Reassessment of species limits and diversification process in the Cape grass genus *Ehrharta* Thunb.

By Lara Wootton

Dissertation presented in fulfilment of the degree of Master of Science specialising in Biological Sciences

Under supervision by:

Assoc. Prof. G.A. Verboom (Department of Biological Sciences, University of Cape Town)

Dr F. Forest (Royal Botanic Gardens, Kew)

Department of Biological Sciences, University of Cape Town

September 2020

Abstract

The Greater Cape Floristic Region (GCFR) of South Africa has one of the most biodiverse floras in the world. Although ecological speciation is thought to be a primary mechanism behind diversification in the GCFR, it has recently been hypothesised that non-ecological speciation may have an influential role in driving diversification in the montane "sky islands" of the GCFR, potentially resulting in cryptic species. This work seeks to test the relative importance of ecological versus non-ecological processes in powering speciation at different elevations in the GCFR, and to assess the existence of multiple, range-restricted cryptic species at high elevations. The Cape grass genus *Ehrharta* is an ideal system in which to investigate these processes, as it contains low-, mid-, and high-elevation lineages, with a previously documented adaptive radiation in the succulent karoo. Population-level phylogenetic analyses using targeted enrichment sequencing data show that *E. rupestris* and *E. setacea*, as currently defined, are polyphyletic, and reveal multiple distinct monophyletic lineages within the Ramosa, Rehmannii and Setacea clades of genus *Ehrharta*. Analysis of genotyping-by-sequencing and morphological data, together with evidence of sympatry, confirm these lineages to be distinct entities, resulting in 13 to 16 putative new species, of which several can be considered cryptic. The crown node of *Ehrharta* is found to have originated 28 Ma, which substantially antedates previous age estimates. Comparisons of morphological evolutionary rates, as well as rates of nonsynonymous to synonymous sequence evolution (ω), provide little evidence to support the hypothesis that non-ecological processes have been more important at high elevations. Both the low- and high-elevation clades show evidence of divergent ecological selection, with the Lowlands clade exhibiting high functional trait variance, and the high-elevation Setacea clade showing subtle ecological differentiation and accelerated rates of morphological evolution and ω relative to the rest of Cape *Ehrharta*. It is instead

suggested that diversification in the Cape *Ehrharta* is triggered by a combination of intrinsic and extrinsic factors unique to each clade, thus corroborating a growing body of research arguing that it is simplistic to generalise radiations as 'adaptive' or 'non-adaptive'. Instead, mountain radiations require an integrated approach to untangle the subtle interaction of geographic, ecological and biological factors that drive diversification.

Acknowledgements

I would like to thank my supervisors, Tony Verboom and Félix Forest, for comprehensive support and guidance throughout this project, for providing timeous and detailed feedback, and maintaining excellent mentorship even through the challenges of a global pandemic. I also received excellent all-round help from my labmates. In particular, I thank Klara Beckerling, Ruan van Mazijk and Seth Musker for assistance with fieldwork and analyses, and Thaabiet Parker for extensive help with labwork and GBS analysis. At Jodrell Laboratory, Kew, I gratefully thank Robyn Cowen, Penny Malakasi and László Csiba for teaching me, and Alina Höwener for learning alongside of me. I owe a special debt of gratitude to Bruce Murphy, who taught me bioinformatics in less than a week. I would also like to acknowledge the help of UCT support staff, including Boitumelo Marope, Desmond Barnes, Marilyn Krige and Andrew Lewis. Computations were performed using the facilities provided by the University of Cape Town's ICTS High Performance Computing team. I thank the National Research Fund, the South African Association of Botanists, the Dorothy Cameron scholarship and the Vice Chancellor's research scholarship for providing generous personal financial support. I would also like to thank Matthew Greenwood for being my genetics consultant, for assistance with labwork, fieldwork, and for unwavering support. Finally, I am deeply grateful to my family, who have helped and supported me immeasurably throughout.

Chapter 1: Introduction

Accurate species delimitation is important, since it underpins a diverse range of other fields, from conservation biology and ecology to medicine. For example, the categorisation and protection of endangered species is often based on species' ranges and abundance. Incorrect species delimitation may thus result in poor conservation decisions, and may even determine whether management interventions prevent population decline and extinction (Chenuil et al., 2019). Human health and safety also depends on accurate species delimitation, with poor species delimitation reducing our ability to prevent disease and develop new medicines. For instance, recent genetic studies have found that several malaria-carrying mosquito species actually comprise of complexes of species with behavioural differences which warrant unique health interventions (Stevenson et al., 2012; Lobo et al., 2015). Similarly, the taxonomic instability of venomous snake genera hinders our ability both to create anti-venoms and to explore the potential therapeutic properties of snake venom (Wüster, Golay & Warrell, 1996; Carrasco et al., 2016). Species concepts also underpin our evolutionary hypotheses, our understanding of ecological interactions, and our study of ecosystem function. Until recently, it has been thought that 8.7 million eukaryote species exist on the planet (Mora et al., 2011), but the application of genetic techniques for species delimitation suggests that the true number may actually be much greater, predominantly owing to the discovery of cryptic species (Scheffers et al., 2012; Adams et al., 2014; Struck et al., 2018).

Cryptic species are species which are morphologically similar, but genetically distinct (Struck et al., 2018). Often, cryptic species are initially classified as a single taxon, only to be found later to consist of two or more unique species (Bickford et al., 2007; Pfenninger & Schwenk, 2007). Despite similarities in their appearance, cryptic species have been shown to play different functional and ecological roles (De Meester et al., 2016), preferring different microhabitats (Guden et al., 2018),

and having contrasting environmental tolerances (De Meester et al., 2011). Taxonomists have traditionally described species on the basis of morphology, differences in morphology being taken as evidence of reproductive isolation, an important criterion for species delimitation under the Biological Species Concept (Mayr, 1942). However, the genetic distinctness of morphologically similar taxa implies that mate recognition attibutes, such as calls, smells, and colour spectra which we are unable to see, may be equally important in reproductive isolation. In recent years, cryptic species have been discovered in abundance across all taxa and biomes (Beheregaray & Caccone, 2007; Pfenninger & Schwenk, 2007; Janzen et al., 2017). For example, nearly 60% of recently described mammalian species are from cryptic complexes (Ceballos & Ehrlich, 2009), even in charismatic taxa such as giraffes and elephants (Roca et al., 2001; Brown et al., 2007). However, there is a bias in cryptic species research, with many studies being conducted on marine meiofauna, but few in plants and even less in fungi and insects (Bickford et al., 2007).

The origin of cryptic species is evolutionarily intriguing, but understudied. Nonetheless, several different mechanisms that could cause cryptic speciation have been postulated to date. A commonly cited explanation for cryptic species is that such species are young (Knowlton, 1993; Reidenbach et al., 2012). If two sister species have recently diverged, insufficient time may have passed for them to have accumulated noticeable phenotypic differences (Struck & Oliveira, 2019). Alternatively, cryptic species may be generated under convergent or parallel evolution (Tang et al., 2019), where unrelated species evolve a similar morphology due to shared environmental pressures. For example, Swift, Daglio & Dawson (2016) found that jellyfish species from distantly related ancestors converged on the same morphology after moving from a marine environment to a lake environment. Finally, species may undergo stasis, where their morphology remains unchanged for millions of years despite substantial genetic divergence (Wada, Kameda & Chiba, 2013; Struck & Oliveira, 2019). Stasis due to stabilising selection can occur when strong

environmental drivers cause a single phenotype to be strongly advantageous. Morphological stasis may also be a consequence of niche conservatism (Wiens, 2004). Species exhibiting niche conservatism track suitable habitat, rather than adapting to environmental heterogeneity and climate-induced change. Such environmental changes may thus result in habitat fragmentation, population isolation and, ultimately, vicariant speciation (Kozak, Weisrock & Larson, 2006). This process is known as non-ecological speciation because species split not in response to divergent selection, but as a consequence of neutral processes such as genetic drift (Rundell & Price, 2009).

The Greater Cape Floristic Region (GCFR) is a hotspot of biological diversity, containing over 11,500 plant species, in an area of under 200,000 km^2 (Myers et al., 2000; Goldblatt & Manning, 2002; Snijman, 2013). The richness of the Cape flora can be attributed to three main factors. The first is the environmental heterogeneity of the GCFR, which describes variability in both soil characteristics (Cramer et al., 2019) and rainfall regime, as well as a rugged topography, with the Cape Fold mountains extending into the north and east of the GCFR, and rising to a maximum elevation of 2249 m above sea level (Manning & Goldblatt, 2012). This heterogeneity may facilitate speciation both by providing gradients along which ecological speciation can occur (Linder, 1985), and by providing a mosaic of habitats within which differentiation by non-ecological divergence can take place (Verboom et al., 2015). Secondly, a comparatively stable Pleistocene climate is thought to have promoted diversity by reducing extinction (Dynesius & Jansson, 2000), this effect being particularly pronounced in the southwestern GCFR, which shows both the highest overall plant species richness and an over-representation of narrow-range endemics (Cowling & Lombard, 2002; Cowling et al., 2017; Linder, 2019; Wüest et al., 2019; Colville et al., 2020). A third factor potentially influencing the richness of the Cape flora is the onset of increasingly arid and seasonal climate during the late Miocene, this being attributed to a combination of factors, including the development of the Benguela Upwelling system, changes in

Antarctic glaciation and the tectonic uplift along the eastern rim of southern Africa (Siesser, 1980; Zachos et al., 2001; Dupont et al., 2011; Rommerskirchen et al., 2011). By completely transforming the lowland habitats of the Cape (Hoffmann, Verboom & Cotterill, 2015), such change may have created opportunities for adaptive radiation (Richardson et al., 2001; Verboom, Linder & Stock, 2003). However, it may also have prompted isolation of montane habitats, thereby providing a stimulus for non-ecological speciation (Forest et al., 2007; Britton, Hedderson & Verboom, 2014).

The high-elevation region of the GCFR lends itself to non-ecological speciation as it is archipelagic in nature, the mountain peaks functioning as climatically- and geologically-isolated islands surrounded by a 'sea' of lowland habitat. Many of the GCFR species that occur at high-elevations have small ranges, and sister species display strong range exclusivity (Verboom et al., 2015), which is characteristic of species generated through non-ecological speciation (Kozak, Weisrock & Larson, 2006; Boucher, Zimmermann & Conti, 2016; Czekanski-Moir & Rundell, 2019). High-elevation clades in the GCFR have also been shown to diversify within their elevational bands (Verboom et al., 2015), and not solely as a result of the Late-Miocene aridification. Consequently, it has been suggested that the relative importance of ecological and non-ecological speciation processes may vary with elevation in the GCFR, with primarily ecological speciation occurring at low- to mid-elevations, and with non-ecological speciation dominating at high-elevations (Verboom et al., 2015). If this is correct, then low-elevation clades that diversified via adaptive radiations should show higher rates of phenotypic and molecular evolution than high-elevation clades. In addition, high-elevation lineages should exhibit similar morphologies, small ranges, and phylogenetic niche conservatism (Kozak, Weisrock & Larson, 2006; Czekanski-Moir & Rundell, 2019). There are already case studies illustrating the potential prevalence of non-ecological speciation in the montane regions. For example, Britton, Hedderson

& Verboom (2014) found that the high-elevation Cape sedge, *Tetraria triangularis*, consisted of multiple, ecologically similar cryptic species. Similarly, topographically isolated white *Protea* populations exhibit limited gene flow between populations, potentially qualifying as distinct evolutionary species (Prunier & Holsinger, 2010). If non-ecological speciation is indeed the dominant mode of speciation at high-elevations, it is possible that the GCFR is even more diverse than previously thought, with a host of cryptic species potentially waiting to be discovered within the montane regions.

The Cape grass genus *Ehrharta* Thunb. represents an ideal model system within which to test whether speciation process varies with elevation in the GCFR. The genus currently consists of approximately 45 taxa (species and subspecies) and has a Gondwanan distribution, with species occurring in Australasia, New Zealand, Malesia, Reunion, Madagascar and Africa. However, a substantial proportion of *Ehrharta* species (22 in total), are GCFR endemics, comprising of a core low-elevation clade that is subtended by a grade of mid- to high-elevation lineages (Gibbs-Russell & Ellis, 1987; Verboom, Linder & Stock, 2003). This low-elevation clade, which associates primarily with summer-arid renosterveld and succulent karoo scrub habitats, has undergone an adaptive radiation in response to the Late Miocene aridification, and has been interpreted as the primary radiation in *Ehrharta* (Verboom, Linder & Stock, 2003, 2004). Nevertheless, although the mid- and high- elevation lineages of Cape *Ehrharta* comprise just six species in total, four of these species contain multiple subspecies. In addition, field observations suggest that there may be more subtle morphological differentiation in the high-elevation clades than current taxonomy suggests, hinting at the existence of cryptic species.

Therefore, this work takes advantage of the opportunity *Ehrharta* offers, both to assess the existence of multiple, range restricted cryptic species at high elevations, and to test explicitly the

relative potential importance of neutral versus adaptive processes in powering speciation at low and high elevations in the GCFR. In Chapter 2, I present a population-level phylogenetic analysis of the Cape *Ehrharta*, focusing on the mid- and high-elevation clades which are comprehensively sampled across their ranges (Fig. 1). I then employ an integrative taxonomic approach (Dayrat, 2005) to evaluate whether the mid- and high-elevation lineages of *Ehrharta* contain potential cryptic species, on the basis of genome-wide SNP and targeted-enrichment sequencing, morphological and ecological datasets. Chapter 3 presents a time-calibrated tree of the Cape *Ehrharta*, which is used to test whether rates of morphological and molecular evolution differ between clades, therefore expanding our understanding of the role that ecological versus non-ecological speciation plays in generating the biodiversity of GCFR.

References

Adams, M., Raadik, T.A., Burridge, C.P. & Georges, A. 2014. Global biodiversity assessment and hyper-cryptic species complexes: More than one species of elephant in the room? *Systematic Biology*. 63(4):518–533. DOI: 10.1093/sysbio/syu017.

Beheregaray, L.B. & Caccone, A. 2007. Cryptic biodiversity in a changing world. *Journal of Biology*. 6(4):1–5. DOI: doi.org/10.1186/jbiol60.

Bickford, D., Lohman, D.J., Sodhi, N.S., Ng, P.K.L., Meier, R., Winker, K., Ingram, K.K. & Das, I. 2007. Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution*. 22(3):148–155. DOI: 10.1016/j.tree.2006.11.004.

Boucher, F.C., Zimmermann, N.E. & Conti, E. 2016. Allopatric speciation with little niche divergence is common among alpine Primulaceae. *Journal of Biogeography*. 43(3):591–602. DOI:

10.1111/jbi.12652.

Britton, M.N., Hedderson, T.A. & Verboom, G.A. 2014. Topography as a driver of cryptic speciation in the high-elevation Cape sedge *Tetraria triangularis* (Boeck.) C. B. Clarke (Cyperaceae: Schoeneae). *Molecular Phylogenetics and Evolution*. 77:96–109. DOI: 10.1016/j.ympev.2014.03.024.

Brown, D.M., Brenneman, R.A., Koepfli, K., Pollinger, J.P., Milá, B., Georgiadis, N.J., Louis, E.E., Grether, G.F., et al. 2007. Extensive population genetic structure in the giraffe. *BMC biology*. 5(1):57.

Carrasco, P.A., Venegas, P.J., Chaparro, J.C. & Scrocchi, G.J. 2016. Nomenclatural instability in the venomous snakes of the *Bothrops* complex: Implications in toxinology and public health. *Toxicon*. 119:122–128. DOI: 10.1016/j.toxicon.2016.05.014.

Ceballos, G. & Ehrlich, P.R. 2009. Discoveries of new mammal species and their implications for conservation and ecosystem services. *Proceedings of the National Academy of Sciences*. 106(10):3841–3846.

Chenuil, A., Cahill, A.E., Délémontey, N., du Luc, E.D.S. & Fanton, H. 2019. Problems and questions posed by cryptic species. A framework to guide future studies. In *From assessing to conserving biodiversity*. ed. Springer, Cham. 77–106.

Colville, J.F., Beale, C.M., Forest, F., Altwegg, R., Huntley, B. & Cowling, R.M. 2020. Plant richness, turnover, and evolutionary diversity track gradients of stability and ecological opportunity in a megadiversity center. *Proceedings of the National Academy of Sciences*. 117(33):20027–20037.

Cowling, R.M., Bradshaw, P.L., Colville, J.F. & Forest, F. 2017. Levyns' Law: Explaining the evolution of a remarkable longitudinal gradient in Cape plant diversity. *Transactions of the Royal Society of South Africa*. 72(2):184–201.

Cowling, R. & Lombard, A. 2002. Heterogeneity, speciation/extinction history and climate: Explaining regional plant diversity patterns in the Cape Floristic Region. *Diversity and Distributions*. 8(3):163–179.

Cramer, M.D., Wootton, L.M., van Mazijk, R. & Verboom, G.A. 2019. New regionally modelled soil layers improve prediction of vegetation type relative to that based on global soil models. *Diversity and Distributions*. 25(11):1736–1750.

Czekanski-Moir, J.E. & Rundell, R.J. 2019. The ecology of nonecological speciation and nonadaptive radiations. *Trends in ecology & evolution*. 34(5):400–415.

Dayrat, B. 2005. Towards integrative taxonomy. *Biological Journal of the Linnean Society*. 85(3):407–415. DOI: 10.1111/j.1095-8312.2005.00503.x.

De Meester, N., Derycke, S., Bonte, D. & Moens, T. 2011. Salinity effects on the coexistence of cryptic species: A case study on marine nematodes. *Marine Biology*. 158(12):2717–2726.

De Meester, N., Gingold, R., Rigaux, A., Derycke, S. & Moens, T. 2016. Cryptic diversity and ecosystem functioning: A complex tale of differential effects on decomposition. *Oecologia*. 182(2):559–571.

Dupont, L.M., Linder, H.P., Rommerskirchen, F. & Schefuß, E. 2011. Climate-driven rampant speciation of the Cape flora. *Journal of Biogeography*. 38(6):1059–1068.

Dynesius, M. & Jansson, R. 2000. Evolutionary consequences of changes in species' geographical distributions driven by Milankovitch climate oscillations. *Proceedings of the National Academy of Sciences*. 97(16):9115–9120.

Forest, F., Grenyer, R., Rouget, M., Davies, T.J., Cowling, R.M., Faith, D.P., Balmford, A., Manning, J.C., et al. 2007. Preserving the evolutionary potential of floras in biodiversity hotspots. *Nature*. 445(7129):757–760.

Gibbs-Russell, G.E. & Ellis, R.P. 1987. Species groups in the genus *Ehrharta* (Poaceae) in southern Africa. *Bothalia*. 17(1):51–65.

Goldblatt, P. & Manning, J.C. 2002. Plant diversity of the Cape region of Southern Africa. *Annals of the Missouri Botanical Garden*. 89(2):281–302. DOI: 10.2307/3298566.

Guden, R.M., Vafeiadou, A., De Meester, N., Derycke, S. & Moens, T. 2018. Living apart-together: Microhabitat differentiation of cryptic nematode species in a saltmarsh habitat. *PloS one*. 13(9):e0204750.

Hoffmann, V., Verboom, G.A. & Cotterill, F.P.D. 2015. Dated plant phylogenies resolve Neogene climate and landscape evolution in the Cape Floristic Region. *PLoS One*. 10(9):e0137847.

Janzen, D.H., Burns, J.M., Cong, Q., Hallwachs, W., Dapkey, T., Manjunath, R., Hajibabaei, M., Hebert, P.D., et al. 2017. Nuclear genomes distinguish cryptic species suggested by their DNA barcodes and ecology. *Proceedings of the National Academy of Sciences*. 114(31):8313–8318.

Knowlton, N. 1993. Sibling species in the sea. *Annual Review of Ecology and Systematics*. 24(1):189–216.

Kozak, K.H., Weisrock, D.W. & Larson, A. 2006. Rapid lineage accumulation in a non-adaptive radiation: Phylogenetic analysis of diversification rates in eastern North American woodland salamanders (Plethodontidae: *Plethodon*). *Proceedings of the Royal Society B: Biological Sciences*. 273(1586):539–546. DOI: 10.1098/rspb.2005.3326.

Linder, H.P. 1985. Gene flow, speciation, and species diversity patterns in a species-rich area: The Cape Flora. *Species and speciation*. 4:53–7.

Linder, H.P. 2019. Rare species, Restionaceae, and the Cape flora. *Journal of Biogeography*. 46(12):2637–2650.

Lobo, N.F., Laurent, B.S., Sikaala, C.H., Hamainza, B., Chanda, J., Chinula, D., Krishnankutty, S.M., Mueller, J.D., et al. 2015. Unexpected diversity of *Anopheles* species in Eastern Zambia: Implications for evaluating vector behavior and interventions using molecular tools. *Scientific reports*. 5(1):17952. DOI: 10.1038/srep17952.

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

Mayr, E. 1942. *Systematics and the origin of species*. 1st ed. New York: Columbia University Press.

Mora, C., Tittensor, D.P., Adl, S., Simpson, A.G.B. & Worm, B. 2011. How many species are there on earth and in the ocean? *PLoS biology*. 9(8):1–8. DOI: 10.1371/journal.pbio.1001127.

Myers, N., Mittermeier, R.A., Mittermeier, C.G., Fonseca, G.A.B. da & Kent, J. 2000. Biodiversity hotspots for conservation priorities. *Nature*. 403:853–858. DOI: https://doi.org/10.1038/35002501.

Pfenninger, M. & Schwenk, K. 2007. Cryptic animal species are homogeneously distributed among taxa and biogeographical regions. *BMC evolutionary biology*. 7(1):121–127. DOI: 10.1186/1471-2148-7-121.

Prunier, R. & Holsinger, K.E. 2010. Was it an explosion? Using population genetics to explore the dynamics of a recent radiation within *Protea* (Proteaceae L.). *Molecular ecology*. 19(18):3968–3980.

Reidenbach, K.R., Neafsey, D.E., Costantini, C., Sagnon, N., Simard, F., Ragland, G.J., Egan, S.P., Feder, J.L., et al. 2012. Patterns of genomic differentiation between ecologically differentiated M and S forms of *Anopheles gambiae* in West and Central Africa. *Genome biology and evolution*. 4(12):1202–1212.

Richardson, J.E., Weitz, F.M., Fay, M.F., Cronk, Q.C., Linder, H.P., Reeves, G. & Chase, M.W. 2001. Rapid and recent origin of species richness in the Cape flora of South Africa. *Nature*. 412(6843):181–183.

Roca, A.L., Georgiadis, N., Pecon-Slattery, J. & O'Brien, S.J. 2001. Genetic evidence for two species of elephant in Africa. *Science*. 293(5534):1473–1477.

Rommerskirchen, F., Condon, T., Mollenhauer, G., Dupont, L. & Schefuß, E. 2011. Miocene to Pliocene development of surface and subsurface temperatures in the Benguela Current system. *Paleoceanography*. 26(3).

Rundell, R.J. & Price, T.D. 2009. Adaptive radiation, nonadaptive radiation, ecological speciation and nonecological speciation. *Trends in Ecology & Evolution*. 24(7):394–399. DOI: 10.1016/j.tree.2009.02.007.

Scheffers, B.R., Joppa, L.N., Pimm, S.L. & Laurance, W.F. 2012. What we know and don't know about Earth's missing biodiversity. *Trends in ecology & evolution*. 27(9):501–510. DOI: 10.1016/j.tree.2012.05.008.

Siesser, W.G. 1980. Late Miocene origin of the Benguela upswelling system off northern Namibia. *Science*. 208(4441):283–285.

Snijman, D.A. 2013. *Plants of the Greater Cape Floristic Region, vol. 2: The Extra Cape flora.* 1st ed. Strelitzia 29. South African National Biodiversity Institute, Pretoria.

Stevenson, J., St Laurent, B., Lobo, N.F., Cooke, M.K., Kahindi, S.C., Oriango, R.M., Harbach, R.E., Cox, J., et al. 2012. Novel vectors of malaria parasites in the western highlands of Kenya. *Emerging infectious diseases*. 18(9):1547–1549. DOI: 10.3201/eid1809.120283.

Struck, T.H., Feder, J.L., Bendiksby, M., Birkeland, S., Cerca, J., Gusarov, V.I., Kistenich, S., Larsson, K.-H., et al. 2018. Finding evolutionary processes hidden in cryptic species. *Trends in Ecology & Evolution*. 33(3):153–163. DOI: 10.1016/j.tree.2017.11.007.

Struck, T.H. & Oliveira, J.C.D. 2019. Cryptic species and their evolutionary significance. *eLS*. Available: http://hdl.handle.net/10852/76096.

Swift, H.F., Daglio, L.G. & Dawson, M.N. 2016. Three routes to crypsis: Stasis, convergence, and parallelism in the *Mastigias* species complex (Scyphozoa, Rhizostomeae). *Molecular Phylogenetics and Evolution*. 99:103–115. DOI: 10.1016/j.ympev.2016.02.013.

Tang, Y., Li, C., Wanghe, K., Feng, C., Tong, C., Tian, F. & Zhao, K. 2019. Convergent evolution misled taxonomy in schizothoracine fishes (Cypriniformes: Cyprinidae). *Molecular Phylogenetics and Evolution*. 134:323–337. DOI: 10.1016/j.ympev.2019.01.008.

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 Phytologist*. 207(2):368–376. DOI: 10.1111/nph.13342.

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(5):1008–1021. DOI: 10.1111/j.0014-3820.2003.tb00312.x.

Verboom, G.A., Linder, H.P. & Stock, W.D. 2004. Testing the adaptive nature of radiation: Growth form and life history divergence in the African grass genus *Ehrharta* (Poaceae: Ehrhartoideae). *American Journal of Botany*. 91(9):1364–1370.

Wada, S., Kameda, Y. & Chiba, S. 2013. Long-term stasis and short-term divergence in the phenotypes of microsnails on oceanic islands. *Molecular Ecology*. 22(18):4801–4810. DOI: 10.1111/mec.12427.

Wiens, J.J. 2004. Speciation and ecology revisited: Phylogenetic niche conservatism and the origin of species. *Evolution*. 58(1):193–197. DOI: 10.1554/03-447.

Wüest, R.O., Boucher, F.C., Bouchenak-Khelladi, Y., Karger, D.N. & Linder, H.P. 2019. Dissecting biodiversity in a global hotspot: Uneven dynamics of immigration and diversification within the Cape Floristic Region of South Africa. *Journal of Biogeography*. 46(9):1936–1947.

Wüster, W., Golay, P. & Warrell, D.A. 1996. Synopsis of recent developments in venomous snake systematics. *Toxicon*. 35(3):319–340. DOI: 10.1016/S0041-0101(96)00152-3.

Zachos, J., Pagani, M., Sloan, L., Thomas, E. & Billups, K. 2001. Trends, rhythms, and aberrations in global climate 65 ma to present. *Science*. 292(5517):686–693.

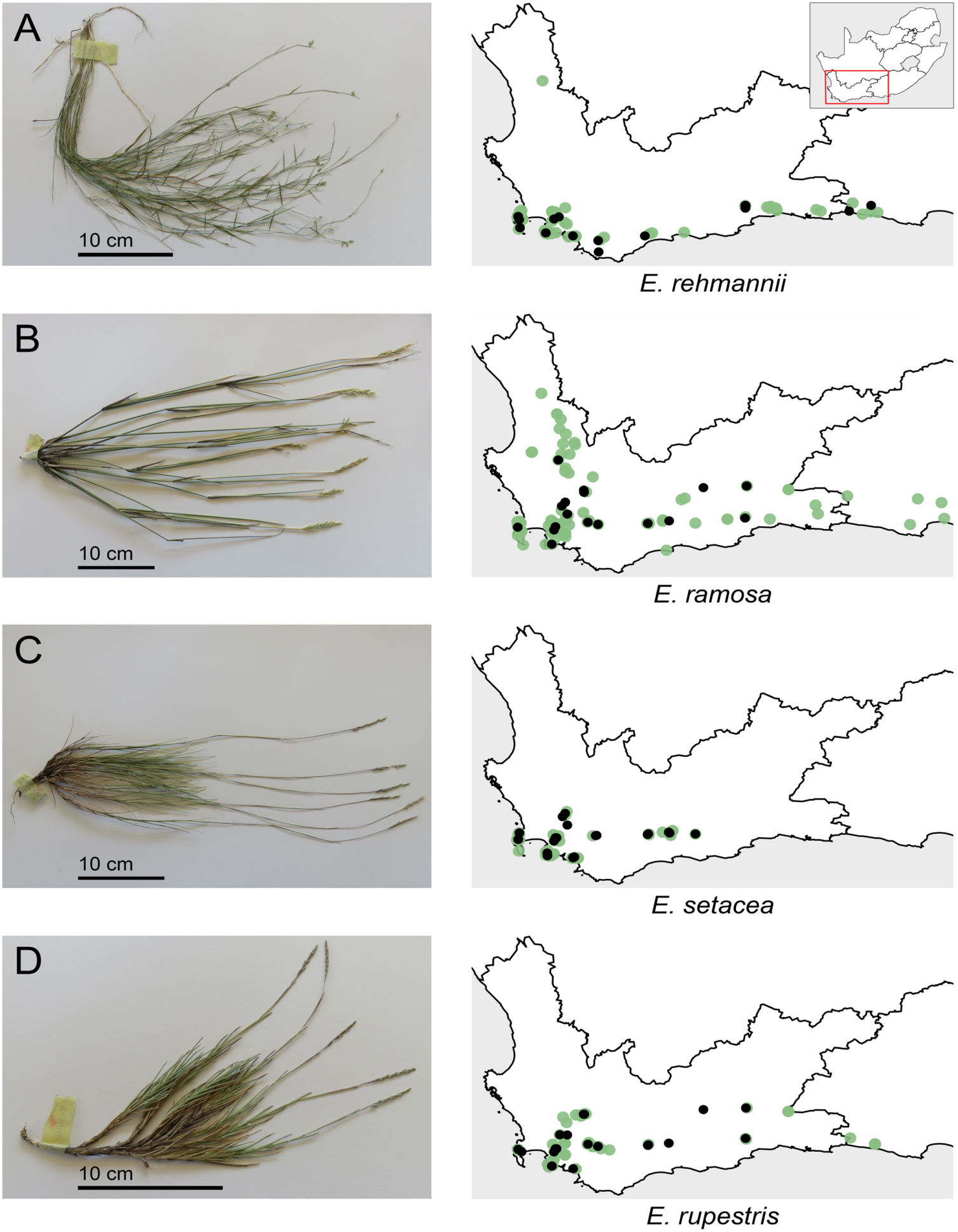

Figure 1. Species distributions based on herbarium specimens (green points, Hoffmann et al., 2015) and the localities sampled in the study (black points) for the four mid- and high-elevation species that are the focus of this work. A) *E. rehmannii*, illustrated by *E. rehmannii* subsp. *filiformis*, B) *E. ramosa*, illustrated by *E. ramosa* subsp. *aphylla*, C) *E. setacea*, illustrated by *E. setacea* subsp. *setacea*, D) *E. rupestris*, illustrated by *E. rupestris* subsp. *rupestris*. Inset shows map of South Africa with the study area delineated in red.

Chapter 2: Species limits in the mid- and high-elevation clades of Cape *Ehrharta*

Introduction

The definition of species has been a controversial topic in biodiversity science, largely because there is no single criterion that satisfactorily defines a species across all living organisms (Mayden, 1997; de Queiroz, 1998, 2005). De Queiroz (2007) realised that the dissonance between species definitions was due to the speciation process itself. As a lineage bifurcates, it accumulates differences in morphology, genetic composition, ecology and degree of reproductive isolation. However, depending on the taxon, the order in which these features appear can vary, with some features occasionally not arising at all. This creates a grey zone in the speciation process, in which it can be ambiguous as to whether or not incipient species have effectively differentiated (de Queiroz, 1998). While most contemporary species concepts accept that a species is an independently evolving lineage (de Queiroz, 2007), determining when speciation has occurred during the gradual divergence of two lineages can often be difficult.

It has become clear that the heterogeneity in the speciation process renders the use of single criteria to delimit species insufficient (Wiley, 1978; Frost & Kluge, 1994; de Queiroz, 2005; Padial et al., 2009). Therefore, a new holistic paradigm of taxonomy, termed integrative taxonomy (Dayrat, 2005; Padial et al., 2010), has been proposed to inform more robust species delimitation. Integrative taxonomy emphasises that multiple lines of evidence, which can include genetic, morphological, phylogenetic and ecological information, should be examined before delimiting and describing a species (Dayrat, 2005; Padial et al., 2010). Despite the extensive implementation of

Dayrat's (2005) integrative taxonomic approach (Dayrat (2005) has been cited over 1400 times to date), and the application of similar integrative approaches by numerous others, morphology remains one of the most important criteria for species delimitation, especially for plant taxa (Rouhan & Gaudeul, 2014; Treurnicht et al., 2017). Cryptic plant species therefore present a challenge to delimit due to the absence of defining morphological features. Molecular barcoding was initially touted as a solution to circumscribe cryptic species, and was proven to be efficient in many cases, especially in animals (Hebert et al., 2004). However, species delimitation using barcoding has several issues, including taxonomic decisions made on arbitrary genetic differences, the inappropriate resolution of barcoding genes for certain taxa, and difficulties of linking new DNA sequences to existing type specimens (Tautz et al., 2003; Moritz & Cicero, 2004; DeSalle, Egan & Siddall, 2005; Goulding & Dayrat, 2016). Furthermore, there are concerns that over-reliance on genetic data will lead to taxonomic inflation (Isaac, Mallet & Mace, 2004; Tattersall, 2007; Zachos et al., 2013; Coates, Byrne & Moritz, 2018), even though the use of genetic information has also led to a reduction of species number in some groups (Burbrink, Lawson & Slowinski, 2000; Devey et al., 2008).

Several recent studies have, therefore, employed integrative taxonomic protocols for the delimitation of cryptic species, and these share several common features (Barley et al., 2013; Jörger & Schrödl, 2013; Fišer, Robinson & Malard, 2018; Singhal et al., 2018; Chaplin et al., 2020). Firstly, all of these protocols acknowledge that extensive population sampling along the entirety of a species complex's range is needed to get a clear idea of genetic relationships, gene flow and geographic range limits (Bernardo, 2011). This prevents the inference of spurious splits between populations which, despite appearing genetically isolated, actually represent points along a continuum of genetic variation (Mason et al., 2020). Next, strong genetic differentiation needs to be demonstrated between putative cryptic species. This differentiation should be shown using

data that incorporate independently evolving regions of the genome, to ensure that variation in gene histories from different loci is represented (Fujita et al., 2012). Commonly used ways of testing genetic differentiation include distance-based methods, where species are delimited only if intraspecific genetic variation is less than interspecific variation (Puillandre et al., 2012), tree-based methods (e.g. Zhang et al., 2018), and coalescent methods (Leaché et al., 2014). Ideally, more than one of these approaches should be used to evaluate hypothesised species boundaries. Following genetic analysis, morphology should be re-examined for the purpose of identifying potentially diagnostic, although subtle, trait differences that corroborate the genetic findings (Barley et al., 2013; Chaplin et al., 2020). Finally, if possible given the distribution of the species complex, reproductive isolation between taxa should ideally be demonstrated through the absence of gene flow between sympatric or parapatric populations (Singhal et al., 2018).

By reducing the reliance on morphology, integrative protocols provide a framework for reassessing the status of infraspecific taxa such as subspecies, especially those described before genetic data was widely available. Subspecies are typically defined as a collection of morphologically variant, and geographically distinct populations (Mayr, 1982), a definition which makes their evolutionary status unclear, resulting in several interpretations (Hamilton & Reichard, 1992). For example, subspecies can be conceptualised as having no evolutionary status, and can be understood merely as an attempt to describe clinal variation in a genetically coherent species (Amadon, 1949; Patten & Unitt, 2002). More commonly, subspecies are interpreted as incipient species (Mallet, 2007), where the rank is considered a stage in the evolutionary divergence of a broader species (Tattersall, 2007; Hawlitschek, Nagy & Glaw, 2012). A further possibility is that subspecies represent evolutionarily-distinct species which have not been recognised as such, on the basis of morphological indistinctness or insufficient data (Manier, 2004). By considering genetic data, in conjunction with other lines of evidence, an integrative taxonomic approach allows for

discrimination between these alternative scenarios.

The most recent infrageneric taxonomy of the Cape *Ehrharta* is that of Gibbs-Russell & Ellis (1987), who divided the genus into seven informal species groups. Most of these groups occur predominately in the arid, lower-elevation, renosterveld and succulent karoo vegetation regions of the GCFR. Within these lowlands groups, there is little taxonomic confusion as the species are clearly morphologically distinct, with characters that easily distinguish the species. Conversely, within two of the three higher-elevation, fynbos-restricted groups, Setacea and Ramosa, the species limits are less clear. The Setacea group consists of two species, *E. setacea* and *E. rupestris*, the former comprising four subspecies (*disticha, setacea, uniflora* and *scabra*), while the latter contains three (*rupestris, tricostata* and *dodii*). As noted by Gibbs-Russell & Ellis (1987), these species and subspecies are morphologically similar and difficult to distinguish. While the subspecies in both groups tend to have distinct ranges, many of them occur in sympatry in the Caledon area, where intermediates also occur. Species limits are even less clear within the Ramosa group, which contains two species, *E. ramosa* with two subspecies (*aphylla* and *ramosa*), and *E. rehmannii* with three subspecies (*rehmannii, filiformis* and *subspicata*). Gibbs-Russell & Ellis (1989) observe that "the five taxa comprising the Ramosa group intergrade and few consistently reliable characters adequately separate the species and subspecies". As these taxa were delimited using only morphological and anatomical characters, it is possible that the application of genetic data may bring clarity to species boundaries in this group.

The archipelagic character of the Cape high-elevation zone lends itself to non-ecological vicariant or peripatric speciation, with the consequence that this zone potentially harbours many undiscovered cryptic species (Britton, Hedderson & Verboom, 2014; Verboom et al., 2015). The high- and mid-elevation Setacea and Ramosa groups of *Ehrharta* are strong candidates for cryptic

speciation, since their distributions encompass multiple mountain ranges between which the presence of deep valleys may restrict gene flow. Moreover, they contain multiple taxonomic entities which, by virtue of their weak morphological differentiation and relatively distinct geographical distribution, are currently conceptualised as subspecies, but may potentially represent distinct species. In this chapter, therefore, I employ an integrative taxonomic approach (Dayrat, 2005) to test the hypothesis that subspecies within the mid- to high-elevation Ramosa and Setacea groups represent cryptic species that constitute distinct evolutionary lineages. In order to evaluate the monophyly of these groups and so provide context for their segregation into species, I first present a multilocus phylogenetic tree of the Cape *Ehrharta* based on genome-wide targeted enrichment data (Lemmon, Emme & Lemmon, 2012), with population-level sampling of the Setacea and Ramosa groups. In addition to insights obtained from these data, I employ the variation in single nucleotide polymorphisms (SNPs) generated through genotyping-by-sequencing (Elshire et al., 2011), and in morphological traits to assess the evolutionary distinctiveness of putative species within these groups.

Methods

Field sampling

Field sampling was conducted between August 2019 and January 2020. For taxa in the Ramosa and Setacea groups, the number of sampled localities per taxon varied between two, for the range restricted *E. setacea* subsp. *disticha* and *E. setacea* subsp. *uniflora*, and twelve for the widespread *E. ramosa* subsp. *ramosa*. Between four and six individuals were sampled from each locality, with the sampled individuals being selected as far as possible to be >30 m apart. For taxa included in the remaining species groups, the number of sampled localities varied between one and two, with a single individual being sampled from each locality. Young leaf and culm tissue was collected from each accession, and dried in silica, before being stored in an airtight container at -20°C to prevent DNA degradation. In addition, a flowering specimen of each accession was pressed as a voucher, to be deposited at the Bolus Herbarium at the University of Cape Town. Finally, *Microlaena stipoides* and *Tetrarrhena laevis* were sampled from the living collection and tissue bank of the Royal Botanic Gardens, Kew, respectively.

DNA extraction

DNA extraction was carried out at the Systematics Laboratory at the Department of Biological Sciences, University of Cape Town, with the exception of *M. stipoides* and *T. laevis*, which were extracted at the Royal Botanic Gardens, Kew. For each sample, 100 mg of frozen, silica-dried leaf and culm tissue was finely ground in liquid nitrogen using a mortar and pestle. Sterilised silica and polyvinyl pyrrolidone was added to each sample before grinding, to increase friction and decrease polysaccharide extraction respectively. DNA samples were then extracted in 1.5 ml

Eppendorf tubes using the cetyl-trimethylammonium bromide (CTAB) method (Doyle & Doyle, 1987) with three 70% ethanol washes. The resulting extracts were treated with RNAse A (Omega Bio-tek, Norcross, GA, USA), and purified using AMPure XP paramagnetic beads (Beckman Coulter, High Wycombe, UK) following the manufacturers' instructions. The concentration and purity of each extract was checked using a NanoDrop ND-1000 spectrophotometer (ThermoFisher Scientific, Waltham, MA, USA). Samples with concentrations 50 ng μl^{-1}, and 260/280 nm absorbance ratios between 1.7 and 2, were considered suitable for genotyping-by-sequencing (GBS), while those with concentrations 25 ng μl^{-1} were considered suitable for targeted enrichment sequencing.

Targeted enrichment sequencing

Targeted enrichment sequencing was conducted in order to infer a multilocus phylogenetic hypothesis for the Cape *Ehrharta* (Lemmon, Emme & Lemmon, 2012). In general, each sampled population was represented by a single accession, with few exceptions where two accessions from a population were sequenced to ratify quality between sequencing runs. In total, there were 106 Cape *Ehrharta* samples, and one sample from each of *M. stipoides* and *T. laevis*. Library preparation and hybridisation for targeted enrichment sequencing was carried out at the Sackler Phylogenomic Laboratory, within the Jodrell Laboratory at the Royal Botanic Gardens, Kew (Richmond, Surrey, UK). The DNA concentration present in the extracts was measured with a QuantusTM fluorometer (Promega corporation, Madison USA) using the using QuantiFluorR dsDNA System kit (Promega corporation, Madison USA). The samples were diluted to 200 ng DNA in 26 µl with Milli-Q water. The DNA was then sheared to fragments with an average of length of 350 bp by sonication for 75 s with a Covaris Focused-ultrasonicator M220 (Woburn,

Massachusetts 01801, www.covarisinc.com)

Library preparation was performed using NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs), using half the recommended volumes to maximise use of reagents and indices. AMPure paramagnetic beads (Beckman Coulter, High Wycombe, UK) were used for size selection of DNA fragments. Libraries were indexed with the NEBNext Multiplex Oligos for Illumina (Dual Index Primer Set 1, New England Biolabs), and enriched with the following thermocycler conditions: initial denaturation for 30s at 98°C, seven cycles comprising denaturation for 10 s at 98°C and annealing and extension for 75 s at 65°C, followed by a final extension for 5 mins at 65°C. The resulting enriched libraries were then cleaned with AMPure paramagnetic beads (Beckman Coulter, High Wycombe, UK), and the quality of the libraries were checked using Agilent Technologies 4200 Tapestation (Agilent Technologies, Palo Alto, California, USA). A library was considered good quality if it had a concentration of greater than 4 ng μ^{-1} and average fragment size of 450 bp to 500 bp. If a library failed the quality check, it was either re-amplified, underwent size selection again, or was entirely repeated.

Libraries were normalised to a concentration of 10 nM, and 10 µl per sample was pooled into batches of 12 or 24 samples per reaction. The pools were vacuum dried using a miVac Duo Concentrator (Genevac, UK) and subsequently eluted in 8 µl of 10 mM tris buffer. The pools were enriched using the Angiosperms-353 v.1 target capture kit (Johnson et al., 2019), following the manufacturers instructions (Arbor Biosciences, Ann Arbor, Michigan, USA). Hybridisation took place for 16 hrs at 65°C, and the resulting product was amplified with 12 cycles of PCR. The reaction quality was checked using an Agilent Technologies 4200 Tapestation (Agilent Technologies, Palo Alto, California, USA). Reactions were normalised to 4mM and pooled. Sequencing took place either on a Illumina® HiSeq at Macrogen (Seoul, South Korea) or on an

Illumina® MiSeq at Royal Botanic Gardens, Kew, producing 2 x 150 bp reads.

Read trimming, filtering and assembly

The raw reads were cleaned using Trimmomatic v.0.39 (Bolger, Lohse & Usadel, 2014). Illuminaclip palindrome mode, with the reference file TruSeq3-PE.fa, was used to remove Illumina sequencing adaptors. Palindrome mode had the following settings: SeedMismatches = 1, palindromeClipThreshold = 30, simpleClipThreshold = 6, minAdapterLength = 2; keepBothReads = TRUE. Poor quality reads were then trimmed with the Maxinfo option. Maxinfo balances the utility of retaining reads of an informative length against the costs of retaining incorrect bases. Initially, Maxinfo is lenient towards potential error calls at a shorter read length, becoming stricter as read length increases. The settings used were targetLength = 40 and strictness = 85. Finally, reads of <36 bp were removed using MinLength. Overall read quality and adaptor content of the remaining reads were checked using FastQC v.0.11.8 (Andrews, 2010).

Reads were assembled using the pipeline HybPiper v.1.3.1 (Johnson et al., 2016). HybPiper consists of a series of python scripts that use SPAdes (Bankevich et al., 2012) and other bioinfomatic tools to assemble reads to a target file. The pipeline first sorts and groups together individual reads for a gene by mapping each read to the gene sequence in the target file. Once the reads are sorted by gene, they are *de novo* assembled using SPAdes, and then realigned to the reference sequence. If multiple contigs are assembled per gene, the contig with either the greatest depth or greatest similarity to the reference file is chosen for the consensus sequence. HybPiper gives the option of two different mapping algorithms, BLASTX (Altschul et al., 1990) and BWA (Li & Durbin, 2010). I ran the pipeline using BLASTX, as it is a less strict algorithm than BWA, and so maximises the number of reads mapped and the lengths of the final sequences. A coverage

of four reads was considered sufficient to generate a consensus sequence. Reads were mapped to a multi-species, amino acid target file matching the probes from the Angiosperms-353 baitset (Johnson et al., 2019). In the case of a multi-species target file, such as the one used here, HybPiper can determine which target sequence for a gene is the most suitable given the sample. The Hybpiper pipeline was also used to assemble the cleaned reads of a full genome of an *Oryza sativa* Japonica cultivar downloaded from Genbank (SRX7089781).

HypPiper was also used to identify potential paralogs. Where HybPiper recovers two or more contigs that cover >85% of the coding region's reference sequence, it flags the region as containing potential paralogs. The presence of two contigs of similar length at a gene locus may, however, alternatively indicate the presence of multiple alleles, a situation which can be differentiated using gene tree inference. In these analyses, HybPiper flagged 50 genes with paralog warnings. Within each gene, the number of samples flagged with paralogs varied between one and 39 samples. I disregarded genes (N = 32) with two or fewer flagged samples, as these were more likely to be alleles rather than paralogs. For the remaining 18 genes, the script paralog_retriever.py was used to extract the main and secondary contigs for a gene into a fasta file, which was then aligned with MAFFT v.7.427 (Katoh & Standley, 2013), followed by gene tree construction using IQTREE v.1.6.11 (Nguyen et al., 2015). The resulting gene trees were inspected in Geneious (https://www.geneious.com) in order to assess the presence of paralogs. On this basis, a total of 11 genes were deemed paralogous or potentially paralogous, and were removed from the dataset.

Phylogenetic inference

Two sets of loci were selected for phylogenetic analysis. The first, referred to as the "full coverage" dataset, comprised genes having sequences for all 106 Cape *Ehrharta* accessions, *M. stipoides* and

T. laevis and *O. sativa*. This dataset comprised a total of 235 genes. The second dataset, referred to as the "90% coverage" dataset, contained genes that were present in at least 90% of the samples, the cut off selected as a compromise between data missingness and the desire to use as many genes as possible. This dataset comprised a total of 307 genes. Both datasets yielded exons, introns, and supercontigs (comprising both exons and introns), giving six datasets in total. Individual genes were aligned using MAFFT with the settings –localpair –maxiterate 1000, the resulting alignments then being trimmed of gappy and poorly aligned regions with TrimAL v.1.2 using the -automated1 option (Capella-Gutiérrez, Silla-Martínez & Gabaldón, 2009). Alignment statistics were calculated using AMAS (Borowiec, 2016).

The individual gene alignments were concatenated using AMAS (Borowiec, 2016), before being subjected to mixed-model, maximum likelihood (ML) phylogenetic analysis. For this purpose, the optimal model of sequence evolution for each locus was determined using IQTREE. Loci sharing a common model of sequence evolution were also identified and merged using IQTREE, which can implement Lanfear's Partitionfinder2 (Lanfear et al., 2012; Chernomor, Haeseler & Minh, 2016). Due to the high computational burden of evaluating partitioning schemes, the relaxed hierarchical clustering algorithm (-rcluster) (Lanfear et al., 2014), which examines only the top n% partition merging schemes, was used. The exon supermatrices were partitioned using -rcluster 10, while the intron and supercontig supermatrices were run with -rcluster 2 due to computational time limits. The locus limits were used as the initial partitions. I then reconstructed ML trees on the merged-partition datasets using IQTREE. Node support was estimated using 1000 ultrafast bootstrap replicates (-bb 1000, UFBoot, Hoang et al., 2018), with a bootstrap percentage (BS) $\geq$ 95 being interpreted as good support

Since the evolutionary histories of individual genes may differ from the history of the species tree

owing to the effects of incomplete lineage sorting (ILS) and horizontal gene transfer, the datasets were also subjected to species tree inference using ASTRAL-III v5.6.3 (Zhang et al., 2018), which employs a coalescent-based quartets approach to infer the species tree from the underlying gene trees. Individual gene trees were constructed from the loci alignments using IQTREE, with ASTRAL then being used to infer the species tree from the individual gene trees, with the local posterior probability (LPP) calculated as measure of support for each node. Since coalescent-based methods are less sensitive to missing taxa (Hosner et al., 2016; Sayyari & Mirarab, 2016; Molloy & Warnow, 2018), I used only the 90% coverage datasets for this analysis. All phylogenetic analyses were run on the University of Cape Town's High Performance Computing Facility.

Genotyping-by-sequencing

To identify patterns of shared gene pools in the Setacea and Ramosa groups, genotyping-by-sequencing (GBS, Elshire et al., 2011) was employed to sample the genome-wide SNP variation across these groups. Four DNA extracts per population were sent to Novogene Genome Sequencing Company Ltd. in Beijing, China for library preparation and sequencing following Novogene's protocols. In brief, each DNA sample was digested with restriction enzymes, and the resulting fragments were ligated to barcoded adaptors. The samples were then PCR amplified, pooled and size-selected. A Qubit® 2.0 fluorometer and an Agilent® 2100 bioanalyzer were used to check the concentration and insert sizes of the completed libraries. The samples were sequenced using paired-end sequencing on an Illumina platform sequencer, with a read length of 144 bp. The raw data were then filtered by Novogene to remove paired reads where either read was adaptor contaminated, had ambiguous nucleotides accounting for >10% of the read, or where

>50% of either read consisted of low quality nucleotides.

The dDocent pipeline (Puritz et al., 2014) was used to map and call single nucleotide polymorphisms (SNPs). The pipeline was run separately on three well-supported clades resolved by the phylogenetic analyses, namely the Rehmannii, Ramosa and Setacea clades (Fig. 1, Fig. 2). The reads were first trimmed using Trimmomatic to remove any reads with remaining adaptor contamination, bases having a quality score of <20, or having a 5 base window with an average quality score <10. The reads were then mapped to an *O. sativa* genome (NCBI accessions AP008207 - AP008218) using the MEM algorithm of BWA with the default settings (-A 1 -B 4 -O 6). SNPs were called using using FreeBayes v.1.2.0 (Garrison & Marth, 2012), which is a Bayesian-based variant detection program. The SNPs were concatenated into a variant call file (VCF) using VCFtools v.3.0 (Danecek et al., 2011), and filtered using the dDocent pipeline recommendations with the following exceptions. Only variants that had been genotyped in $\geq$ 10% of individuals were retained, and individuals with >30% missing data were removed. Loci were also filtered out if they were missing >10% of individuals within five or more populations. A within-population Hardy-Weinberg Equilibrium filter was applied at a significance level of 0.001 to remove erroneous variant calls and reduce the dataset to only biallelic SNPs.

Individuals within each clade were then assigned to potential ancestral gene pools with a sparse non-negative matrix factorisation (sNMF) algorithm (Frichot et al., 2014) using the R package *LEA* (Frichot & François, 2015). The sNMF algorithm has a similar level of accuracy to likelihood-based inferences of population structure, such as STRUCTURE (Pritchard, Stephens & Donnelly, 2000) and ADMIXTURE (Tang et al., 2005; Alexander, Novembre & Lange, 2009), but has a faster computing time (Frichot et al., 2014). One hundred ancestry coefficient matrices (Q matrices) were generated for each value of K between K = 1 and K = 20 for the Ramosa and

Rehmannii clades, and between K = 1 and K = 25 for the Setacea clade. K represents the number of potential ancestral gene pools. The fit of each value of K given the data was evaluated using the average entropy criterion of 100 iterations. The entropy criterion evaluates the fit of the statistical model using a cross-validation technique, with lower values representing a better fit (Frichot & François, 2015). However, I also used the value of K that corresponded to the number of putative species within each clade as suggested by the phylogenetic analyses. CLUMPAK (cluster Markov packager across K, Kopelman et al. (2015)) was used to summarise the Q matrices for the relevant K values, and the results visualised using the R package *pophelper* (Francis, 2017). Finally, principal coordinate analyses (PCoA) of the filtered SNP dataset were generated with Euclidian distance matrices using the function "gl.pcoa" in the R package *dartR* (Gruber & Georges, 2019; R Core Team, 2019).

Morphological measurements and analysis

Morphological traits of four flowering individuals per population in the Ramosa and Setacea groups were measured using a ruler, precise to 5 mm, digital vernier callipers, precise to 1 µm or a dissection microscope, precise to equipped with an eyepiece graticule, precise to 1 µm. In total, 24 floral and vegetative traits, comprising 17 measurements and seven potentially informative measurement ratios, were scored (Table 1). Of these traits, only 17 were used in downstream analyses to reduce autocorrelation.

Analysis of the morphological data was informed by the phylogenetic and SNP-based results, as follows. First, the morphological data were analysed separately for three clades (the Ramosa, Rehmannii and Setacea clades) which were resolved, with good support, by the phylogenetic analysis. Second, although multivariate patterns of morphological trait variation were initially

explored using principal components analysis, strong congruence between the clades resolved by the phylogenetic analyses and the population clusters revealed by analyses of the SNP data (i.e. sNMF and PCoA) allowed for the application of linear discriminant analysis (LDA) in order to assess morphological support for these groups. Linear discriminant analyses were conducted in R, using the package *MASS* (Venables & Ripley (2002)). In addition to these multivariate analyses, variation in individual morphological traits was compared between putative species for the purpose of identifying diagnostic traits.

Results

Sequence data

Targeted-enrichment sequencing yielded 353 loci across 108 accessions, to which *O. sativa* was added as an outgroup. Following read trimming, an average of 228 907 reads per sample were mapped to the target loci. Of the 353 loci, an average of 200 loci per sample had reads mapped to at least 75% of their length. Phylogenetic trees inferred from the 90% coverage and full coverage datasets were very similar, both topologically and in terms of their branch lengths, whether they were based on exons, introns or supercontigs. Therefore, only results obtained from the 90% coverage supercontig dataset are presented here. This dataset contained 307 loci, each represented in a minimum of 98 of the total 109 accessions. The concatenated alignment used for ML analysis was 698 832 bp long and contained 12.5% gap characters. Of these sites, 388 704 (55.6%) were variable, and 226 897 (32.5%) were potentially parsimony informative. The percent GC content was 39.4%. Partitionfinder reduced the number of partitions from 307 to 84.

While the majority of GBS libraries were sequenced successfully with good data yields, a single library of an *E. setacea* subsp. *disticha* accession (1589) failed quality control and was not sequenced. The full GBS dataset was split into three clades for analysis. The Setacea clade, represented by 95 individuals and 24 populations, originally consisted of 572 345 variants, but this decreased to 3753 variants after filtering. The Rehmannii and Ramosa clades were each represented by 48 individuals and 12 populations. Following filtering, the number of variants in the Rehmannii clade decreased from 450 635 to 3544, and in the Ramosa clade from 443 722 to 3430 variants.

Major clade relationships

Maximum-likelihood and ASTRAL yielded well-supported, topologically-similar trees (Fig. 1, Fig. 2). In both trees, six major clades, (here termed the Australian, Rehmannii, Ramosa, Setacea, Dura and Lowlands clades), are resolved with strong support, the relationships of these clades also being consistently well supported. The branches subtending these major clades are long, with branch length in the ML tree reflecting sequence divergence and that in the ASTRAL tree reflecting number of coalescent units (generations/effective population size). In the Setacea, Ramosa and Rehmannii clades, internal branches are short relative to both the stem and terminal branches, resulting in a "broom and handle" morphology (Crisp & Cook (2009)) (Fig. 1, Fig. 2). In contrast, the Lowlands clade has generally longer branches, with branch lengths being more evenly distributed throughout the clade.

The Australian clade, comprising *M. stipoides* and *T. laevis* (BS = 100, LPP = 1.00), is consistently resolved as sister to the predominantly-montane Setacea clade (BS = 100, LPP = 1.00) with strong support (BS = 100, LPP = 1.00), this pair being sister to the rest of *Ehrharta* (Fig. 1, Fig. 2). The Setacea clade itself is composed of two species, *E. setacea* and *E. rupestris*, neither of which appear be monophyletic.

The montane Dura clade (BS = 100, LPP = 1.00), comprising *E. dura* and *E. microlaena*, is resolved with strong support (BS = 100, LPP = 1.00) as sister to a well-supported clade (BS = 100, LPP = 1.00) comprising the Rehmannii, Ramosa and Lowlands clades. Within this clade, the Rehmannii (BS = 100, LPP = 1.00) and Ramosa clades (BS = 100, LPP = 1.00) are resolved as sisters with strong support (BS = 100, LPP = 1.00). The monophyly of the Lowlands clade, which comprises species occurring predominately at low to mid elevations in the GCFR, is also

well supported (BS = 100, LPP = 1.00), with the two analyses resolving the internal relationships of this clade identically and with high support (BS = 100, LPP = 1.00) on most nodes (Fig. 1, Fig. 2). Relationships within the Lowlands clade are largely consistent with those reported by Verboom, Linder & Stock (2003).

Within-clade relationships and species delimitation

Rehmannii clade

The ML and ASTRAL analyses resolve three groups within the Rehmannii clade (Filiformis, Subspicata and Rehmannii groups) which correspond broadly to the existing subspecies of *E. rehmannii*, namely subsp. *filiformis*, subsp. *subspicata*, and subsp. *rehmannii* (Fig. 3). Where the ML analysis identifies all three groups as monophyletic with moderate to strong support (BS = 99, 77, 95, respectively), with the the Filiformis and Subspicata subgroups being sister lineages (BS = 97, Fig. 3A), the ASTRAL analysis identifies only the Filiformis and Subspicata groups as monophyletic (LPP = 0.83, 0.96, respectively), the Rehmannii group being paraphyletic (LPP = 0.97, 0.85, Fig.3B). A sample collected near Agulhas, initially identified as *E. rehmannii* subsp. *filiformis* (1658), is included in the Subspicata clade by both analyses.

Both the sNMF analysis with K = 3 and the PCoA based on GBS-based SNP data revealed three main population clusters, which correspond to the three subclades/grades resolved by the phylogenetic analysis (Fig. 4A, Fig. 5A, Fig. 6). All three clusters are dominated by a different ancestral gene pool and, although there is some evidence of potential admixture, this is typically insufficient to erode the inference of three genetically-distinct clusters (Fig. 5). For example, while Filiformis shows evidence of potential admixture with Subspicata on the Cape Peninsula (LW1)

and Subspicata shows evidence of potential admixture with Filiformis at Pringle Bay (1590), this is limited in scale and insufficient to erase the dominant genetic affinities of these populations (Fig. 6, Fig. 7). A Subspicata population (1600) from the Agulhas Plain is, however, slightly anomalous in reflecting the Filiformis and Subspicata gene pools in approximately equal measure. Although accession 1658 was originally identified as *E. rehmannii* subsp. *filiformis*, both the phylogenetic analyses and the sNMF include this accession in the Subspicata group (Fig. 3, Fig. 5).

The cross-entrophy criterion identified K = 10 as optimal for the Rehmannii clade (Fig. 4A). At this value of K there is strong population-level genetic differentiation, with almost every locality representing a distinct ancestral gene pool (Fig. 5B). This is particularly apparent in the Filiformis cluster. Surprisingly, within the Subspicata cluster, Agulhas population 1659 shares a gene pool with Cape Peninsula population LW3, but not with the other Agulhas populations (1658, 1659, 1660, Fig. 5B). Of the three clades, Rehmannii shows the greatest internal cohesion, with the inference of two gene pools that show evidence of admixture at Nature's Valley (1652, Fig. 5B). The major mode genetic structure presented for K = 10, however, reflects only 39% of the sNMF replicates.

A LDA of the morphological data separates the Rehmannii clade into three fairly discrete clusters corresponding to the putative species groups, with the exception of a few Rehmannii group individuals which were misclassified as Subspicata (Fig. 8). The Subspicata accession initially identified as *E. rehmannii* subsp. *filiformis* (1658) clusters with the other Subspicata accessions. Univariate comparisons reveal that Rehmannii plants are generally taller than those of Subspicata, their inflorescences having more spikelets and protrude higher above the leaves (Fig. 9). Conversely, Filiformis plants are more delicate than either Rehmannii or Subspicata, with finer leaves, finer culms, and smaller and fewer spikelets (Fig. 9).

Ramosa clade

In contrast to the existing taxonomy, which describes two subspecies within *E. ramosa*, both the ML and ASTRAL trees resolve three principal subclades within the Ramosa clade (Fig. 10). The composition of these subclades, however, differs subtly between the two analyses. The Eastern Ramosa group (from the Swartberg and Langeberg) is monophyletic within the ASTRAL tree (LPP = 0.88, Fig. 10B), but consists of a grade in the ML tree (Fig. 10A). The converse is true for the Western ramosa group (Boland Mountains to Cederberg), which is monophyletic in the ML tree (BS = 83), but paraphyletic in the ASTRAL tree (LPP = 0.71). In contrast, the Aphylla group (Cape Peninsula to Betty's Bay) is monophyletic with strong support in both trees (BS = 100, LPP = 0.95). The *E. ramosa* subsp. *aphylla* and *E. ramosa* subsp. *ramosa* specimens sampled from sites of sympatry at Milner Peak and Limietberg are not recovered as sisters in either tree, potentially indicating the presence of multiple taxa within the Western Ramosa clade.

The admixture results showed limited support for three ancestral populations (K = 3, Fig. 11A) that correspond to the recovered phylogenetic groups. While all Aphylla individuals were assigned to the same ancestral population with high probability, most Western Ramosa individuals shared a large proportion of genomic information with the Aphylla group. A population from Milner Peak (1624) is exceptional, as all individuals from this locality were assigned to a genetic cluster largely distinct to that of Aphylla. Although the dominant gene pool in the Aphylla cluster (blue in Fig. 11A) also occurs in the Eastern Ramosa cluster, a high frequency of a third gene pool (red in Fig. 11A) distinguishes the latter.

The cross-entrophy criterion identifies K = 8 as optimal for the Ramosa clade (Fig. 4B) . At this level, the Aphylla group is dominated by a single ancestral gene pool that is largely absent from

the Western and Eastern Ramosa groups, thereby corroborating the phylogenetic distinctness of Aphylla (Fig. 11 B). In contrast, the Western and Eastern Ramosa clades both show strong structure at K = 8, with most populations representing a distinct ancestral gene pool (Fig. 11 B). The exceptions are the Milner Peak (1622) and Limietberg (1609) populations of the Western Ramosa group, which share an ancestral gene pool, and are resolved as sister taxa in the phylogenetic trees (Fig. 10, Fig. 11B, Fig. 13), as well as the Swartberg (1656) and Seweweekspoort (1654) populations of the Eastern Ramosa group, which again are resolved as sisters in the phylogenetic tree (Fig. 10). Interestingly, the sympatric Milner Peak populations (1622 and 1624) do not share genetic information, despite being collected within a few kilometres of each other (Fig. 13). The strong genetic structure of the Western Ramosa group is also evident in the PCoA (Fig. 12), especially in the PC3 vs PC4 biplot. By contrast, the Eastern Ramosa cluster appears much more coherent in the PCoA (Fig. 12).

An LDA based on trait data confirms the morphological distinctness of the three Ramosa groups (Fig. 14) . Consistent with the genetic data, the Western Ramosa group tends to show greater variability in most traits, often encompassing much of the trait range of the other two groups (Fig. 15). In general, Aphylla group plants tend to have longer leaves, and longer glumes relative to the spikelets, than plants in the Eastern and Western Ramosa groups. On the other hand, Eastern Ramosa plants have generally longer inflorescences and spikelets which are longer relative to their width (Fig. 13).

Setacea clade

The phylogeny of the Setacea clade is characterised by extremely short internal branch lengths which are often poorly supported (Fig. 16). Nonetheless, it is clear that both *E. setacea* and *E.*

rupestris, as currently defined, are polyphyletic. Four major clades are recovered with good support by the ML and ASTRAL analyses. These are: (i) the Eastern Rupestris clade (BS = 100, LPP = 1.0); (ii) the Western Rupestris clade (BS = 100, LPP = 1.0); (iii) a clade comprising the Fernkloof, Uniflora, and Restioid Tricostata clades (BS = 99, LPP = 0.74), with Uniflora and Restioid tricostata recovered as sisters in both trees (BS = 100, LPP = 0.72); and (iv) a clade consisting of the Setacea, Dodii, Leafy Tricostata and Scabra clades (BS = 97, LPP = 0.68). There is some incongruence at this level, however, with the ML analysis embedding the Western Rupestris clade within clade (iv), and so rendering the latter paraphyletic (Fig. 16).

Within clade (iii), the monophyly of both the Uniflora clade (BS = 100, LPP = 1.00), composed of all accessions of *E. setacea* subsp. *uniflora*, and the Restioid Tricostata clade (BS = 100, LPP = 0.97), comprising *E. rupestris* subsp. *tricostata* accessions from the Cape Peninsula, Hottentots Holland to Fernkloof area, are well supported, and these clades are retrieved as sister with good support (BS = 100, LPP = 0.72, Fig. 16). However, where the ML analysis retrieves a weakly supported (BS = 59) Fernkloof clade, comprising *E. setacea* subsp. *setacea* and *E. setacea* subsp. *disticha* plants from the Fernkloof Reserve in Hermanus and *E. setacea* subsp. *disticha* plants from the Buffelstalberg near Betty's Bay, this assemblage is paraphyletic (LPP = 0.37) in the ASTRAL tree (Fig. 16). Both trees, however, reflect good support for the monophyly of a clade comprising two Fernkloof accessions of *E. setacea* subsp. *setacea* (1595 and 1600) sampled from boggy depressions (BS = 100; LPP = 1.00) and a clade consisting of two Fernkloof accessions of *E. setacea* subsp. *disticha* (1597 and 1598) plus two Fernkloof accessions of *E. setacea* subsp. *setacea* (1593 and 1599) sampled from dry, rocky slopes (BS = 98, LPP = 0.88).

Although clade (iv), which encompasses most accessions of *E. setacea* subsp. *setacea*, *E. setacea* subsp. *scabra, E. rupestris* subsp. *dodii* and eastern accessions of *E. rupestris* subsp. *tricostata*, is

monophyletic in the ASTRAL tree, it is paraphyletic in the ML tree owing to the inclusion of the Western Rupestris clade (Fig. 16). Within clade (iv), there are high levels of incongruence due to poorly supported internal relationships. However, both trees reflect good support for the monophyly of the Scabra clade, comprising all accessions of *E. setacea* subsp. *scabra* (BS = 100, LPP = 1.00), and the Dodii clade, comprising all accessions of *E. rupestris* subsp. *dodii* from the western GCFR (BS = 96, LPP = 0.95). In addition, the ASTRAL tree resolves both a monophyletic Leafy Tricostata clade, comprising inland *E. rupestris* subsp. *tricostata* accessions from Ceres to Robinson's Pass (LPP = 0.70), and a monophyletic Setacea clade, comprising all non-Fernkloof accessions of *E. setacea* subsp. *setacea*, albeit with weak support (LPP = 0.56, Fig. 16B). However, the ML analysis retrieves neither of these clades, with both assemblages being paraphyletic in the ML tree (Fig. 16A). Finally, two accessions, originally designated as *E. rupestris* subsp. *tricostata* (1640 and 1657) occupy isolated positions within clade (iv). Although accession 1640 is retrieved as sister to the Scabra clade in both trees with good support (BS = 100, LPP = 1.00), the position of accession 1657 in the ASTRAL tree is uncertain, while in the ML tree it is embedded in the Setacea grade.

The sNMF analysis with K = 9 largely corroborates the genetic coherence of the nine named clades revealed by the phylogenetic analyses (Fig. 17A). Although the Dodii, Leafy Tricostata, Restioid Tricostata and Setacea clades show some evidence of between-group admixture, the genomic composition of most clades is highly distinct (Fig. 17A). This is also apparent from the PCoA, in which each group forms a well-defined cluster (Fig. 18). The exceptions to this pattern are the Fernkloof (1589 and 1597) and Scabra clades, which share a very similar and highly admixed genomic profile under K = 9, although the PCoA resolves these clades as distinct (Fig. 17A, Fig. 18). Beyond the nine named clades, the sNMF with K = 9 assigns the Wemmershoek population represented by accession 1657 to a unique ancestral population. In addition, within the

Fernkloof clade, the genomic profile of accession 1600 (*E. setacea* subsp. *setacea* from Fernkloof) is distinct from that of the other Fernkloof clade accessions included in the analysis (Fig. 19).

The cross-entrophy criterion identifies K = 19 as optimal within the Setacea clade (Fig. 4C). Under K = 19, almost every clade shows substructure, although there is evidence of within-group admixture in the Setacea, Restioid Tricostata, Leafy Tricostata, Dodii and Eastern Rupestris clades (Fig. 17B). Within the Fernkloof clade, the two *E. setacea* subsp. *disticha* (1589 and 1597) populations share some genomic information with each other, but not with 1600. Based on the results of the phylogenetic and sNMF analyses, the Fernkloof group was split into two for the morphological analysis, with Fernkloof A consisting of accessions assigned to *E. setacea* subsp. *disticha* (1589 and 1597) and accessions from dry, rocky slopes assigned to *E. setacea* subsp. *setacea* (1593 and 1599), and Fernkloof B containing the accessions from flat, boggy sites assigned to *E. setacea* subsp. *setacea* (1595 and 1600).

A LDA of the morphological data shows that both the Uniflora and Restioid Tricostata groups are morphologically divergent from the remainder of the Setacea clade (Fig. 20A & B). Where Restioid Tricostata plants are characterised by long culm internodes which are barely covered by the leaf sheaths, short pedicels and a small glume to spikelet length ratio, Uniflora plants are tall and gracile, having exceptionally thin culms and fine leaves. In addition, the inflorescences of Uniflora plants almost always contain two spikelets, in which the glumes considerably exceed the florets (Fig. 21). The LDA, particularly with the Restioid Tricostata and Uniflora groups omitted, also reveals clear pairwise separation between most groups at least along one of the discriminant axes (Fig. 20C & D). Exceptions to this pattern are the Dodii and Leafy Tricostata groups, which show considerable overlap on all discriminant axes, and potentially the Eastern Rupestris and Western Rupestris groups which, despite being morphologically close, show minimal overlap (Fig.

20C & D). Despite their lack of separation in multivariate space, the Dodii and Leafy Tricostata groups can be clearly differentiated on the basis of spikelet number per inflorescence (Fig. 21). Likewise, Western Rupestris plants are distinguishable from Eastern Rupestris plants on the basis of their more robust character, having thicker culms, generally larger leaves (leaf length and width), and relatively more elongate spikelets (spikelet length:width ratio; Fig. 21). Significantly, although both contain accessions initially assigned to *E. setacea* subsp. *setacea*, the LDA classifies the Fernkloof A and Fernkloof B groups into different clusters, these groups being distinguishable primarily on the basis of floral characteristics, such as the position of the inflorescence above the leaves, spikelet length and the ratio of lemma lengths (Fig. 20, Fig. 21). In addition, the LDA identifies both these groups as being distinct from the Setacea group, Setacea plants being distinguishable from Fernkloof A plants on the basis of inflorescence length and from Fernkloof B plants on the basis of shorter spikelets (Fig. 20, Fig. 21).

Discussion

The data presented here suggest that the Rehmannii, Ramosa and Setacea groups of the Cape *Ehrharta* each contain multiple species. In particular, there is evidence that each of the subspecies within the Rehmannii and Ramosa groups warrant species status, and that the Setacea group contains up to eleven taxa. Here, I discuss the evidence for each species hypothesis, using the unified species concept of de Queiroz (2007), which conceptualises a species as a separately evolving metapopulation-level lineage. The operational criteria used for developing species boundaries in this work are evidence of common ancestry inferred from phylogenetic analysis, evidence of genetic isolation inferred from genetic population assignment analysis (sNMF), and evidence of morphological and ecological differentiation. As such, this can be considered an integrative taxonomic approach, with multiple lines of evidence being used to delimit species boundaries. These operational criteria do not necessarily require a species to be monophyletic. Although it is a common taxonomic practice only to recognise monophyletic species, paraphyletic species are a likely outcome of some speciation models (Funk & Omland, 2003; Hörandl & Stuessy, 2010). For example, under peripatric speciation, a population on the periphery of a species' range differentiates to form a new species, thereby rendering the parent species paraphyletic relative to the newly-created daughter species (Mayr & Bock, 2002; Crawford, 2010). Although both taxa will eventually attain genetic monophyly, the rate at which this lineage sorting occurs depends on factors such as population size and the rate of within-species gene flow (Rieseberg & Brouillet, 1994; Funk & Omland, 2003). Rather than the exception, it has been argued that peripatric speciation may be common in plants (Rieseberg & Brouillet, 1994). Therefore, should the two taxa be genetically, morphologically and/or ecologically distinct, it should not be necessary either to sink the daughter species or to split the parent species simply for the purpose of enforcing

species monophyly (Crisp & Chandler, 1996; Hörandl & Stuessy, 2010; Kuchta, Brown & Highton, 2018; Carnicero et al., 2019). It is envisaged that the species hypotheses presented in this chapter will be formalized in a future taxonomic work, a full revision of the Ramosa, Rehmannii and Setacea clades of *Ehrharta* being outside the scope of the present study.

Species hypotheses

Rehmannii clade

Within the Rehmannii clade, three groups were recovered that correspond broadly to the three subspecies of *E. rehmannii* and which can each justifiably be considered a species. While Gibbs-Russell & Ellis (1989) considered the number of morphological intermediates between these groups to be too many to allow their distinction as species, the data presented in this study identify them as distinct. The phylogenetic analyses identify both the Filiformis and Subspicata groups as monophyletic and although the sNMF analyses provide some evidence of potential admixture in sites of sympatry, this is limited in scale and insufficient to erase their dominant genetic affinities. Moreover, a genetically-based PCoA resolves both as distinct clusters. The observation that admixture is limited even where these entities occur in sympatry, as where they form a mixed sward at Pringle Bay (1590, 1591), indicates that they do not hybridize freely in nature and so constitute distinct biological species (Britton, Hedderson & Verboom, 2014). Although Filiformis often grows on moist mountain slopes, it was also collected near coastal streams, in one locality in Pringle Bay (1591) growing interspersed with a Subspicata population (1590). Filiformis is also morphologically distinct from Subspicata, having finer leaves, finer culms and fewer spikelets, the latter also being held within rather than above the foliage. Finally,

although the two species occur sympatrically at some sites, Filiformis most commonly inhabits damp sites on mountain slopes, while Subspicata typically associates with coastal sands. The distribution of the Filiformis group corresponds to that described in Gibbs-Russell & Ellis (1989) for *E. rehmannii* subsp. *filiformis*, extending from the Cape Peninsula to Agulhas, while the distribution of the Subspicata group corresponds to that of *E. rehmannii* subsp. *subspicata*. Accession 1660, however, from a limestone pavement near De Hoop, is somewhat enigmatic. Based on its unique habitat and morphology, corresponding to the 'limestone' form of *E. rehmannii* subsp. *subspicata* described in Gibbs-Russell & Ellis (1989), it is possible that this accession represents a distinct species. However, given its phylogenetic position embedded within Subspicata, and the sNMF analysis with K = 3 assigning this accession with nearly equal probability to the Subspicata and Filiformis clusters, I follow the recommendation of Gibbs-Russell & Ellis (1989), and retain the accession within Subspicata pending further sampling. Although the Rehmannii group is identified as monophyletic by only the ML analysis, the ASTRAL analysis resolving it as paraphyletic with respect to Filifiormis + Subspicata, the sNMF result identifies this entity as being genetically distinct under both K = 3 and K = 10. The evidence of admixture at Nature's Valley (1652) under K = 10 suggests that the group is genetically cohesive. Moreover, the LDA reveals Rehmannii to be morphologically distinct. Though similar to Subspicata, Rehmannii differs in being taller, in having larger inflorescences which also protrude higher above the foliage, and by its relatively shorter glumes. In addition, Rehmannii has a disjunct range, occurring from the Langeberg eastward to the Tsitsikamma Mountains, where it occurs primarily along the margins of Afrotemperate Forest patches.

Ramosa clade

The Ramosa clade shows evidence of three to five species. Although phylogenetic analysis reveals the existence of three principal groupings, the distinction of the Aphylla and Western Ramosa groups is unsupported by the sNMF analysis under K = 3. Rather, the sNMF analysis assigns these groups to a common gene pool and identifies only one of the Milner Peak populations (1624) as genetically distinct. Given the high levels of genetic disparity displayed shown by the Western Ramosa group (Fig. 12), however, it is likely that the sNMF analysis with K = 8 more accurately represents genetic structure within the Ramosa clade. The Aphylla group is especially well supported as a distinct species, being not only clearly monophyletic but also genetically cohesive, with the sNMF analysis under K = 8 assigning all individuals in this group to a single, exclusive gene pool. The Aphylla group is also morphologically divergent from the rest of the Ramosa clade, having spikelets with relatively longer glumes, longer leaves, fewer spikelets per inflorescence and often having striated sterile lemmas. The distribution range of this group as defined here is, however, more confined than that of *E. ramosa* subsp. *aphylla* of Gibbs-Russell & Ellis (1989), consisting only of populations from the Cape Peninsula to Betty's Bay.

The Eastern Ramosa group, which occurs along the Swartberg and Langeberg mountain ranges, is probably best treated as a single species since it is genetically coherent, as shown by strong genetic similarity between populations (Fig. 12), and is recovered as monophyletic in the ASTRAL tree. Moreover, notwithstanding some evidence of potential admixture, the sNMF analysis with K = 3 assigns all Eastern Ramosa individuals with moderate probability to an ancestral gene pool that is absent from either the Aphylla or Western Ramosa groups. While at K = 8 the Langeberg accessions (1635, 1645) show little admixture, this may be an artefact of insufficient sampling density between their populations (Fig. 13, Mason et al., 2020). Although

LDA based on morphology separates Eastern Ramosa from Western Ramosa with the misclassification of just a few individuals, there is no single trait which clearly distinguishes the Eastern Ramosa group from the rest of the Ramosa clade.

Most challenging of all to delimit, the Western Ramosa group is widely distributed in the mountains of the western CFR, where it shows considerable genetic and morphological heterogeneity, and likely comprises of two or more species. Under K = 8, most populations are genetically distinct, even those collected in sympatry. For example, two accessions from Milner Peak (1622, 1624) showed no admixture, despite occurring within a few kilometres of each other, which suggests that they represent distinct biological species. Similarly, the sympatric accessions from Limietberg (1609, 1611) are not recovered as sisters in the phylogenetic analysis. The sampling density employed in this study is, however, insufficient to test the species limits of these potential taxa confidently (Mason et al., 2020), and I therefore err on the side of caution and do not recommend splitting this complex pending further sampling. Instead, I recommend that the Western Ramosa group be provisionally treated as a single species, since it is monophyletic in the ML analysis, and paraphyletic only with respect to the Aphylla group in the ASTRAL analysis. In addition, the LDA reveals that the Western Ramosa group is morphologically distinct from the other Ramosa groups.

Setacea clade

Gibbs-Russell (1987) described two species and seven subspecies within the Setacea clade, the majority of which are found to be polyphyletic in this study. By contrast, the data presented here suggest that the Setacea clade potentially contains between nine and eleven species, largely corresponding to the groups recovered by the phylogenetic analyses. *Ehrharta rupestris* subsp.

rupestris is found to consist of two well-supported lineages, the Eastern and Western Rupestris groups, which are not, however, resolved as sisters in the phylogenetic analyses. The well-supported monophyly of Western Rupestris, which comprises the type form of *E. rupestris* subsp. *rupestris*, and its assignment by sNMF to a single, exclusive gene pool under K = 9, strongly support its promotion to species status. The same is true for Eastern Rupestris, which corresponds to the smaller form of *E. rupestris* subsp. *rupestris* mentioned in Gibbs-Russell (1987). The two Rupestris groups are, however, morphologically similar, the Eastern group having slightly finer leaves and culms. The ecology of the two is subtly different, with the Western Rupestris group occurring on the rocky summit ridges of the Riviersonderend and Langeberg mountains, and the Eastern Rupestris group generally being collected in moist depressions on north-facing slopes of the Klein and Groot Swartberg mountains.

The Uniflora group, corresponding to *E. setacea* subsp. *uniflora*, is monophyletic and has both a distinct genetic composition and a unique morphology. These attributes clearly identify it as a species in its own right. The group is characterised by fine leaves and culms, usually with two long-glumed, small spikelets per inflorescence and grows in damp, boggy areas, often in the understory of dense *Psoralea* stands. Like the Uniflora group, the Restioid Tricostata group is monophyletic and shows little evidence of admixture with other groups in the Rupestris clade. This group also has a distinctive morphology, showing clear separation from the rest of the Rupestris clade in the LDA. Univariate comparisons show that it is characterised by longer internodes, highly reduced pedicels, and leaf sheaths that are long relative to leaf length. The Restioid Tricostata group inhabits seeps from the Cape Peninsula to Hermanus, where it assumes a slightly restioid appearance. Conversely, the Fernkloof group, comprising accessions initially identified as *E. setacea* subsp. *setacea* and *E. setacea* subsp. *disticha* from Hangklip to Fernkloof, likely contains at least two species. One of these potential species, here termed Fernkloof A,

comprises the *E. setacea* subsp. *disticha* accessions from Buffelstalberg (1589) and Fernkloof (1597, 1598), and two further accessions from Fernkloof, one a small-statured plant which is intermediate in form between *E. setacea* subsp. *disticha* and *E. setacea* subsp. *setacea* (1599), and the other a larger-statured plant with a creeping growth habit (1593). The second species, termed Fernkloof B, consists of two accessions (1595 and 1600) which are morphologically similar to plants included in the Setacea group. The primary basis for segregating these groups as distinct species is the lack of admixture between them, despite their co-occurrence within the Fernkloof Nature Reserve. There is also evidence of morphological differentiation, with Fernkloof B having substantially longer spikelets, more robust leaves, and thicker culms. Finally, the two species also occupy contrasting habitats, Fernkloof A associating with seasonally-dry rocky slopes and Fernkloof B inhabiting perennially-damp, boggy depressions.

The single Scabra group population (1631) included in the sNMF analysis appears to reflect high levels of admixture under K = 9, but this effect disappears entirely under K = 19, where it is genetically distinct. Moreover, this population shows little evidence of admixture with a sympatric Setacea group population (1633) under either K = 9 or K = 19. This, paired with its morphological distinctness and strongly supported monophyly, argues for the elevation of the Scabra group to the level of species. This entity, which corresponds to *E. setacea* subsp. *scabra*, grows on mountain slopes along the Langeberg from Swellendam to Garcia's Pass. An accession originally identified as *E. rupestris* subsp. *tricostata* (1640) is recovered as sister to the remainder of the Scabra group and corresponds to the Scabra individuals misclassified in the LDA plot. It is unlikely that this accession should be included within Scabra, yet without its inclusion in the sNMF analysis it is difficult to tell whether it constitutes a unique species, or whether it forms part of a paraphylum with the Dodii and Leafy Tricostata groups as its morphology implies.

The data presented here also suggest that the Setacea group can be considered a distinct species, and corresponds to all *E. setacea* subsp. *setacea* accessions occurring outside of Fernkloof Nature Reserve. The group is monophyletic in the ASTRAL tree and, although most populations reflect admixture of multiple gene pools, these gene pools are largely exclusive to this group. Consequently, the PCoA resolves the Setacea group as a distinct genetic cluster. While the morphology of this group is somewhat variable and overlaps with that of the Fernkloof groups, it nonetheless forms a relatively cohesive cluster within the LDA. Overall, the group is perhaps best distinguished by having glumes of approximately the same length as the lemmas. The Setacea group is one of the most widespread in the Rupestris clade, occurring in moist, boggy habitats from the Cape Peninsula to the Limietberg, and beyond to the Langeberg.

The Leafy Tricostata group comprises accessions initially identified as *E. rupestris* subsp. *tricostata* from the Hex River and Langeberg mountains, while the Dodii group corresponds to *E. rupestris* subsp. *dodii*, albeit with a distribution that extends further east than that described by Gibbs-Russell (1987), with accessions sampled at the Boosmansbos Wilderness Area in the Langeberg mountains. Phylogenetically affiliated to the Setacea group, but differing by their much shorter lemmas and their narrower culms, the Leafy Tricostata and Dodii groups are morphologically similar, distinguishable mainly by the number of spikelets per inflorescence. Nonetheless, genetic data suggest that they are separate species, as both form well supported clades in the ASTRAL analysis, and both the sNMF and PCoA analyses reveal little genetic overlap between the two groups. Both prefer moist habitats, with the Dodii group occurring at the base of cliffs and rocky overhangs, while the Leafy Tricostata group grows near streams or seeps. Finally, the accession 1657, collected from a damp, east-facing slope near Wemmershoek Dam, and initially identified as *E. rupestris* subsp. *tricostata*, is another potential species. Poorly resolved in the phylogenetic analysis, it shows no admixture with any other species groups, and

has a morphology intermediate between that of the Setacea and Leafy Tricostata groups.

Taxonomic crypsis in *Ehrharta*

The proposed species delimitation scheme increases the number of species in the Cape *Ehrharta* by 13, from 22 to 35, with the number of species in the Rehmannii, Ramosa and Setacea clades together now nearly equaling that of the Lowlands clade. Of these, several are cryptic or semi-cryptic taxa. The Setacea clade was expected to show the greatest levels of cryptic diversity, owing to its general association with fragmented high-elevation habitats which foster non-adaptive divergence (Verboom et al., 2015), yet many species in this clade are morphologically distinguishable. The Restioid Tricostata, Uniflora, and Scabra groups are easy to differentiate, implying some role for adaptive divergence in the diversification of this clade, while the Dodii/Leafy Tricostata and Eastern Rupestris/Western Rupestris species pairs might be better described as examples of semi-cryptic differentiation, as their morphological differences are subtle. The most striking example of true taxonomic crypsis in the Setacea clade is provided by the Setacea and Fernkloof B groups, which are clearly genetically and phylogenetically distinct, but are nearly identical with respect to morphology and habitat preference. Perhaps more surprising is the high level of cryptic differentiation revealed within the Ramosa clade. Strong population-level genetic divergence within the Ramosa species implies that gene flow is limited, which is an unexpected result given the greater tolerance of Ramosa for arid conditions (Gibbs-Russell, 1987; Gibbs-Russell & Ellis, 1989) and what this means for its ability to traverse the drier valleys between adjacent mountain ranges. It is unclear, however, whether the strong genetic structure observed within Ramosa is a function of the heterogeneous topography of the Cape, or of a biological characteristic, such as poor propagule dispersibility or limited seed output

compared to clonal growth (Van Rossum et al., 2004), which might compromise rates of gene flow. In the latter case, the observed pattern may represent strong isolation-by-distance, with the apparent genetic discontinuities between populations a function of insufficient sampling density (Perez et al., 2018; Mason et al., 2020).

Cryptic species have been the object of considerable criticism, including that they are artefacts of new species concepts, that genetic data are insufficient to delimit distinct evolutionary units, and that delimiting cryptic species is contributing to taxonomic inflation, disrupting established taxonomy and hindering conservation efforts (Isaac, Mallet & Mace, 2004; Mace, 2004; Tattersall, 2007; Heller et al., 2013). Consistent with these criticisms, it has been argued that morphology should be one of the primary characters for species delimitation, on the basis that it is a proxy for reproductive isolation (Valdecasas, Williams & Wheeler, 2008), and that careful investigation will reveal morphological differences where the species delimitation is valid (Korshunova et al., 2019). Yet, not only can morphology be heavily influenced by environmental conditions, increasing the noisiness of its phylogenetic signal, but genetic data are a more direct measure of reproductive isolation (Lorenzen, Arctander & Siegismund, 2008; Bernardo, 2011), and allow for a clearer understanding of the processes driving historical and contemporary speciation pattens (Georges et al., 2018). Nonetheless, in order to alleviate the fear that genetic data increases Type 1 error rate, where the rank of species is assigned to a population (Cotterill et al., 2014), this study uses two lines of genetic evidence with difference resolutions to delimit species, which taken together, provide a robust picture of species boundaries. Few other studies have conducted such a geographically thorough, population-level genetic evaluation of species boundaries within the GCFR, notable exceptions being the works of Rymer et al. (2010), Prunier & Holsinger (2010), Segarra-Moragues & Ojeda (2010), Britton, Hedderson & Verboom (2014), Lexer et al. (2014), Prunier et al. (2017) and Shaik (2019). Of these studies, several have found evidence of cryptic

species, suggesting that the GCFR may contain substantially more biodiversity than is currently appreciated, particularly within species-lineages that are widespread throughout the mid- and high-elevation zones.

References

Alexander, D.H., Novembre, J. & Lange, K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome research.* 19(9):1655–1664.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. 1990. Basic local alignment search tool. *Journal of Molecular Biology.* 215(3):403–410.

Amadon, D. 1949. The seventy-five per cent rule for subspecies. *The Condor.* 51(6):250–258.

Andrews, S. 2010. FastQC: A quality control tool for high throughput sequence data. Available: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., et al. 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology.* 19(5):455–477.

Barley, A.J., White, J., Diesmos, A.C. & Brown, R.M. 2013. The challenge of species delimitation at the extremes: Diversification without morphological change in Philippine sun skinks. *Evolution.* 67(12):3556–3572.

Bernardo, J. 2011. A critical appraisal of the meaning and diagnosability of cryptic evolutionary diversity, and its implications for conservation in the face of climate change. *Climate change,*

ecology and systematics. 380–438.

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

Borowiec, M.L. 2016. AMAS: A fast tool for alignment manipulation and computing of summary statistics. *PeerJ*. 4:e1660.

Britton, M.N., Hedderson, T.A. & Verboom, G.A. 2014. Topography as a driver of cryptic speciation in the high-elevation Cape sedge *Tetraria triangularis* (Boeck.) C. B. Clarke (Cyperaceae: Schoeneae). *Molecular Phylogenetics and Evolution*. 77:96–109. DOI: 10.1016/j.ympev.2014.03.024.

Burbrink, F.T., Lawson, R. & Slowinski, J.B. 2000. Mitochondrial DNA phylogeography of the polytypic North American rat snake (*Elaphe obsoleta*): A critique of the subspecies concept. *Evolution*. 54(6):2107–2118.

Capella-Gutiérrez, S., Silla-Martínez, J.M. & Gabaldón, T. 2009. TrimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 25(15):1972–1973. DOI: 10.1093/bioinformatics/btp348.

Carnicero, P., Schönswetter, P., Garcia-Jacas, N. & Galbany-Casals, M. 2019. Is there a need for accepting paraphyletic taxa? A case study in the Sardinian endemic *Cymbalaria muelleri* (Plantaginaceae). *Botanical Journal of the Linnean Society*. 191(3):325–338.

Chaplin, K., Sumner, J., Hipsley, C.A. & Melville, J. 2020. An integrative approach using phylogenomics and high-resolution x-ray computed tomography for species delimitation in cryptic taxa. *Systematic Biology*. 69(2):294–307.

Chernomor, O., Haeseler, A.V. & Minh, B.Q. 2016. Terrace aware data structure for phylogenomic inference from supermatrices. *Systematic Biology*. 65(6):997–1008. DOI: 10.1093/sysbio/syw037.

Coates, D.J., Byrne, M. & Moritz, C. 2018. Genetic diversity and conservation units: Dealing with the species-population continuum in the age of genomics. *Frontiers in Ecology and Evolution*. 6:165.

Cotterill, F.P., Taylor, P.J., Gippoliti, S., Bishop, J.M. & Groves, C.P. 2014. Why one century of phenetics is enough: Response to "are there really twice as many bovid species as we thought?". *Systematic Biology*. 63(5):819–832.

Crawford, D.J. 2010. Progenitor-derivative species pairs and plant speciation. *Taxon*. 59(5):1413–1423.

Crisp, M.D. & Chandler, G.T. 1996. Paraphyletic species. *Telopea*. 6(4):813–844.

Crisp, M.D. & Cook, L.G. 2009. Explosive radiation or cryptic mass extinction? Interpreting signatures in molecular phylogenies. *Evolution: International Journal of Organic Evolution*. 63(9):2257–2265.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., et al. 2011. The variant call format and VCFtools. *Bioinformatics*. 27(15):2156–2158.

Dayrat, B. 2005. Towards integrative taxonomy. *Biological Journal of the Linnean Society*. 85(3):407–415. DOI: 10.1111/j.1095-8312.2005.00503.x.

de Queiroz, K. 1998. The general lineage concept of species, species criteria, and the process of

speciation. *Endless forms: species and speciation*.

de Queiroz, K. 2005. A unified concept of species and its consequences for the future of taxonomy. *Proceedings of the California Academy of Sciences*.

de Queiroz, K. 2007. Species concepts and species delimitation. *Systematic Biology*. 56(6):879–886.

DeSalle, R., Egan, M.G. & Siddall, M. 2005. The unholy trinity: Taxonomy, species delimitation and DNA barcoding. *Philosophical transactions of the royal society B: Biological sciences*. 360(1462):1905–1916.

Devey, D.S., Bateman, R.M., Fay, M.F. & Hawkins, J.A. 2008. Friends or relatives? Phylogenetics and species delimitation in the controversial European orchid genus *Ophrys*. *Annals of Botany*. 101(3):385–402.

Doyle, J.J. & Doyle, J.L. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*. (19):11–15.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S. & Mitchell, S.E. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS one*. 6(5):e19379.

Fišer, C., Robinson, C.T. & Malard, F. 2018. Cryptic species as a window into the paradigm shift of the species concept. *Molecular Ecology*. 27(3):613–635. DOI: 10.1111/mec.14486.

Francis, R.M. 2017. Pophelper: An R package and web app to analyse and visualise population structure. *Molecular Ecology Resources*. 17(1):27–32. DOI: 10.1111/1755-0998.12509.

Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G. & François, O. 2014. Fast and efficient estimation of individual ancestry coefficients. *Genetics*. 196(4):973–983.

Frichot, E. & François, O. 2015. LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*. 6(8):925–929.

Frost, D.R. & Kluge, A.G. 1994. A consideration of epistemology in systematic biology, with special reference to species. *Cladistics*. 10(3):259–294.

Fujita, M.K., Leaché, A.D., Burbrink, F.T., McGuire, J.A. & Moritz, C. 2012. Coalescent-based species delimitation in an integrative taxonomy. *Trends in ecology & evolution*. 27(9):480–488.

Funk, D.J. & Omland, K.E. 2003. Species-level paraphyly and polyphyly: Frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology, Evolution, and Systematics*. 34(1):397–423.

Garrison, E. & Marth, G. 2012. Haplotype-based variant detection from short-read sequencing. *arXiv preprint arXiv:1207.3907*.

Georges, A., Gruber, B., Pauly, G.B., White, D., Adams, M., Young, M.J., Kilian, A., Zhang, X., et al. 2018. Genomewide SNP markers breathe new life into phylogeography and species delimitation for the problematic short-necked turtles (Chelidae: *Emydura*) of eastern Australia. *Molecular Ecology*. 27(24):5195–5213.

Gibbs-Russell, G.E. 1987. Taxonomy of the genus *Ehrharta* (Poaceae) in southern Africa: The Setacea group. *Bothalia*. 17(1):67–73.

Gibbs-Russell, G.E. & Ellis, R.P. 1987. Species groups in the genus *Ehrharta* (Poaceae) in

southern Africa. *Bothalia*. 17(1):51–65.

Gibbs-Russell, G.E. & Ellis, R.P. 1989. Taxonomy and leaf anatomy of the genus *Ehrharta* (Poaceae) in southern Africa: The Ramosa group. *Bothalia*. 19(2):189–207.

Goulding, T.C. & Dayrat, B. 2016. Integrative taxonomy: Ten years of practice and looking into the future. 54:116–133.

Gruber, B. & Georges, A. 2019. DartR: Importing and analysing SNP and silicodart data generated by genome-wide restriction fragment analysis. Available: https://CRAN.R-project.org/package=dartR.

Hamilton, C.W. & Reichard, S.H. 1992. Current practice in the use of subspecies, variety, and forma in the classification of wild plants. *Taxon*. 41(3):485–498.

Hawlitschek, O., Nagy, Z.T. & Glaw, F. 2012. Island evolution and systematic revision of Comoran snakes: Why and when subspecies still make sense. *PLoS One*. 7(8):e42970.

Hebert, P.D., Penton, E.H., Burns, J.M., Janzen, D.H. & Hallwachs, W. 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences*. 101(41):14812–14817.

Heller, R., Frandsen, P., Lorenzen, E.D. & Siegismund, H.R. 2013. Are there really twice as many bovid species as we thought? *Systematic biology*. 62(3):490–493.

Hoang, D.T., Chernomor, O., Haeseler, A. von, Minh, B.Q. & Vinh, L.S. 2018. UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular biology and evolution*. 35(2):518–522. DOI: 10.1093/molbev/msx281.

Hosner, P.A., Faircloth, B.C., Glenn, T.C., Braun, E.L. & Kimball, R.T. 2016. Avoiding missing data biases in phylogenomic inference: An empirical study in the landfowl (Aves: Galliformes). *Molecular biology and evolution*. 33(4):1110–1125. DOI: 10.1093/molbev/msv347.

Hörandl, E. & Stuessy, T.F. 2010. Paraphyletic groups as natural units of biological classification. *Taxon*. 59(6):1641–1653.

Isaac, N.J.B., Mallet, J. & Mace, G.M. 2004. Taxonomic inflation: Its influence on macroecology and conservation. *Trends in Ecology & Evolution*. 19(9):464–469. DOI: 10.1016/j.tree.2004.06.004.

Johnson, M.G., Gardner, E.M., Liu, Y., Medina, R., Goffinet, B., Shaw, A.J., Zerega, N.J. & Wickett, N.J. 2016. HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Applications in plant sciences*. 4(7):1600016.

Johnson, M.G., Pokorny, L., Dodsworth, S., Botigue, L.R., Cowan, R.S., Devault, A., Eiserhardt, W.L., Epitawalage, N., et al. 2019. A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-medoids clustering. *Systematic Biology*. 68(4):594–606.

Jörger, K.M. & Schrödl, M. 2013. How to describe a cryptic species? Practical challenges of molecular taxonomy. *Frontiers in zoology*. 10(1):59. DOI: 10.1186/1742-9994-10-59.

Katoh, K. & Standley, D.M. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular biology and evolution*. 30(4):772–780.

Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A. & Mayrose, I. 2015. Clumpak: A program for identifying clustering modes and packaging population structure inferences across K.

Molecular ecology resources. 15(5):1179–1191.

Korshunova, T., Picton, B., Furfaro, G., Mariottini, P., Pontes, M., Prkić, J., Fletcher, K., Malmberg, K., et al. 2019. Multilevel fine-scale diversity challenges the 'cryptic species' concept. *Scientific reports*. 9(1):6732. DOI: 10.1038/s41598-019-42297-5.

Kuchta, S.R., Brown, A.D. & Highton, R. 2018. Disintegrating over space and time: Paraphyly and species delimitation in the Wehrle's Salamander complex. *Zoologica Scripta*. 47(3):285–299.

Lanfear, R., Calcott, B., Ho, S.Y.W. & Guindon, S. 2012. PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular biology and evolution*. 29(6):1695–1701. DOI: 10.1093/molbev/mss020.

Lanfear, R., Calcott, B., Kainer, D., Mayer, C. & Stamatakis, A. 2014. Selecting optimal partitioning schemes for phylogenomic datasets. *BMC evolutionary biology*. 14(1):82. DOI: 10.1186/1471-2148-14-82.

Leaché, A.D., Fujita, M.K., Minin, V.N. & Bouckaert, R.R. 2014. Species delimitation using genome-wide SNP data. *Systematic Biology*. 63(4):534–542. DOI: 10.1093/sysbio/syu018.

Lemmon, A.R., Emme, S.A. & Lemmon, E.M. 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. *Systematic Biology*. 61(5):727–744. DOI: 10.1093/sysbio/sys049.

Lexer, C., Wüest, R.O., Mangili, S., Heuertz, M., Stölting, K.N., Pearman, P.B., Forest, F., Salamin, N., et al. 2014. Genomics of the divergence continuum in an African plant biodiversity hotspot, i: Drivers of population divergence in *Restio capensis* (Restionaceae). *Molecular Ecology*. 23(17):4373–4386. DOI: 10.1111/mec.12870.

Li, H. & Durbin, R. 2010. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics*. 26(5):589–595.

Lorenzen, E.D., Arctander, P. & Siegismund, H.R. 2008. High variation and very low differentiation in wide ranging plains zebra (*Equus quagga*): Insights from mtDNA and microsatellites. *Molecular ecology*. 17(12):2812–2824.

Mace, G.M. 2004. The role of taxonomy in species conservation. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*. 359(1444):711–719.

Mallet, J. 2007. Subspecies, semispecies, superspecies. *Encyclopedia of biodiversity*. 5:523–526.

Manier, M.K. 2004. Geographic variation in the long-nosed snake, *Rhinocheilus lecontei* (Colubridae): Beyond the subspecies debate. *Biological Journal of the Linnean Society*. 83(1):65–85.

Mason, N.A., Fletcher, N.K., Gill, B.A., Funk, W.C. & Zamudio, K.R. 2020. Coalescent-based species delimitation is sensitive to geographic sampling and isolation by distance. *Systematics and Biodiversity*. 18(3):269–280.

Mayden, R.L. 1997. A hierarchy of species concepts: The denouement in the saga of the species problem. *Systematics Association Special Volume*. 54:381–424.

Mayr, E. 1982. Speciation and macroevolution. *Evolution*. 36(6):1119–1132.

Mayr, E. & Bock, W.J. 2002. Classifications and other ordering systems. *Journal of Zoological Systematics and Evolutionary Research*. 40(4):169–194.

Molloy, E.K. & Warnow, T. 2018. To include or not to include: The impact of gene filtering on

species tree estimation methods. *Systematic Biology*. 67(2):285–303. DOI: 10.1093/sysbio/syx077.

Moritz, C. & Cicero, C. 2004. DNA barcoding: Promise and pitfalls. *PLoS Biol*. 2(10):1529–1531.

Nguyen, L., Schmidt, H.A., Haeseler, A. von & Minh, B.Q. 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular biology and evolution*. 32(1):268–274. DOI: 10.1093/molbev/msu300.

Padial, J.M., Castroviejo-Fisher, S., Köhler, J., Vilà, C., Chaparro, J.C. & De la Riva, I. 2009. Deciphering the products of evolution at the species level: The need for an integrative taxonomy. *Zoologica Scripta*. 38(4):431–447. DOI: 10.1111/j.1463-6409.2008.00381.x.

Padial, J.M., Miralles, A., De la Riva, I. & Vences, M. 2010. The integrative future of taxonomy. *Frontiers in zoology*. 7(1):16.

Patten, M.A. & Unitt, P. 2002. Diagnosability versus mean differences of Sage Sparrow subspecies. *The Auk*. 119(1):26–35.

Perez, M.F., Franco, F.F., Bombonato, J.R., Bonatelli, I.A., Khan, G., Romeiro-Brito, M., Fegies, A.C., Ribeiro, P.M., et al. 2018. Assessing population structure in the face of isolation by distance: Are we neglecting the problem? *Diversity and Distributions*. 24(12):1883–1889.

Pritchard, J.K., Stephens, M. & Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics*. 155(2):945–959. Available: http://www.genetics.org/cgi/content/abstract/155/2/945.

Prunier, R., Akman, M., Kremer, C.T., Aitken, N., Chuah, A., Borevitz, J. & Holsinger, K.E. 2017. Isolation by distance and isolation by environment contribute to population differentiation

in *Protea repens* (Proteaceae L.), a widespread South African species. *American Journal of Botany*. 104(5):674–684.

Prunier, R. & Holsinger, K.E. 2010. Was it an explosion? Using population genetics to explore the dynamics of a recent radiation within *Protea* (Proteaceae L.). *Molecular ecology*. 19(18):3968–3980.

Puillandre, N., Lambert, A., Brouillet, S. & Achaz, G. 2012. ABGD, automatic barcode gap discovery for primary species delimitation. *Molecular ecology*. 21(8):1864–1877.

Puritz, J.B., Matz, M.V., Toonen, R.J., Weber, J.N., Bolnick, D.I. & Bird, C.E. 2014. Demystifying the RAD fad. *Molecular ecology*. 23(24):5937–5942.

R Core Team. 2019. *R: A language and environment for statistical computing*. ed. Vienna, Austria: R Foundation for Statistical Computing. Available: https://www.R-project.org/.

Rieseberg, L.H. & Brouillet, L. 1994. Are many plant species paraphyletic? *Taxon*. 43(1):21–32.

Rouhan, G. & Gaudeul, M. 2014. Plant taxonomy: A historical perspective, current challenges, and perspectives. 1st ed. (Molecular plant taxonomy). Springer. 1–37.

Rymer, P.D., Manning, J.C., Goldblatt, P., Powell, M.P. & Savolainen, V. 2010. Evidence of recent and continuous speciation in a biodiversity hotspot: A population genetic approach in southern African gladioli (*Gladiolus*; Iridaceae). *Molecular Ecology*. 19(21):4765–4782. DOI: 10.1111/j.1365-294X.2010.04794.x.

Sayyari, E. & Mirarab, S. 2016. Fast coalescent-based computation of local branch support from quartet frequencies. *Molecular biology and evolution*. 33(7):1654–1668. DOI:

10.1093/molbev/msw079.

Segarra-Moragues, J.G. & Ojeda, F. 2010. Postfire response and genetic diversity in *Erica coccinea*: Connecting population dynamics and diversification in a biodiversity hotspot. *Evolution: International Journal of Organic Evolution.* 64(12):3511–3524.

Shaik, Z. 2019. Species delimitation and speciation process in the *Seriphium plumosum* L. complex (Gnaphalieae: Asteraceae) in South Africa. Master's thesis. University of Cape Town.

Singhal, S., Hoskin, C.J., Couper, P., Potter, S. & Moritz, C. 2018. A framework for resolving cryptic species: A case study from the lizards of the Australian wet tropics. *Systematic Biology.* 67(6):1061–1075. DOI: 10.1093/sysbio/syy026.

Tang, H., Peng, J., Wang, P. & Risch, N.J. 2005. Estimation of individual admixture: Analytical and study design considerations. *Genetic Epidemiology: The Official Publication of the International Genetic Epidemiology Society.* 28(4):289–301.

Tattersall, I. 2007. Madagascar's lemurs: Cryptic diversity or taxonomic inflation? *Evolutionary Anthropology: Issues, News, and Reviews: Issues, News, and Reviews.* 16(1):12–23.

Tautz, D., Arctander, P., Minelli, A., Thomas, R.H. & Vogler, A.P. 2003. A plea for DNA taxonomy. *Trends in ecology & evolution.* 18(2):70–74.

Treurnicht, M., Colville, J.F., Joppa, L.N., Huyser, O. & Manning, J. 2017. Counting complete? Finalising the plant inventory of a global biodiversity hotspot. *PeerJ.* 5:e2984.

Valdecasas, A.G., Williams, D. & Wheeler, Q.D. 2008. "Integrative taxonomy" then and now: A response to Dayrat (2005). *Biological Journal of the Linnean Society.* 93(1):211–216.

Van Rossum, F., Bonnin, I., Fenart, S., Pauwels, M., Petit, D. & Saumitou-Laprade, P. 2004. Spatial genetic structure within a metallicolous population of *Arabidopsis halleri*, a clonal, self-incompatible and heavy-metal-tolerant species. *Molecular Ecology*. 13(10):2959–2967.

Venables, W.N. & Ripley, B.D. 2002. *Modern applied statistics with S*. Fourth ed. New York: Springer. Available: http://www.stats.ox.ac.uk/pub/MASS4.

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 Phytologist*. 207(2):368–376. DOI: 10.1111/nph.13342.

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(5):1008–1021. DOI: 10.1111/j.0014-3820.2003.tb00312.x.

Wiley, E.O. 1978. The evolutionary species concept reconsidered. *Systematic zoology*. 27(1):17–26.

Zachos, F.E., Apollonio, M., Bärmann, E.V., Festa-Bianchet, M., Göhlich, U., Habel, J.C., Haring, E., Kruckenhauser, L., et al. 2013. Species inflation and taxonomic artefacts—a critical comment on recent trends in mammalian classification. *Mammalian Biology*. 78(1):1–6.

Zhang, C., Rabiee, M., Sayyari, E. & Mirarab, S. 2018. ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC bioinformatics*. 19(Suppl 6):153. DOI: 10.1186/s12859-018-2129-y.

Table 1: Descriptions of the morphological characters examined in this study. Column 3 shows characters that were used in the linear discriminant analyses. For each character, a single measurement was made from each specimen.

Trait	Trait description	Used in analyses
Plant height	Length from rhizome to inflorescence tip	No
Vegetative height	Length from rhizome to second highest leaf (i.e. excluding flag leaf)	Yes
Leaf width	Width at the widest point of a representative leaf	Yes
Leaf length	Length from base to tip of a representative leaf	Yes
Sheath length	Length of sheath subtending measured leaf	No
Internode distance	Length between two nodes, measured in lowest 3rd of specimen, unless other internodes judged more representative.	Yes
Culm diameter	Measured in lowest 3rd of specimen	Yes
Inflorescence length	Length of a representative inflorescence	Yes
Spikelet number	Count of spikelets in a representative inflorescence	Yes
Pedicel length	Length of pedicel	Yes
Glume length outer	Length from base of spikelet to the tip of outer glume	No
Glume length inner	Length from base of spikelet to the tip of inner glume	No
Spikelet width	Measured a third of the way up the spikelet, just before the glumes flare. Glumes are included, except for spp. such as tricostata where the glumes are reduced	No
Spikelet length	Length from the base of spikelet to the tip of longest lemma	Yes
Lemma length outer	Length from base of spikelet with glumes removed, to tip of outer lemma	No
Lemma length inner	Length from base of spikelet with glumes removed, to tip of inner lemma	No
Striation number	Count of striations on inner lemma	Yes
Relative sheath length	Sheath length/Internode distance, measures extent to which the internode is enclosed by the sheath	Yes
Relative leaf length	Leaf length/Sheath length	Yes
Glume enclosure	Glume length outer/Spikelet length, measures the extent to which the glume encloses the spikelet	Yes
Glume ratio	Glume length outer/Glume length inner, measures the equality of the glume lengths	Yes
Spikelet shape	Spikelet length/Spikelet width	Yes
Lemma ratio	Outer lemma length/Inner lemma length, measures the equality of the lemma lengths	Yes
Inflorescence position	Plant height/Veg height, measures extent to which the inflorescence extends above the foliage	Yes

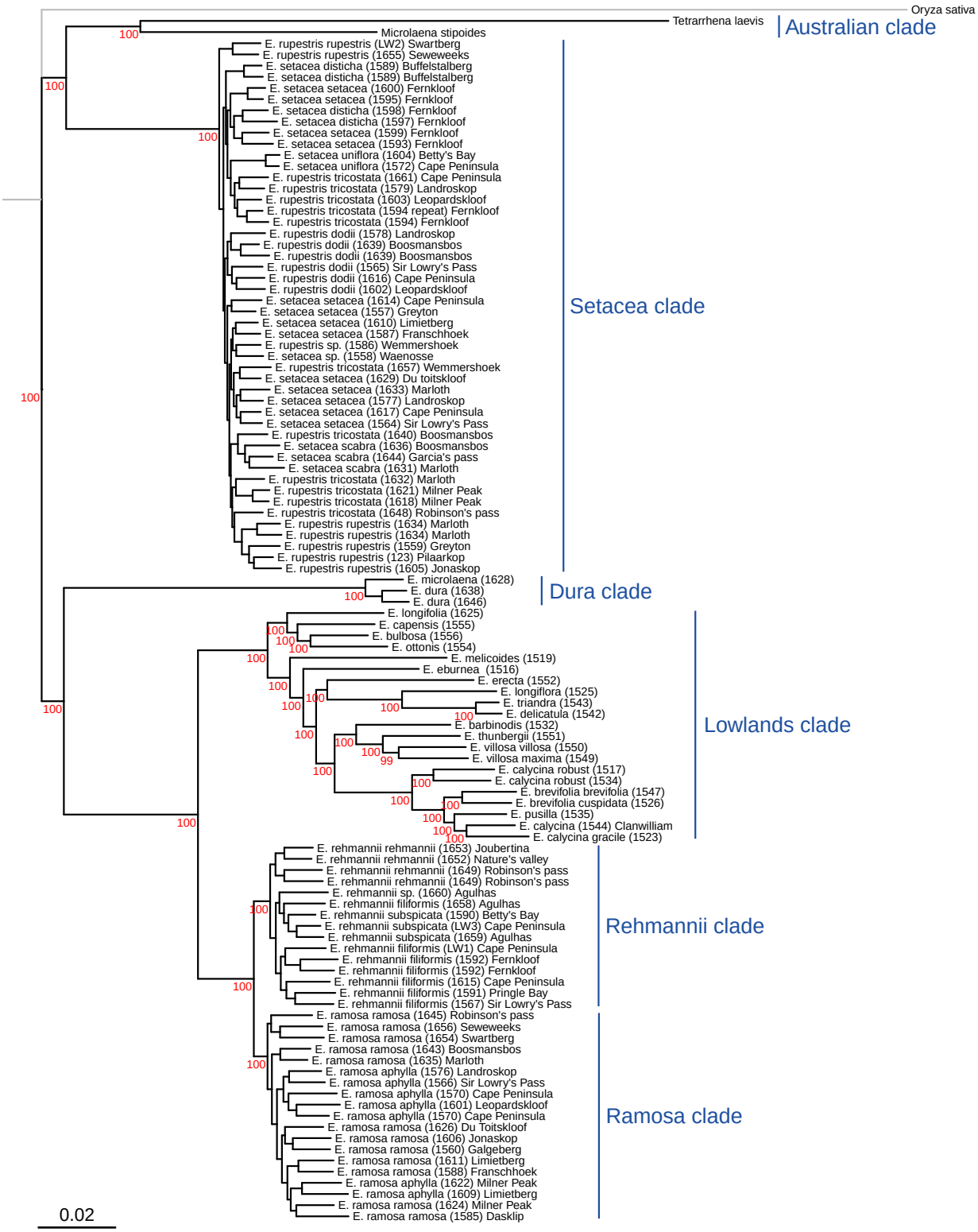

Figure 1: Maximum-likelihood tree constructed using IQTREE with Partitionfinder from 307 concatenated supercontig loci. Red numbers show ultrafast bootstraps. Branch lengths represent number of substitutions. The outgroup branch length, coloured grey, has been shortened by a factor of 2.5 for clarity. Branch support for the Rehmannii, Ramosa and Setacea clades is shown in Figs. 3, 10 and 16 respectively.

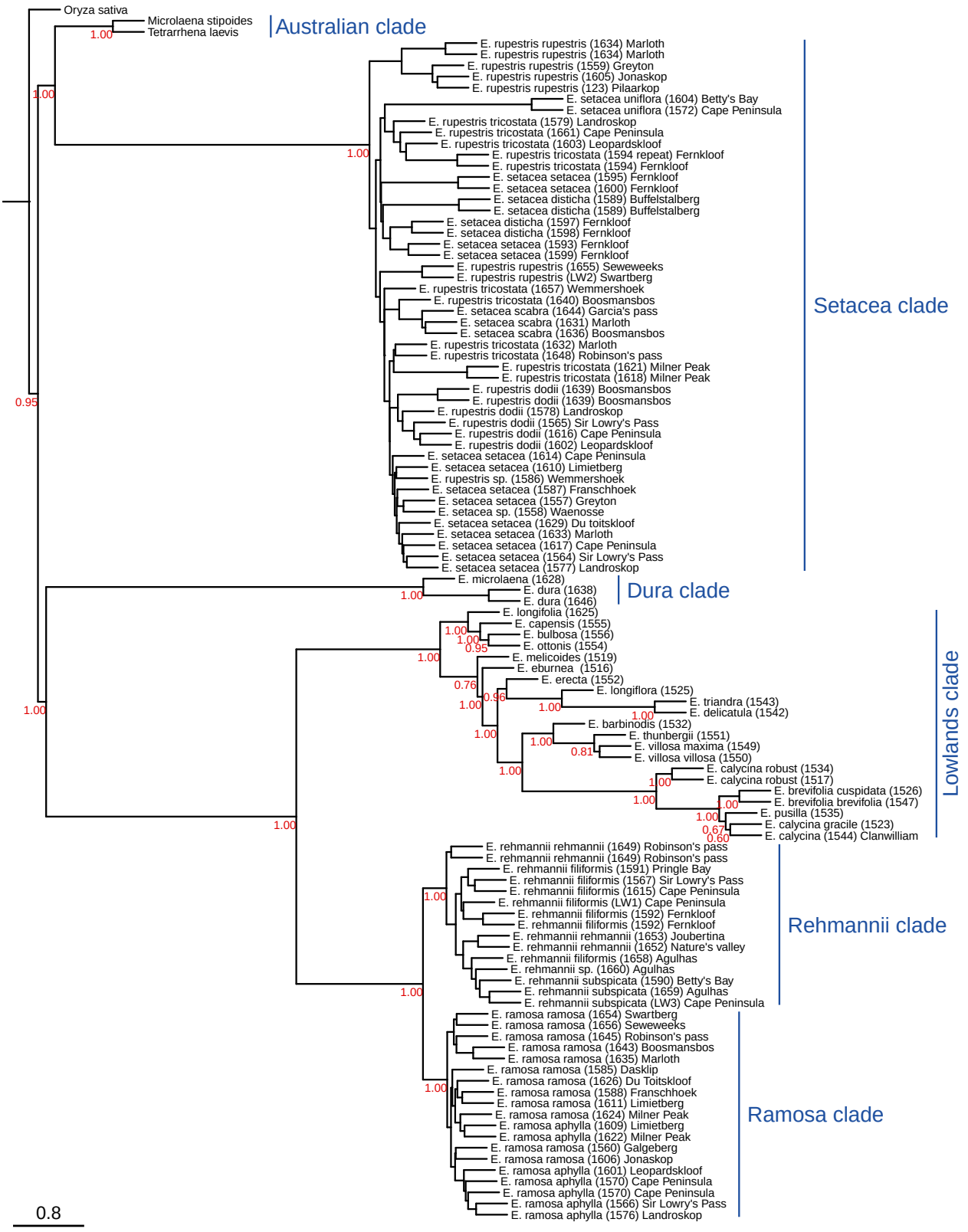

Figure 2: ASTRAL tree constructed using IQTREE gene trees from 307 supercontig loci. Branch lengths represent coalescent units (generations/effective population size). Tip branches are an arbitrary length not calculated by ASTRAL. Red numbers show local posterior probabilities. Branch support for the Rehmannii, Ramosa and Setacea clades is shown in Figs. 3, 10 and 16 respectively.

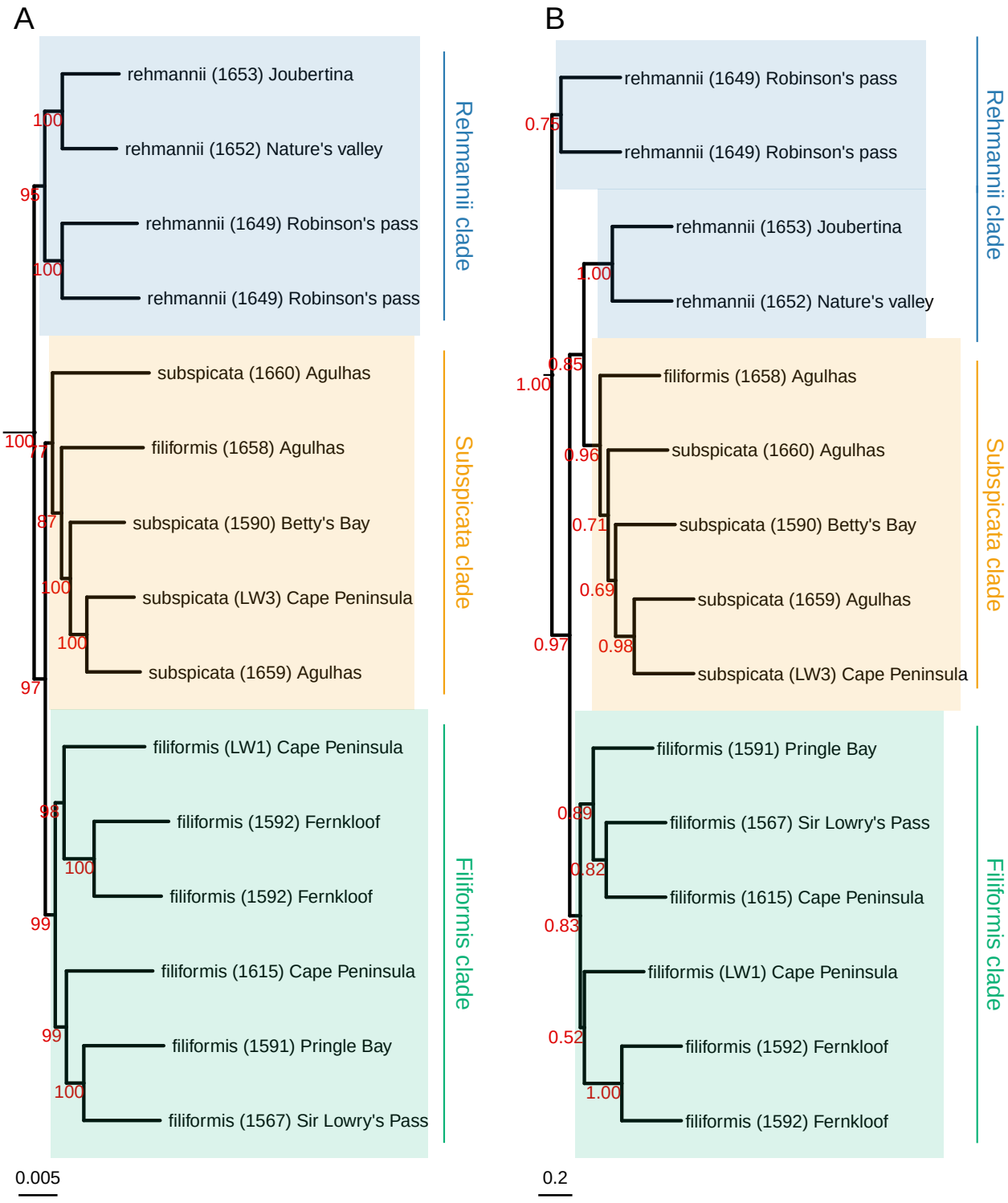

Figure 3: Phylogenetic relationships of the Rehmannii clade. A) Maximum-likelihood tree constructed using IQTREE with Partitionfinder from 307 concatenated supercontig loci. Red numbers show ultrafast bootstraps. Branch lengths represent number of substitutions. B) ASTRAL tree constructed using IQTREE gene trees from 307 supercontig loci. Branch lengths represent coalescent units (generations/effective population size). Tip branches are an arbitrary length not calculated by ASTRAL. Red numbers represent local posterior probabilities.

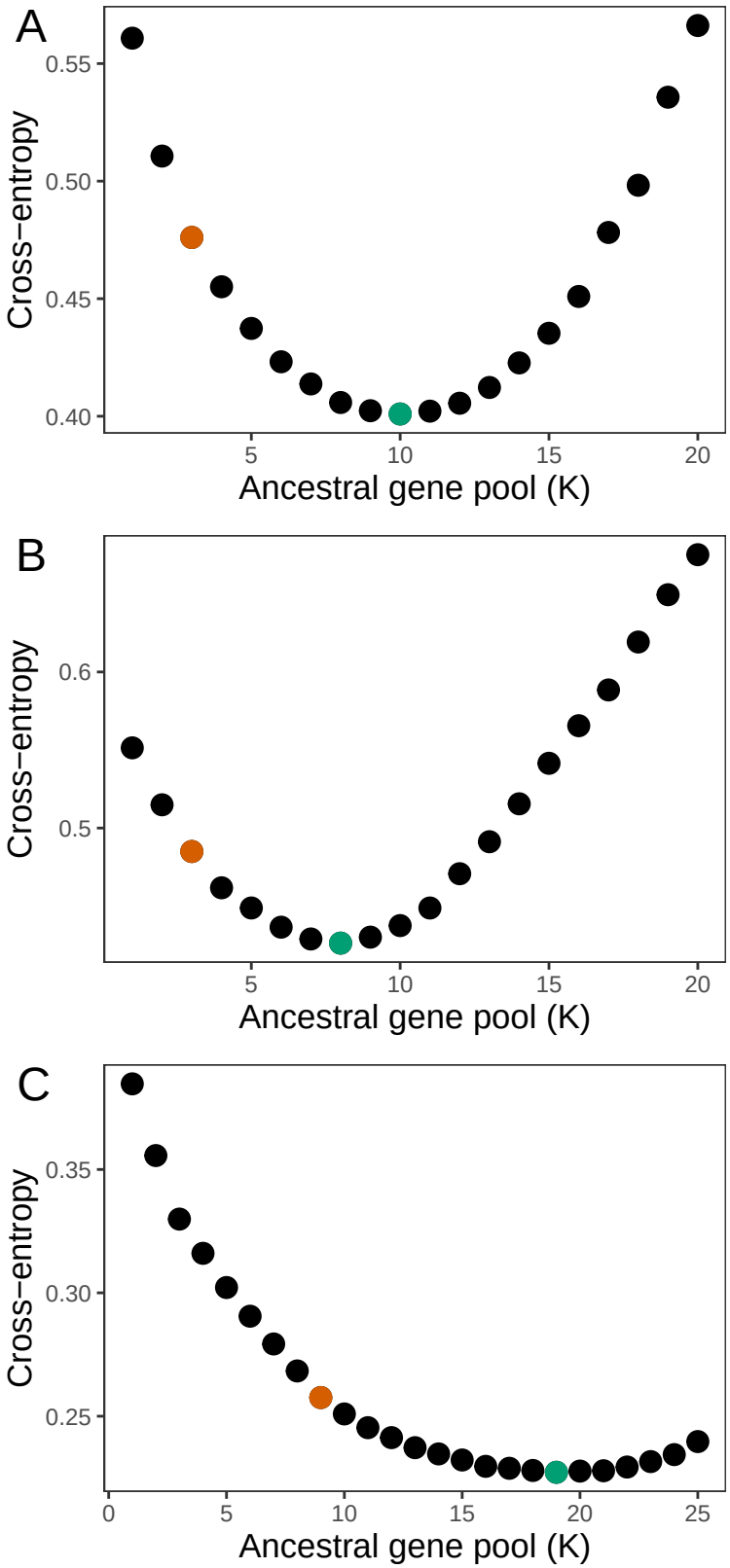

Figure 4: The cross-entropy criterion plotted for each value of K between K = 1 and K = 20 for A) the Rehmannii clade and B) the Ramosa clade, and for each value of K between K = 1 and K = 25 for C) the Setacea clade. Orange points show K corresponding to number of recovered phylogenetic groups, green points show the K with the lowest cross-entropy criterion.

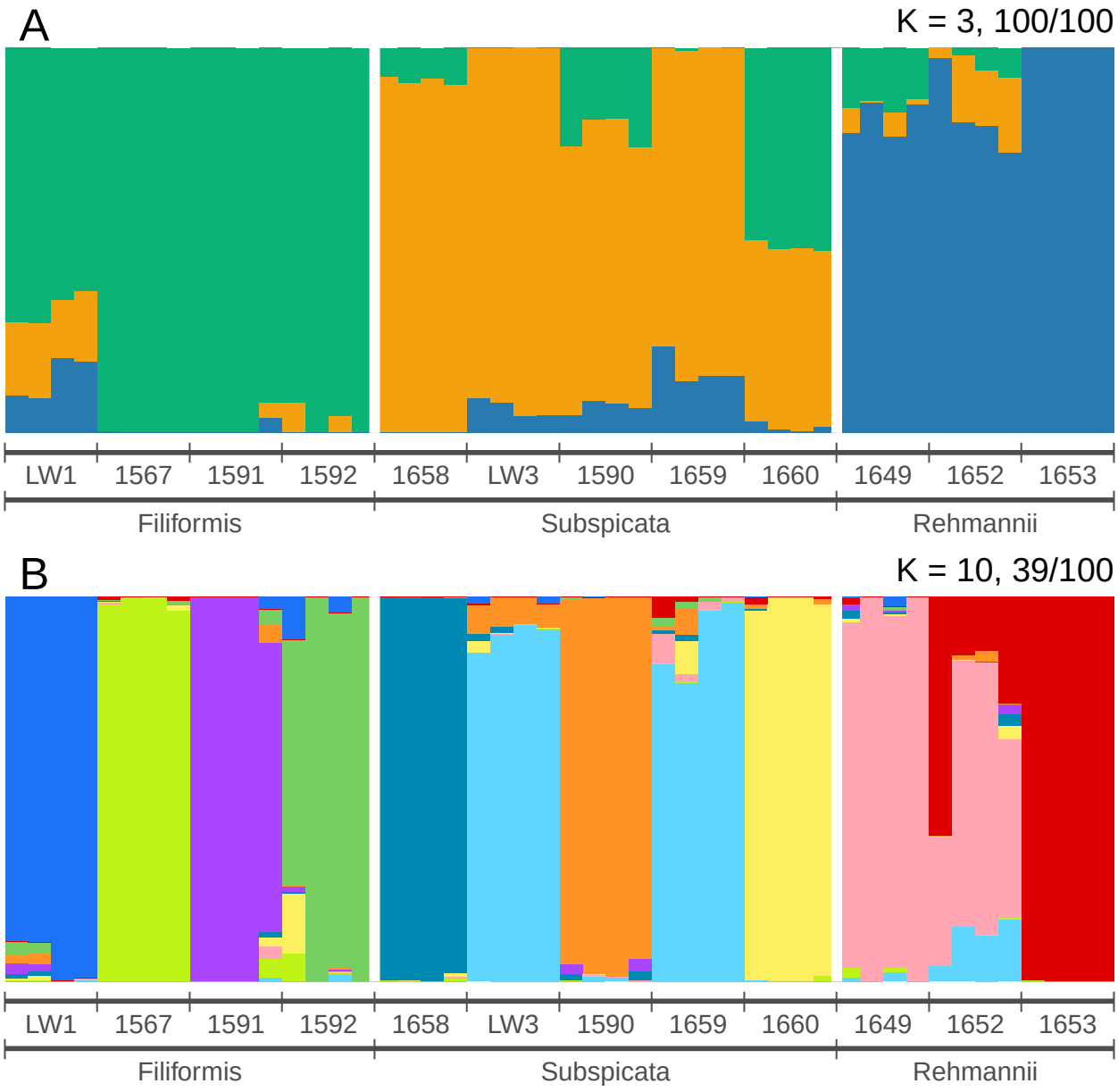

Figure 5: Plots of sNMF genomic assignment of individuals within the Rehmannii clade for A) K = 3, and B) K = 10. Fractions displayed at the top right of the graphs show the number of times a given sNMF repetition recovered the displayed assignments. The bars below the plots identify population accession numbers and taxon groups.

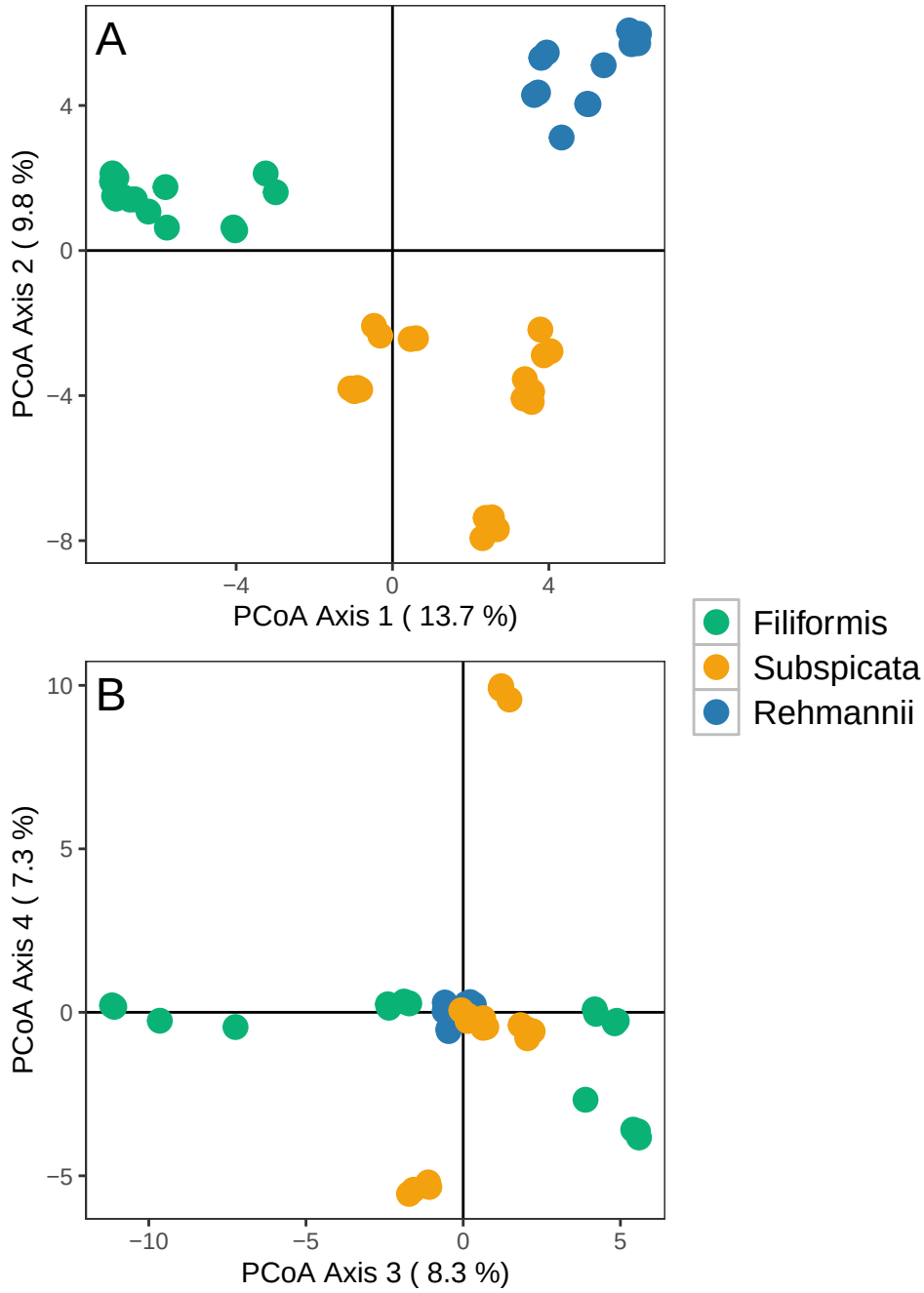

Figure 6: Representation of genetic similarity between individuals in the Rehmannii clade using a principal coordinates analysis of SNP variants. A) Principal coordinates axes 1 and 2, B) principal coordinates axes 3 and 4.

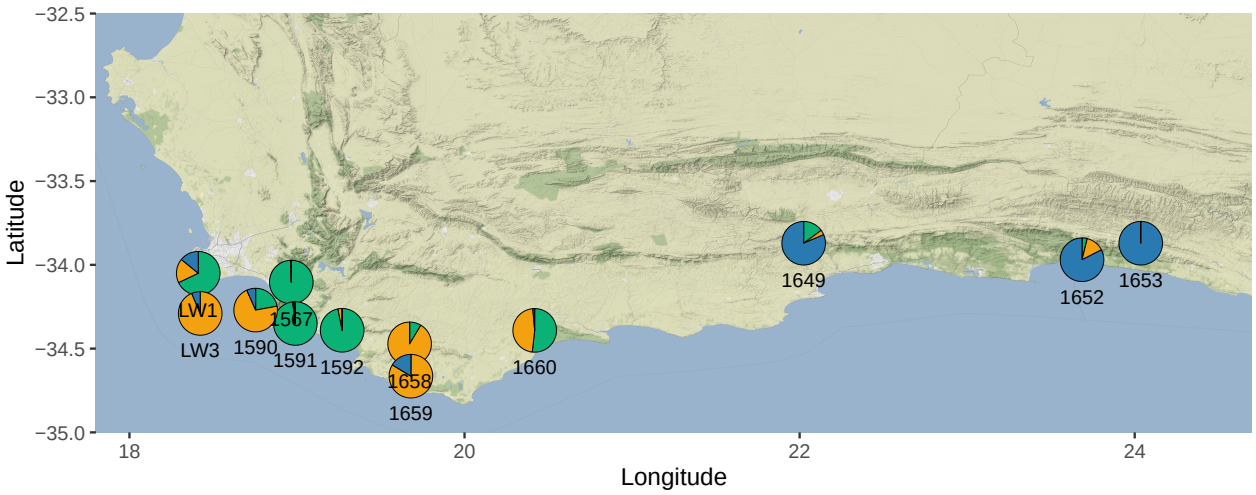

Figure 7: Map showing average sNMF assignments per population at K = 9 for the Rehmannii clade. Colours correspond to those of Figure 5A.

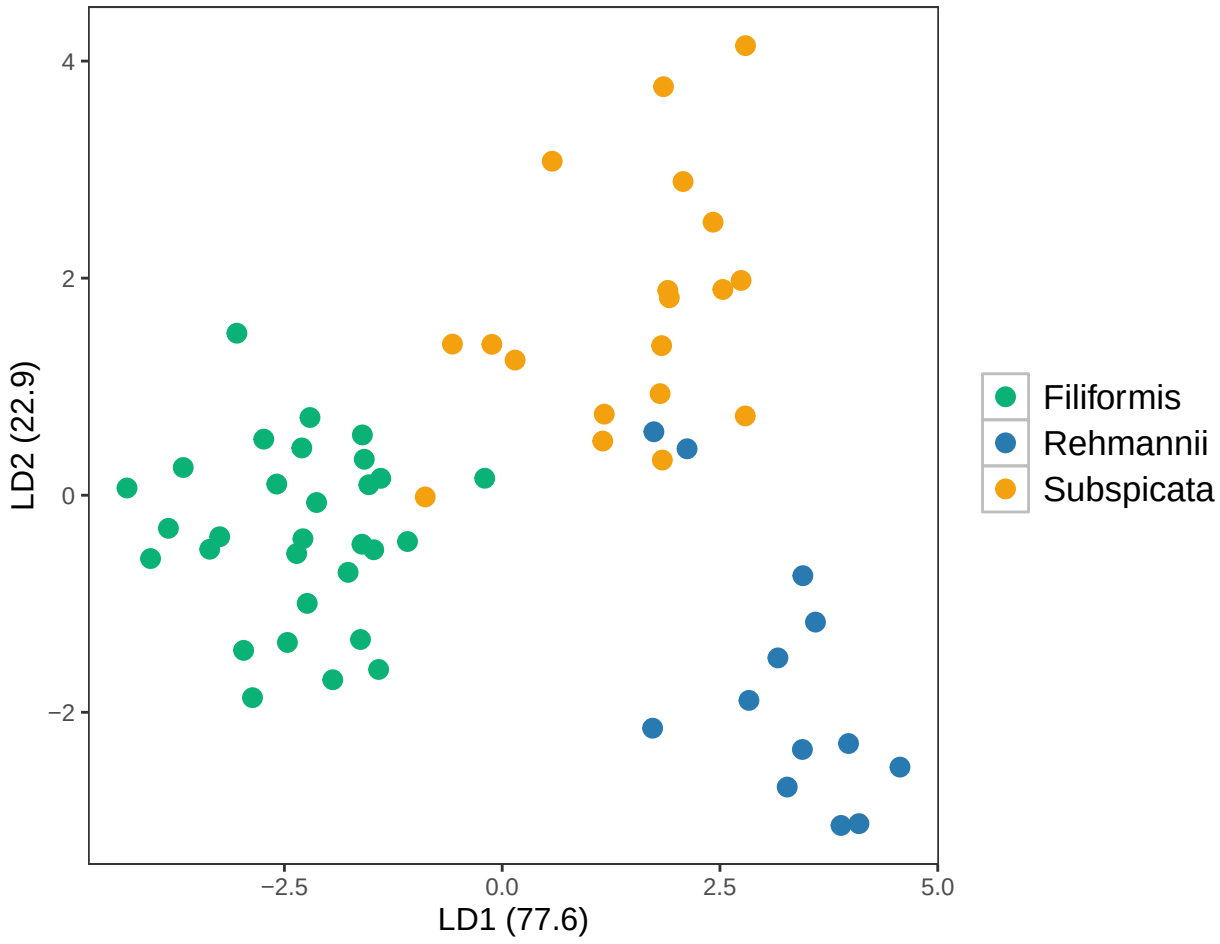

Figure 8: A linear discriminant analysis of morphological data for the Rehmannii clade. The 17 included traits are described in Table 1. The brackets show the percentage of explained between-group variance.

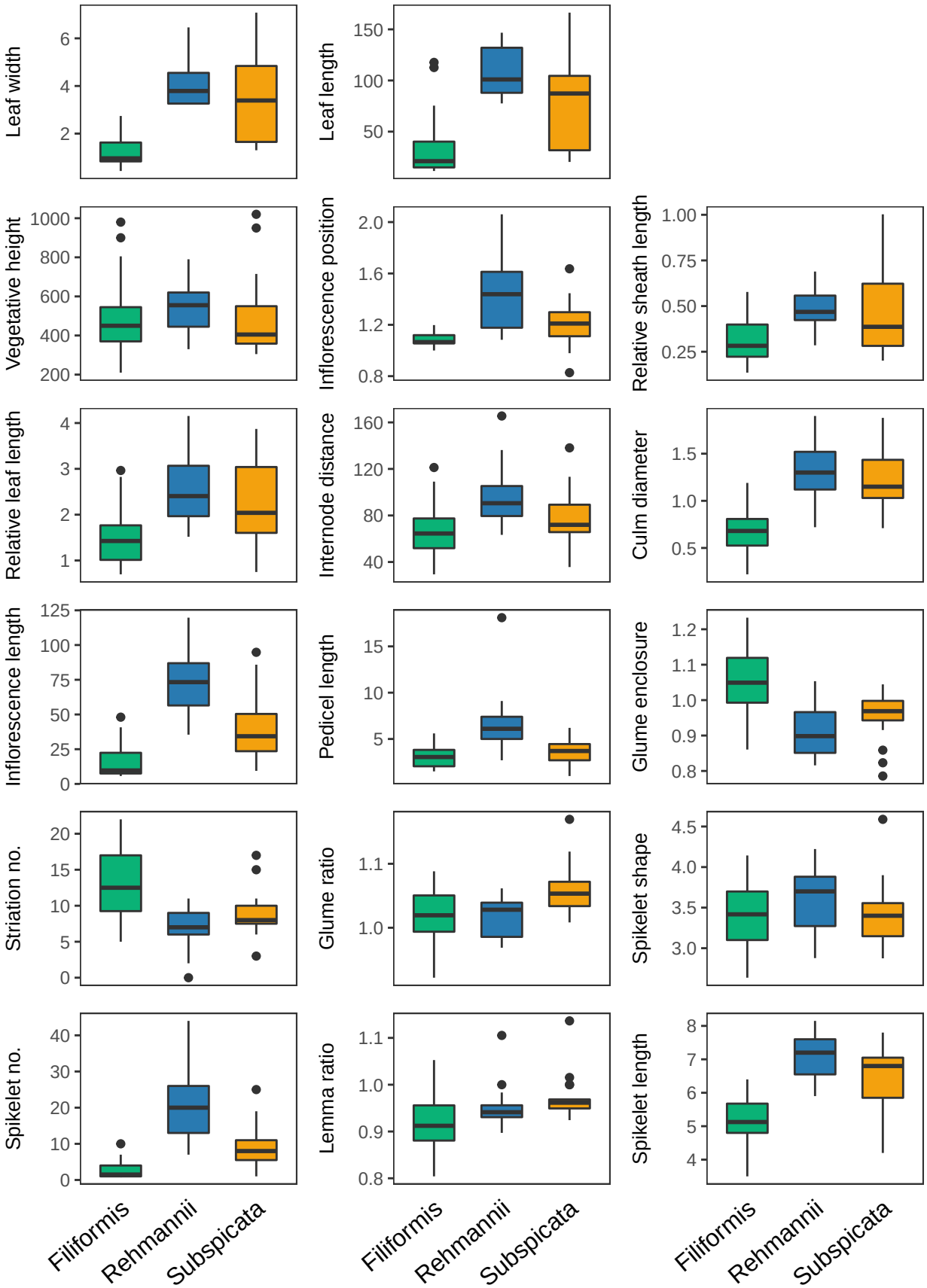

Figure 9: Boxplots of morphological traits used in the linear discriminant analysis for the Rehmannii clade. See Table 1 for trait descriptions. Where applicable, measurements are in mm.

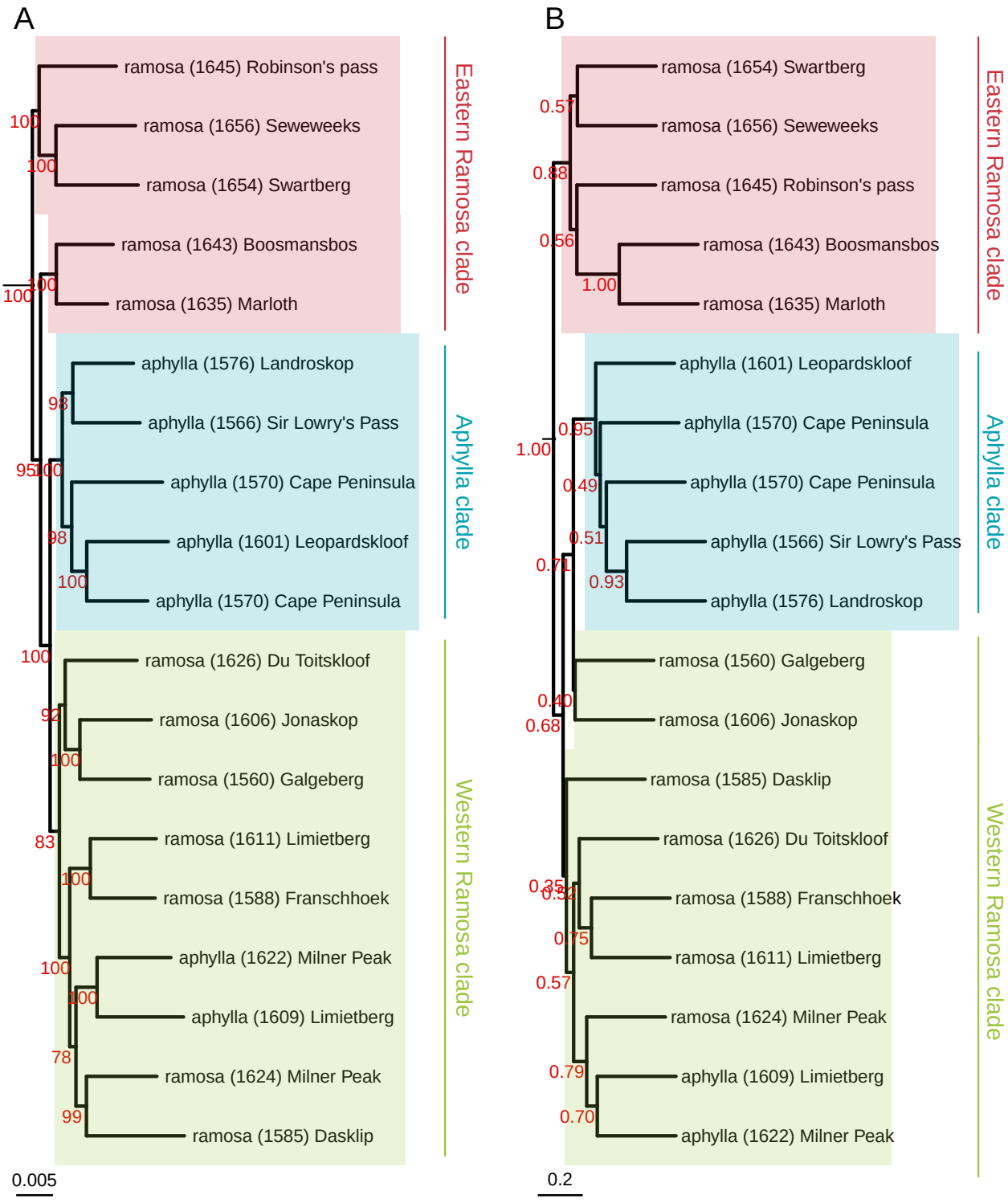

Figure 10: Phylogenetic relationships of Ramosa clade. A) Maximum-likelihood tree constructed using IQTREE with Partitionfinder from 307 concatenated supercontig loci. Red numbers show ultrafast bootstraps. Branch lengths represent number of substitutions. B) ASTRAL tree constructed using IQTREE gene trees from supercontig 307 loci. Branch lengths represent coalescent units (generations/effective population size). Tip branches are an arbitrary length not calculated by ASTRAL III. Red numbers represent local posterior probabilities.

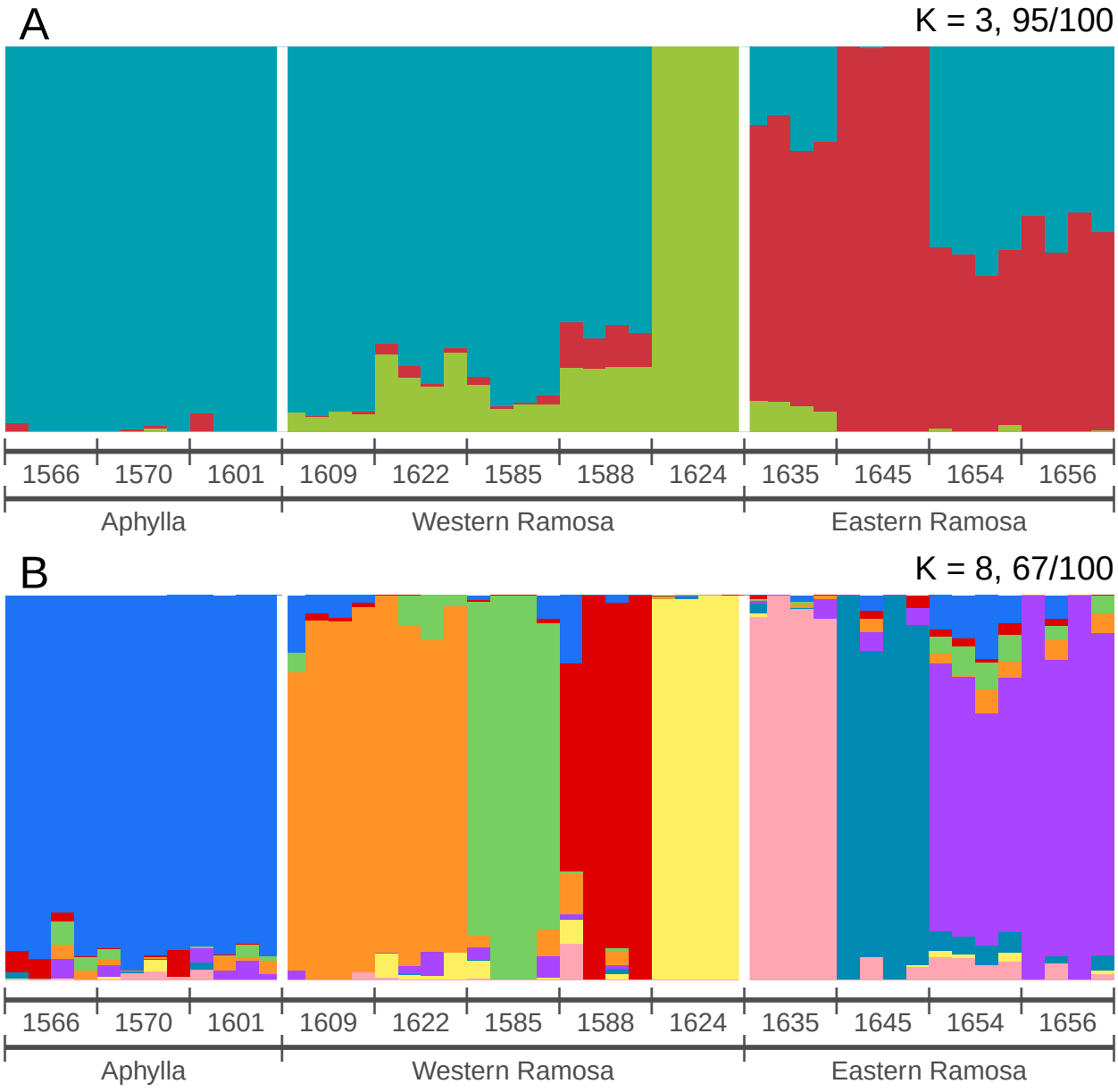

Figure 11: Plots of sNMF genomic assignment of individuals within the Ramosa clade for A) K = 3, and B) K = 8. Fractions displayed at the top right of the graphs show the number of times a given sNMF repetition recovered the displayed assignments. The bars below the plots identify population accession numbers and taxon groups.

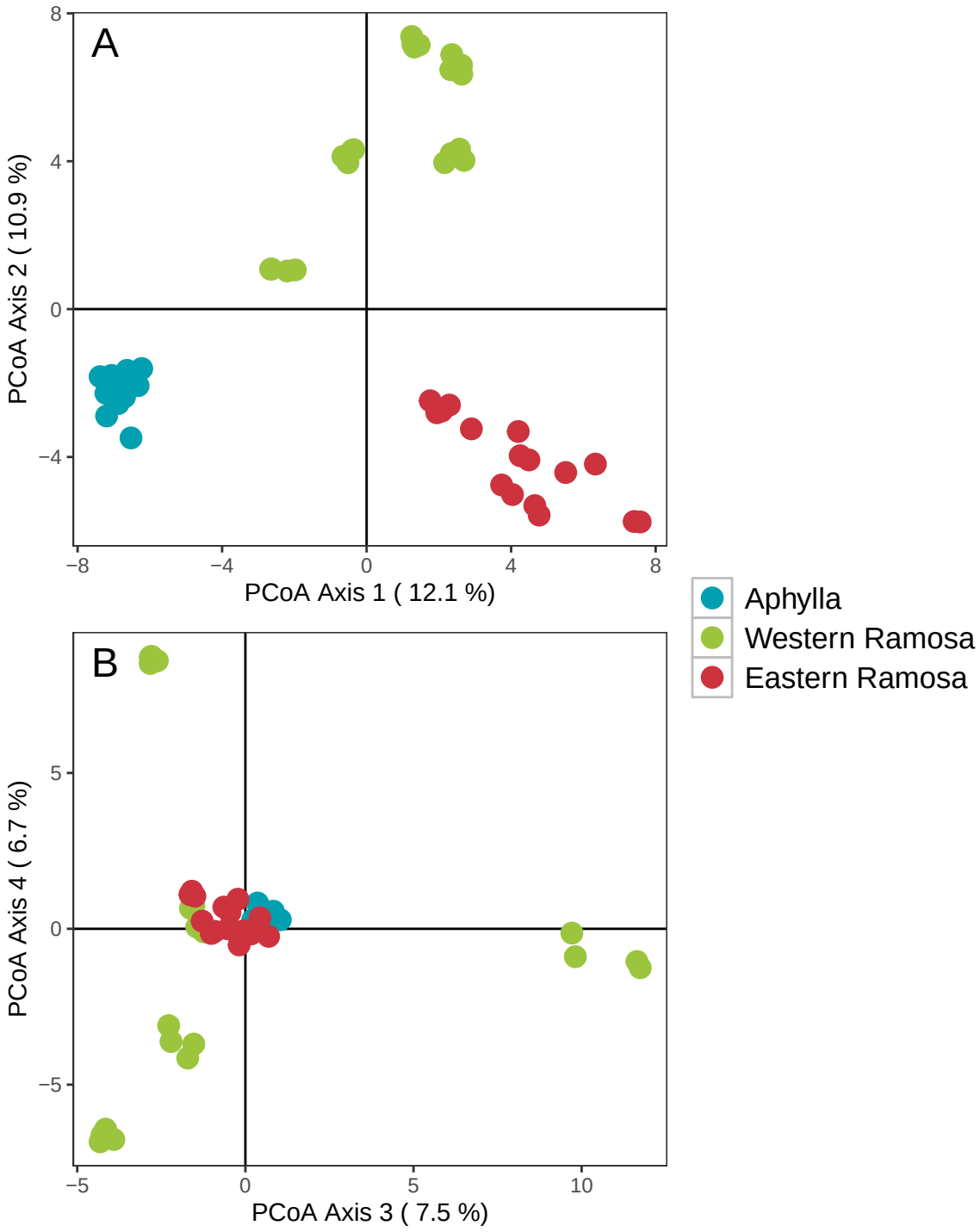

Figure 12: Representation of genetic similarity between individuals in the Ramosa clade using a principal coordinates analysis of SNP variants. A) Principal coordinates axes 1 and 2, B) principal coordinates axes 3 and 4.

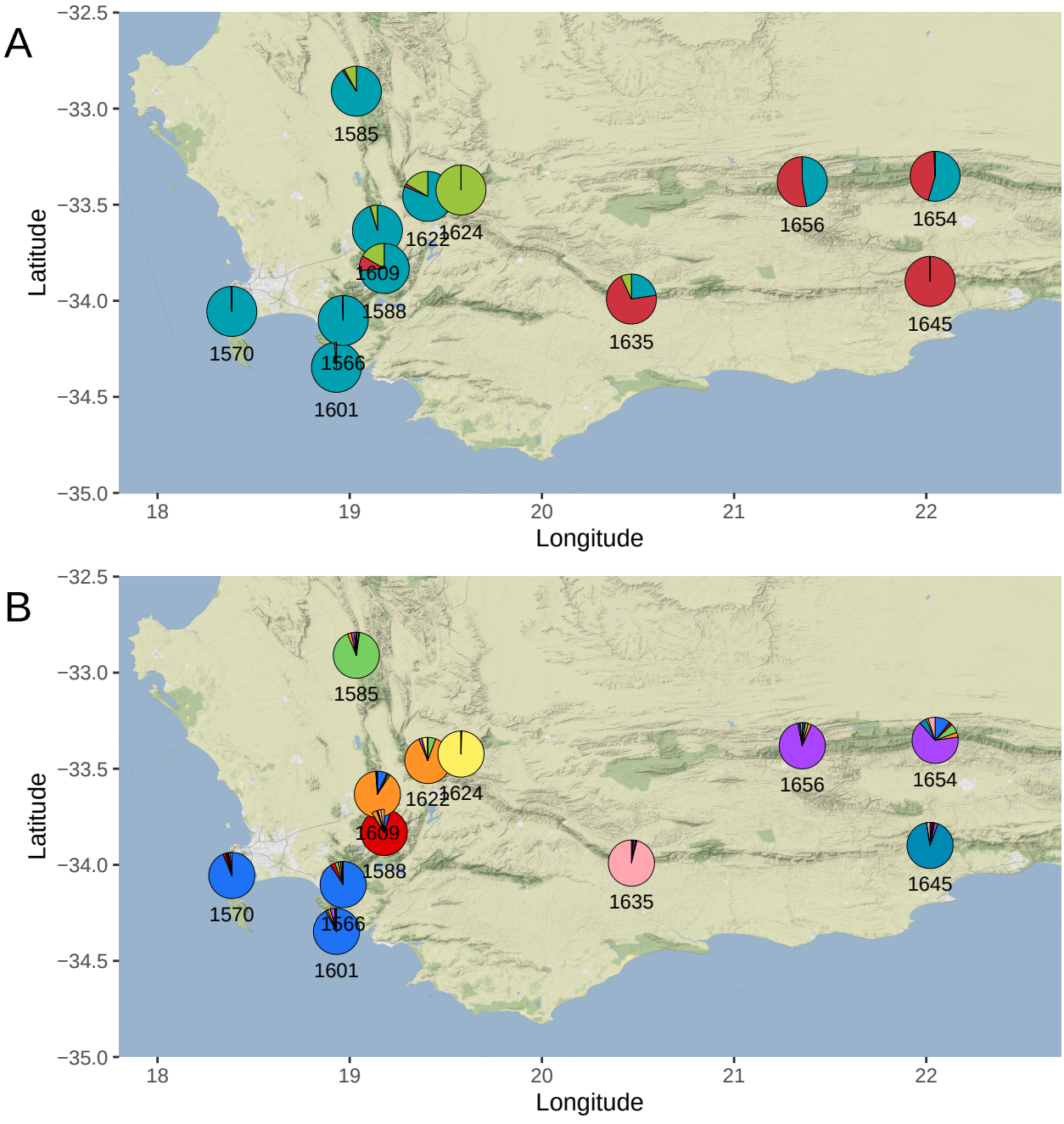

Figure 13: Map showing average sNMF assignments per population for the Ramosa clade at A) K = 3, B) K = 8. Colours correspond to those of Figure 11.

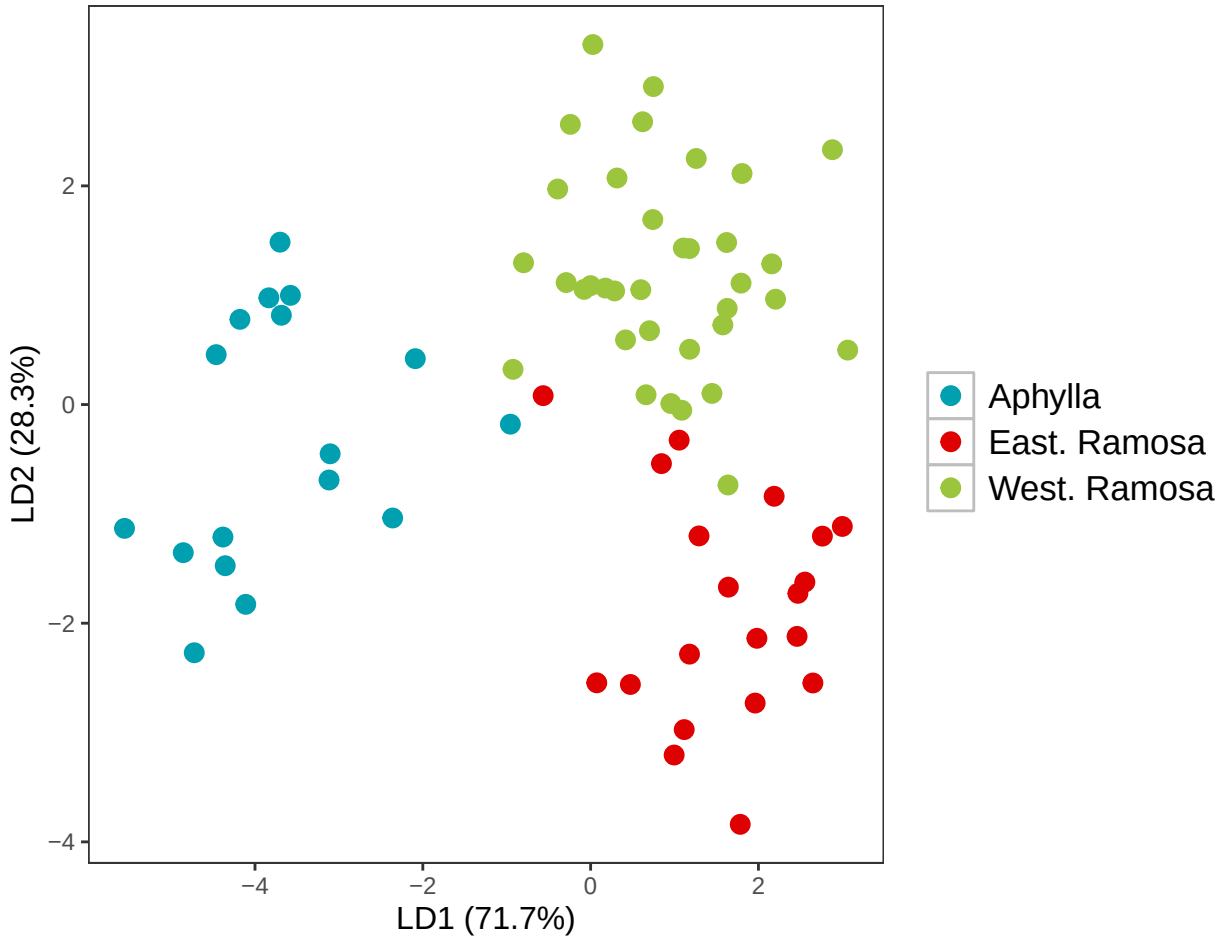

Figure 14: A linear discriminant analysis of morphological data for the Ramosa clade. The 17 included traits are described in Table 1. The brackets show the percentage of explained between-group variance.

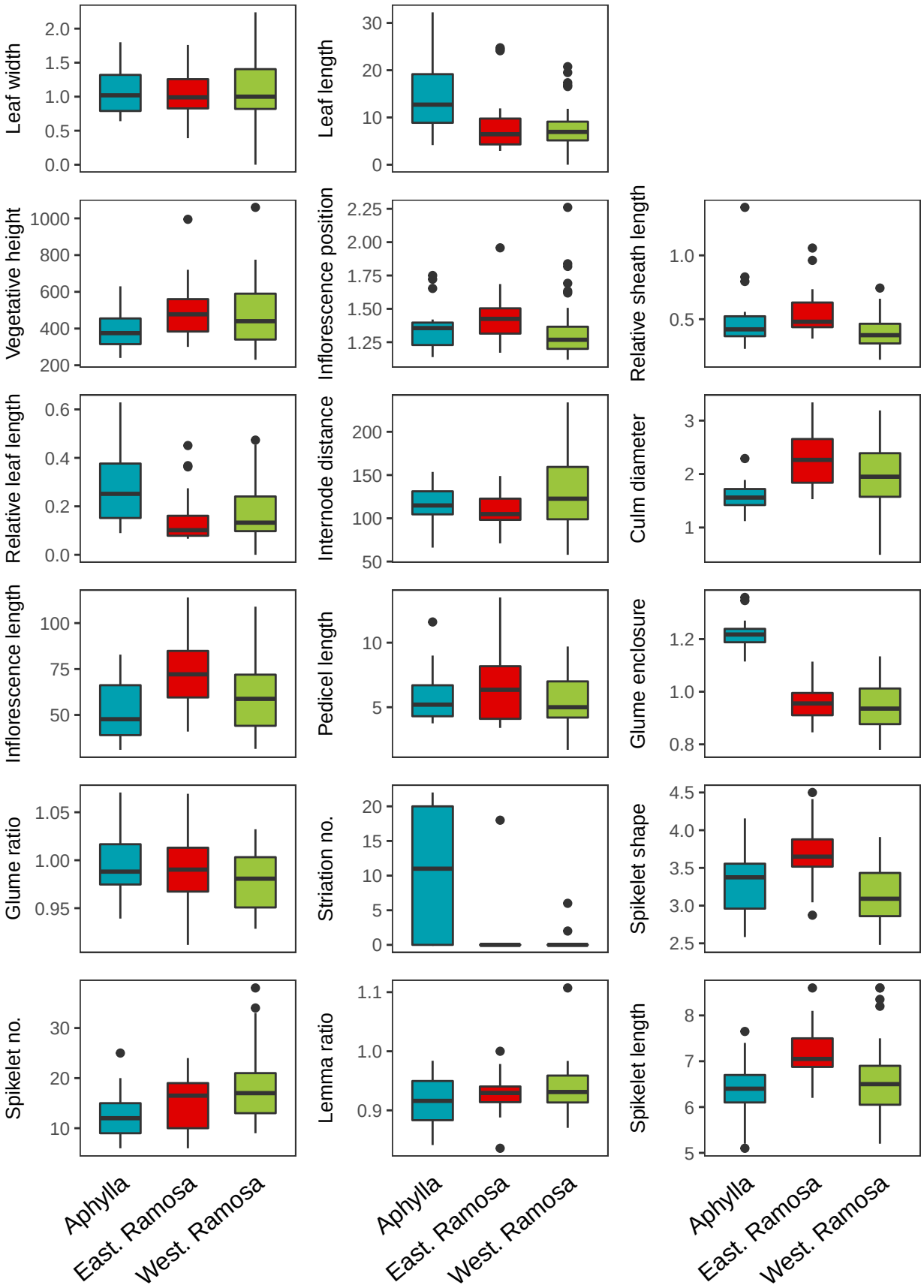

Figure 15: Boxplots of morphological traits used in the linear discriminant analysis for the Ramosa clade. See Table 1 for trait descriptions. Where applicable, measurements are in mm.

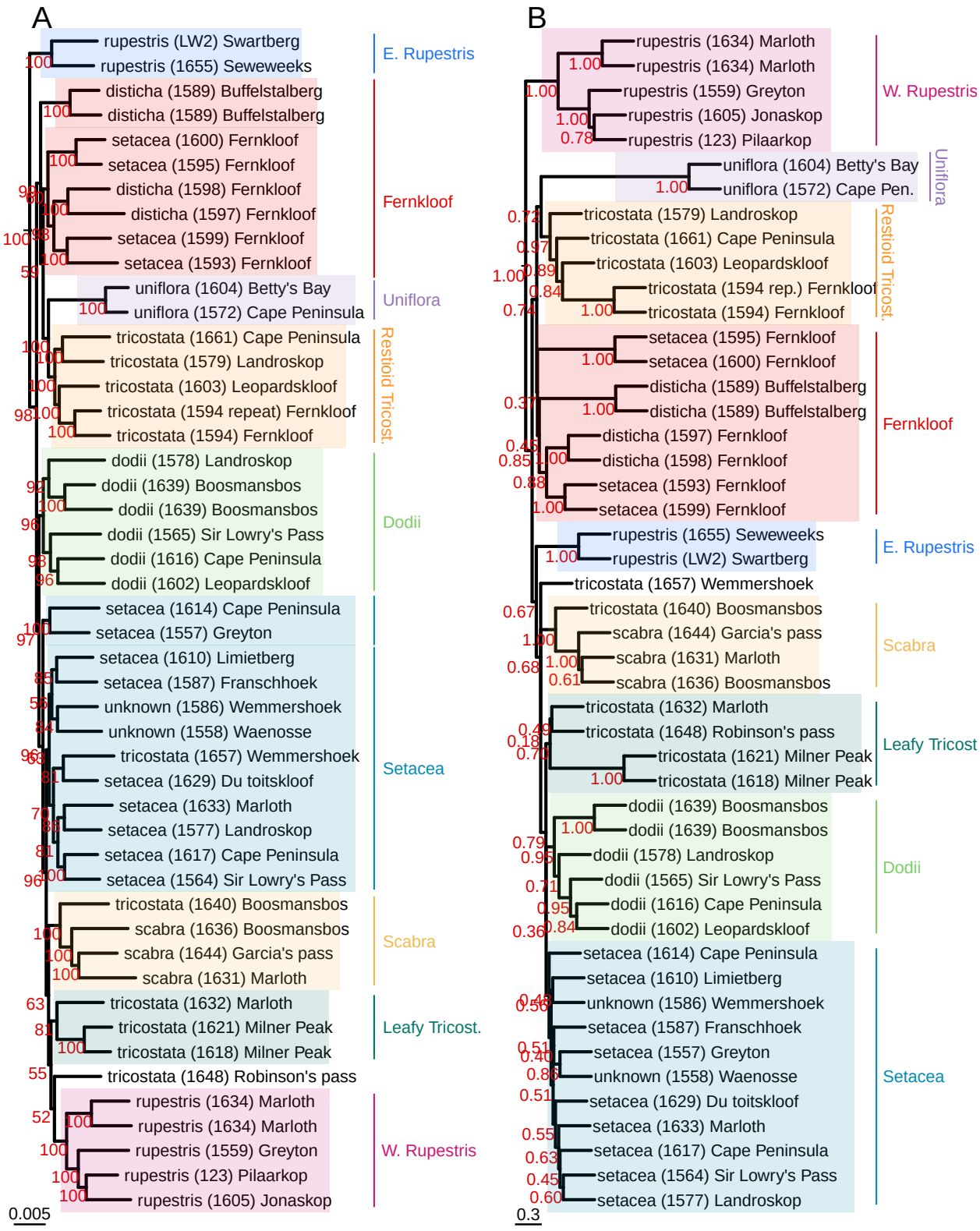

Figure 16: Phylogenetic relationships of the Setacea clade. A) Maximum-likelihood tree constructed using IQTREE with Partitionfinder from 307 concatenated supercontig loci. Red numbers show ultrafast bootstraps. Branch lengths represent number of substitutions. B) ASTRAL tree constructed using IQTREE gene trees from 307 supercontig loci. Branch lengths represent coalescent units (generations/effective population size). Tip branches are an arbitrary length not calculated by ASTRAL. Red numbers represent local posterior probabilities. Leafy Tricost. = Leafy Tricostata, Restioid Tricost. = Restioid Tricostata.

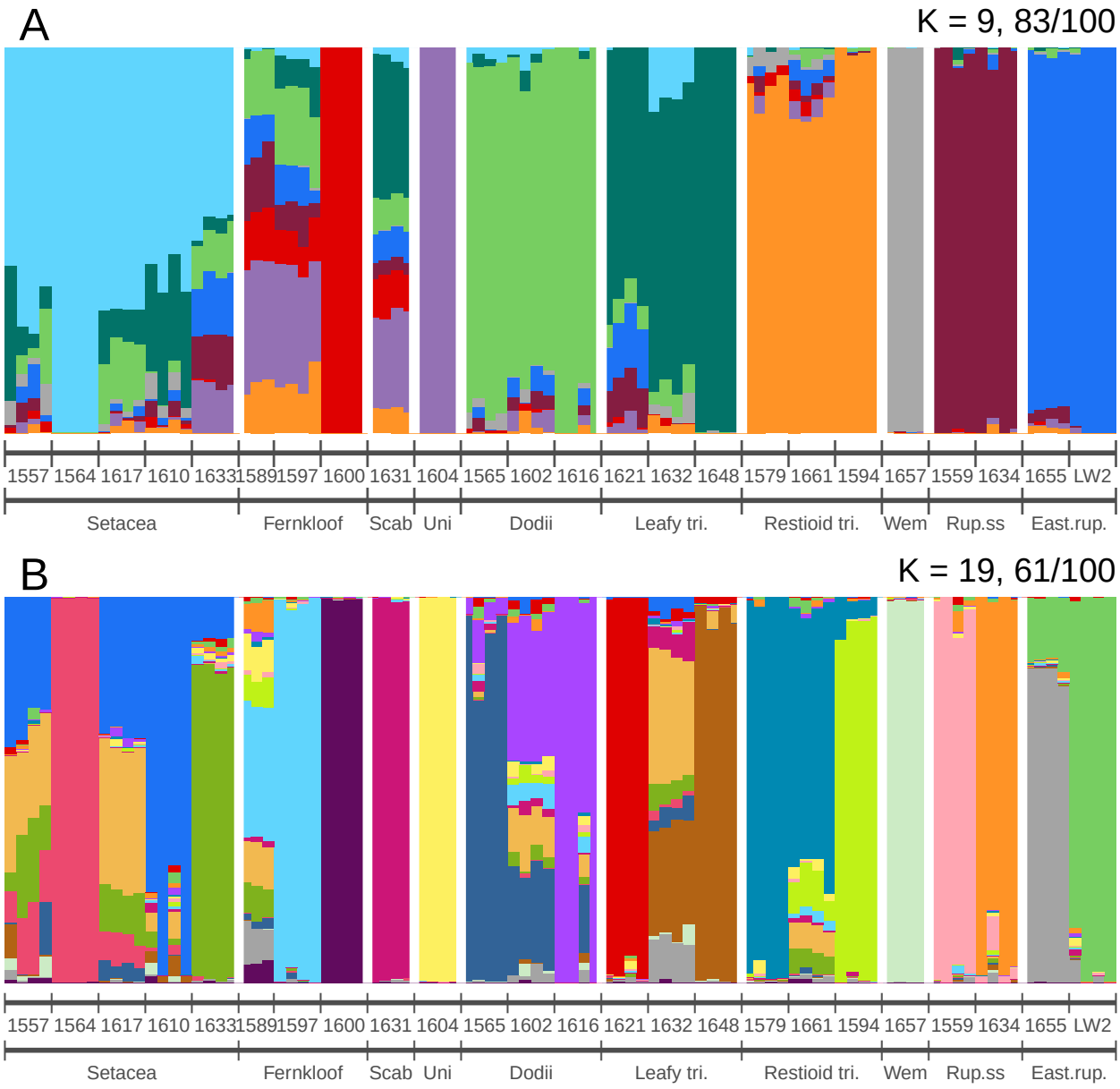

Figure 17: Plots of sNMF genomic assignment of individuals within the Setacea clade for A) K = 9, and B) K = 19. Fractions displayed at the top right of the graphs show the number of times a given sNMF repetition recovered the displayed assignments. Bars below the plots identify population accession numbers and taxon groups.

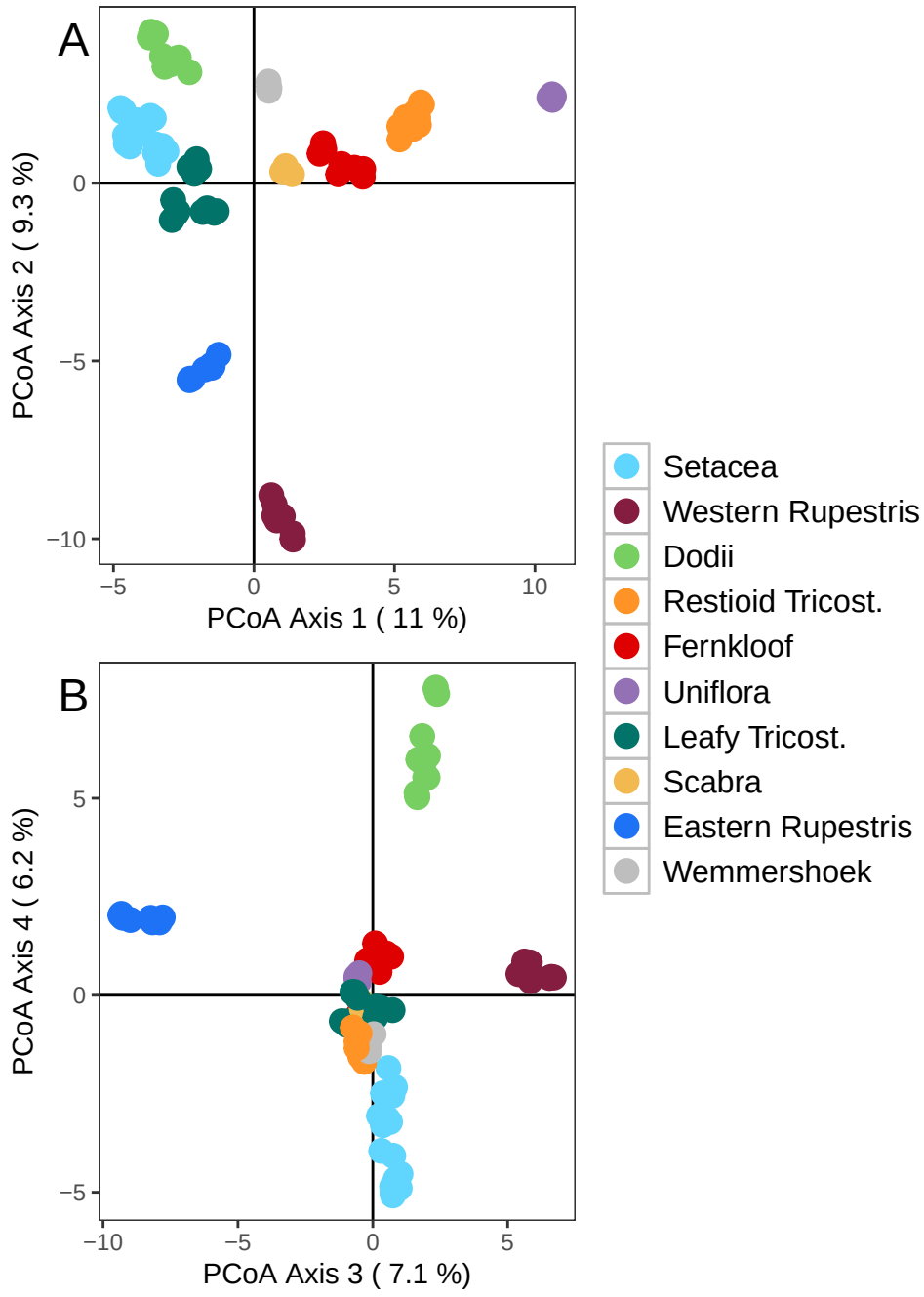

Figure 18: Representation of genetic similarity between individuals in the Setacea clade using a principal coordinates analysis of SNP variants. A) Principal coordinates axes 1 and 2, B) principal coordinates axes 3 and 4. Leafy Tricost. = Leafy Tricostata, Restioid Tricost. = Restioid Tricostata.

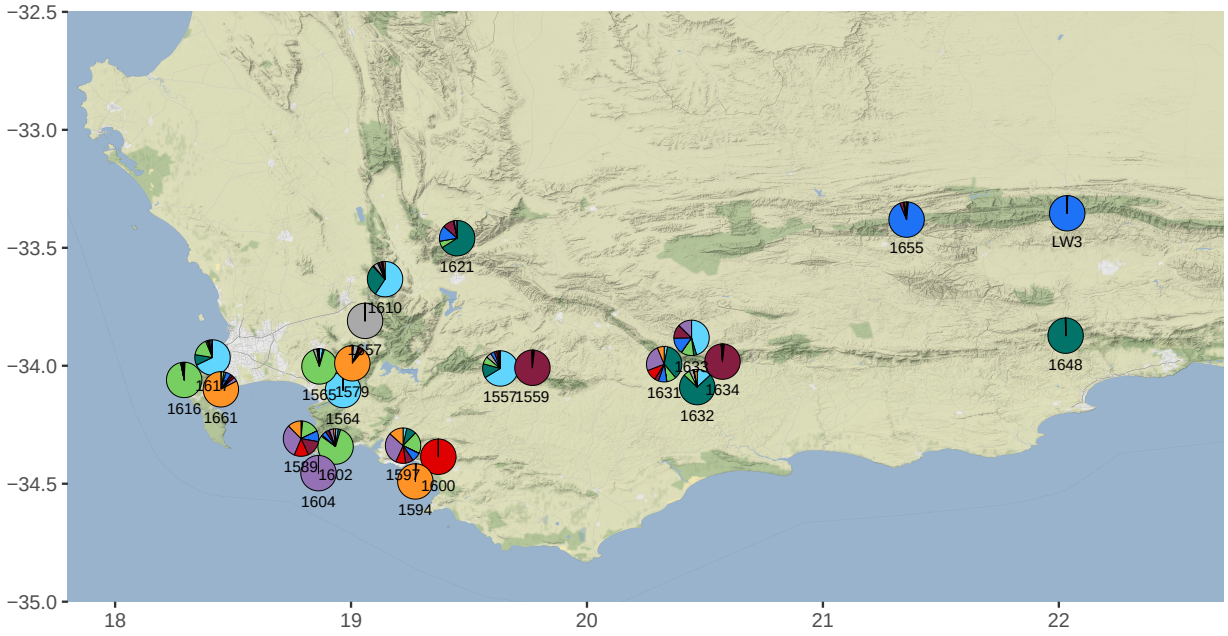

Figure 19: Map showing average sNMF assignments per population at K = 9 for the Setacea clade. Colours correspond to those of Figure 17A.

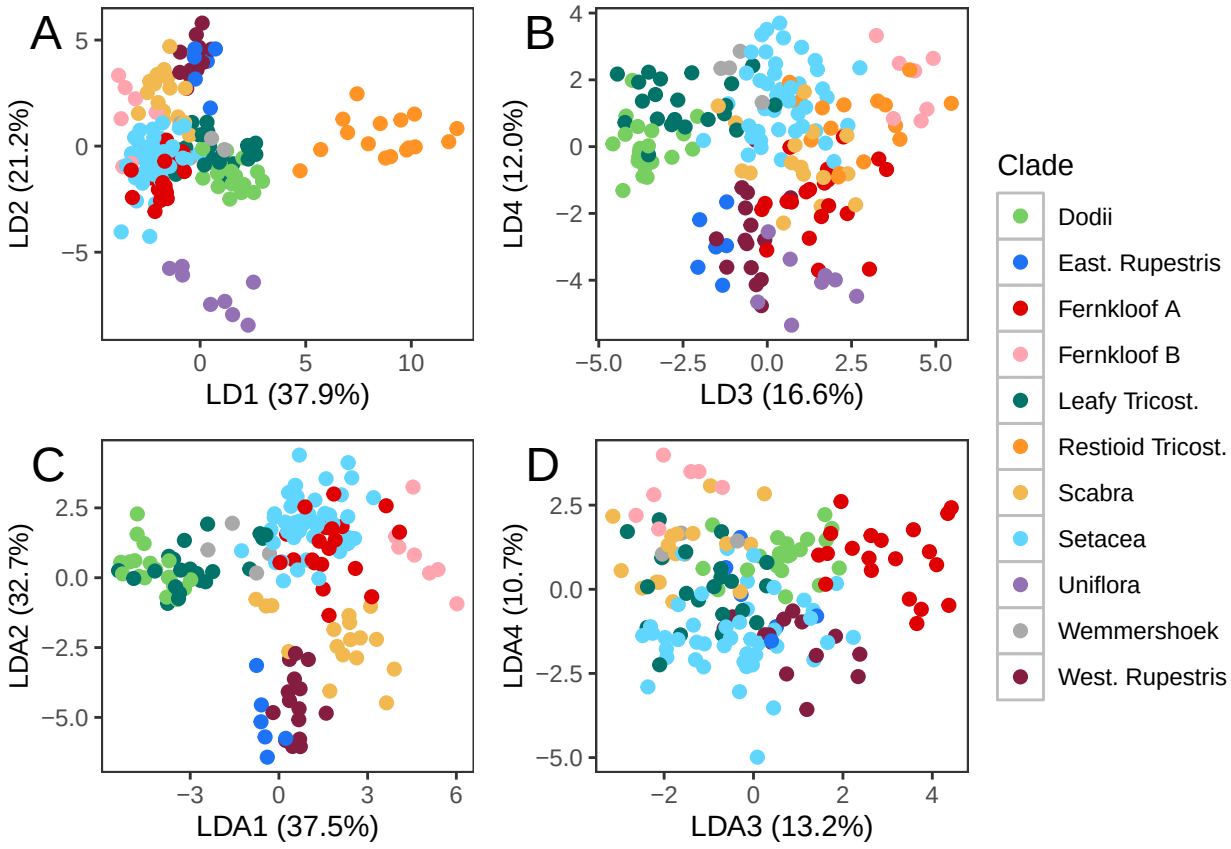

Figure 20: A linear discriminant analysis of the morphological data for the Setacea clade. Traits included are described in Table 1. For the full dataset, plot A) shows axes PC1 and PC2, B) shows PC3 and PC4. Restioid tricostata and Uniflora are removed from the analysis in C) and D). The brackets show the percentage of explained between-group variance.

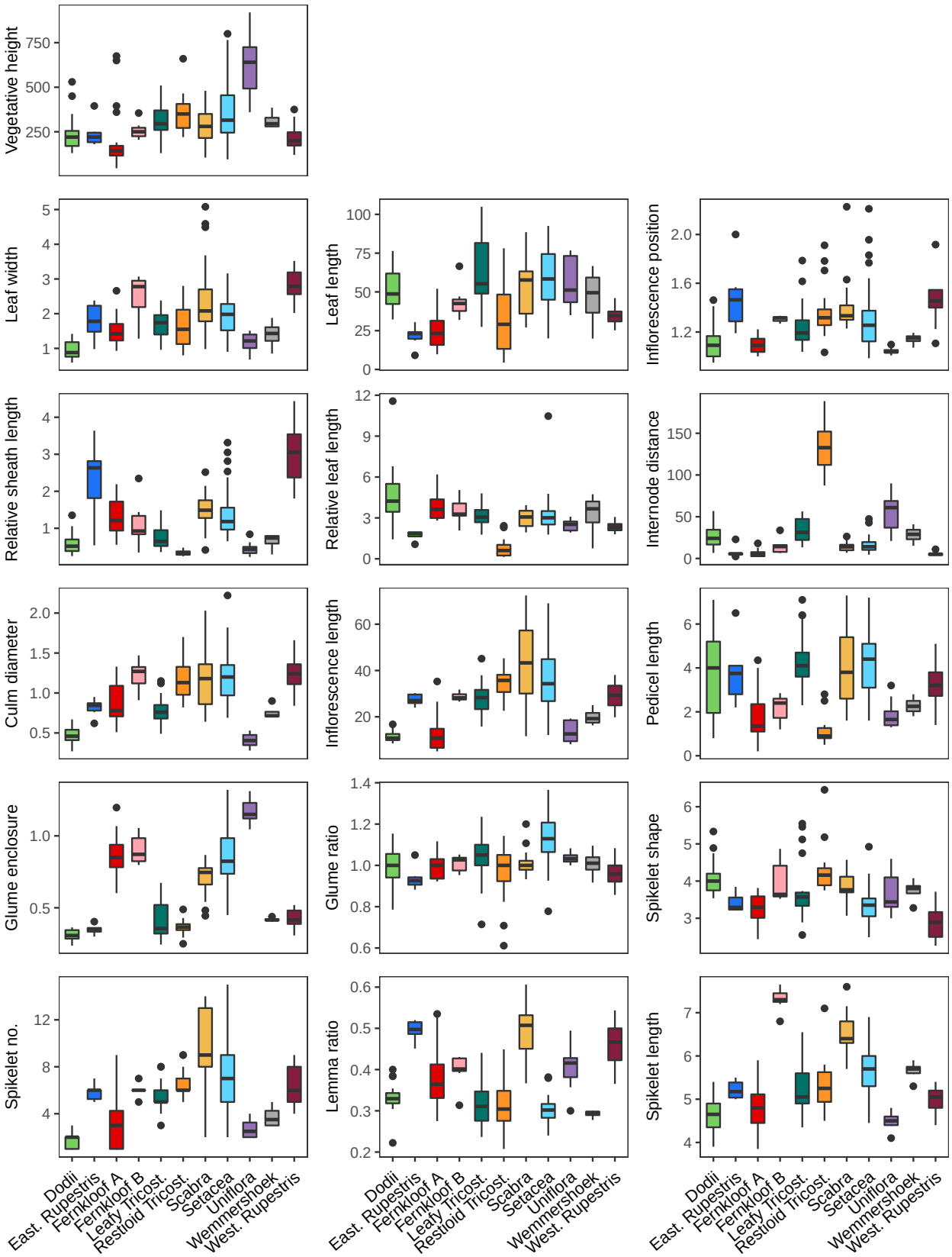

Figure 21: Boxplots of morphological traits used in the linear discriminant analysis for the Setacea clade. See Table 1 for trait descriptions.

Chapter 3: Investigating whether diversification processes are elevation-specific in Cape *Ehrharta*

Introduction

Lineage diversification is often accompanied by an increase in ecological diversity (Givnish, 1997, 2015), with speciation arising as a consequence of ecologically-based divergent selection (i.e. "ecological speciation"; Rundle & Nosil, 2005; Schluter, 2009). Where this is the case, and diversification is substantial, a lineage may be said to have undergone an "adaptive radiation" (Schluter, 2000). Although adaptive radiation has been widely invoked to explain the diversity of species-rich lineages, including Darwin's finches (Grant & Grant, 2008), African cichlids (Seehausen, 2006), and Hawaiian silverswords (Witter & Carr, 1988; Givnish et al., 2009), it is increasingly recognised that not all radiations are adaptive, with many radiation events requiring a more nuanced understanding (Losos & Ricklefs, 2009; Olson & Arroyo-Santos, 2009; Simões et al., 2016). Beyond adaptive divergence, a variety of both intrinsic and extrinsic factors, such as climate change, geographic uplift, pollinator-mediated isolation, dispersal limitation, allopolyploidy and hybridisation, among others, have the potential to influence gene flow patterns and so determine whether a lineage is likely to radiate (Wen et al., 2014; Bouchenak-Khelladi et al., 2015). Of particular interest in montane regions are geographic, or non-adaptive radiations (Kadereit, Griebeler & Comes, 2004; Hughes & Eastwood, 2006; Verboom et al., 2015; Boucher, Zimmermann & Conti, 2016; Ebersbach et al., 2017), in which diversification is triggered by a physical barriers to gene flow (Rundell & Price, 2009; Simões et al., 2016). Speciation in non-adaptive radiations is a consequence of genetic divergence powered by neutral processes such as genetic drift, which play out when conspecific populations become isolated (Rundell & Price,

2009). The resulting taxa generally show little niche differentiation, and have allopatric or parapatric distributions (Gittenberger, 1991; Czekanski-Moir & Rundell, 2019).

Adaptive and non-adaptive radiation events are expected to leave contrasting morphological and genetic signatures within their descendent lineages (Gavrilets & Losos, 2009). Owing to the role of natural selection, lineages diversifying via adaptive radiation typically have substantial morphological variation, as well as elevated rates of morphological and genetic change (e.g. Givnish, 2015; Givnish et al., 2009; Schluter, 2009; Lerner et al., 2011). Conversely, non-adaptive radiation events often result in species that are poorly morphologically differentiated, and that have approximately neutral rates of genetic evolution (Rundell & Price, 2009; Gaut et al., 2011). For the purpose of distinguishing between the two radiation types, rates of morphological change can be quantified and compared between clades, either by hypothesising *a priori* where a rate shift occurs in the tree (O'Meara et al., 2006), or by using model comparisons to locate rate shifts in a way that does not require any *a priori* hypothesis (e.g. Eastman et al., 2011). In addition, the strength and direction of natural selection between clades can be evaluated by comparing the ratio of non-synonymous (dN) to synonymous (dS) substitutions ($dN/dS = \omega$) in protein-coding genes (Yang, 1998; Bielawski & Yang, 2004). A non-synonymous substitution is a mutation that changes an amino acid within a protein, and can therefore experience selection, while a synonymous substitution does not alter the amino acid sequence and so is hidden from selection. Critically, where synonymous substitutions are expected to accumulate at the same rate as the overall genomic mutation rate (Kimura, 1968; but see Chamary, Parmley & Hurst, 2006), the rate of non-synonymous substitutions can be elevated or depressed by divergent and stabilising selection respectively, changing the value of ω. However, determining ω is a non-trivial matter, as it can vary over time, between sites and between lineages (Bielawski, 2013; Yang, 2019).

Adaptive radiation and ecological speciation have undoubtedly played an important role in generating the high floristic richness of the GCFR. The rugged topography of the Cape underpins strong environmental gradients, which together with fine-scale edaphic mosaics, provide the context needed for local adaption and, potentially, ecological speciation (Linder, 1985). Ecological differentiation is thus frequent between closely related plant species (van der Niet & Johnson, 2009), and many of the iconic GCFR radiations, including those of *Erica*, *Protea*, *Phylica*, *Muraltia*, *Pelargonium* and Aizoaceae, are thought to have an ecological component, whether this involves shifts in climate or edaphic preference, pollinator syndrome or fire-survival strategy (Richardson et al., 2001; Klak, Reeves & Hedderson, 2004; Forest et al., 2007; Schnitzler et al., 2011; van der Niet et al., 2014; Pirie et al., 2016). Early researchers, however, envisaged an altogether different role for mountains in fostering diversification of the Cape flora. Adamson (1958) and Goldblatt (1978) speculated that high elevation plant communities were relics of a previously widespread flora that contracted into higher-elevation refuges in the face of an increasingly arid climate, thereby precipitating widespread vicariant speciation. Although variation in the timing of radiation between high-elevation lineages (e.g. Linder & Hardy, 2004; Verboom et al., 2015) contradicts the idea of a synchronous vicariant speciation event, the hypothesis of non-adaptive speciation, driven by isolation-by-distance, in explaining high-elevation diversity has recently regained traction (Britton, Hedderson & Verboom, 2014; Verboom et al., 2015; Shaik, 2019). Clades of montane plants in the Cape often show small and strongly allopatric ranges, with sister species occupying comparable ecological niches (Verboom et al., 2015), which is characteristic of non-adaptive radiations (Rundell & Price, 2009). Crucially, there is also evidence for diversification within ecologically-similar elevational bands, suggesting that factors other than ecological divergence are contributing to the species richness of the Cape (Verboom et al., 2015). Together with evidence of niche equivalency within clades (Latimer et al.,

2009), there is reason to believe that non-adaptive radiations have also played a role in generating the species richness of the GCFR.

The Cape-centred *Ehrharta* clade is an ideal system in which to investigate elevation-specific drivers of diversification in the GCFR. The Lowlands clade is believed to have experienced accelerated speciation during the Late Miocene, this radiation having a strong ecological component and being associated with the lineage's transition into the Succulent Karoo (Verboom, Linder & Stock, 2003, 2004). While the early-diverging lineages of Cape *Ehrharta* are restricted to Table Mountain sandstone, Succulent Karoo taxa have adapted to a variety of soils, including granite (*E. barbinodis* and *E. pusilla*), coastal dunes (*E. villosa*) and shale (e.g. *E. melicoides* and *E. eburnea*). As expected in an adaptive radiation, the Lowlands clade exhibits substantial trait variation, particularly in floral morphology, growth form, and life history (Gibbs-Russell & Ellis, 1987; Verboom, Linder & Stock, 2004). The Lowlands radiation was considered to be the only diversification event within the Cape *Ehrharta*, tying in with the narrative that the majority of GCFR radiations are based on ecological speciation (Verboom, Linder & Stock, 2004). However, the present study provides evidence for multiple new species in the mid- and high-elevation Ramosa-Rehmannii and Setacea clades (Chapter 2), which suggests that there may have been more than one diversification event in the Cape *Ehrharta*. Moreover, the cryptic to semi-cryptic nature of many of the putative new species delimited in Chapter 2 hints at the possibility of a non-adaptive radiation driven by geographic isolation on the archipelagic peaks of the Cape Fold mountains.

The aim of this chapter, then, is to uncover whether different processes are driving diversification events in the low- and high-elevation regions of the GCFR, using the Cape *Ehrharta* as a model system. Incorporating the sequence data generated in Chapter 2, I present a dated, species-level

phylogenetic hypothesis of the Cape *Ehrharta* and use this as a framework for estimating and comparing diversification rates, ω, and phenotypic evolution between high-, mid- and low-elevation clades. Following Verboom et al. (2015), it is hypothesised that stronger population isolation paired with a more uniform selective regime increases the relative influence of neutral processes as a driver of differentiation at high elevations. As such, I predict lower rates of phenotypic evolution and lower ω in the high-elevation Setacea clade relative to those in the Ramosa-Rehmannii and Lowlands clades.

Methods

Dating analyses

Molecular dating made use of the targeted enrichment dataset compiled in Chapter 2, but with reduced taxon sampling. Analyses included a single accession per species in the Ramosa, Rehmannii and Setacea clades, as delimited in Chapter 2, and one accession per taxon (species and subspecies) in the Lowlands and Dura clades. In addition, *O. sativa* was included as an outgroup taxon and *Microlaena stipoides* was included to represent the Australasian clade. The selection of representative accessions was based on sequencing coverage and completeness of morphological data. The possibility of multiple species within the Western Ramosa clade warranted the inclusion of two accessions from this group, while three accessions of *E. calycina*, corresponding to forms ("robust", "gracile" and "Clanwilliam") that likely represent distinct species (Musker, 2013), and accessions of all subspecies of both *E. villosa* and *E. brevifolia* were included, resulting in a total of 42 taxa (Table 1).

Owing to the computational demands of dating methods, running an analysis on the full number of loci was not a viable way forward. Instead, ten sets of 25 supercontig loci were drawn randomly from a pool consisting of the 308 supercontig loci from the 90% coverage dataset of Chapter 2, following the methods of Stubbs et al. (2018). In addition, the gene "shopping" SortaDate package (Smith, Brown & Walker, 2018) was used to choose the 25 "best" and 25 "worst" loci, by generating gene trees for each locus with IQTREE v.1.6.11 (Nguyen et al., 2015), and then ranking the gene trees first by their concordance with the ASTRAL species tree, then by the extent to which their evolution is clock-like, and finally their length (-order 3,1,2). The loci within each set were individually aligned using MAFFT v.7.427 (Katoh & Standley, 2013), with TrimAL

v.1.2 being used to remove gappy regions under the -automated1 option (Capella-Gutiérrez, Silla-Martínez & Gabaldón, 2009). Thereafter, the trimmed loci were concatenated and partitioned using AMAS (Borowiec, 2016). IQTREE was used to find which of the substitution models incorporated in BEAST v1.10.4 (Drummond & Rambaut, 2007; Suchard et al., 2018) were suitable for each locus with the option -mset JC, HKY, GTR, TN93. BEAUTi v1.10.4 (Suchard et al., 2018) was used to configure xml files for input into BEAST. All loci were parameterised using unlinked substitution models corresponding to the best model found by IQTREE, while clock and tree models were linked across the loci. Taxon relationships were constrained to correspond to the topology of the ASTRAL tree in Chapter 2. Two calibration points with normal priors were used based on the fossil-calibrated tree of Prasad et al. (2011), one constraining the split between *O. sativa* and *Ehrharta* at 72.5 Ma (sd = 5.61, 95% CI = [66.89, 78.11]), and the other between *M. stipoides* and the Setacea clade at 20 Ma (sd = 7.65, 95% CI = [12.35, 27.65]). BEAST analyses were run using a birth-death tree prior and relaxed uncorrelated lognormal clock model, sampling parameters every 1000 generations. Four replicate BEAST runs of 50 million generations per loci set were used to confirm convergence and stationarity through visualisation in TRACER v1.7.1 (Rambaut et al., 2018), and a fifth replicate of 250 million generations was run to ensure that the Estimated Sample Size (ESS) for each parameter had an approximate value of >100. The five runs based on each set of loci were combined using LogCombiner v1.10.4 (Suchard et al., 2018) with the first 10 million generations per replicate excluded as burn-in. An exception was the Sortadate 'best' loci set, which took longer to reach stationarity and required 20 million generations per replicate to be discarded as burn-in. TreeAnnotator v1.10.4 (Suchard et al., 2018) was used to construct maximum clade credibility trees and summary statistics. Analyses were conducted either using CIPRES Science Gateway (Miller, Pfeiffer & Schwartz, 2010) or the facilities of the University of Cape Town's ICTS High Performance Computing unit.

Diversification analyses

To test whether rates of diversification differed between clades, BAMM v2.5.0 (Rabosky, 2014) was used to model potential rate shifts. As the mid-elevation Ramosa and Rehmannii lineages are monophyletic, they were treated as a single clade, hereafter referred to as the Ramosa-Rehmannii clade. Additionally, for the diversification analyses, I evaluated a fourth clade (Succulent Karoo clade), consisting of a subset of Lowlands taxa (Table 1), in the attempt to replicate the finding of a diversification rate shift that Verboom, Linder & Stock (2003) associated with the lineage's transition to the Succulent Karoo. Although BAMM can account for incomplete taxon sampling within a tree (Rabosky, 2014), this study was explicitly geared towards assessing diversification patterns within the Cape *Ehrharta*, and I therefore excluded unsampled *Ehrharta* species occurring outside the Cape, such as *E. longiglumis*, as well as *O. sativa* and *M. stipoides* were excluded from the analysis. Priors for each of the twelve BEAST trees were generated using the function "setBAMMpriors" from the R package *BAMMTools* (Rabosky et al., 2014). For all trees, the expected number of diversification rate shifts was set to one, based on evidence for an adaptive radiation within the Lowlands clade (Verboom, Linder & Stock, 2003, 2004). BAMM was run for 3 million generations, sampling every 1000 generations, with run stationarity and ESS being checked using the R package *Coda* (Plummer et al., 2006). The 95% credible set of rate shift configurations was extracted with the function "credibleShiftSet", with the marginal posterior-to-prior odds ratio for detecting a rate shift set to five. In addition, the mean rates of diversification within each of the Lowlands, Succulent Karoo, Ramosa and Setacea clades for each BEAST tree were calculated using the function "getCladeRates" (Rabosky et al., 2014), and significant differences between clades were tested using a Kruskal-Wallis test. To investigate the influence of the postulated new taxa outlined in Chapter 2, the analysis was then repeated with

the BEAST trees trimmed to contain only the set of taxa considered by Verboom, Linder & Stock (2003). For this purpose, only one accession corresponding to each of *E. setacea*, *E. rupestris*, *E. rehmannii*, *E. calycina*, *E. villosa*, and *E. brevifolia* were retained. The BAMM results were corroborated using MEDUSA (model = "mixed", Alfaro et al., 2009), in the R package *geiger* (Pennell et al., 2014).

Rates of morphological change

The non-censored model of O'Meara et al. (2006) was used in order to evaluate whether rates of morphological evolution in the Lowlands, Ramosa and Setacea clades differed from that in the the rest of the tree. This method was chosen as it allows for specific, *a priori* hypothesis testing, rather than searching for rate shifts anywhere on the tree (e.g. Eastman et al., 2011). For each of the twelve BEAST trees, the rate of evolution of a single trait (log-transformed to better fit Brownian motion assumptions (O'Meara et al., 2006)) within one of the three clades was compared to the evolutionary rate within the rest of the tree using the function "transformPhylo.ML" and the model = "clade" argument from the R package *MOTMOT* (Puttick et al., 2020). Sixteen different morphological traits, as described in Chapter 2 and measured on the same individuals as represented in the BEAST trees, were included in the analysis. The variation in selected traits across the tree was visualised using the function "contmap" implemented in the R package *phytools* (Revell, 2012), and a Levene's test for equal variance (Levene, 1960), followed by a Tukey HSD post-hoc test, was used to evaluate the extent to which trait variance differed between clades.

Rate of non-synonymous to synonymous change

The codeml program from the PAML package (v4.9, Yang, 2007) was used to infer synonymous (dS) and non-synonymous (dN) rates of change, and the ratio (dN/dS = ω) between them. First, the loci contained in the 90% coverage exon dataset from Chapter 2 for the Cape *Ehrharta* included in the BEAST analysis were aligned using the codon-aware alignment program MACSE v2.03 (Ranwez et al., 2011). Summary statistics for each alignment were then calculated using AMAS (Borowiec, 2016), and only loci containing <25% gap characters, >20% variable sites, and having no missing taxa were retained for the analysis, leaving a total of 44 loci. The loci were concatenated into a single alignment, and columns containing >5% gaps were removed using TrimAl v.1.2 under the setting -gt 95 (Capella-Gutiérrez, Silla-Martínez & Gabaldón, 2009). As codeml is insensitive to initial branch lengths, a single unrooted BEAST tree was used for all codeml analyses (Yang, 2007). The free-ratios model was implemented to estimate dN, dS and ω for each branch (model = 1, NSsites = 0). The dN, dS and ω of terminal branches were then grouped by clade, and an ANOVA, followed by a Tukey HSD post-hoc test, was used to test for significant differences between clades. In addition, two-rate models were fitted that allowed the ω of foreground branches, in this case all branches within a given clade, to differ from the remaining background branches (model = 2, NSsites = 0, Yang, 1998; Yang & Nielsen, 1998). A likelihood-ratio test was then used to evaluate whether the two-rate models fitted the data significantly better than a the one-rate branch model (model = 0, NSsites = 0), which estimates a single ω across the tree.

Results

Based on the median age estimates obtained from each of the twelve 25-locus sets used for dating (i.e. the 10 randomly sampled sets, plus the 'best' and 'worst' Sortadate sets), the mean age of the split between *Oryza* and *Ehrharta* was estimated at 68.4 Ma, with estimates varying between 67.9 and 69.3 Ma (Fig. 1, Table 2). For the crown node of *Ehrharta*, the mean age estimate was 28.4 Ma, with a range of 26.5 to 29.7 Ma. The crown age of the Lowlands clade was always inferred to be older than those of the Ramosa-Rehmannii and Setacea clades, having a mean of 15.8 Ma, but with substantial variation in the estimates obtained from the different locus sets (12.9 to 18.9 Ma, Table 2). On average, the crown node of the Succulent Karoo clade was 1.4 Myr younger than that of the Lowlands clade (Fig. 1). The Ramosa-Rehmannii clade, originating at 6.2 Ma (range: 3.7 to 9.0 Ma), was, on average, 1.1 Myr older than the Setacea clade, but in two trees (trees 6 and 10) was reconstructed as the younger. Overall, Setacea was the most recent clade, emerging 5.1 Ma (range: 2.9 to 8.9 Ma). On average, the dates inferred using the Sortadate 'worst' locus set were older than those obtained using the other locus sets, although not extremely so. On the other hand, the Sortadate 'best' locus set yielded the youngest or second youngest estimates for all nodes, except for the split between *Oryza* and *Ehrharta*, for which the Sortadate 'best' locus set produced the oldest estimate (Table 2).

With the inclusion of accessions representing each of the taxa delimited in Chapter 2, and one accession per subspecies or species in the Lowlands and Dura clades, BAMM provided little evidence of a significant shift in diversification rate in Cape *Ehrharta* (Fig. 2). For all twelve BEAST trees, the 95% credible set of shift configurations contained no evidence that a shift in diversification rate had occurred, with 100% of samples from the posterior distribution containing zero shifts. Although several trees suggested a slight elevation in diversification rate for the

Setacea clade (e.g. Fig. 2, Fig. 3) a Kruskal-Wallis test revealed no significant between-clade differences ($\chi^2 = 2.74$, df = 3, p = 0.43). Although we had expected (cf. Verboom, Linder & Stock, 2003) the exclusion of all but one accession representing each of *E. setacea*, *E. rupestris*, *E. rehmannii*, *E. calycina*, *E. villosa* and *E. brevifolia* as currently defined (Gibbs-Russell & Ellis, 1987) to reveal a significant acceleration in the Lowlands clade, this was not the case (Fig. 4). As before, none of the trees based on the 12 locus sets yielded a 95% credible set of shift configurations containing any evidence of a rate shift. The MEDUSA results found diversification rate shifts in three of the twelve BEAST trees, all located at the split between the Western Rupestris and the remainder of the Setacea clade (Fig. 5).

The Lowlands clade tended to have a greater morphological trait variance than the other two clades (Fig. 6). Of the nine traits that showed a significant difference in variance between clades, the variance in the Lowlands clade was greater than that in the Setacea and Ramosa-Rehmannii clades in eight and five of the traits respectively (Fig. 6). Mapping trait values on a sample phylogeny revealed greater across-species variation in most traits in the Lowlands and Setacea clades than in the Ramosa-Rehmannii clade (Fig. 7), implying greater evolutionary lability in the former. An analysis of morphological evolutionary rates, in which branch lengths are taken into account, revealed that variation in evolutionary rates is strongly clade specific and independent of actual trait variance (Fig. 8). Relative to the rest of the tree, the Setacea clade showed elevated rates of evolution in the majority of the measured traits, while the converse was true for the Ramosa-Rehmannii clade. The Lowlands clade displayed slow to moderate rates of evolution relative to the rest of the tree for all traits, with the exception of the strongly elevated rates for inner lemma length, outer glume length and spikelet length. Overall, the inferred rates were consistent across the twelve BEAST trees (Fig. 8).

The terminal branches of the Lowlands clade had the highest rates of estimated non-synonymous (dN) and synonymous (dS) changes, followed by those of the Ramosa-Rehmannii and Setacea clades (Fig. 9). Conversely, the terminal branches of Setacea clade had a significantly greater ω than the Lowlands clade ($p < 0.05$, Fig. 9), although the estimated ω for all clades was low at <0.5. For each of the three clades, a branch model allowing the ω of foreground branches to differ from background ω fit the data better than a model with a single ω across the tree (Table 3). Both the Ramosa-Rehmannii and Setacea clades showed a higher ω than the background rate, while Lowlands clade had a lower ω compared to the rest of the tree (Table 3).

Discussion

The data presented here find little support for the hypothesis that the influence of neutral processes in driving diversification in Cape *Ehrharta* becomes progressively more important with increasing elevation. There is no indication of accelerated diversification within the low-elevation Succulent Karoo or Lowlands clades, while there is limited evidence for elevated diversification rate within the high-elevation Setacea clade. In addition, despite the Lowlands clade showing considerable absolute trait variation, morphological evolutionary rates show the opposite pattern to that predicted, with the Setacea clade, rather than in the Lowlands clade, showing most evidence of accelerated trait evolution. Similarly, evidence of positive selection at the nucleotide level is greatest in the Setacea clade. Overall, it appears that diversification in both the Lowlands and Setacea clades has followed an ecological model of speciation (Nosil, 2012), being also accompanied by significant niche shifts, while non-ecological speciation has been more important in the Ramosa-Rehmannii clade.

Establishing age estimates for lineages within Poaceae has been hindered by a scarcity of macrofossils (Strömberg, 2011), leading Prasad et al. (2011) to make use of phytolith microfossils to calibrate their phylogenetic tree of Poaceae. The resulting age estimates of Prasad et al. (2011), which were used here to calibrate both the *Oryza-Ehrharta* (72.5 Ma) and *Ehrharta-Microlaena* splits (20 Ma), generally exceed previously published age estimates for these nodes (see below). Consequently, the age estimates presented in this study for nodes within *Ehrharta*, lie at the older end of the range of possibilities suggested by previous higher-level dating analyses. For example, Vicentini et al. (2008), Bouchenak-Khelladi et al. (2010), and Hoffmann, Verboom & Cotterill (2015) dated the *Oryza-Ehrharta* split at 35, 50 and 25 Ma respectively, in contrast to the 72.5 Ma estimate of Prasad et al. (2011), while Bouchenak-Khelladi et al. (2010) also estimated the

Ehrharta-Microlaena split at 10 Ma. The older date estimates, suggested by the Prasad et al. (2011) calibrations, are difficult to align with current understanding of the climatic, geological and vegetation history of the GCFR. In particular, the results place the origin of the Succulent Karoo clade in the Mid-Miocene, which antedates the onset of a summer-arid climate during the Pliocene (Dupont et al., 2011; Hoffmann, Verboom & Cotterill, 2015) and the origin of the Succulent Karoo vegetation type itself (Verboom et al., 2009). Rescaling the trees obtained in this study, using the above alternative dates for the *Oryza-Ehrharta* divergence, substantially alters the ages of key clades within *Ehrharta* (Table 4), which has implications for understanding the context in which their diversification took place. For example, the rescaled dates place the crown age of the Succulent Karoo clade between 10.6 and 5.3 Ma, thereby coinciding with the suggested timing of the radiation of the succulent karoo flora (Verboom et al., 2009). Under this scenario, the crown age of the Setacea clade is dated to between 3.7 and 1.9 Ma, which coincides more closely with the onset of the Pleistocene. Given the difficulty of assigning grass microfossils to particular lineages (Christin et al., 2014), and the lack of concordance with previous studies, the absolute dates presented in this work should therefore be interpreted with caution. However, the relative consistency of the divergence time estimates, based on the 12 independent 50-locus data sets, suggests that the relative divergence time presented here are remarkably robust.

Overall, diversification within Cape *Ehrharta* appears to have occurred at a relatively constant rate, albeit with some evidence of accelerated diversification within the Setacea clade. This inference holds both for trees containing all species- and subspecies-level taxa included in Cape *Ehrharta* and for trees pruned to contain only those taxa considered by Verboom, Linder & Stock (2003). The uniform diversification rate scenario is inconsistent with the findings of Verboom, Linder & Stock (2003) and Hoffmann, Verboom & Cotterill (2015), who identified a rapid increase in diversification rate on the branch leading to the Succulent Karoo clade. The inconsistency may

be an artefact of the DNA marker choice, since the dating inferences of both Verboom, Linder & Stock (2003) and Hoffmann, Verboom & Cotterill (2015) were based on just two markers, ITS and *trnL-trnF*, in contrast to the inferences presented here which are based on multiple marker sets, each comprising 50 loci. Alternatively, failure to retrieve a diversification rate shift at the base of the Succulent Karoo clade, may reflect the omission of the non-Africa genera, *Microlaena*, *Tetrarrhena* and *Petriella*, which may have reduced the power to detect rate shifts across the tree (e.g. Guyer & Slowinski, 1993; Agapow & Purvis, 2002). However, uniform or even decreasing diversification rate is a common phenomenon in the Greater Cape flora, with clades such as Restionaceae (Linder & Hardy, 2004; Bouchenak-Khelladi & Linder, 2017) and *Leucadendron* (Hoffmann, Verboom & Cotterill, 2015), showing a steady accumulation of lineages over time, and others, such as *Protea*, *Moraea* and *Tetraria* showing diversification slowdowns (Schnitzler et al., 2011; Slingsby, Britton & Verboom, 2014; Hoffmann, Verboom & Cotterill, 2015). These patterns are generally attributed to low extinction rates (Cowling & Lombard, 2002; Linder, 2005, 2008; Valente et al., 2010), associated with climatic buffering during the Pleistocene glacial cycles (Dynesius & Jansson, 2000; Colville et al., 2020). The southwestern GCFR, where the majority of Lowlands and Succulent Karoo clade taxa occur, is thought to have been particularly stable for long periods of time (Colville et al., 2020), potentially facilitating the slow and constant accumulation of species in the Lowlands clade. In the mid- to high-elevation lineages of Cape *Ehrharta* (Setacea, Ramosa-Rehmannii and Dura clades), however, the relative recentness and sudden emergence of contemporary species diversity, which together produce a "broom and handle" tree shape, is suggestive of high rates of extinction (Nee et al., 1994; Crisp & Cook, 2009)), and it is possible that these lineages, of which almost all extant taxa are moisture dependent, underwent substantial extinction with the onset of drier conditions in the Late Miocene-Pliocene (Buerki et al., 2012).

Adaptive radiations are expected to generate high levels of morphological variance between taxa, as the lineages differentiate to fill available niche space (Schluter, 2000). The Lowlands clade, predicted to display the strongest signature of adaptive evolution, did indeed show considerable trait variation, especially in comparison to the Setacea clade. Despite the difficulty of directly inferring a causal link between adaptive processes and trait variation (Reich et al., 2003; Mitchell et al., 2015), Verboom, Linder & Stock (2004) were able to show that the trait values in the Lowlands clade do indeed reflect adaptation to heterogeneous environmental conditions. This result is corroborated by elevated rates of functional trait evolution (leaf length, plant height, lemma length and spikelet size, the latter being a proxy for seed size) in the Lowlands clade in comparison to the rest of the tree, reflecting divergent selection on traits necessary for survival (Cornelissen et al., 2003; Jardine et al., 2020). Conversely, the lower levels of trait variation in the Ramosa-Rehmannii and Setacea clades is consistent with non-adaptive speciation (Rundell & Price, 2009). Absolute trait variance, however, does not take into account evolutionary time, and it is possible that young clades under strong selection may not have had sufficient time to differentiate, potentially hiding support for ecological speciation. This may apply to the Setacea clade which, despite showing generally low trait variance, exhibits substantially higher rates of evolution for most phenotypic traits when compared to the rest of the tree. In contrast, the Ramosa-Rehmannii clade showed striking levels of conservatism in almost all traits, relative to the rest of the tree, suggesting that this clade is under strong stabilising selection or at least has not responded to strong diversifying selection (Losos, 2008). This interpretation of the morphological rate results is, however, contingent on the assumption that trait divergence happens gradually and continuously. Should evolution in these clades follow the punctuated equilibrium model of speciation (Gould & Eldredge, 1993), where pulses of morphological change coincide with speciation events and are followed by long periods of stasis, then the comparisons of absolute trait

variance may be the more appropriate measure of adaptive divergence (McEntee et al., 2018).

None of the clades displayed a signature of positive selection ($\omega > 1$). This is unsurprising, however, since the branch models used in this study estimate ω as an average rate across all sites (Yang, 2007) and the loci sampled are likely to be conserved (Johnson et al., 2019). Nonetheless, it is still informative to use such models to compare ω between clades in order to evaluate relative strength of selection (Buschiazzo et al., 2012; De La Torre et al., 2017). The Setacea clade displayed the highest value of ω (Setacea clade $\omega = 0.23$, background $\omega = 0.18$), therefore showing the strongest evidence for divergent selection. Although this contradicts the hypothesis of speciation being non-adaptive at high-elevations, it is in accordance with the elevated rates of phenotypic evolution observed in the Setacea clade. The Ramosa-Rehmannii clade also showed a slightly higher value of ω compared to the rest of the tree ($\omega = 0.21$ vs $\omega = 0.19$), but this was likely the consequence of a large positive outlier corresponding to Eastern Ramosa (Fig. 10). Surprisingly, although the Lowlands clade displayed higher rates of sequence evolution (dS and dN) than the Setacea and Ramosa-Rehmannii clades, possibly as a consequence of a prevalence of short-lived annual and weak perennial species in the clade (Smith & Donoghue, 2008; Verboom et al., 2012), it displayed the weakest evidence of positive selection, yielding the lowest estimates of ω. This result is unexpected given the high phenotypic variance of this clade, as well as the inferred adaptive mode of its radiation (Verboom, Linder & Stock, 2004). While it is possible that there has been very little divergent selection within the clade, and that the observed trait variance is due to neutral divergence over long periods of time, this seems unlikely given the affinity of Lowlands taxa to their local substrates (Verboom, Linder & Stock, 2004). Alternatively, the analyses used here are insufficiently sensitive or comprehensive enough to find evidence of divergent selection in the clade. For example, selection pressures act heterogeneously across the genome (Wolf et al., 2009; Künstner et al., 2010), and the 44 loci sampled here are most likely not

representative of the genome as a whole. Additionally, divergent selection on just one or two amino acid sites can make a substantial difference to protein function, and this signal is easily lost when using averaged values of ω (Yang, 2007; Weadick & Chang, 2012).

Therefore, the results of this study suggest that, while the radiation mode of Cape *Ehrharta* does vary between clades, the relative influence of divergent selection does not decline as a simple function of increasing elevation as hypothesised by Verboom et al. (2015). While the Lowlands clade does appear to have diversified via an adaptive radiation, with changes in life history and growth form (e.g. annuals, geophytes) providing access to arid environments previously unavailable to *Ehrharta* (Verboom, Linder & Stock, 2004), this diversification was not explosive (Givnish, 2015). Rather, the Lowlands clade seems to have diversified over an extended period, with new species evolving episodically to fill the diversity of niches at low elevations, and low rates of extinction fostering the high standing diversity of the clade (Dynesius & Jansson, 2000). Under some definitions, this is nonetheless considered adaptive radiation (Givnish, 2015). In contrast, the mid-elevation Ramosa-Rehmannii clade is the radiation least likely to have been powered by adaptive divergence. Species in the Ramosa clade occur in similar habitats, consisting of dry, rocky slopes in the fynbos, which, together with limited morphological differentiation, suggests a minimal role for divergent adaptation in their diversification. Instead, low levels of seed set (*pers. obs.*), potentially limited dispersal ability and reliance on clonal reproduction may make the precipitous Cape mountains and deep shale valleys an effective barrier to gene flow in this group (Givnish, 2010; Kisel & Barraclough, 2010). Similarly, although Rehmannii clade species have undergone some ecological differentiation, with Subspicata favouring coastal habitats, Rehmannii forest margin habitats and Filiformis damp sites in fynbos, the limited range overlap shown by these species suggests that speciation may originally have been allopatric and potentially non-ecological. Finally, the high rates of phenotypic and genetic change shown by the

high-elevation Setacea clade suggest that the radiation of this clade has almost certainly involved a substantial ecological component, with the semicryptic nature of some species potentially being a consequence of the recentness of this radiation. There also appears to be strong ecological differentiation within the clade, for example, Western Rupestris occurs on rocky ridges on mountains peaks, Setacea favours seepages, Uniflora is almost hydrophytic in swamps, Dodii grows beneath wet cliffs, and Fernkloof A prefers rocky slopes. It is difficult to pinpoint the exact trigger of diversification in the Setacea clade given the uncertainty in the age estimates. However, it is possible that climate oscillations during the Pleistocene may have been a contributing factor, as many montane clades have undergone substantial diversification during this epoch, with the fluctuating climate causing suitable habitat to expand and contract along mountain slopes, altering population connectivity and gene flow patterns (Mairal et al., 2017; Flantua & Hooghiemstra, 2018; Flantua et al., 2019; Muellner-Riehl et al., 2019).

Substantial effort has gone into understanding and predicting patterns of radiation in the GCFR (e.g. Linder & Hardy, 2004; Linder, 2008; van der Niet & Johnson, 2009; Schnitzler et al., 2011; Verboom et al., 2015; Bouchenak-Khelladi & Linder, 2017). It is clear that the evolutionary history of the Cape flora is complex, with a range of climatic, geological and historical events influencing diversification (Cowling & Lombard, 2002; Verboom et al., 2014; Linder & Verboom, 2015). This study provides corroboration, demonstrating that even within a single lineage of moderate size, multiple intrinsic and extrinsic factors have powered diversification. In so doing, this work contributes to a growing body of research which argues that it is no longer sufficient to characterise a radiation simplistically as 'adaptive' or 'non-adaptive' (Bouchenak-Khelladi et al., 2015; Soulebeau et al., 2015; Simões et al., 2016), especially those occurring in montane regions, where the biotic and abiotic drivers of diversification form particularly complex interactions (Boucher, Zimmermann & Conti, 2016; Lagomarsino et al., 2016; Ebersbach et al., 2017;

Muellner-Riehl et al., 2019).

References

Adamson, R. 1958. The Cape as an ancient African flora. *The Advancement of Science*. 58:1–10.

Agapow, P.M. & Purvis, A. 2002. Power of eight tree shape statistics to detect nonrandom diversification: A comparison by simulation of two models of cladogenesis. *Systematic biology*. 51(6):866–872.

Alfaro, M.E., Santini, F., Brock, C., Alamillo, H., Dornburg, A., Rabosky, D.L., Carnevale, G. & Harmon, L.J. 2009. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proceedings of the National Academy of Sciences*. 106(32):13410–13414.

Bielawski, J.P. 2013. Detecting the signatures of adaptive evolution in protein-coding genes. *Current protocols in molecular biology*. 101(1):19.1. 1–19.1. 21.

Bielawski, J.P. & Yang, Z. 2004. A maximum likelihood method for detecting functional divergence at individual codon sites, with application to gene family evolution. *Journal of Molecular Evolution*. 59(1):121–132.

Borowiec, M.L. 2016. AMAS: A fast tool for alignment manipulation and computing of summary statistics. *PeerJ*. 4:e1660.

Bouchenak-Khelladi, Y., Onstein, R.E., Xing, Y., Schwery, O. & Linder, H.P. 2015. On the complexity of triggering evolutionary radiations. *New Phytologist*. 207(2):313–326.

Bouchenak-Khelladi, Y., Verboom, G.A., Savolainen, V. & Hodkinson, T.R. 2010. Biogeography

of the grasses (Poaceae): A phylogenetic approach to reveal evolutionary history in geographical space and geological time. *Botanical Journal of the Linnean Society*. 162(4):543–557.

Bouchenak-Khelladi, Y. & Linder, H.P. 2017. Frequent and parallel habitat transitions as driver of unbounded radiations in the Cape flora. *Evolution*. 71(11):2548–2561.

Boucher, F.C., Zimmermann, N.E. & Conti, E. 2016. Allopatric speciation with little niche divergence is common among alpine Primulaceae. *Journal of Biogeography*. 43(3):591–602. DOI: 10.1111/jbi.12652.

Britton, M.N., Hedderson, T.A. & Verboom, G.A. 2014. Topography as a driver of cryptic speciation in the high-elevation Cape sedge *Tetraria triangularis* (Boeck.) C. B. Clarke (Cyperaceae: Schoeneae). *Molecular Phylogenetics and Evolution*. 77:96–109. DOI: 10.1016/j.ympev.2014.03.024.

Buerki, S., Jose, S., Yadav, S.R., Goldblatt, P., Manning, J.C. & Forest, F. 2012. Contrasting biogeographic and diversification patterns in two Mediterranean-type ecosystems. *PLoS One*. 7(6):e39377.

Buschiazzo, E., Ritland, C., Bohlmann, J. & Ritland, K. 2012. Slow but not low: Genomic comparisons reveal slower evolutionary rate and higher dN/dS in conifers compared to angiosperms. *BMC evolutionary biology*. 12(1):8.

Capella-Gutiérrez, S., Silla-Martínez, J.M. & Gabaldón, T. 2009. TrimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 25(15):1972–1973. DOI: 10.1093/bioinformatics/btp348.

Chamary, J., Parmley, J.L. & Hurst, L.D. 2006. Hearing silence: Non-neutral evolution at

synonymous sites in mammals. *Nature Reviews Genetics*. 7(2):98–108.

Christin, P.-A., Spriggs, E., Osborne, C.P., Strömberg, C.A., Salamin, N. & Edwards, E.J. 2014. Molecular dating, evolutionary rates, and the age of the grasses. *Systematic biology*. 63(2):153–165.

Colville, J.F., Beale, C.M., Forest, F., Altwegg, R., Huntley, B. & Cowling, R.M. 2020. Plant richness, turnover, and evolutionary diversity track gradients of stability and ecological opportunity in a megadiversity center. *Proceedings of the National Academy of Sciences*. 117(33):20027–20037.

Cornelissen, J., Lavorel, S., Garnier, E., Dı'az, S., Buchmann, N., Gurvich, D., Reich, P., ter Steege, H., et al. 2003. A handbook of protocols for standardised and easy measurement of plant functional traits worldwide. *Australian journal of Botany*. 51(4):335–380.

Cowling, R. & Lombard, A. 2002. Heterogeneity, speciation/extinction history and climate: Explaining regional plant diversity patterns in the Cape Floristic Region. *Diversity and Distributions*. 8(3):163–179.

Crisp, M.D. & Cook, L.G. 2009. Explosive radiation or cryptic mass extinction? Interpreting signatures in molecular phylogenies. *Evolution: International Journal of Organic Evolution*. 63(9):2257–2265.

Czekanski-Moir, J.E. & Rundell, R.J. 2019. The ecology of nonecological speciation and nonadaptive radiations. *Trends in ecology & evolution*. 34(5):400–415.

De La Torre, A.R., Li, Z., Van de Peer, Y. & Ingvarsson, P.K. 2017. Contrasting rates of molecular evolution and patterns of selection among gymnosperms and flowering plants.

Molecular Biology and Evolution. 34(6):1363–1377.

Drummond, A.J. & Rambaut, A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC evolutionary biology*. 7(1):1–8.

Dupont, L.M., Linder, H.P., Rommerskirchen, F. & Schefuß, E. 2011. Climate-driven rampant speciation of the Cape flora. *Journal of Biogeography*. 38(6):1059–1068.

Dynesius, M. & Jansson, R. 2000. Evolutionary consequences of changes in species' geographical distributions driven by Milankovitch climate oscillations. *Proceedings of the National Academy of Sciences*. 97(16):9115–9120.

Eastman, J.M., Alfaro, M.E., Joyce, P., Hipp, A.L. & Harmon, L.J. 2011. A novel comparative method for identifying shifts in the rate of character evolution on trees. *Evolution: International Journal of Organic Evolution*. 65(12):3578–3589.

Ebersbach, J., Schnitzler, J., Favre, A. & Muellner-Riehl, A. 2017. Evolutionary radiations in the species-rich mountain genus *Saxifraga* L. *BMC evolutionary biology*. 17(1):119.

Flantua, S.G., O'Dea, A., Onstein, R.E., Giraldo, C. & Hooghiemstra, H. 2019. The flickering connectivity system of the north Andean páramos. *Journal of Biogeography*. 46(8):1808–1825.

Flantua, S.G. & Hooghiemstra, H. 2018. Historical connectivity and mountain biodiversity. In *Mountains, climate and biodiversity*. 1st ed. C. Hoorn, A. Perrigio, & A. Antonelli, Eds. Wiley-Blackwell Oxford, UK. 171–185.

Forest, F., Nänni, I., Chase, M.W., Crane, P.R. & Hawkins, J.A. 2007. Diversification of a large genus in a continental biodiversity hotspot: Temporal and spatial origin of *Muraltia*

(Polygalaceae) in the Cape of South Africa. *Molecular phylogenetics and evolution.* 43(1):60–74.

Gaut, B., Yang, L., Takuno, S., Eguiarte, L.E. & others. 2011. The patterns and causes of variation in plant nucleotide substitution rates. *Annual Review of Ecology, Evolution and Systematics.* 42:245–266.

Gavrilets, S. & Losos, J.B. 2009. Adaptive radiation: Contrasting theory with data. *Science.* 323(5915):732–737.

Gibbs-Russell, G.E. & Ellis, R.P. 1987. Species groups in the genus *Ehrharta* (Poaceae) in southern Africa. *Bothalia.* 17(1):51–65.

Gittenberger, E. 1991. What about non-adaptive radiation? *Biological Journal of the Linnean Society.* 43(4):263–272.

Givnish, T. 1997. Adaptive radiation and molecular systematics: Issues and approaches. In *Molecular evolution and adaptive radiation.* 1st ed. Cambridge University Press. 1–54.

Givnish, T.J. 2010. Ecology of plant speciation. *Taxon.* 59(5):1326–1366.

Givnish, T.J. 2015. Adaptive radiation versus 'radiation'and "explosive diversification": Why conceptual distinctions are fundamental to understanding evolution. *New Phytologist.* 207(2):297–303.

Givnish, T.J., Millam, K.C., Mast, A.R., Paterson, T.B., Theim, T.J., Hipp, A.L., Henss, J.M., Smith, J.F., et al. 2009. Origin, adaptive radiation and diversification of the Hawaiian lobeliads (Asterales: Campanulaceae). *Proceedings of the Royal Society B: Biological Sciences.* 276(1656):407–416.

Goldblatt, P. 1978. An analysis of the flora of southern Africa: Its characteristics, relationships, and orgins. *Annals of the Missouri Botanical Garden*. 369–436.

Gould, S.J. & Eldredge, N. 1993. Punctuated equilibrium comes of age. *Nature*. 366(6452):223–227.

Grant, P.R. & Grant, B.R. 2008. *How and why species multiply: The radiation of Darwin's finches*. 1st ed. Princeton University Press.

Guyer, C. & Slowinski, J.B. 1993. Adaptive radiation and the topology of large phylogenies. *Evolution*. 47(1):253–263.

Hoffmann, V., Verboom, G.A. & Cotterill, F.P.D. 2015. Dated plant phylogenies resolve Neogene climate and landscape evolution in the Cape Floristic Region. *PLoS One*. 10(9):e0137847.

Hughes, C. & Eastwood, R. 2006. Island radiation on a continental scale: Exceptional rates of plant diversification after uplift of the Andes. *Proceedings of the National Academy of Sciences*. 103(27):10334–10339.

Jardine, E.C., Thomas, G.H., Forrestel, E.J., Lehmann, C.E. & Osborne, C.P. 2020. The global distribution of grass functional traits within grassy biomes. *Journal of Biogeography*. 47(3):553–565.

Johnson, M.G., Pokorny, L., Dodsworth, S., Botigue, L.R., Cowan, R.S., Devault, A., Eiserhardt, W.L., Epitawalage, N., et al. 2019. A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-medoids clustering. *Systematic Biology*. 68(4):594–606.

Kadereit, J.W., Griebeler, E.M. & Comes, H.P. 2004. Quaternary diversification in European alpine plants: Pattern and process. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*. 359(1442):265–274. DOI: 10.1098/rstb.2003.1389.

Katoh, K. & Standley, D.M. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular biology and evolution*. 30(4):772–780.

Kimura, M. 1968. Evolutionary rate at the molecular level. *Nature*. 217(5129):624–626.

Kisel, Y. & Barraclough, T.G. 2010. Speciation has a spatial scale that depends on levels of gene flow. *The American Naturalist*. 175(3):316–334.

Klak, C., Reeves, G. & Hedderson, T. 2004. Unmatched tempo of evolution in Southern African semi-desert ice plants. *Nature*. 427(6969):63–65.

Künstner, A., Wolf, J.B., Backström, N., Whitney, O., Balakrishnan, C.N., Day, L., Edwards, S.V., Janes, D.E., et al. 2010. Comparative genomics based on massive parallel transcriptome sequencing reveals patterns of substitution and selection across 10 bird species. *Molecular ecology*. 19:266–276.

Lagomarsino, L.P., Condamine, F.L., Antonelli, A., Mulch, A. & Davis, C.C. 2016. The abiotic and biotic drivers of rapid diversification in Andean bellflowers (Campanulaceae). *New Phytologist*. 210(4):1430–1442.

Latimer, A.M., Silander, J.A., Rebelo, A.G. & Midgley, G.F. 2009. Experimental biogeography: The role of environmental gradients in high geographic diversity in Cape Proteaceae. *Oecologia*. 160(1):151–162. DOI: 10.1007/s00442-009-1275-3.

Lerner, H.R., Meyer, M., James, H.F., Hofreiter, M. & Fleischer, R.C. 2011. Multilocus resolution of phylogeny and timescale in the extant adaptive radiation of Hawaiian honeycreepers. *Current Biology*. 21(21):1838–1844.

Levene, H. 1960. Robust tests for equality of variances. In *Contributions to probability and statistics*. 1st ed. I. Olkin, Ed. Stanford University Press.

Linder, H.P. 1985. Gene flow, speciation, and species diversity patterns in a species-rich area: The Cape Flora. *Species and speciation*. 4:53–7.

Linder, H.P. 2005. Evolution of diversity: The Cape flora. *Trends in Plant Science*. 10(11):536–541. DOI: 10.1016/j.tplants.2005.09.006.

Linder, H.P. 2008. Plant species radiations: Where, when, why? *Philosophical Transactions of the Royal Society B: Biological Sciences*. 363(1506):3097–3105.

Linder, H.P. & Hardy, C.R. 2004. Evolution of the species–rich Cape flora. *Philosophical Transactions of the Royal Society of London.Series B: Biological Sciences*. 359(1450):1623–1632.

Linder, H.P. & Verboom, G.A. 2015. The evolution of regional species richness: The history of the southern African flora. *Annual Review of Ecology, Evolution, and Systematics*. 46:393–412.

Losos, J.B. 2008. Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. *Ecology letters*. 11(10):995–1003.

Losos, J.B. & Ricklefs, R.E. 2009. Adaptation and diversification on islands. *Nature*. 457(7231):830–836.

Mairal, M., Sanmartín, I., Herrero, A., Pokorny, L., Vargas, P., Aldasoro, J.J. & Alarcón, M. 2017. Geographic barriers and Pleistocene climate change shaped patterns of genetic variation in the Eastern Afromontane biodiversity hotspot. *Scientific Reports*. 7:45749.

McEntee, J.P., Tobias, J.A., Sheard, C. & Burleigh, J.G. 2018. Tempo and timing of ecological trait divergence in bird speciation. *Nature ecology & evolution*. 2(7):1120–1127.

Miller, M.A., Pfeiffer, W. & Schwartz, T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *2010 gateway computing environments workshop (GCE)*. 1–8.

Mitchell, N., Moore, T.E., Mollmann, H.K., Carlson, J.E., Mocko, K., Martinez-Cabrera, H., Adams, C., Silander Jr, J.A., et al. 2015. Functional traits in parallel evolutionary radiations and trait-environment associations in the Cape Floristic Region of South Africa. *The American Naturalist*. 185(4):525–537.

Muellner-Riehl, A.N., Schnitzler, J., Kissling, W.D., Mosbrugger, V., Rijsdijk, K.F., Seijmonsbergen, A.C., Versteegh, H. & Favre, A. 2019. Origins of global mountain plant biodiversity: Testing the "mountain-geobiodiversity hypothesis". *Journal of Biogeography*. 46(12):2826–2838.

Musker, S. 2013. Polyploid speciation in the Greater Cape Floristic Region: Species limits within *Ehrharta calycina*. Honours thesis. University of Cape Town.

Nee, S., Holmes, E.C., May, R.M. & Harvey, P.H. 1994. Extinction rates can be estimated from molecular phylogenies. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*. 344(1307):77–82.

Nguyen, L., Schmidt, H.A., Haeseler, A. von & Minh, B.Q. 2015. IQ-TREE: A fast and effective

stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular biology and evolution*. 32(1):268–274. DOI: 10.1093/molbev/msu300.

Nosil, P. 2012. *Ecological speciation*. 1st ed. Oxford University Press.

Olson, M.E. & Arroyo-Santos, A. 2009. Thinking in continua: Beyond the "adaptive radiation" metaphor. *BioEssays*. 31(12):1337–1346.

O'Meara, B.C., Ané, C., Sanderson, M.J. & Wainwright, P.C. 2006. Testing for different rates of continuous trait evolution using likelihood. *Evolution*. 60(5):922–933.

Pennell, M.W., Eastman, J.M., Slater, G.J., Brown, J.W., Uyeda, J.C., FitzJohn, R.G., Alfaro, M.E. & Harmon, L.J. 2014. Geiger v2. 0: An expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. *Bioinformatics*. 30(15):2216–2218.

Pirie, M., Oliver, E., Kuppler, A.M. de, Gehrke, B., Le Maitre, N., Kandziora, M. & Bellstedt, D. 2016. The biodiversity hotspot as evolutionary hot-bed: Spectacular radiation of *Erica* in the Cape Floristic Region. *BMC evolutionary biology*. 16(1):190.

Plummer, M., Best, N., Cowles, K. & Vines, K. 2006. CODA: convergence diagnosis and output analysis for mcmc. *R News*. 6(1):7–11. Available: https://journal.r-project.org/archive/.

Prasad, V., Strömberg, C., Leaché, A., Samant, B., Patnaik, R., Tang, L., Mohabey, D., Ge, S., et al. 2011. Late Cretaceous origin of the rice tribe provides evidence for early diversification in Poaceae. *Nature communications*. 2(1):1–9.

Puttick, M.N., Ingram, T., Clarke, M. & Thomas, G.H. 2020. MOTMOT: Models of trait macroevolution on trees (an update). *Methods in Ecology and Evolution*. 11(3):464–471.

Rabosky, D.L. 2014. Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PloS one*. 9(2):e89543.

Rabosky, D., Grundler, M., Anderson, C., Title, P., Shi, J., Brown, J., Huang, H. & Larson, J. 2014. BAMMtools: An R package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods in Ecology and Evolution*. 5:701–707.

Rambaut, A., Drummond, A.J., Xie, D., Baele, G. & Suchard, M.A. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic biology*. 67(5):901.

Ranwez, V., Harispe, S., Delsuc, F. & Douzery, E.J. 2011. MACSE: Multiple Alignment of Coding SEquences accounting for frameshifts and stop codons. *PloS one*. 6(9):e22594.

Reich, P.B., Wright, I., Cavender-Bares, J., Craine, J., Oleksyn, J., Westoby, M. & Walters, M. 2003. The evolution of plant functional variation: Traits, spectra, and strategies. *International Journal of Plant Sciences*. 164(S3):S143–S164.

Revell, L.J. 2012. Phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*. 3:217–223.

Richardson, J.E., Weitz, F.M., Fay, M.F., Cronk, Q.C., Linder, H.P., Reeves, G. & Chase, M.W. 2001. Rapid and recent origin of species richness in the Cape flora of South Africa. *Nature*. 412(6843):181–183.

Rundell, R.J. & Price, T.D. 2009. Adaptive radiation, nonadaptive radiation, ecological speciation and nonecological speciation. *Trends in Ecology & Evolution*. 24(7):394–399. DOI: 10.1016/j.tree.2009.02.007.

Rundle, H.D. & Nosil, P. 2005. Ecological speciation. *Ecology Letters*. 8(3):336–352.

Schluter, D. 2000. *The ecology of adaptive radiation*. 1st ed. Oxford University Press.

Schluter, D. 2009. Evidence for ecological speciation and its alternative. *Science*. 323(5915):737–741.

Schnitzler, J., Barraclough, T.G., Boatwright, J.S., Goldblatt, P., Manning, J.C., Powell, M.P., Rebelo, T. & Savolainen, V. 2011. Causes of plant diversification in the Cape biodiversity hotspot of South Africa. *Systematic Biology*. 60(3):343–357.

Seehausen, O. 2006. African cichlid fish: A model system in adaptive radiation research. *Proceedings of the Royal Society B: Biological Sciences*. 273(1597):1987–1998.

Shaik, Z. 2019. Species delimitation and speciation process in the *Seriphium plumosum* L. complex (Gnaphalieae: Asteraceae) in South Africa. Master's thesis. University of Cape Town.

Simões, M., Breitkreuz, L., Alvarado, M., Baca, S., Cooper, J.C., Heins, L., Herzog, K. & Lieberman, B.S. 2016. The evolving theory of evolutionary radiations. *Trends in ecology & evolution*. 31(1):27–34.

Slingsby, J.A., Britton, M.N. & Verboom, G.A. 2014. Ecology limits the diversity of the cape flora: Phylogenetics and diversification of the genus *Tetraria*. *Molecular Phylogenetics and Evolution*. 72:61–70.

Smith, S.A., Brown, J.W. & Walker, J.F. 2018. So many genes, so little time: A practical approach to divergence-time estimation in the genomic era. *PloS one*. 13(5):e0197433.

Smith, S.A. & Donoghue, M.J. 2008. Rates of molecular evolution are linked to life history in

flowering plants. *Science*. 322(5898):86–89.

Soulebeau, A., Aubriot, X., Gaudeul, M., Rouhan, G., Hennequin, S., Haevermans, T., Dubuisson, J. & Jabbour, F. 2015. The hypothesis of adaptive radiation in evolutionary biology: Hard facts about a hazy concept. *Organisms Diversity & Evolution*. 15(4):747–761.

Strömberg, C.A. 2011. Evolution of grasses and grassland ecosystems. *Annual review of Earth and planetary sciences*. 39:517–544.

Stubbs, R.L., Folk, R.A., Xiang, C.-L., Soltis, D.E. & Cellinese, N. 2018. Pseudo-parallel patterns of disjunctions in an Arctic-alpine plant lineage. *Molecular phylogenetics and evolution*. 123:88–100.

Suchard, M.A., Lemey, P., Baele, G., Ayres, D.L., Drummond, A.J. & Rambaut, A. 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus evolution*. 4(1):vey016.

Valente, L.M., Reeves, G., Schnitzler, J., Mason, I.P., Fay, M.F., Rebelo, T.G., Chase, M.W. & Barraclough, T.G. 2010. Diversification of the African genus *Protea* (Proteaceae) in the Cape biodiversity hotspot and beyond: Equal rates in different biomes. *Evolution: International Journal of Organic Evolution*. 64(3):745–760.

van der Niet, T., Pirie, M.D., Shuttleworth, A., Johnson, S.D. & Midgley, J.J. 2014. Do pollinator distributions underlie the evolution of pollination ecotypes in the Cape shrub *Erica plukenetii*? *Annals of botany*. 113(2):301–316.

van der Niet, T. & Johnson, S.D. 2009. Patterns of plant speciation in the Cape floristic region. *Molecular Phylogenetics and Evolution*. 51(1):85–93. DOI: 10.1016/j.ympev.2008.11.027.

Verboom, G.A., Archibald, J.K., Bakker, F.T., Bellstedt, D.U., Conrad, F., Dreyer, L.L., Forest, F., Galley, C., et al. 2009. Origin and diversification of the Greater Cape flora: Ancient species repository, hot-bed of recent radiation, or both? *Molecular Phylogenetics and Evolution*. 51(1):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 Phytologist*. 207(2):368–376. DOI: 10.1111/nph.13342.

Verboom, G.A., Linder, H.P., Forest, F., Hoffmann, V., Bergh, N.G., Cowling, R.M., Allsopp, N. & Colville, J.F. 2014. Cenozoic assembly of the Greater Cape flora. In *Fynbos: Ecology, evolution, and conservation of a megadiverse region*. 1st ed. N. Allsopp, J.F. Colville, & G.A. Verboom, Eds. Oxford Press.

Verboom, G.A., Moore, T.E., Hoffmann, V. & Cramer, M.D. 2012. The roles of climate and soil nutrients in shaping the life histories of grasses native to the Cape Floristic Region. *Plant and soil*. 355(1-2):323–340.

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(5):1008–1021. DOI: 10.1111/j.0014-3820.2003.tb00312.x.

Verboom, G.A., Linder, H.P. & Stock, W.D. 2004. Testing the adaptive nature of radiation: Growth form and life history divergence in the African grass genus *Ehrharta* (Poaceae: Ehrhartoideae). *American Journal of Botany*. 91(9):1364–1370.

Vicentini, A., Barber, J.C., Aliscioni, S.S., Giussani, L.M. & Kellogg, E.A. 2008. The age of the

grasses and clusters of origins of C4 photosynthesis. *Global Change Biology*. 14(12):2963–2977.

Weadick, C.J. & Chang, B.S. 2012. Complex patterns of divergence among green-sensitive (RH2a) African cichlid opsins revealed by Clade model analyses. *BMC evolutionary biology*. 12(1):206.

Wen, J., Zhang, J., Nie, Z.-L., Zhong, Y. & Sun, H. 2014. Evolutionary diversifications of plants on the Qinghai-Tibetan Plateau. *Frontiers in genetics*. 5:4.

Witter, M.S. & Carr, G.D. 1988. Adaptive radiation and genetic differentiation in the Hawaiian silversword alliance (Compositae: Madiinae). *Evolution*. 42(6):1278–1287.

Wolf, J.B., Künstner, A., Nam, K., Jakobsson, M. & Ellegren, H. 2009. Nonlinear dynamics of nonsynonymous (dN) and synonymous (dS) substitution rates affects inference of selection. *Genome biology and evolution*. 1:308–319.

Yang, Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Molecular biology and evolution*. 15(5):568–573.

Yang, Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular biology and evolution*. 24(8):1586–1591.

Yang, Z. 2019. Adaptive molecular evolution. In *Handbook of statistical genomics: Two volume set*. 1st ed. Wiley Online Library. 369–396.

Yang, Z. & Nielsen, R. 1998. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *Journal of molecular evolution*. 46(4):409–418.

Table 1: Accession numbers, original identification, the name that the accession is referred to in this chapter, and the clade membership of the 42 taxa used in the BEAST and downstream analyses. SK = Succulent Karoo. Accession numbers are prefixed by 'GAV' unless otherwise indicated.

Accession no.	Original ID	Name	Clade
1516	E. eburnea	E. eburnea	Lowlands/SK
1519	E. melicoides	E. melicoides	Lowlands/SK
1523	E. calycina 'gracile'	E. gracile	Lowlands/SK
1525	E. longiflora	E. longiflora	Lowlands/SK
1526	E. brevifolia subsp. cuspidata	E. cuspidata	Lowlands/SK
1532	E. barbinodis	E. barbinodis	Lowlands/SK
1534	E. calycina 'robust'	E. robust	Lowlands/SK
1535	E. pusilla	E. pusilla	Lowlands/SK
1542	E. delicatula	E. delicatula	Lowlands/SK
1543	E. triandra	E. triandra	Lowlands/SK
1544	E. calycina 'clanwilliam'	E. clanwilliam	Lowlands/SK
1547	E. brevifolia subsp. brevifolia	E. brevifolia	Lowlands/SK
1549	E. villosa subsp. maxima	E. maxima	Lowlands/SK
1550	E. villosa subsp. villosa	E. villosa	Lowlands/SK
1551	E. thunbergii	E. thunbergii	Lowlands/SK
1552	E. erecta subsp. erecta	E. erecta	Lowlands/SK
1554	E. ottonis	E. ottonis	Lowlands
1555	E. capensis	E. capensis	Lowlands
1556	E. bulbosa	E. bulbosa	Lowlands
1625-u	E. longifolia	E. longifolia	Lowlands
1622-1	E. ramosa subsp. aphylla	Milner	Ramosa-Rehmannii
1576-1	E. ramosa subsp. aphylla	Aphylla	Ramosa-Rehmannii
1592-2	E. rehmannii subsp. filiformis	Filiformis	Ramosa-Rehmannii
1611-2	E. ramosa subsp. ramosa	Western ramosa	Ramosa-Rehmannii
1635-2	E. ramosa subsp. ramosa	Eastern ramosa	Ramosa-Rehmannii
1649-1	E. rehmannii subsp. rehmannii	Rehmannii	Ramosa-Rehmannii
LW3-1	E. rehmannii subsp. subspicata	Subspicata	Ramosa-Rehmannii
1572-2	E. setacea subsp. uniflora	Uniflora	Setacea
1594-2	E. rupestris subsp. tricostata	Restioid tricostata	Setacea
1597-2	E. setacea subsp. disticha	Fernkloof A	Setacea
1600-2	E. setacea subsp. setacea	Fernkloof B	Setacea
1602-3	E. rupestris subsp. dodii	Dodii	Setacea
1605-3	E. rupestris subsp. rupestris	Western rupestris	Setacea
1621-4	E. rupestris subsp. tricostata	Leafy tricostata	Setacea
1629-3	E. setacea subsp. setacea	Setacea	Setacea
1644-2	E. setacea subsp. scabra	Scabra	Setacea
1655-1	E. rupestris subsp. rupestris	Eastern rupestris	Setacea
1657-2	E. rupestris subsp. tricostata	Wemmershoek	Setacea
1628-3	E. microlaena	E. microlaena	
1646	E. dura	E. dura	
	Microlaena stipoides	M. stipoides	
	Oryza sativa	O. sativa	

Table 2: Median node ages (Ma) for the most recent common ancestor of *Oryza* and *Ehrharta* (MRCA), as well as for the crown nodes of the Setacea, Ramosa-Rehmannii (RR), Succulent Karoo (SK) and Lowlands clades for BEAST trees either each constructed from 25 randomly chosen supercontig loci from the ninety coverage dataset, or from the gene 'shopping' method of Sortadate (Smith et al., 2018).

Tree	MRCA	*Ehrharta*	*Microlaena*	Lowlands	SK	RR	Setacea
1	68.0	28.3	25.7	16.0	14.6	6.2	4.9
2	68.8	26.9	22.4	15.3	13.4	5.7	3.2
3	68.1	29.4	19.6	14.8	13.7	3.7	3.0
4	68.2	29.3	20.9	16.6	14.6	7.2	4.4
5	68.7	27.5	26.1	16.3	15.4	4.8	3.9
6	68.4	28.4	23.6	15.2	13.7	6.1	7.5
7	67.9	29.6	24.6	18.9	16.4	7.5	5.9
8	68.1	28.6	26.0	14.9	13.8	7.5	6.1
9	68.4	28.6	23.9	15.2	14.3	6.1	4.2
10	68.0	29.6	25.2	16.0	14.5	6.3	9.0
Sortadate 'best'	69.3	26.6	19.1	13.0	11.2	3.7	3.0
Sortadate 'worst'	68.4	28.4	24.1	17.2	16.1	9.0	6.4
Mean age	68.4	28.4	23.4	15.8	14.3	6.2	5.1

Table 3: Parameter estimates, log-likelihood scores, and likelihood ratio test (LRT) p-values obtained from branch models implemented in codeml, based on 44 concatenated exons. The null model fits a single ω across the tree, while the remaining models estimate ω for both foreground (the clade of interest) and background branches

Model	No. parameters	Background ω	Foreground ω	ln L	χ^2 stat.	LRT p-value
Null	79	0.19	NA	-142965.039	NA	
Lowlands	80	0.20	0.18	-142960.303	9.47	$p = 0.002$
Ramosa-Rehmannii	80	0.19	0.21	-142962.883	4.31	$p = 0.038$
Setacea	80	0.18	0.23	-142955.888	18.30	$p < 0.001$

Table 4: The date estimates (Ma) of the present study, which are based on the calibrations of Prasad et al., 2011, rescaled using the date estimates of previous works. Ory-Ehr = the split between *O. sativa* and *Ehrharta*, SK = Succulent Karoo, RR = Ramosa-Rehmannii.

Author	Ory-Ehr	Lowlands	SK	RR	Setacea
Current work	68	15.8	14.5	6.2	5.1
Bouchenak-Khelladi et al., 2010	50	11.5	10.6	4.5	3.7
Vicentini et al., 2008	35	8.1	7.4	3.2	2.6
Hoffmann et al., 2015	25	5.8	5.3	2.3	1.9

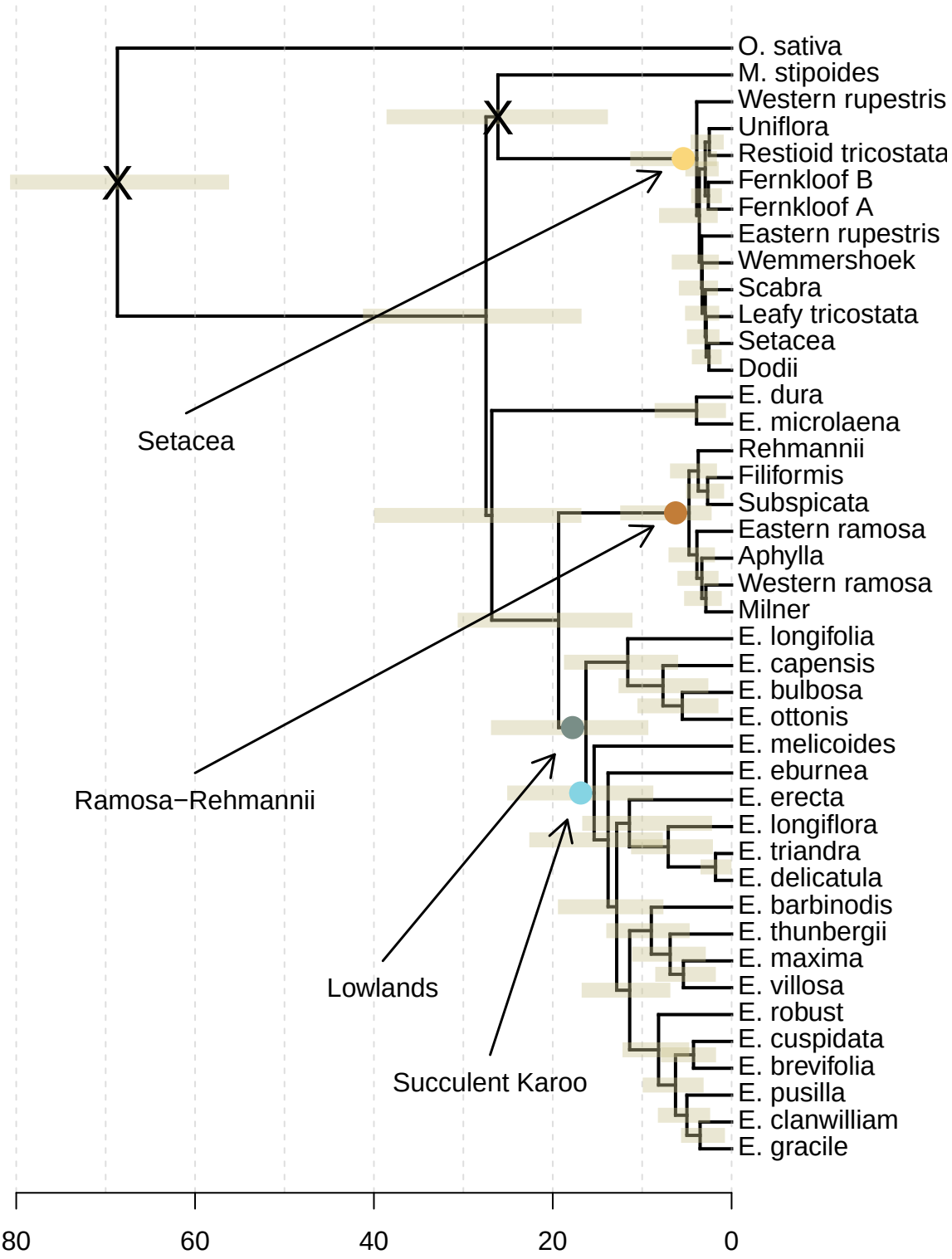

Figure 1: A representative time-calibrated BEAST tree of *Ehrharta* (tree 5) constructed using 25 supercontig loci, where the x-axis is years before present (Ma). Beige bands show the 95% HPD interval for each node. Calibration points from Prasad et al. 2011 are represented by 'X'. Relevant clades are subtended by colourful points and labeled with arrows.

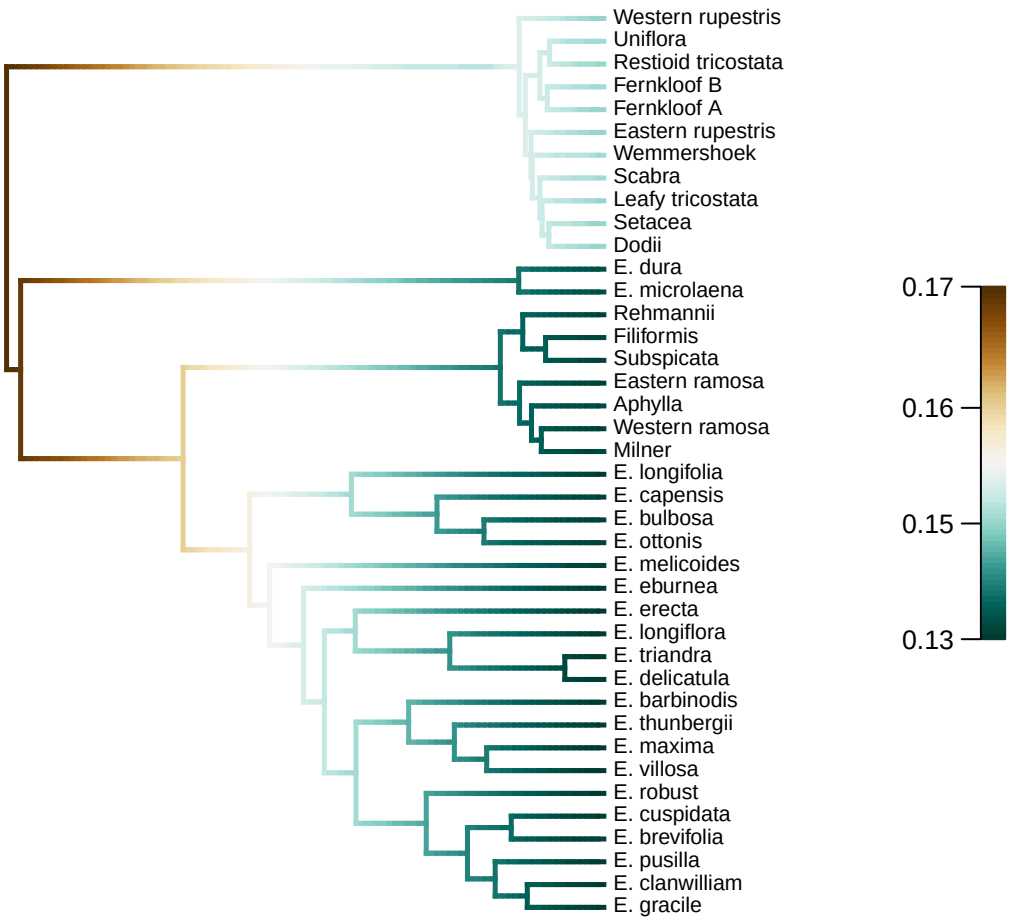

Figure 2: Diversification rates (lineages per Myr) within a representative BEAST tree, calculated with BAMM. None of the twelve BEAST trees contained significant rate shifts.

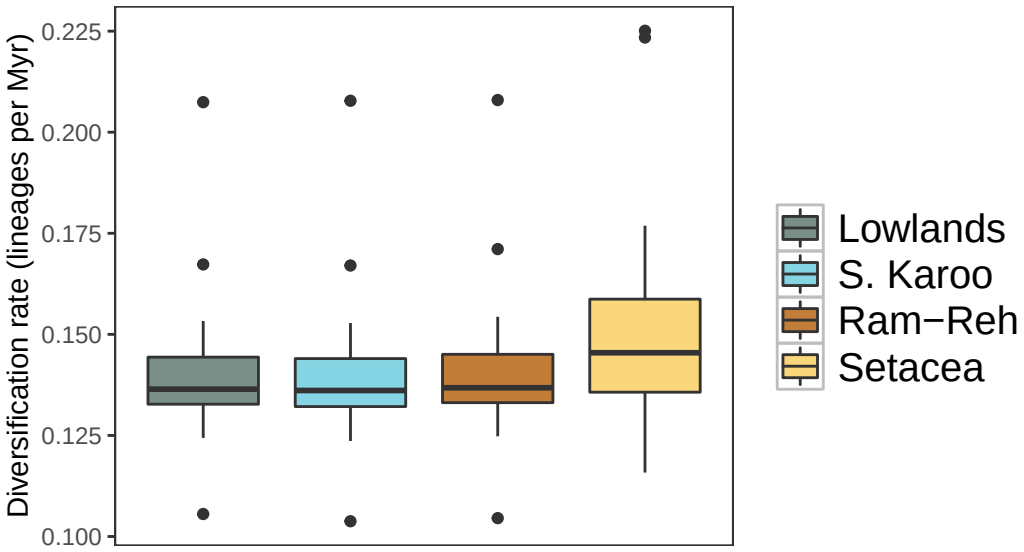

Figure 3: Variation in diversification rates (lineages per Myr) within the Lowlands, Succulent Karoo, Ramosa-Rehmannii (Ram-Reh) and Setacea clades across the twelve BEAST trees, where diversification rates were inferred using BAMM. A Kruskal-Wallis test found no significant differences between clades ($\chi^2 = 2.74$, df = 3, p = 0.43).

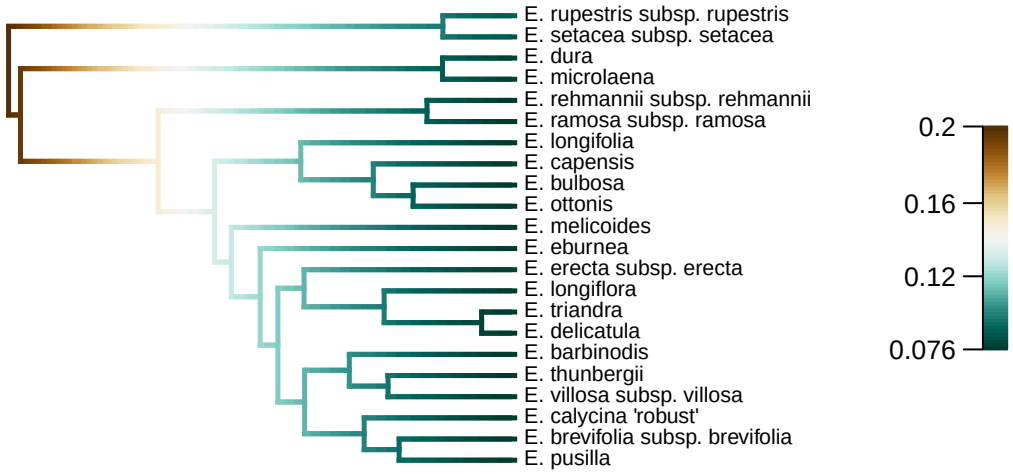

Figure 4: Diversification rates (lineages per Myr) within a representative BEAST tree containing only the species included in Verboom, Linder & Stock (2003), calculated with BAMM. None of the twelve BEAST trees contained significant rate shifts.

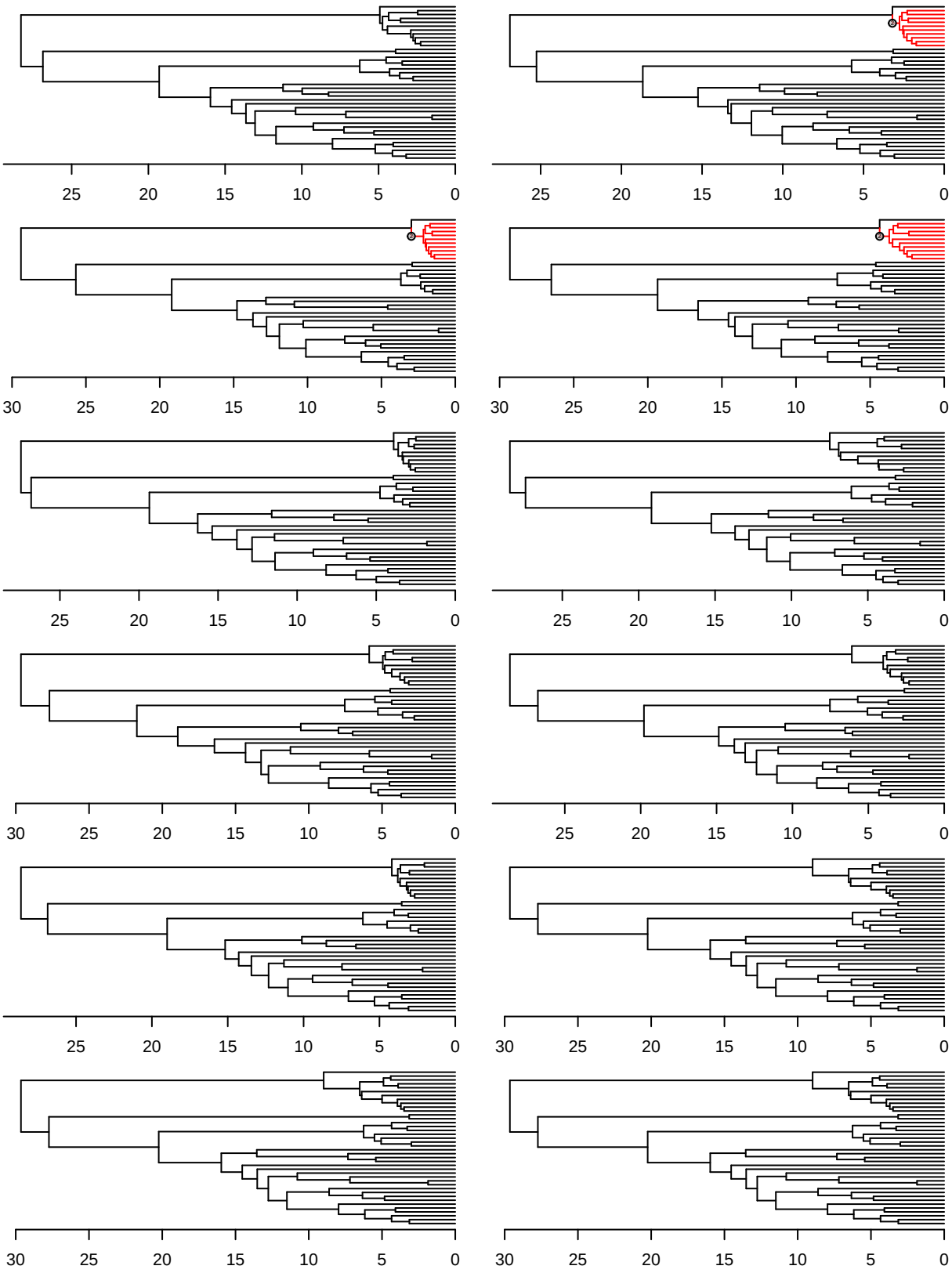

Figure 5: Diversification rate shifts as modeled by MEDUSA, for all twelve BEAST trees. Red branches show where a diversification rate shift has occurred, x-axis scale is in Ma.

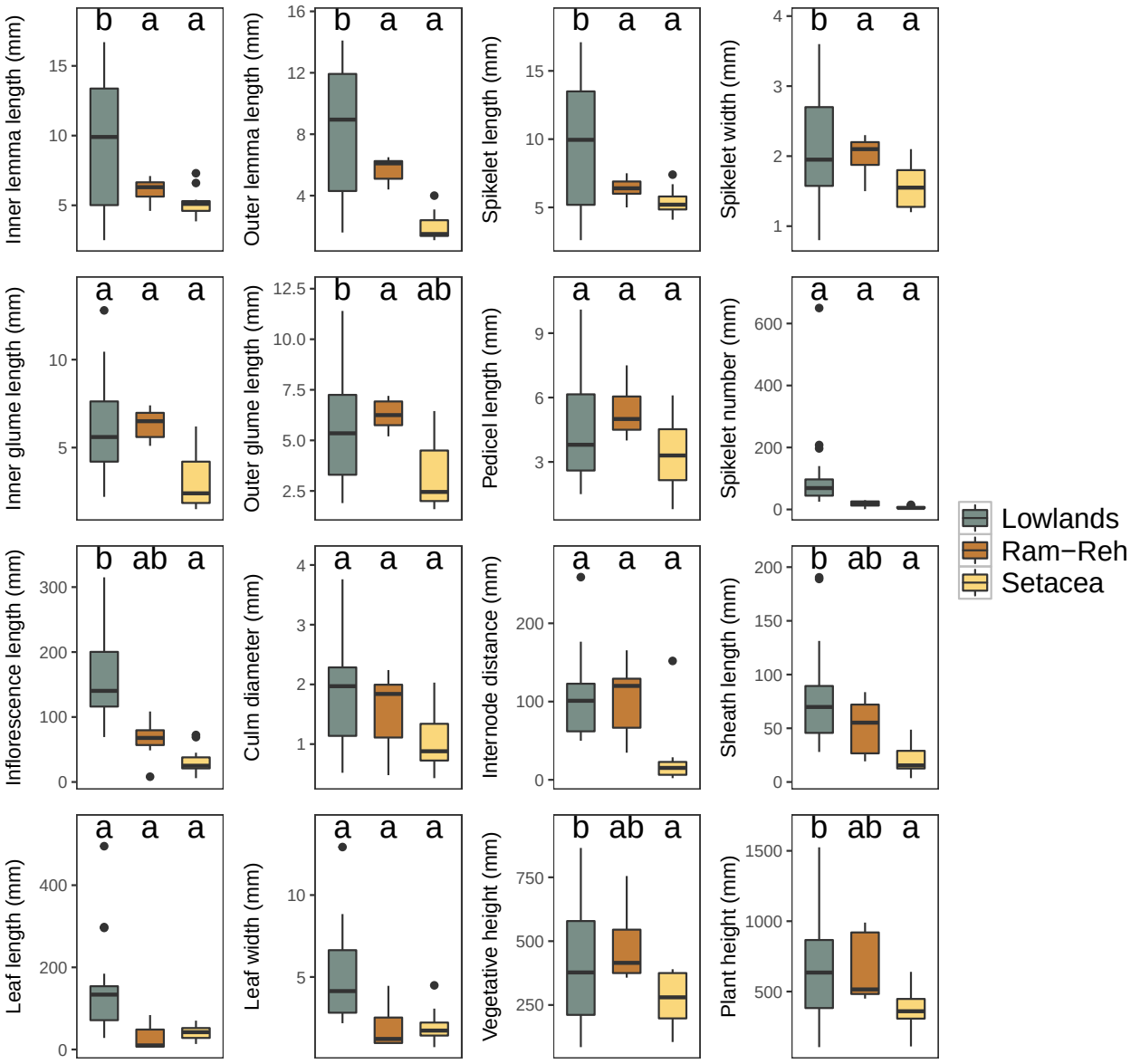

Figure 6: Morphological trait variance within clades. Letters signify significant differences based on an ANOVA and a post-hoc Tukey HSD test. Ram-Reh = Ramosa-Rehmannii.

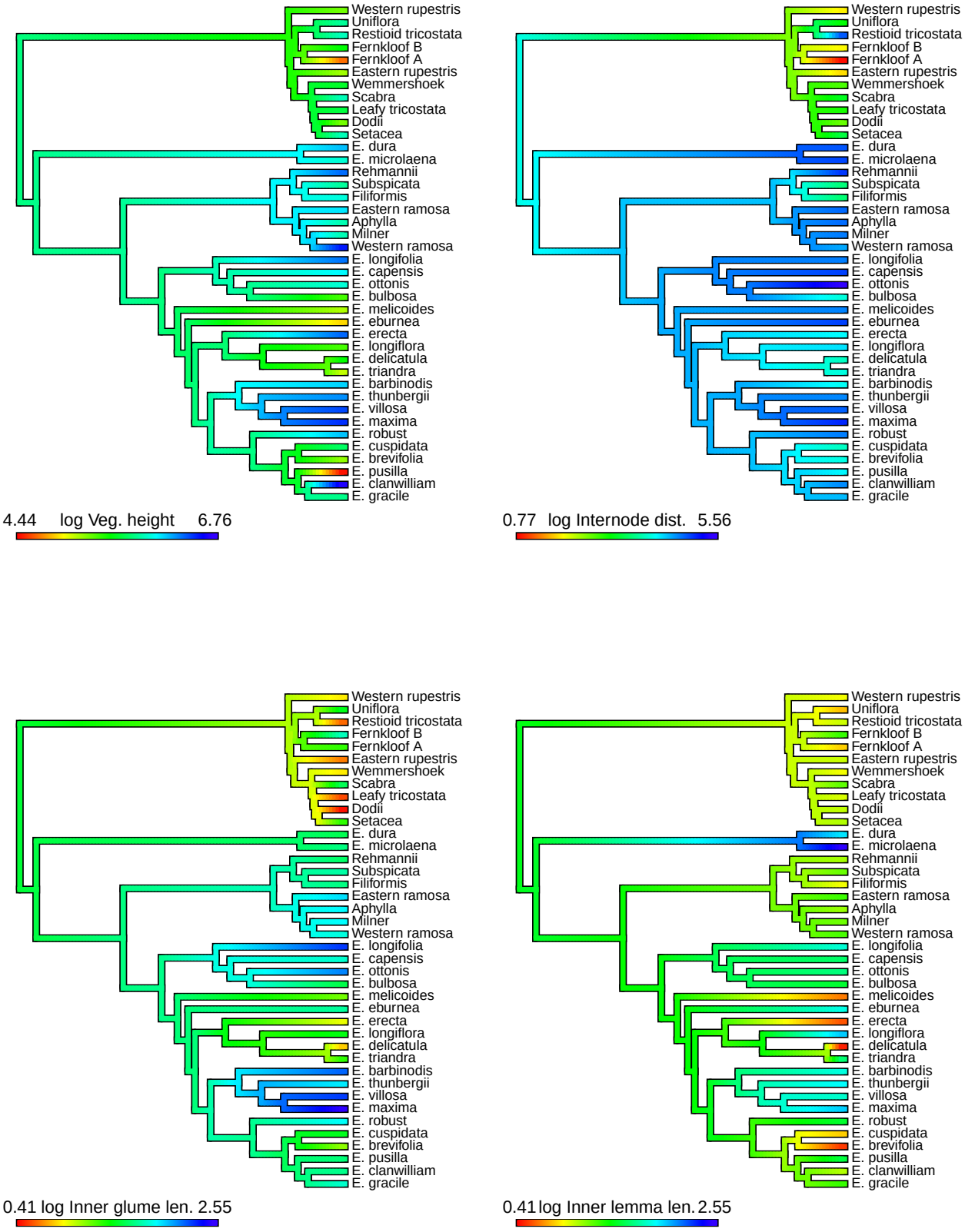

Figure 7: The variation in four representative traits (log mm) mapped to a sample BEAST tree. Warmer colours represent smaller values.

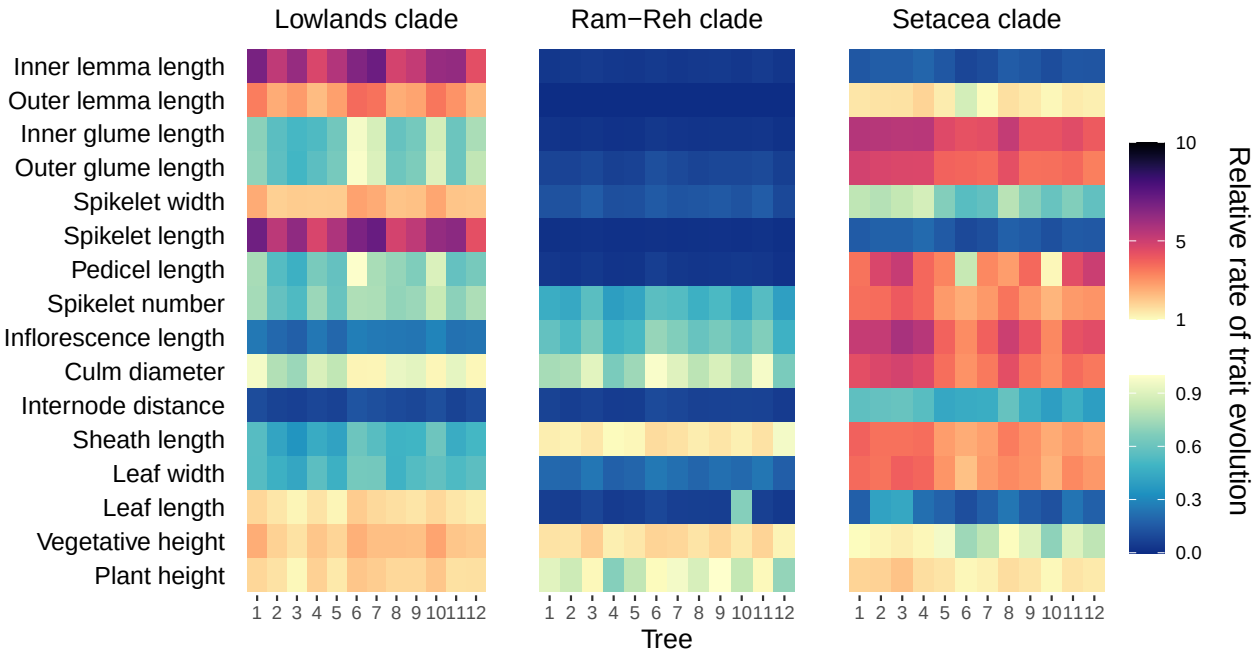

Figure 8: Maximum likelihood estimates of the relative rate of morphological evolution of 16 traits, calculated separately for each of the Lowlands, Ramosa and Setacea clades relative to the remainder of the tree, using the non-censored model of O'Meara et al., 2006. Each column represents a BEAST tree. For ease of visualisation, rates less than and greater than one have different colour scales, with the former in shades of blue and the latter in warm colours.

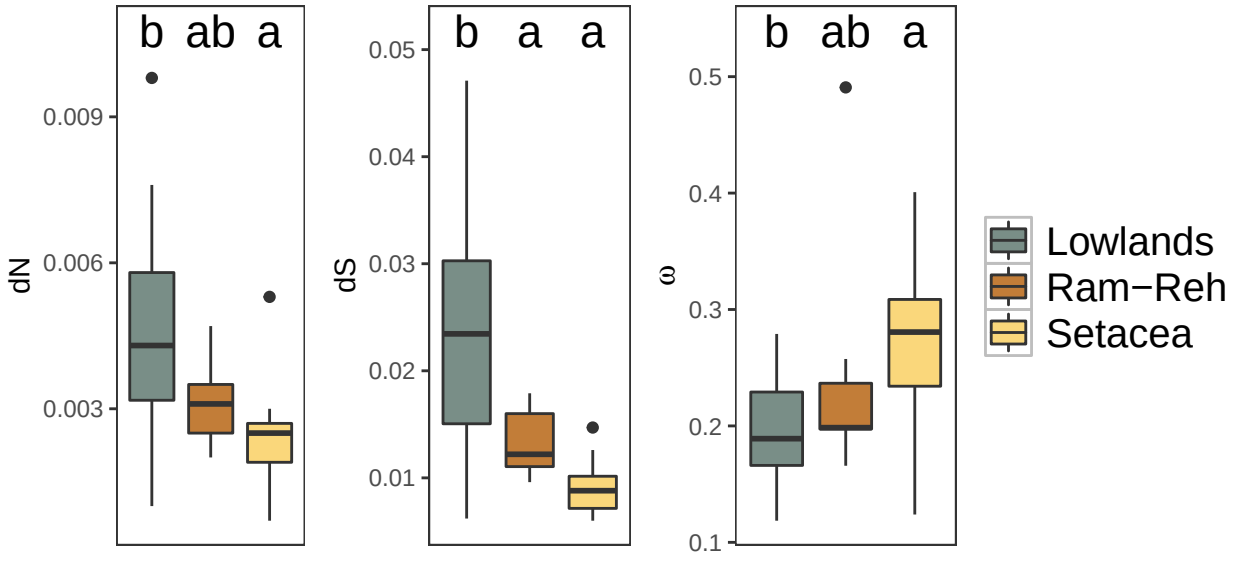

Figure 9: The differences in number of non-synonymous (dN), synonymous (dS) and substitution rate ratio (ω) for the terminal branches of each clade, estimated with the free-ratios model in codeml on 44 concatenated exons (model = 1, NSsites = 0). Letters signify significant differences based on a Levene's test for equal variance and a post-hoc Tukey HSD test. Ramosa-Rehmannii = Ram-Reh.

Chapter 4: Synthesis

The GCFR has been the subject of sustained taxonomic research for over 250 years (Linnaeus, 1753), making it one of the best-characterised biological hotspots in the world (Joppa et al., 2011). It is thought that <1% of the flora remains undiscovered in the Core Cape Region (Treurnicht et al., 2017), although a comparable estimate has not been determined for the Extra Cape Flora (Snijman, 2013). Most taxonomic work conducted to date has, however, focused on morphologically-based characters for delimitation (Treurnicht et al., 2017), consequently biasing species discovery towards morphologically diagnosable taxa, and undercounting cryptic species (Adams et al., 2014). The present study documents multiple new species at higher elevations, some of which are barely distinguishable on the basis of morphology, but are genetically distinct (Chapter 2). Should such a pattern hold generally for plant lineages in the Cape, as is suggested by other recent discoveries of semi-cryptic or cryptic species (Britton, Hedderson & Verboom, 2014; Shaik, 2019), it is possible that a significant portion of the montane Cape flora is yet to be described. In particular, species with large distributions spanning multiple mountain ranges are likely to contain hidden diversity (Britton, Hedderson & Verboom, 2014). To locate and accurately delimit such hidden diversity, however, requires the implementation of robust, integrative taxonomic practices which incorporate multiple sources of evidence (Dayrat, 2005; Padial et al., 2010). Moreover, this study demonstrates the value of using genome-wide data at different resolutions. The targeted enrichment dataset employed here, which consists of loci that are relatively conserved across the angiosperms (Johnson et al., 2019), revealed the polyphyly of *E. setacea* and *E. rupestris* as currently defined, while the GBS dataset, consisting of neutral markers, uncovered the genetic boundaries between closely related sympatric taxa (e.g. at Fernkloof and Milner Peak). This work also reiterates the need for dense population-level sampling to gain confidence in hypothesised species limits (Bernardo, 2011;

Singhal et al., 2018). In particular, future studies should concentrate sampling in the southwestern GCFR, where the strong genetic structure found within the Ramosa clade, together with the prevalence of range-restricted species in the region (Cowling, Holmes & Rebelo, 1992; Bradshaw, Colville & Linder, 2015), implies that diversification processes are operating at especially fine scales.

The Cape Fold mountains have long been considered an influential force in driving diversification in the GCFR (Linder, 1985). The steep mountain slopes create strong environmental gradients, increasing potential niche space and allowing ecological speciation to occur (Rundle & Nosil, 2005; Schwery et al., 2015). Yet, the isolated peaks also form "sky islands" that promote non-ecological speciation (Rundell & Price, 2009; Britton, Hedderson & Verboom, 2014), prompting the hypothesis that the relative importance of ecological and non-ecological speciation may vary with elevation (Verboom et al., 2015). The present study, however, provides little support for this hypothesis. Instead, speciation processes are found to be clade specific, with the radiation of each clade triggered by a distinct set of intrinsic and extrinsic factors (Bouchenak-Khelladi et al., 2015). For example, the Lowlands clade evolved a diversity of life histories that allowed the clade to radiate adaptively onto the varied substrates of the arid succulent karoo and renosterveld regions (Verboom, Linder & Stock, 2004), while extrinsic barriers to gene flow likely played a more influential role in the radiation of the Ramosa-Rehmannii clade. *Ehrharta* is not alone in highlighting the complexity of plant radiations, especially in montane regions (Wen et al., 2014; Muellner-Riehl, 2019). Other studies of montane taxa have similarly revealed drivers of radiation to be clade-specific (Bentley, Verboom & Bergh, 2014; Lagomarsino et al., 2016). For example, diversification of subgroups of *Saxifraga* in the Hengduan Mountains was shown to variously be modulated by the effects of key innovation, extrinsic opportunities, or a combination thereof (Ebersbach et al., 2017). It has also been shown that differing trait-environment associations, linked to different life histories, can provoke unique evolutionary responses in lineages that exist under the

same environmental conditions (Mitchell et al., 2015). Together, these factors suggest that even if historical and environmental conditions are well known, as they are in the Cape (Cowling & Lombard, 2002; Verboom et al., 2014; Linder & Verboom, 2015), it is difficult to understand triggers of diversification without sufficient natural history knowledge of the lineage in question.

Yet, despite the idiosyncratic nature of lineage diversification, montane regions contain a much greater proportion of the Earth's biodiversity than is expected given their relatively small global area (Rahbek, Borregaard, Colwell, et al., 2019), and the Cape mountains are no exception (Cowling & Lombard, 2002; Thuiller et al., 2006). This implies that there must be biome-level processes unique to these regions that facilitate the accumulation of species richness (Rahbek, Borregaard, Antonelli, et al., 2019). Current biogeographical hypotheses however, struggle to explain this phenomenon, with statistical models of large-scale patterns of species richness generally under-predicting montane diversity (Rahbek, Borregaard, Colwell, et al., 2019). Nonetheless, a new framework, known as the 'mountain geo-biodiversity hypothesis' (MGH) has recently been proposed to explain montane species richness in the context of Tibeto-Himalayan mountain region (Mosbrugger et al., 2018), which could potentially be applied to the Cape Fold Mountains. The MGH takes a holistic approach to explaining species richness by intertwining biological, geological and historical factors, and states that there are three conditions which, if present in a given system, maximise regional montane biodiversity. These conditions are i) full elevational zonation (e.g. lowland, montane and alpine zones), which provides for both local adaptation and the immigration of diverse pre-adapted lineages; ii) high-relief terrain, which functions to provide climatic refugia and to reduce migratory/dispersal distances under climate change; and iii) historical climatic fluctuations which create a 'species pump' effect in the context of high-relief terrain (Mosbrugger et al., 2018; Muellner-Riehl et al., 2019). The Cape Fold mountains meet at least the first two conditions as they have strong elevational, moisture and temperature gradients

and, while the Cape alpine zone is not as extensive as that of the Tibeto-Himalayan region, high-elevation plants do require some tolerance of subzero conditions (Manning & Goldblatt, 2012). In addition, there is evidence of refugia within the Cape Fold mountains, especially in the Kogelberg region (Grant & Samways, 2007; Linder, 2019).

The third condition of the MGH, that of fluctuating climate, perhaps applies less obviously to the GCFR. Glacial cycles during the Pleistocene had a large impact on global species dynamics, causing both wide-scale extinction and extensive radiations events (Willis & Niklas, 2004; Hughes & Eastwood, 2006; Gill et al., 2009). In the Cape region, however, the Pleistocene climate was relatively stable (Dynesius & Jansson, 2000), with the species richness of the GCFR being partially attributed to limited extinction during this time (Cowling & Lombard, 2002; Linder, 2008; Linder & Verboom, 2015). Nonetheless, climate-induced fluctuations in species distributions did occur (Dynesius & Jansson, 2000; Scott & Woodborne, 2007; Huntley et al., 2016), potentially at remarkably fine geographic scales (Chase et al., 2019). Given the poor dispersal ability of many Cape plants (Goldblatt, 1997), this may have been sufficient to create 'flickering connectivity' between populations (Flantua & Hooghiemstra, 2018), in which the migration of species up and down elevational bands in response to climate change alters the degree of population isolation (Mairal et al., 2017; Flantua et al., 2019). Flickering connectivity is thought promote speciation via a recently proposed mixing-isolation-mixing (MIM) model of speciation (He et al., 2019). Here, speciation involves recurrent cycles of gene flow between genetically distinct populations followed by geographic isolation, until total genetic isolation eventually occurs (He et al., 2019).

In the context of this study, the Setacea clade may be an example of the MIM model of speciation. Firstly, the near simultaneous origin of this clade's taxa suggests a large-scale climatic trigger, such as the onset of Pleistocene glacial cycles (Janssens et al., 2009), rather than gradual divergence into

new niche space as seen in the Lowlands clade. Secondly, there is some indication of past gene flow events between taxa in the clade. For example, the short branch lengths and poor phylogenetic support in the ASTRAL tree are indicative of gene tree discordance (Sayyari & Mirarab, 2016; Zhang et al., 2018), potentially resulting from gene flow (Solís-Lemus, Yang & Ané, 2016; Thom et al., 2018). Additionally, there is evidence of between-species admixture (e.g. in the Fernkloof, Setacea, Leafy Tricostata groups), potentially reflecting the historical mixing of previously differentiated gene pools (Qu et al., 2012; Martin et al., 2013). These concepts need to be tested explicitly, however, by modelling Pleistocene climate fluctuations and species connectivity at fine spatiotemporal scales (Chase et al., 2019; Flantua et al., 2019) and relating them to historical gene flow events (Sousa & Hey, 2013; Peñalba, Joseph & Moritz, 2019). The MIM model of speciation can result in the rapid accumulation of species in areas where transient geographical barriers allow cyclical patterns of population isolation and mixing (He et al., 2019). Should this be the case for the GCFR, the MIM may better explain the rapid diversification of many Cape clades during the Pleistocene than allopatric or ecological models of speciation alone (Abbott, 2019), and should be assessed in conjunction with the MGH as a potential mechanism behind the montane biodiversity in the GCFR.

References

Abbott, R.J. 2019. A mixing–isolation–mixing model of speciation can potentially explain hotspots of species diversity. *National Science Review.* 6(2):290–291.

Adams, M., Raadik, T.A., Burridge, C.P. & Georges, A. 2014. Global biodiversity assessment and hyper-cryptic species complexes: More than one species of elephant in the room? *Systematic Biology.* 63(4):518–533. DOI: 10.1093/sysbio/syu017.

Bentley, J., Verboom, G.A. & Bergh, N.G. 2014. Erosive processes after tectonic uplift stimulate vicariant and adaptive speciation: Evolution in an Afrotemperate-endemic paper daisy genus. *BMC Evolutionary Biology*. 14(1):27.

Bernardo, J. 2011. A critical appraisal of the meaning and diagnosability of cryptic evolutionary diversity, and its implications for conservation in the face of climate change. *Climate change, ecology and systematics*. 380–438.

Bouchenak-Khelladi, Y., Onstein, R.E., Xing, Y., Schwery, O. & Linder, H.P. 2015. On the complexity of triggering evolutionary radiations. *New Phytologist*. 207(2):313–326.

Bradshaw, P.L., Colville, J.F. & Linder, H.P. 2015. Optimising regionalisation techniques: Identifying centres of endemism in the extraordinarily endemic-rich Cape Floristic Region. *PloS one*. 10(7):e0132538.

Britton, M.N., Hedderson, T.A. & Verboom, G.A. 2014. Topography as a driver of cryptic speciation in the high-elevation Cape sedge *Tetraria triangularis* (Boeck.) C. B. Clarke (Cyperaceae: Schoeneae). *Molecular Phylogenetics and Evolution*. 77:96–109. DOI: 10.1016/j.ympev.2014.03.024.

Chase, B.M., Boom, A., Carr, A.S., Chevalier, M., Quick, L.J., Verboom, G.A. & Reimer, P.J. 2019. Extreme hydroclimate response gradients within the western Cape Floristic Region of South Africa since the Last Glacial Maximum. *Quaternary Science Reviews*. 219:297–307.

Cowling, R.M., Holmes, P.M. & Rebelo, A.G. 1992. Plant diversity and endemism. In *The ecology of fynbos: Nutrients, fire and diversity*. 1st ed. R.M. Cowling, Ed. Oxford University Press: Cape Town. 62–112.

Cowling, R. & Lombard, A. 2002. Heterogeneity, speciation/extinction history and climate:

Explaining regional plant diversity patterns in the Cape Floristic Region. *Diversity and Distributions*. 8(3):163–179.

Dayrat, B. 2005. Towards integrative taxonomy. *Biological Journal of the Linnean Society*. 85(3):407–415. DOI: 10.1111/j.1095-8312.2005.00503.x.

Dynesius, M. & Jansson, R. 2000. Evolutionary consequences of changes in species' geographical distributions driven by Milankovitch climate oscillations. *Proceedings of the National Academy of Sciences*. 97(16):9115–9120.

Ebersbach, J., Schnitzler, J., Favre, A. & Muellner-Riehl, A. 2017. Evolutionary radiations in the species-rich mountain genus *Saxifraga* L. *BMC evolutionary biology*. 17(1):119.

Flantua, S.G., O'Dea, A., Onstein, R.E., Giraldo, C. & Hooghiemstra, H. 2019. The flickering connectivity system of the north Andean páramos. *Journal of Biogeography*. 46(8):1808–1825.

Flantua, S.G. & Hooghiemstra, H. 2018. Historical connectivity and mountain biodiversity. In *Mountains, climate and biodiversity*. 1st ed. C. Hoorn, A. Perrigio, & A. Antonelli, Eds. Wiley-Blackwell Oxford, UK. 171–185.

Gill, J.L., Williams, J.W., Jackson, S.T., Lininger, K.B. & Robinson, G.S. 2009. Pleistocene megafaunal collapse, novel plant communities, and enhanced fire regimes in North America. *Science*. 326(5956):1100–1103.

Goldblatt, P. 1997. Floristic diversity in the Cape flora of South Africa. *Biodiversity & Conservation*. 6(3):359–377.

Grant, P.B. & Samways, M.J. 2007. Montane refugia for endemic and Red Listed dragonflies in the

Cape Floristic Region biodiversity hotspot. *Biodiversity and Conservation.* 16(3):787–806.

He, Z., Li, X., Yang, M., Wang, X., Zhong, C., Duke, N.C., Wu, C. & Shi, S. 2019. Speciation with gene flow via cycles of isolation and migration: Insights from multiple mangrove taxa. *National Science Review.* 6(2):275–288.

Hughes, C. & Eastwood, R. 2006. Island radiation on a continental scale: Exceptional rates of plant diversification after uplift of the Andes. *Proceedings of the National Academy of Sciences.* 103(27):10334–10339.

Huntley, B., Collingham, Y.C., Singarayer, J.S., Valdes, P.J., Barnard, P., Midgley, G.F., Altwegg, R. & Ohlemüller, R. 2016. Explaining patterns of avian diversity and endemicity: Climate and biomes of southern Africa over the last 140,000 years. *Journal of Biogeography.* 43(5):874–886.

Janssens, S.B., Knox, E.B., Huysmans, S., Smets, E.F. & Merckx, V.S. 2009. Rapid radiation of *Impatiens* (Balsaminaceae) during Pliocene and Pleistocene: Result of a global climate change. *Molecular Phylogenetics and Evolution.* 52(3):806–824.

Johnson, M.G., Pokorny, L., Dodsworth, S., Botigue, L.R., Cowan, R.S., Devault, A., Eiserhardt, W.L., Epitawalage, N., et al. 2019. A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-medoids clustering. *Systematic Biology.* 68(4):594–606.

Joppa, L.N., Roberts, D.L., Myers, N. & Pimm, S.L. 2011. Biodiversity hotspots house most undiscovered plant species. *Proceedings of the National Academy of Sciences.* 108(32):13171–13176.

Lagomarsino, L.P., Condamine, F.L., Antonelli, A., Mulch, A. & Davis, C.C. 2016. The abiotic and biotic drivers of rapid diversification in Andean bellflowers (Campanulaceae). *New Phytologist.*

210(4):1430–1442.

Linder, H.P. 1985. Gene flow, speciation, and species diversity patterns in a species-rich area: The Cape Flora. *Species and speciation.* 4:53–7.

Linder, H.P. 2008. Plant species radiations: Where, when, why? *Philosophical Transactions of the Royal Society B: Biological Sciences.* 363(1506):3097–3105.

Linder, H.P. 2019. Rare species, Restionaceae, and the Cape flora. *Journal of Biogeography.* 46(12):2637–2650.

Linder, H.P. & Verboom, G.A. 2015. The evolution of regional species richness: The history of the southern African flora. *Annual Review of Ecology, Evolution, and Systematics.* 46:393–412.

Linnaeus, C. 1753. *Species plantarum: Exhibentes plantas rite cognitas, ad genera relates, cum differentiis specificis, nominibus trivialibus, synonymis selectis, locis natalibus, secundum systema sexuale digestas.* 1st ed. V. 2. Holmiae: Impensis Laurentii Salvii.

Mairal, M., Sanmartín, I., Herrero, A., Pokorny, L., Vargas, P., Aldasoro, J.J. & Alarcón, M. 2017. Geographic barriers and Pleistocene climate change shaped patterns of genetic variation in the Eastern Afromontane biodiversity hotspot. *Scientific Reports.* 7:45749.

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

Martin, S.H., Dasmahapatra, K.K., Nadeau, N.J., Salazar, C., Walters, J.R., Simpson, F., Blaxter, M., Manica, A., et al. 2013. Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome research.* 23(11):1817–1828.

Mitchell, N., Moore, T.E., Mollmann, H.K., Carlson, J.E., Mocko, K., Martinez-Cabrera, H., Adams, C., Silander Jr, J.A., et al. 2015. Functional traits in parallel evolutionary radiations and trait-environment associations in the Cape Floristic Region of South Africa. *The American Naturalist.* 185(4):525–537.

Mosbrugger, V., Favre, A., Muellner-Riehl, A.N., Päckert, M. & Mulch, A. 2018. Cenozoic evolution of geo-biodiversity in the Tibeto-Himalayan region. In *Mountains, climate, and biodiversity.* 1st ed. C. Hoorn, A. Perrigio, & A. Antonelli, Eds. Wiley-Blackwell Chichester.

Muellner-Riehl, A.N. 2019. Mountains as evolutionary arenas: Patterns, emerging approaches, paradigm shifts, and their implications for plant phylogeographic research in the Tibeto-Himalayan Region. *Frontiers in plant science.* 10:195.

Muellner-Riehl, A.N., Schnitzler, J., Kissling, W.D., Mosbrugger, V., Rijsdijk, K.F., Seijmonsbergen, A.C., Versteegh, H. & Favre, A. 2019. Origins of global mountain plant biodiversity: Testing the "mountain-geobiodiversity hypothesis". *Journal of Biogeography.* 46(12):2826–2838.

Padial, J.M., Miralles, A., De la Riva, I. & Vences, M. 2010. The integrative future of taxonomy. *Frontiers in zoology.* 7(1):16.

Peñalba, J.V., Joseph, L. & Moritz, C. 2019. Current geography masks dynamic history of gene flow during speciation in northern Australian birds. *Molecular ecology.* 28(3):630–643.

Qu, Y., Zhang, R., Quan, Q., Song, G., Li, S.H. & Lei, F. 2012. Incomplete lineage sorting or secondary admixture: Disentangling historical divergence from recent gene flow in the Vinous-throated parrotbill (*Paradoxornis webbianus*). *Molecular Ecology.* 21(24):6117–6133.

Rahbek, C., Borregaard, M.K., Antonelli, A., Colwell, R.K., Holt, B.G., Nogues-Bravo, D.,

Rasmussen, C.M., Richardson, K., et al. 2019. Building mountain biodiversity: Geological and evolutionary processes. *Science*. 365(6458):1114–1119.

Rahbek, C., Borregaard, M.K., Colwell, R.K., Dalsgaard, B., Holt, B.G., Morueta-Holme, N., Nogues-Bravo, D., Whittaker, R.J., et al. 2019. Humboldt's enigma: What causes global patterns of mountain biodiversity? *Science*. 365(6458):1108–1113.

Rundell, R.J. & Price, T.D. 2009. Adaptive radiation, nonadaptive radiation, ecological speciation and nonecological speciation. *Trends in Ecology & Evolution*. 24(7):394–399. DOI: 10.1016/j.tree.2009.02.007.

Rundle, H.D. & Nosil, P. 2005. Ecological speciation. *Ecology Letters*. 8(3):336–352.

Sayyari, E. & Mirarab, S. 2016. Fast coalescent-based computation of local branch support from quartet frequencies. *Molecular biology and evolution*. 33(7):1654–1668. DOI: 10.1093/molbev/msw079.

Schwery, O., Onstein, R.E., Bouchenak-Khelladi, Y., Xing, Y., Carter, R.J. & Linder, H.P. 2015. As old as the mountains: The radiations of the Ericaceae. *New Phytologist*. 207(2):355–367.

Scott, L. & Woodborne, S. 2007. Vegetation history inferred from pollen in Late Quaternary faecal deposits (hyraceum) in the Cape winter-rain region and its bearing on past climates in South Africa. *Quaternary Science Reviews*. 26(7-8):941–953.

Shaik, Z. 2019. Species delimitation and speciation process in the *Seriphium plumosum* L. complex (Gnaphalieae: Asteraceae) in South Africa. Master's thesis. University of Cape Town.

Singhal, S., Hoskin, C.J., Couper, P., Potter, S. & Moritz, C. 2018. A framework for resolving

cryptic species: A case study from the lizards of the Australian wet tropics. *Systematic Biology.* 67(6):1061–1075. DOI: 10.1093/sysbio/syy026.

Snijman, D.A. 2013. *Plants of the Greater Cape Floristic Region, vol. 2: The Extra Cape flora.* 1st ed. Strelitzia 29. South African National Biodiversity Institute, Pretoria.

Solís-Lemus, C., Yang, M. & Ané, C. 2016. Inconsistency of species tree methods under gene flow. *Systematic biology.* 65(5):843–851.

Sousa, V. & Hey, J. 2013. Understanding the origin of species with genome-scale data: Modelling gene flow. *Nature Reviews Genetics.* 14(6):404–414.

Thom, G., Amaral, F.R.D., Hickerson, M.J., Aleixo, A., Araujo-Silva, L.E., Ribas, C.C., Choueri, E. & Miyaki, C.Y. 2018. Phenotypic and genetic structure support gene flow generating gene tree discordances in an Amazonian floodplain endemic species. *Systematic biology.* 67(4):700–718.

Thuiller, W., F. Midgley, G., Rougeti, M. & M. Cowling, R. 2006. Predicting patterns of plant species richness in megadiverse South Africa. *Ecography.* 29(5):733–744.

Treurnicht, M., Colville, J.F., Joppa, L.N., Huyser, O. & Manning, J. 2017. Counting complete? Finalising the plant inventory of a global biodiversity hotspot. *PeerJ.* 5:e2984.

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 Phytologist.* 207(2):368–376. DOI: 10.1111/nph.13342.

Verboom, G.A., Linder, H.P., Forest, F., Hoffmann, V., Bergh, N.G., Cowling, R.M., Allsopp, N. & Colville, J.F. 2014. Cenozoic assembly of the Greater Cape flora. In *Fynbos: Ecology, evolution,*

and conservation of a megadiverse region. 1st ed. N. Allsopp, J.F. Colville, & G.A. Verboom, Eds. Oxford Press.

Verboom, G.A., Linder, H.P. & Stock, W.D. 2004. Testing the adaptive nature of radiation: Growth form and life history divergence in the African grass genus *Ehrharta* (Poaceae: Ehrhartoideae). *American Journal of Botany*. 91(9):1364–1370.

Wen, J., Zhang, J., Nie, Z.-L., Zhong, Y. & Sun, H. 2014. Evolutionary diversifications of plants on the Qinghai-Tibetan Plateau. *Frontiers in genetics*. 5:4.

Willis, K.J. & Niklas, K.J. 2004. The role of Quaternary environmental change in plant macroevolution: The exception or the rule? *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*. 359(1442):159–172.

Zhang, C., Rabiee, M., Sayyari, E. & Mirarab, S. 2018. ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC bioinformatics*. 19(Suppl 6):153. DOI: 10.1186/s12859-018-2129-y.

Appendix

Table 1: Accession numbers, original and revised names, and the GPS coordinates for all accessions used for genetic or morphological data in this work. Accession numbers are prefixed by 'GAV' unless otherwise indicated.

Accession no.	Species	Revised Name	Latitude	Longitude
1516	E. eburnea		-31.35000	19.13333
1517	E. calycina robust		-31.35000	19.13333
1519	E. melicoides		-31.35000	19.13333
1523	E. calycina gracile		-31.36667	19.01667
1525	E. longiflora		-31.36667	19.01667
1526	E. brevifolia cuspidata		-31.75000	18.65000
1532	E. barbinodis		-30.08333	17.91667
1534	E. calycina robust		-30.08333	17.91667
1535	E. pusilla		-29.45000	17.73333
1542	E. delicatula		-29.68333	17.70000
1543	E. triandra		-29.91667	17.68333
1544	E. calycina		-32.06667	18.80000
1547	E. brevifolia brevifolia		-33.60000	18.41667
1549	E. villosa maxima		-33.76667	18.46667
1550	E. villosa villosa		-33.33898	18.19957
1551	E. thunbergii		-33.37040	18.37362
1552	E. erecta		-33.95943	18.46463
1554	E. ottonis		-34.49598	19.91020
1555	E. capensis		-34.46918	19.85632
1556	E. bulbosa		-34.41998	19.79557
1557	E. setacea setacea	Setacea	-34.01140	19.63280
1558	E. setacea sp.	Setacea	-34.00790	19.64710
1559	E. rupestris rupestris	Western Rupestris	-34.00842	19.66875
1560	E. ramosa ramosa	Western Ramosa	-34.00842	19.66875
1563	E. rupestris rupestris	Western Rupestris	-34.00842	19.66875
1564	E. setacea setacea	Setacea	-34.10277	18.96608
1565	E. rupestris dodii	Dodii	-34.10277	18.96608
1566	E. ramosa aphylla	Aphylla	-34.10277	18.96608
1567	E. rehmannii filiformis	Filiformis	-34.10277	18.96608
1568	E. setacea setacea	Setacea	-34.10277	18.96608
1569	E. setacea setacea	Setacea	-34.10277	18.96608
1570	E. ramosa aphylla	Aphylla	-34.05535	18.38662
1571	E. rehmannii filiformis	Filiformis	-34.05958	18.39315
1572	E. setacea uniflora	Uniflora	-34.07373	18.39585
1573	E. rupestris dodii	Dodii	-34.05958	18.39315
1576	E. ramosa aphylla	Aphylla	-34.04757	18.98912
1577	E. setacea setacea	Setacea	-34.04795	18.99062
1578	E. rupestris dodii	Dodii	-34.04795	18.99062
1579	E. rupestris tricostata	Restioid Tricostata	-34.04938	19.00538
1580	E. setacea setacea	Setacea	-34.04938	19.00538
1582	E. setacea setacea	Setacea	-34.04853	19.00652
1583	E. rehmannii filiformis	Filiformis	-34.06648	19.04478
1585	E. ramosa ramosa	Western Ramosa	-32.90930	19.03453
1586	E. rupestris sp.	Leafy Tricostata	-33.81410	19.17835
1587	E. setacea setacea	Setacea	-33.83168	19.17950

1588	E. ramosa ramosa	Western Ramosa	-33.83168	19.17950
1589	E. setacea disticha	Fernkloof A	-34.30883	18.85788
1590	E. rehmannii subspicata	Subspicata	-34.34018	18.83532
1591	E. rehmannii filiformis	Filiformis	-34.34018	18.83532
1592	E. rehmannii filiformis	Filiformis	-34.39050	19.26952
1593	E. setacea setacea	Fernkloof A	-34.39183	19.27177
1594	E. rupestris tricostata	Restioid Tricostata	-34.38903	19.27247
1595	E. setacea setacea	Fernkloof B	-34.38903	19.27247
1597	E. setacea disticha	Fernkloof A	-34.37863	19.29052
1598	E. setacea disticha	Fernkloof A	-34.37690	19.29173
1599	E. setacea setacea	Fernkloof A	-34.37690	19.29173
1600	E. setacea setacea	Fernkloof B	-34.38563	19.27005
1601	E. ramosa aphylla	Aphylla	-34.34630	18.93113
1602	E. rupestris dodii	Dodii	-34.34245	18.93437
1603	E. rupestris tricostata	Restioid Tricostata	-34.34407	18.93048
1604	E. setacea uniflora	Uniflora	-34.35525	18.86197
1605	E. rupestris rupestris	Western Rupestris	-33.97105	19.50848
1606	E. ramosa ramosa	Western Ramosa	-33.97057	19.50752
1609	E. ramosa aphylla	Western Ramosa	-33.63283	19.14413
1610	E. setacea setacea	Setacea	-33.63283	19.14413
1611	E. ramosa ramosa	Western Ramosa	-33.62825	19.14570
1614	E. setacea setacea	Setacea	-34.07743	18.39373
1615	E. rehmannii filiformis	Filiformis	-34.06005	18.39265
1616	E. rupestris dodii	Dodii	-34.06005	18.39265
1617	E. setacea setacea	Setacea	-33.96505	18.41278
1618	E. rupestris tricostata	Leafy Tricostata	-33.46028	19.43952
1621	E. rupestris tricostata	Leafy Tricostata	-33.45923	19.44975
1622	E. ramosa aphylla	Western Ramosa	-33.45683	19.44610
1624	E. ramosa ramosa	Western Ramosa	-33.42360	19.44280
1625	E. longifolia		-33.42002	19.43742
1626	E. ramosa ramosa	Western Ramosa	-33.69532	19.08932
1628	E. microlaena		-33.68852	19.10035
1629	E. setacea setacea	Setacea	-33.68850	19.10298
1631	E. setacea scabra	Scabra	-33.99345	20.46272
1632	E. rupestris tricostata	Leafy Tricostata	-33.99035	20.46742
1633	E. setacea setacea	Setacea	-33.98173	20.47318
1634	E. rupestris rupestris	Western Rupestris	-33.98243	20.47443
1635	E. ramosa ramosa	Eastern Ramosa	-33.99095	20.46572
1636	E. setacea scabra	Scabra	-33.96827	20.81500
1638	E. dura		-33.96608	20.80160
1639	E. rupestris dodii	Dodii	-33.95352	20.80165
1640	E. rupestris tricostata	Scabra	-33.95352	20.80165
1641	E. setacea scabra	Scabra	-33.95352	20.80165
1642	E. setacea scabra	Scabra	-33.95352	20.80165
1643	E. ramosa ramosa	Eastern Ramosa	-33.94778	20.80722
1644	E. setacea scabra	Scabra	-33.98415	21.22685
1645	E. ramosa ramosa	Eastern Ramosa	-33.89910	22.01970
1646	E. dura		-33.89910	22.01970
1647	E. rehmannii rehmannii	Rehmannii	-33.87240	22.02882
1648	E. rupestris tricostata	Leafy Tricostata	-33.87240	22.02882
1649	E. rehmannii rehmannii	Rehmannii	-33.87107	22.02417
1653	E. rehmannii rehmannii	Rehmannii	-33.87087	24.03602
1654	E. ramosa ramosa	Eastern Ramosa	-33.35222	22.04607
1655	E. rupestris rupestris	Eastern Rupestris	-33.38050	21.35470

1656	E. ramosa ramosa	Eastern Ramosa	-33.38050	21.35470
1657	E. rupestris tricostata	Wemmershoek	-33.80962	19.05945
1658	E. rehmannii filiformis	Subspicata	-34.47162	19.67233
1659	E. rehmannii subspicata	Subspicata	-34.66448	19.67995
1660	E. rehmannii subspicata.	Subspicata	-34.39110	20.41868
1661	E. rupestris tricostata	Restioid Tricostata	-34.09987	18.44823
LW1	E. rehmannii filiformis	Filiformis	-34.11457	18.41063
LW2	E. rupestris rupestris	Eastern Rupestris	-33.35365	22.03552
LW3	E. rehmannii subspicata	Subspicata	-34.25153	18.42237
LW4	E. rehmannii filiformis	Filiformis	-34.39130	19.26955
SM123	E. rupestris rupestris	Western Rupestris	-34.08788	19.85271