Molecular Phylogenetics and Evolution 51 (2009) 31–43

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Using fossils and molecular data to reveal the origins of the Cape proteas (subfamily Proteoideae)

Hervé Sauquet [a,b,c,*], Peter H. Weston [b], Nigel P. Barker [d], Cajsa Lisa Anderson [e], David J. Cantrill [c,f], Vincent Savolainen [a,g]

[a] Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond TW9 3DS, UK
[b] National Herbarium of New South Wales, Botanic Gardens Trust, Mrs Macquaries Road, Sydney, NSW 2000, Australia
[c] Department of Palaeobotany, Swedish Museum of Natural History, P.O. Box 50007, 104 05 Stockholm, Sweden
[d] Department of Botany, Rhodes University, Grahamstown 6140, South Africa
[e] Department of Systematic Biology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, SE-752 36 Uppsala, Sweden
[f] National Herbarium of Victoria, Royal Botanic Gardens, Melbourne, Private Bag 2000, South Yarra, Vic. 3141, Australia
[g] Imperial College London, Silwood Park Campus, Buckhurst Road, Ascot, Berkshire SL5 7PY, UK

ARTICLE INFO

Article history:
Received 4 January 2008
Revised 10 December 2008
Accepted 11 December 2008
Available online 24 December 2008

Keywords:
Molecular dating
Bayesian dating
Penalized likelihood
Fossil calibration
Cape Floristic Region
Cape clades
Proteaceae

ABSTRACT

The angiosperm family Proteaceae is a distinct component of the Cape Floristic Region biodiversity hotspot with 330 endemic species. Phylogenetic analyses of subfamily Proteoideae using sequence data from one nuclear and six plastid loci show that most of this diversity is contained in two distinct Cape floral clades. Molecular dating analyses, using Bayesian and penalized likelihood methods and four phylogenetically supported fossil age constraints, reveal contrasting histories for these two clades. The genus *Protea* belongs to a lineage that may have been in Africa since the Late Cretaceous but began to diversify in the Cape only 5–18 Myr ago. In contrast, the Leucadendrinae clade presumably arrived in the region no earlier than 46 Myr ago by long-distance dispersal from an Australian ancestor and the extant members of this clade began to diversify in the Cape 22–39 Myr ago. These results join a growing number of case studies that challenge the commonly accepted view that most of the Cape flora radiated synchronously in the Late Miocene and Early Pliocene when a Mediterranean climate settled in the region.

1. Introduction

With more than 6000 endemic plant species, the Cape Floristic Region (CFR) of South Africa is both one of the most species-rich regions in the World and one under serious threat from human impact, which has led to its recognition as one of 34 biodiversity hotspots (Myers et al., 2000; Mittermeier et al., 2004). Although its extraordinary diversification is thought to be linked with the onset of a Mediterranean climate in the Late Miocene (Linder et al., 1992; Goldblatt and Manning, 2000), the age and origins of its floral components remain poorly known overall. For instance, it is still unclear whether all the largest clades in the Cape have diversified at the same time, when their ancestors first appeared in the region, and where they came from. Central to these questions is how to obtain reliable time estimates for important events in the phylogenetic history of the Cape flora. Excellent progress has been made recently in developing molecular dating methods that do not assume a strict molecular clock (e.g., Thorne and Kishino, 2002; Sanderson, 2002) and these have been applied to a number of plant clades in the Cape (Richardson et al., 2001; Goldblatt et al., 2002; Linder and Hardy, 2004; Edwards and Hawkins, 2007; Forest et al., 2007). However, partly due to the lack of a suitable fossil record for these groups, these studies have often relied on a single, distant calibration point, sometimes a secondary age estimate derived from a previous, higher-level dating study (e.g., Goldblatt et al., 2002). Since the use of multiple calibration points is now widely considered to be critical for divergence time estimation when molecular rates vary (Near and Sanderson, 2004; Benton and Donoghue, 2007; Rutschmann et al., 2007), it is desirable to apply these methods to at least one plant group with a good fossil record.

Proteoideae, a subfamily of Proteaceae, appear to be an excellent candidate for several reasons. First, they are abundant in the Cape, comprising 13 genera and 331 species, 99% of which are endemic (Table 1; Cowling and Lamont, 1998). Second, the group has left a rich and widespread track in the fossil record throughout Southern Gondwana. This has the potential to provide many calibration points across the family. Last, Proteoideae are flagship members of the Cape flora. They are diagnostic of the 'fynbos', oc-

* Corresponding author. Address: Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond TW9 3DS, UK. Fax: +44 2083325310.
E-mail address: herve.sauquet@gmail.com (H. Sauquet).

1055-7903/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2008.12.013

Table 1
Proteaceae genera in South Africa and the Cape Floristic Region (classification according to Weston and Barker, 2006; distribution and species numbers after Mabberley, 1997; Rourke, 1998, and Rebelo, 2001).

Genus	Distribution	Total spp.	CFR spp.
Grevilleoideae			
Macadamieae: Macadamiinae			
Brabejum L.	CFR	1	1
Proteoideae			
Leucadendreae: Leucadendrinae			
Diastella Salisb. ex Knight	CFR	7	7
Leucadendron R.Br.	CFR (and E South Africa)	85	84
Leucospermum R.Br.	CFR (and E South Africa to Zimbabwe)	48	45
Mimetes Salisb.	CFR	13	13
Orothamnus Pappe ex Hook.	CFR	1	1
Paranomus Salisb.	CFR	19	19
Serruria Salisb.	CFR	54	54
Sorocephalus R.Br.	CFR	11	11
Spatalla Salisb.	CFR	20	20
Vexatorella Rourke	CFR	4	4
Petrophileae			
Aulax Bergius	CFR	3	3
Proteeae			
Faurea Harv.	South Africa to tropical Africa (and Madagascar)	15	1
Protea L.	South Africa to tropical Africa	112	69
Total Proteaceae		393	332

cupy a wide range of CFR habitats, and may thus be regarded as a typical component of the Cape. Tribe Proteeae (sensu Johnson and Briggs, 1975) was listed by Linder (2003) as one of 33 Cape floral clades to foster more research on the timing and patterns of plant radiation in the region. Phylogenetic analyses have shown that this group is polyphyletic and composed of three different clades nested in subfamily Proteoideae, each with a sister group outside Africa (Hoot and Douglas, 1998; Barker et al., 2002, 2007; Weston and Barker, 2006; Table 1). An additional member of Proteaceae occurring in the Cape is *Brabejum*, a monotypic genus in subfamily Grevilleoideae.

This paper aims to: (1) reconstruct a phylogeny of subfamily Proteoideae with all of the genera sampled and sequence data from seven loci; (2) identify the biogeographic origins of the three Proteoideae components of the Cape flora; (3) date the origin of these three clades using multiple fossil age constraints and methods that do not assume a strict molecular clock.

2. Materials and methods

2.1. Taxon sampling

Taxon sampling for this study was designed so as to maximize character sampling over the seven loci selected, while including at least one representative of each genus of Proteoideae. This led us to combine sequence data from different species of the same genus where necessary (Table 2). While species-level phylogenies are scarce in the family, we avoided, to the best of our knowledge, combining sequence data from species of genera whose monophyly has been questioned (e.g., Barker et al., 2002), sampling in one specific subclade of such genera as much as possible. *Protea* is the only genus for which enough data were available to be represented by more than one terminal taxon. The lumping of *Protea repens* and *Protea roupelliae* was supported by the species-level phylogeny of Barraclough and Reeves (2005), which showed that these two species belong to the same subclade of *Protea*, to the exclusion of the other two *Protea* representatives in our analysis (*Protea cynaroides* and *Protea neriifolia*). Specific attention was also paid not to use the original *atpB* and *atpB–rbcL* sequences of *Protea* from Hoot and Douglas (1998), which have turned out to be produced from a misidentified specimen of *Leucadendron* (see Barker et al., 2007).

In addition to representatives of each genus of Proteoideae, we have included both genera of Symphionematoideae, the putative sister group of Proteoideae (Weston and Barker, 2006), as well as six representatives of Grevilleoideae, one representative of Persoonioideae, and *Bellendena*. Furthermore, the two closest outgroups of Proteaceae (*Platanus* and *Nelumbo*) as well as the putative sister group of Proteales (Sabiaceae) and a more distant outgroup (Buxaceae) were also represented by one terminal each in the analysis to ensure proper rooting and calibration of the molecular dating procedure. This is based on congruent evidence from numerous molecular studies of angiosperm and eudicot phylogeny (Hoot et al., 1999; Qiu et al., 2000, 2005; Soltis et al., 2000; Hilu et al., 2003; Kim et al., 2004; Worberg et al., 2007).

2.2. DNA sequencing

New ITS sequences were generated by N.P.B. and P.H.W. for this study following the protocols outlined in Barker et al. (2002). New *rbcL* sequences were generated by N.P.B. as in Barker et al. (2007). New *matK* data were produced by C.L.A. as follows. DNA was extracted from herbarium or silica gel-dried leaf material using DNeasy plant mini kit (Qiagen), following the manufacturer's instructions. The complete *matK* gene and the 3′-end of the *trnK* intron were amplified by PCR using a combination of primers specifically designed for Proteaceae (Table 3), using either Accu Power PCR PreMix (Bioneer) or puRETaq Ready-To-Go PCR Beads kit (Amersham biosciences) and conditions as follows: 3 min at 95 °C; 35 cycles of 30 s at 95 °C, 1 min at 50 °C, and 2 min at 72 °C; and 10 min at 72 °C. The PCR products were purified with the QIAquick PCR Purification (Qiagen) or the Milllipore purification system and sequenced by Macrogen Inc., Korea (protocols at www.macrogen.com). New *rbcL* and *trnL* intron/*trnL–trnF* spacer sequences by H.S. were produced mostly from samples of the Kew DNA Bank or silica gel-dried leaf material extracted following a modified version of the CTAB protocol (Doyle and Doyle, 1987). *rbcL* was amplified using a combination of two universal and two Proteaceae-specific primers (Table 3) and PCR as follows: 2 min at 94 °C; 30 cycles of 1 min at 94 °C, 1 min at 48 °C, and 1.5 min at 72 °C; and 4 min at 72 °C. The *trnL* intron and *trnL–trnF* spacer were amplified using the universal primers of Taberlet et al. (1991) and PCR as follows: 3 min at 94 °C; 35 cycles of 1 min at 94 °C, 1 min at 52 °C, and 2 min at 72 °C; and 10 min at 72 °C. PCR products were purified using QIAquick columns and sequencing reactions were conducted using 26 cycles of 10 s at 96 °C, 5 s at 50 °C, and 4 min at 60 °C. Individual sequences were read by an automated ABI 3100 capillary sequencer from Applied Biosystems Inc and assembled using Sequencher version 4.5 (Gene Codes Corporation, Ann Arbor, MI, USA).

2.3. Alignment and data set assembly

New sequences as well as selected sequences imported from GenBank were aligned for each locus using BioEdit version 7.0.9 (Hall, 1999). A global alignment was first produced using the ClustalW algorithm as implemented in this software and then checked and edited by eye where necessary. Regions of noncoding DNA that could not be aligned unambiguously due to many overlapping insertion–deletions of different length were excluded from phylogenetic analyses (mainly *atpB–rbcL* spacer and ITS). In addition, the particu-