measure of allele frequency distributions; Peter, 2016) on the same dataset, indicates the likely presence of a 'ghost lineage' contributing to the origin of the *abietina* + *diabolis* clade. Taken together, our results exemplify a highly dynamic system of lineage divergence, where the concerted effects of spatial isolation, morphological innovations, edaphic adaptation, and past and ongoing hybridization and introgression contribute to ongoing lineage diversification (Musker *et al.*, 2023) (Musker et al., 2024b).

Experience in methods, reuse options, outreach, and public relations

The 'Erica285' target set will be the new standard for phylogenomic investigations in *Erica* and will also prove useful in the related genus *Rhododendron* and beyond Ericaceae in more global phylogenetic analyses on flowering plants (compatible with Angiosperm353 data sets). It will be made available OA (Musker *et al.*, in prep.-b) (Musker et al, 2024a), as is the *Erica*-optimised extraction protocol that has been shown to produce genomic DNA of sufficient quality and quantity for high throughput sequencing approaches (Musker, 2023; Musker *et al.*, in prep.-b) (Musker et al, 2024a), and that we already shared with researchers in the frame of the DFG SPP 1991 TaxonOmics. In summary, we demonstrated the benefit of Hyb-Seq approaches using our 'Erica285' target set (1) in identifying polyploid species by paralog frequencies (Musker, 2023; Fig. 2.8), and, more importantly, (2) in resolving rapid, recent radiations in *Erica* (Musker *et al.*, in prep.-a; Musker *et al.*, in prep.-b) (Musker 2023; Musker et al, 2024b), which is not possible with classical approaches (Fig. 2-A, B). Phylogenetic hypotheses and time-calibrated trees will be made openly available in supplementary material and/or public repositories (Dryad, Figshare). We followed international standards (e.g., Darwin Code) for databasing, also contributing to the updated WFO Plant List for *Erica* (https://wfoplantlist.org). Remaining results from the project that are not yet published and publicly available will be disseminated in the OA theme collection "Systematics, natural history, and conservation of *Erica* (Ericaceae)" in *PhytoKeys* (Musker et al, 2024a).

***Erica* symposium on systematics, evolution, ecology, and conservation 2019**: On the 22nd of March 2019 we held a first meeting on *Erica* conservation and research at SANBI's research centre at Kirstenbosch Botanic Garden in Cape Town. There were around

25 people attending, including researchers from South Africa and Europe, students from the UCT, and *Erica* enthusiasts from the general public. The current stage of research on *Erica* was presented and future directions discussed. The meeting established networks between naturalists and scientists interested in *Erica* in the Cape.

***Erica* research and conservation symposia at SANBI 2022**: On the 7th of October 2022 we held a follow-up *Erica* conservation and research meeting, again at SANBI. There were 17 participants in person, with at least a further 18 online including several from botanic gardens and universities in Italy, Spain, Switzerland, and the UK. Talks covered topics from conservation, pollination ecology, phylogeography, and anatomy, and are openly available: https://youtube.com/playlist?list=PL_LcKJotoHz6dX6aGFYj6UJOAem1b1ix-. An extended discussion with active participation also from our online attendees allowed us to pinpoint several future challenges and opportunities for progress combining research and conservation goals.

Musker presented a **poster at Botany 2021 on TargetVet**, and gave a talk at a **South African Society for Systematic Biology (SASSB) webinar** "Studying macro- and microevolution" entitled "The genomics of dynamic diversification on a continental archipelago – the case of *Erica abietina* from the Cape Floristic Region: Or 'What to call the things I'm calling things'". The online format transformed this series into an international platform combining prominent keynote speakers with select junior researchers.

Musker and Kandziora led a two-day **workshop on sequence capture** analysis on 27th to 28th April 2022, at the University of Bergen, with attendees from the University Museum and the faculty of Biology, and the University of Bayreuth, combining the trip with a research visit to Pirie.

We have sought to disseminate ongoing work to targeted members of the general public, particularly **stakeholders in biodiversity conservation and citizen scientists in South Africa and visitors to the University Gardens in Bergen**. Two popular articles were published in *Veld & Flora* (Bellstedt *et al.*, 2020) and *Årringen* (the OA periodical of the Bergen University Gardens; Pirie *et al.*, 2022a) and presented through the webpages of UiB and GCC *Erica* and shared through our media channels.

4. Publicly Accessible Project Results

4.1 Publications with scientific quality assurance

Musker, S.D., Pirie, M.D., & Nürk, N.M. (2024b). Pollinator shifts despite hybridisation in the Cape's hyperdiverse heathers (Erica, Ericaceae). *Molecular Ecology* 33: e17505. doi: 10.1111/mec.17505. Open Access

Elliott, A.C., Bester, S.P., Klopper, R.R., Nelson, E.C., & **Pirie, M.D.** (2024). Curating an online checklist for *Erica* L. (Ericaceae): contributing to and supporting global conservation through the World Flora Online. *PhytoKeys* 243: 121–135. doi: 10.3897/phytokeys.243.121555. Open Access

Nelson, E.C., **Pirie, M.D.**, & Bellstedt, D.U. (2024). Redefining the megagenus *Erica* L. (Ericaceae): the contributions of E. G. H. Oliver and I. M. Oliver (née Nitzsche) to

taxonomy and nomenclature. *PhytoKeys* 244: 39–55. doi: 10.3897/phytokeys.244.121705. Open Access

Oliver, E.G.H., Forshaw, N., Oliver, I.M., Volk, F., Schumann, A.W.S., Dorr, L.J., Hoekstra, R.D., **Musker, S.D., Nürk, N.M., Pirie, M.D.**, & Rebelo, A.G. 2024. Genus Erica: An identification aid version 4.00. *PhytoKeys* 241: 143–154. doi: 10.3897/phytokeys.241.117604. Open Access

Pirie, M.D., Bellstedt, D.U., Bouman, R.W., Fagúndez, J., Gehrke, B., Kandziora, M., Maitre, N.C.L., **Musker, S.D.**, Newman, E., **Nürk, N.M.**, Oliver, E.G.H., Pipins, S., Niet, T. van der, & Forest, F. (2024). Spatial decoupling of taxon richness, phylogenetic diversity and threat status in the megagenus Erica (Ericaceae). *PhytoKeys* 244: 127–150. doi: 10.3897/phytokeys.244.124565. Open Access

Nelson, E.C., Oliver, E.G.H., & **Pirie, M.D.** (2023). *Erica* L. (Ericaceae): homonyms amongst published names for African species and proposed replacement names. *PhytoKeys* 236: 157–178. doi: 10.3897/phytokeys.236.110498. Open Access

Pirie, M.D., Blackhall-Miles, R., Bourke, G., Crowley, D., Ebrahim, I., Forest, F., Knaack, M., Koopman, R., Lansdowne, A., **Nürk**, N.M., Osborne, J., Pearce, T.R., Rohrauer, D., Smit, M., & Wilman, V. (2022). Preventing species extinctions: A global conservation consortium for *Erica*. *Plants, People, Planet* 4: 335–344. doi: 10.1002/ppp3.10266. Open Access

Nelson, E.C., & **Pirie**, M. (2022). Where have all the heathers gone? *Sibbaldia: The International Journal of Botanic Garden Horticulture*. doi: 10.24823/Sibbaldia.2022.1887. Open Access

Le Maitre, N.C., **Pirie**, M.D., & Bellstedt, D.U. (2019a). An approach to determining anthocyanin synthesis enzyme gene expression in an evolutionary context: an example from *Erica plukenetii*. *Annals of Botany* 1–9. doi: 10.1093/aob/mcz046. Open Access

Le Maitre, N.C., **Pirie**, M.D., & Bellstedt, D.U. (2019b). Floral Color, Anthocyanin Synthesis Gene Expression and Control in Cape *Erica* Species. *Frontiers in Plant Science* 10: 1–9. doi: 10.3389/fpls.2019.01565. Open Access

Pirie, M.D., Kandziora, M., **Nürk**, N.M., Le Maitre, N.C., Mugrabi de Kuppler, A., Gehrke, B., Oliver, E.G.H., & Bellstedt, D.U. (2019). Leaps and bounds: geographical and ecological distance constrained the colonisation of the Afrotemperate by *Erica*. *BMC Evolutionary Biology* 19: 222. doi: 10.1186/s12862-019-1545-6. Open Access

4.2 Other publications and published results

Musker, S.D., **Nürk**, N.M., & **Pirie**, M.D. (2024a). Worth the fuss? Maximising informativeness for target capture-based phylogenomics in *Erica* (Ericaceae). Pensoft Publishers. https://doi.org/10.3897/arphapreprints.e138316 Open Access

Hackel, J., Razafimanantsoa, A.H.I., Porcher, V., & **Pirie, M.** 2024, Heathers (*Erica*, Ericaceae) of Madagascar: taxonomy, evolution, ecology and uses. Pensoft Publishers. https://doi.org/10.3897/arphapreprints.e141580 Open Access

Musker, S.D. (2023). Phylogenomic Studies in Heathers (*Erica* L.). Doctoral thesis, University of Bayreuth, Faculty of Biology, Chemistry and Earth Sciences Bayreuth. doi: 10.15495/EPub_UBT_00007069, available at: https://epub.uni-bayreuth.de/id/eprint/7069. Open Access

Musker, S.D., **Pirie**, M.D., & **Nürk**, N.M. (2023). The dynamics of species divergence in South Africa's megadiverse Cape region. *bioRxiv* 549682. doi: 10.1101/2023.07.19.549682. Open Access

Science communication

Pirie, M.D., Eriksson, L.T., Kandziora, M., & **Nürk**, N.M. (2023). In search of Cape Ericas: – research visit to South Africa, October 2022. Årbok for Universitetsmuseet i Bergen 40: 171–186. https://hdl.handle.net/11250/3113124 Open Access

Pirie, M.D., Blackhall-Miles, R., Bourke, G., Crowley, D., Ebrahim, I., Forest, F., Knaack, M., Koopman, R., Lansdowne, A., **Nürk**, N.M., Osborne, J., Pearce, T.R., Rohrauer, D., Smit, M., & Wilman, V. (2022). Botaniske hager går sammen for å stanse utryddelsen av arter: Global Conservation Consortium for *Erica* (lyng). *Årringen 2020–2021* (24–25): 103–110. available at: https://hdl.handle.net/11250/3036012, https://bora.uib.no/bora-xmlui/handle/11250/3036012. Open Access

Bellstedt, D.U., Coetzee, A., Hitchcock, A., Kandziora, M., van der Niet, T., **Nürk**, N.M., **Musker**, S. & **Pirie**, M.D. (2020). Small differences, big secrets. *Veld & Flora* 106(1): 22–31. doi: 10.10520/EJC-1be3924832, available at: https://journals.co.za/doi/epdf/10.10520/EJC-1be3924832.

Infrastructure

Raw reads from **Musker** *et al.* (2024b) are deposited in NCBI Sequence Read Archive under BioProjects PRJNA1030154 (GBS data) and PRJNA1030750 (WGS data). Other associated data including genotype calls (in VCF format), metadata and code are deposited on FigShare, available at https://doi.org/10.25375/uct.25218470. v1.

Raw reads from **Musker** *et al.* (2024a) are deposited in NCBI Sequence Read Archive under BioProjects PRJNA1164706. Other associated data are deposited on FigShare, available at https://doi.org/0.25375/uct.27134208.v1.

Musker, S. D. (2022) TargetVet, available at https://github.com/SethMusker/TargetVet.

Global Conservation Consortium (GCC) for *Erica*: Both **Pirie** and **Nürk** are members of the steering committee of GCC *Erica*, in collaboration with international partners under the auspices of Botanic Gardens Conservation International. Pirie is coordinating the consortium and leading its phylogenomics and taxonomy working group on behalf of the Bergen University Gardens (https://www.globalconservationconsortia.org/gcc/erica/).

References

Anderson EC. 2008. Bayesian inference of species hybrids using multilocus dominant genetic markers. *Philosophical Transactions of the Royal Society B: Biological Sciences* **363**: 2841–2850. doi: 10.1098/rstb.2008.0043

Bellstedt DU, Coetzee A, Hitchcock A, Kandziora M, van der Niet T, Nürk NM, Musker SD, Pirie MD. 2020. Small differences, big secrets. *Veld & Flora* **106**: 22–31. doi: 10.10520/EJC-1be3924832

Bellstedt DU, Pirie MD, Visser JC, de Villiers MJ, Gehrke B. 2010. A rapid and inexpensive method for the direct PCR amplification of DNA from plants. *American Journal of Botany* **97**: e65-e68. doi: 10.3732/ajb.1000181

Berger SA, Krompass D, Stamatakis A. 2011. Performance, accuracy, and web server for evolutionary placement of short sequence reads under maximum likelihood. *Systematic Biology* **60**: 291–302. doi: 10.1093/sysbio/syr010

Bhatia G, Patterson N, Sankararaman S, Price AL. 2013. Estimating and interpreting FST: The impact of rare variants. *Genome Research* **23**: 1514–1521. doi: 10.1101/gr.154831.113

Elliot AC, Bester SP, Klopper RR, Nelson EC, Pirie MD. 2024. Curating an online checklist for *Erica* L. (Ericaceae): contributing to and supporting global conservation through the World Flora Online. *PhytoKeys* 243: 121–135. doi: 10.3897/phytokeys.243.121555

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS One* **6**: e19379. doi: 10.1371/journal.pone.0019379.s004

Giarla TC, Esselstyn JA. 2015. The challenges of resolving a rapid, recent radiation: Empirical and simulated phylogenomics of Philippine shrews. *Systematic Biology* **64**: 727–740. doi: 10.1093/sysbio/syv029

Hackel J, Razafimanantsoa AHI, Porcher V, Pirie, M. 2024. Heathers (*Erica*, Ericaceae) of Madagascar: taxonomy, evolution, ecology and uses. Pensoft Publishers. https://doi.org/10.3897/arphapreprints.e141580

Humboldt A. 1846. *Cosmos: Essai d'une description physique du monde*. Paris: Sirou et Desquers

Humboldt A, Bonpland A. 1805. *Essai sur la géographie des plantes: Accompagné d'un tableau physique des régions équinoxiales, fondé sur des mesures exécutées, depuis le dixième degré de latitude boréale jusqu'au dixième degré de latitude australe, pendant les an- nées 1799, 1800, 1801, 1802 et 1803*. Paris: Levrault, Schoell et compagnie

Jiang X, Edwards SV, Liu L. 2020. The multispecies coalescent model outperforms concatenation across diverse phylogenomic data sets. *Systematic Biology* **69**: 795–812. doi: 10.1093/sysbio/syaa008

Kadlec M, Bellstedt DU, Le Maitre NC, Pirie MD. 2017. Targeted NGS for species level phylogenomics: “Made to measure” or “one size fits all”? *PeerJ* **5**: e3569. doi: 10.7717/peerj.3569

Linder HP. 2003. The radiation of the Cape flora, southern Africa. *Biological Reviews* **78**: 597–638. doi: 10.1017/s1464793103006171

Manning JC, Goldblatt P. 2012. Plants of the greater cape floristic region. 1: the Core Cape Flora. *Strelitzia* **29**: South African National Biodiversity Institute, Pretoria

McLay TGB, Birch JL, Gunn BF, Ning W, Tate JA, Nauheimer L, Joyce EM, Simpson L, Schmidt-Lebuhn AN, Baker WJ*, et al.* 2021. New targets acquired: Improving locus recovery from the Angiosperms353 probe set. *Applications in Plant Sciences* **9**: e11420. doi: 10.1002/aps3.11420

Musker SD. 2023. *Phylogenomic studies in heathers (Erica L.).* Doctoral thesis [Dissertation, Dr. rer. nat.], University of Bayreuth, Faculty of Biology, Chemistry and Earth Sciences Bayreuth. doi: 10.15495/EPub_UBT_00007069. Availabe at https://epub.uni-bayreuth.de/id/eprint/7069

Musker SD, Nürk NM. in prep. TargetVet: rapid and user-friendly quality control and paralog detection for phylogenomic data.

Musker SD, Nürk NM, Pirie MD. in prep.-a. Rapid diversification and dynamic floral trait evolution in Cape *Erica* (Ericaceae).

Musker SD, Pirie MD, Nürk NM. 2023. The dynamics of species divergence in South Africa's megadiverse Cape region. *bioRxiv* **549682**. doi: 10.1101/2023.07.19.549682

Musker SD, Pirie MD, Nürk NM. in prep.-b. New and improved target sets for phylogenomics in *Erica* and *Rhododendron* (Ericaceae).

Musker SD, Nürk NM, Pirie MD. 2024a. Worth the fuss? Maximising informativeness for target capture-based phylogenomics in Erica (Ericaceae). Pensoft Publishers. doi: 10.3897/arphapreprints.e138316

Musker SD, Pirie MD, Nürk NM. 2024b. Pollinator shifts despite hybridisation in the Cape’s hyperdiverse heathers (Erica, Ericaceae). *Molecular Ecology* 33: e17505. doi: 10.1111/mec.17505

Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J. 2000. Biodiversity hotspots for conservation priorities. *Nature* **403**: 853–858. doi: 10.1038/35002501

Nelson EC, Oliver EGH, Pirie MD. 2023. *Erica* L. (Ericaceae): homonyms amongst published names for African species and proposed replacement names. *PhytoKeys* 236: 157–178. doi: 10.3897/phytokeys.236.110498

Nelson EC, Pirie MD, Bellstedt DU. 2024. Redefining the megagenus *Erica* L. (Ericaceae): the contributions of E. G. H. Oliver and I. M. Oliver (née Nitzsche) to taxonomy and nomenclature. *PhytoKeys* 244: 39–55. doi: 10.3897/phytokeys.244.121705

Oliver EGH, Forshaw N. 2012. Genus *Erica*: An identification aid. *Contributions to the Bolus Herbarium* **22**: Bolus Herbarium, University of Cape Town. Availabe at http://www.proteaatlas.org.za/ericakey.htm

Oliver EGH, Forshaw N, Oliver IM, Volk F, Schumann AWS, Dorr LJ, Hoekstra RD, Musker SD, Nürk NM, Pirie MD, Rebelo AG. 2024. Genus *Erica*: An identification aid version 4.00. *PhytoKeys* 241: 143–154. doi: 10.3897/phytokeys.241.117604

Peter BM. 2016. Admixture, population structure, and F-statistics. *Genetics* **202**: 1485–1501. doi: 10.1534/genetics.115.183913

Pirie MD, Bellstedt DU, Bouman RW, Fagúndez J, Gehrke B, Kandziora M, Le Maitre NC, Musker SD, Newman E, Nürk NM*, et al.* in prep. Phylogenetic diversity and conservation prioritisation of *Erica* (Ericaceae). *PhytoKeys*

Pirie MD, Bellstedt DU, Bouman RW, Fagúndez J, Gehrke B, Kandziora M, Le Maitre NC, Musker SD, Newman E, Nürk NM, Oliver EGH, Pipins S, Niet T van der, Forest F. 2024. Spatial decoupling of taxon richness, phylogenetic diversity and threat status in the megagenus *Erica* (Ericaceae). *PhytoKeys* 244: 127–150. doi: 10.3897/phytokeys.244.124565

Pirie MD, Blackhall-Miles R, Bourke G, Crowley D, Ebrahim I, Forest F, Knaack M, Koopman R, Lansdowne A, Nürk NM*, et al.* 2022a. Botaniske hager går sammen for å stanse utryddelsen av arter: Global Conservation Consortium for *Erica* (lyng). *Årringen* **2020–2021** 103–110. Availabe at https://hdl.handle.net/11250/3036012

Pirie MD, Blackhall-Miles R, Bourke G, Crowley D, Ebrahim I, Forest F, Knaack M, Koopman R, Lansdowne A, Nürk NM*, et al.* 2022b. Preventing species extinctions: A global conservation consortium for *Erica*. *Plants, People, Planet* **4**: 335–344. doi: 10.1002/ppp3.10266

Pirie MD, Kandziora M, Nürk NM, Le Maitre NC, Mugrabi de Kuppler A, Gehrke B, Oliver EGH, Bellstedt DU. 2019. Leaps and bounds: Geographical and ecological distance constrained the colonisation of the Afrotemperate by *Erica*. *BMC Evolutionary Biology* **19**: 1–12. doi: 10.1186/s12862-019-1545-6

Pirie MD, Oliver EGH, Mugrabi de Kuppler A, Gehrke B, Le Maitre NC, Kandziora M, Bellstedt DU. 2016. The biodiversity hotspot as evolutionary hot-bed: Spectacular radiation of *Erica* in the Cape Floristic Region. *BMC Evolutionary Biology* **16**: 190. doi: 10.1186/s12862-016-0764-3

Portik DM, Streicher JW, Blackburn DC, Moen DS, Hutter CR, Wiens JJ, O'Connell M. 2023. Redefining possible: Combining phylogenomic and supersparse data in frogs. *Molecular Biology and Evolution* **40**. doi: 10.1093/molbev/msad109

Schrempf D, Minh BQ, von Haeseler A, Kosiol C, Saitou N. 2019. Polymorphism-aware species trees with advanced mutation models, bootstrap, and rate heterogeneity. *Molecular Biology and Evolution* **36**: 1294–1301. doi: 10.1093/molbev/msz043

Soza VL, Lindsley D, Waalkes A, Ramage E, Patwardhan RP, Burton JN, Adey A, Kumar A, Qiu R, Shendure J*, et al.* 2019. The *Rhododendron* genome and chromosomal organization provide insight into shared whole-genome duplications across the heath family (Ericaceae). *Genome Biology and Evolution* **11**: 3353–3371. doi: 10.1093/gbe/evz245

Steinig EJ, Neuditschko M, Khatkar MS, Raadsma HW, Zenger KR. 2016. NetView P: A network visualization tool to unravel complex population structure using genome-wide SNPs. *Molecular Ecology Resources* **16**: 216–227. doi: 10.1111/1755-0998.12442

Yang F-S, Nie S, Liu H, Shi T-L, Tian X-C, Zhou S-S, Bao Y-T, Jia K-H, Guo J-F, Zhao W*, et al.* 2020. Chromosome-level genome assembly of a parent species of widely cultivated azaleas. *Nature Communications* **11**: 5269. doi: 10.1038/s41467-020-18771-4

Zhang C, Mirarab S, Takahashi A. 2022. Weighting by gene tree uncertainty improves accuracy of quartet-based species trees. *Molecular Biology and Evolution* **39**: msac215. doi: 10.1093/molbev/msac215

Zhang L-G, Li X-Q, Jin W-T, Liu Y-J, Zhao Y, Rong J, Xiang X-G. 2022. Asymmetric migration dynamics of the tropical Asian and Australasian floras. *Plant Diversity* **00**: 2468–2659. doi: 10.1016/j.pld.2022.05.006