Molecular phylogenetics of Euploca (Boraginaceae): homoplasy in many characters, including the C4 photosynthetic pathway

Michael W Frohlich, Rowan F Sage, Lyn A Craven, Sebastian Schuster, Guillaume Gigot, Hartmut H Hilger, Hossein Akhani, Parastoo Mahdavi, Federico Luebert, Maximilian Weigend, et al.

Botanical Journal of the Linnean Society, 2022, **199**, 497–537. With 17 figures.

Molecular phylogenetics of *Euploca* (Boraginaceae): homoplasy in many characters, including the C_4 photosynthetic pathway

MICHAEL W. FROHLICH[1,2,†], ROWAN F. SAGE[3,*], LYN A. CRAVEN[4,†], SEBASTIAN SCHUSTER[1,§], GUILLAUME GIGOT[1,5,§], HARTMUT H. HILGER[6,§], HOSSEIN AKHANI[7,§], PARASTOO MAHDAVI[7,8,§], FEDERICO LUEBERT[9,10], MAXIMILIAN WEIGEND[9], MATS THULIN[11,§], JEFF J. DOYLE[12], JANE L. DOYLE[12], PATRICK VOGAN[3,13,§], ALAN FORREST[14,§], TIMOTHY K. FULCHER[1], DION S. DEVEY[1] and MARK W. CHASE[1,15]

[1]*Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3DS, UK*
[2]*Laboratoire de Reproduction et Développement des Plantes, École Normale Supérieure de Lyon, INRAE, CNRS, 46 Allée d'Italie, 69364 Lyon Cedex 07, France*
[3]*Department of Ecology and Evolutionary Biology, University of Toronto, 25 Willcocks Street, Toronto Ontario M5S 3B2, Canada*
[4]*Australian National Herbarium, Centre for Plant Biodiversity Research, GPO Box 1600, Canberra ACT 2601, Australia*
[5]*Current address: Muséum National d'Histoire Naturelle, UMS 2006 PatriNat, Paris, France*
[6]*Institut für Biologie, Freie Universität Berlin, Systematische Botanik und Pflanzengeographie, Altensteinstraße 6, D-14195, Berlin*
[7]*Halophytes and C_4 Plants Research Laboratory, Department of Plant Sciences, School of Biology, College of Science, University of Tehran, P.O. Box 14155-6455, Tehran, Iran*
[8]*Current address: Institute of Biology and Environmental Science, Vegetation Science & Nature Conservation, University of Oldenburg, Oldenburg, Germany*
[9]*Nees-Institut für Biodiversität der Pflanzen, Rheinische Friedrich-Wilhelms-Universität Bonn, Meckenheimer Allee 170, D-53115 Bonn, Germany*
[10]*Facultad de Ciencias Agronómicas and Departmento de Silvicultura y Conservación de la Naturaleza, Universidad de Chile, Santa Rosa 11315, La Pintana, Santiago 8820000, Chile*
[11]*Systematic Biology, Department of Organismal Biology, EBC, Uppsala University, Norbyvägen 18D, SE-75236 Uppsala, Sweden*
[12]*School of Integrative Plant Science, Section of Plant Breeding & Genetics and Section of Plant Biology, 240 Emerson Hall, Cornell University, Ithaca, New York 14853 USA*
[13]*Current address: NewLeaf Symbiotics, Inc., 1005 North Warson Road, Saint Louis, MO 63132, USA*
[14]*Centre for Middle Eastern Plants, Royal Botanic Garden, Edinburgh, 20A Inverleith Row, Edinburgh, EH3 5LR, UK*
[15]*Department of Environment and Agriculture, Curtin University, Bentley, Western Australia 6102, Australia*

*Corresponding author. E-mail: r.sage@utoronto.ca
[†]Deceased.
[§]These authors contributed by obtaining DNA sequences or by providing dried material for DNA extraction, delta $\delta^{13}C$ determinations and/or seeds.

The co-authors dedicate this paper to the memory of Mike Frohlich, who passed away suddenly in late November, 2021, within a few days of signing off on the page proofs for this paper. This publication was 20 years in the making, and largely came to fruition through the perseverance, networking and insights of M. Frohlich. *Heliotropium*/*Euploca* was Mike's passion, and this work became his magnum opus. For the rest of us, his dedicated effort is now a cherished gift that can be shared far into the future.

Received 8 January 2021; revised 29 July 2021; accepted for publication 5 October 2021

We present a phylogenetic analysis using plastid (*matK*, *rbcL*) and nuclear (nrITS) DNA for diverse *Euploca* spp. (formerly Heliotropium section Orthostachys) from the worldwide distribution of a genus and including species encompassing the wide physiological and morphological diversity of the genus. Our results indicate that some remarkably complex features arose multiple times in parallel in *Euploca*, including attributes of its subsections under section *Orthostachys*, notably plants that, above ground, consist almost entirely of inflorescences. To elucidate in greater detail the distribution of C_4 species in *Euploca* and *Heliotropium s.s.*, we made > 800 $\delta^{13}C$ determinations, including some from the traditional genus *Tournefortia*. We greatly increase the number of proven C_4 species in *Euploca*, but found none outside Euploca. Of the tested *Euploca* spp., *c.* 28% are C_3 or intermediate in carbon fixation pathway. Our phylogenetic results indicate four parallel/convergent acquisitions of C_4 photosynthesis or fewer origins with subsequent loss in some species.

ADDITIONAL KEYWORDS: C_2 photosynthesis – C_3–C_4 intermediate – contagious distribution – desert flora – *Heliotropium* section *Orthostachys* – inflorescence plant – seed set modelling.

INTRODUCTION

Euploca Nutt. (=*Heliotropium* L. section *Orthostachys* R.Br.) (Boraginaceae *sensu* APG IV or Heliotropiaceae) is a widespread genus with *c*. 100–120 species (Frohlich, 1978; Birecka, Frohlich & Glickman, 1983; Luebert *et al.*, 2016). It exhibits remarkable variation in physiological, morphological, chemical, reproductive and life-history features, including various intermediate states, making it a potentially valuable system for evolutionary studies. In particular, C_4 photosynthesis arose in the genus, and a number of species show a range of intermediate phenotypes between the C_3 and C_4 pathways (Frohlich, 1978; Vogan, Frohlich & Sage, 2007; Muhaidat *et al.*, 2011; Sage, Sage & Kocacinar, 2012). To test evolutionary hypotheses regarding how, where and in what sequence the various traits arose, especially C_4 photosynthesis, we present a molecular phylogenetic study to clarify relationships in *Euploca*, with a particular emphasis on phylogenetic nodes uniting C_4 and non-C_4 clades.

The taxonomy of *Euploca* and *Heliotropium s.l.* has been unsettled for decades. The most recent worldwide species-level study of all *Heliotropium s.l.* was published by Alphonse de Candolle (de Candolle, 1845: vol. 9: 529–548, 565–566), but he attributed primary authorship to his father, Augustin Pyramus de Candolle (DC.). *Heliotropium* has long been considered closely related to *Tournefortia* L. in Boraginaceae *s.l.*, and Boraginaceae have long been considered closely related to Hydrophyllaceae. Molecular studies have demonstrated that Hydrophyllaceae and Lennoaceae fall within the broader concepts of Boraginaceae, resulting in the expanded Boraginaceae of APG III and APG IV (APG III, 2009; Reveal & Chase, 2011; APG IV, 2016). To render families more homogeneous, Luebert *et al.* (2016) divided Boraginaceae *sensu* APG into 11 narrower families, and placed in Heliotropiaceae all the species of *Heliotropium* and *Tournefortia*, as traditionally understood, essentially from the time of Linnaeus (1753: 130, 140) through to the end of the 20th century: traditionally, *Heliotropium* and *Tournefortia* were distinguished by the key character, respectively, of dry versus moist fruit. Here we refer to those traditional circumscriptions as *Heliotropium s.l.* and *Tournefortia s.l.*, although the designation '*s.l.*' for the former may be problematic, as noted below.

Heliotropium section *Orthostachys* has been recognized taxonomically since 1810 (Brown, 1810: 493) and contains the few known C_4 species in *Heliotropium s.l.* Johnston (1928) divided section *Orthostachys* into three subsections, two based on the presence (*Bracteata* I.M.Johnst.) or absence (*Ebracteata* I.M.Johnst.) of bracts in the cymose inflorescences. His third subsection, *Axillaria* I.M.Johnst., was described as having 'flowers borne along the leafy stem' (Johnston, 1928: 47). Förther (1998) published a morphologically based generic and section-level study of Heliotropioideae, in which he reinstated *Schleidenia* Endl. for *Heliotropium* section *Orthostachys* subsection *Axillaria* and split *Hilgeria* Förther from section *Orthostachys*.

Based on a survey of Kranz leaf anatomy in Mexican species of *H.* section *Orthostachys*, Frohlich (1978: 58–70, 107–112) showed that C_4 photosynthesis is present in most, but not all species of subsections *Bracteata* and *Axillaria,* and that some of those species are intermediate, wherea C_4 is absent in subsection *Ebracteata*; thus, Kranz anatomy does not define subsections *Bracteata* plus *Axillaria*. However, these interpretations relied on a limited number of species without the context of a modern phylogenetic framework.

Recent molecular studies (Diane, Förther & Hilger, 2002; Hilger & Diane, 2003; Luebert *et al.*, 2011a; Nazaire & Hufford, 2012; Weigend *et al.*, 2014) indicated that *Heliotropium s.l.* includes species of *Tournefortia s.l.* embedded in widely separated places, so neither genus is monophyletic. A major split is evident in *Heliotropium s.l.*, with one clade consisting of *H.* section *Orthostachys* and *Tournefortia*

s.l. section *Cyphocyema* I.M.Johnst. and the other containing the remaining sections of *Heliotropium s.l.* (subgenus *Heliotropium*) and *Tournefortia s.l.* The South American shrub *Ixorhea* Fenzl appears to be sister to *Heliotropium* section *Orthostachys* plus *Tournefortia* section *Cyphocyema* (Hilger & Diane, 2003). *Schleidenia* and *Hilgeria* fall in *Heliotropium* section *Orthostachys*.

The type species of *Heliotropium* (*H. europaeum* L.) is not in section *Orthostachys*, so one resolution would be to recognize section *Orthostachys* as the genus *Euploca* Nutt., for which the type, dating from 1837, is *E. convolvulacea* Nutt. (Nuttall, 1837). The small sister clade, formerly *Tournefortia s.l.* section *Cyphocyema*, would then be recognized as the genus *Myriopus* Small (Small, 1933). Alternatively, section *Orthostachys* could remain in *Heliotropium*, with *Heliotropium* expanded to encompass all of *Tournefortia* (Craven, 2005) and perhaps also *Ixorhea*. Either resolution results in many name changes, and the different grammatical genders of *Heliotropium* (neuter) versus *Euploca* and *Tournefortia* (feminine) cause further complications, especially for computerized searches for specific epithets. We agree with Hilger & Diane (2003) that the less disruptive resolution is to recognize *Euploca* and *Myriopus*. New combinations in *Euploca* required for this paper have been made (Frohlich, Thulin & Chase, 2020). As a result, *Heliotropium s.l.* loses *Euploca*, and gains most of *Tournefortia s.l.* and some small genera (Diane *et al.*, 2016) are subsumed into *Heliotropium s.s.* The designations 'wide' versus 'narrow' are not fully accurate, as not all of *Heliotropium s.s.* is contained in *Heliotropium s.l.*, but for convenience we use them here.

Euploca is most diverse in moderately dry to arid regions of warm-temperate to tropical latitudes, although some species occur in moist, evenly or seasonally wet habitats (Frohlich, 1978; Craven, 1996), often on alkaline substrates. Biogeographically, *Euploca* has two major diversity centres: tropical and subtropical North, Central and South America, including the Caribbean, and Australia (Frohlich, 1978; Craven, 1996; Förther, 1998; Diane *et al.*, 2016). There are several species in the Horn of Africa and adjacent parts of Arabia (Thulin, 2005, 2006), but the large dry regions of south-western and southern Asia, northern, southern and south-western Africa have few, but widespread, species. In contrast, *Heliotropium s.s.* is diverse in south-western Asia and South America, but not in North America, the Caribbean or Australia (Johnston, 1928; Akhani & Förther, 1994; Craven, 1996; Luebert, Hilger & Weigend, 2011b; Akhani, 2007). *Euploca* spp. exhibit great diversity in life-history strategy, including long-lived shrubs of semi-deserts, geophytes and short-lived annuals of seasonal habitats (e.g. receding shores of dry-season lakes and rivers or ephemerally moist habitats). Some species occur in disturbed habitats, and a few are weeds. Reproductive strategies are diverse, including species that are dioecious/insect pollinated, perfect-flowered/insect pollinated, chasmogamous/apparently obligately self-pollinated and cleistogamous, and there is even a putative apomict (Frohlich, 1978; Craven, 1996; Frohlich, unpublished).

In addition to containing all known C_4 species in *Heliotropium s.l.*, *Euploca* also contains C_3 species and at least some species with intermediate photosynthetic physiologies and leaf structure (Frohlich, 1978: 58–70, 107–112; Vogan *et al.*, 2007; Muhaidat *et al.*, 2011). The phylogenetic relationships between the C_3, C_4 and C_3–C_4 intermediate species are not clear, hampering interpretations of evolutionary change in the structural and physiological characters of photosynthesis. A principal objective here is to characterize the distribution of C_4 photosynthesis in *Heliotropium s.l.*, especially in section *Orthostachys* (=*Euploca*), specifically by mapping its C_4 and intermediate species onto the phylogenetic tree. In addition, the ecological and biogeographic distribution of the various forms in *Euploca* could be useful for identifying environments that favour the evolution of C_4 and the consequences of C_4 for the diversification of specific clades. Sage *et al.* (2018) hypothesized that harsh soils (e.g. sand dunes, rocky outcrops and limestone glades) where C_3 and C_3–C_4 intermediate forms of *Euploca* occur would favour C_4 evolution by causing higher rates of photorespiration. This would especially apply to summer annuals, which experience high summer but not lower winter temperatures. Because of the variation in photosynthetic types, coupled with diversity in life form and reproductive strategy, *Euploca* represents an excellent group for investigating the evolutionary ecology of C_4 photosynthesis, its potential relationship to life form and life history and its consequences for long-distance dispersal and speciation.

Large clades often exhibit multiple origins of C_4 photosynthesis; e.g. grasses have > 20 independent clades with C_4 photosynthesis, and in eudicots there are at least 14 independent origins in Amaranthaceae (including the old Chenopodiaceae) (GPWG II, 2012; Kadereit, Ackerly & Pirie, 2012; Sage, 2016). Even in small clades there can be multiple origins, as shown for *Flaveria* Juss., with only 22 species but two independent origins of C_4 photosynthesis (McKown, Moncalvo & Dengler, 2005; Stata, Sage & Sage, 2019). In contrast, certain large clades, notably the species-rich *Euphorbia* L. subgenus *Chamaesyce* (Gray) House (Euphorbiaceae, > 350 species) have a single origin followed by extensive radiation that generates substantial diversity of form (Yang & Berry, 2011). *Flaveria* has dominated research on C_4 evolution, largely because it has more confirmed intermediates

than any other C_3/C_4 clade (Sage *et al.*, 2014). However, to have a broad understanding of C_4 evolution, there is a widespread desire to identify additional clades that can independently clarify when, where, how and why C_4 photosynthesis evolved. *Euploca* has been identified as a potentially valuable system for examining C_4 evolution due to the significant diversity of physiological and structural character states associated with C_3/C_4 diversity (Muhaidat *et al.*, 2011). Two categories of C_3/C_4 intermediacy occur in *Euploca*. The first is C_2 photosynthesis, which is a CO_2-concentrating mechanism that shuttles photorespiratory glycine from mesophyll cells, where it is produced, to bundle sheath cells, where it is metabolized to CO_2 and serine by glycine decarboxylase (Rawsthorne, 1992; Sage *et al.*, 2014). Bundle-sheath chloroplasts positioned alongside mitochondria rapidly refix this CO_2 and in doing so, operate at greater photosynthetic efficiency than chloroplasts in the mesophyll cells. The second category, termed 'proto-Kranz', occurs in physiologically functional C_3 plants when the bundle sheath mitochondria and some chloroplasts cluster along the vascular tissues, forming a pattern that resembles an early version of the intense organelle clustering found in C_2 and C_4 plants (Sage *et al.*, 2012). Proto-Kranz species often exhibit slightly reduced CO_2 compensation points of photosynthesis, which with their structural features indicate proto-Kranz could be the first recognizable stage toward C_2 and C_4 evolution (Sage, Khoshravesh & Sage, 2014). At least five C_2 intermediates have been identified in *Euploca* (*E. convolvulacea*, *E. cremnogena*, *E. greggi*, *E. lagoens*is and *E. racemosa*). Three species exhibit proto-Kranz phenotypes (*E. karwinsky*i, *E. procumbens* and *E. ovalifolia*; Frohlich, 1978: 65–70; Vogan *et al.*, 2007; Muhaidat *et al.*, 2011). *Euploca* also has more C_3 species than *Flaveria*, indicating greater potential to understand the ecological and evolutionary context that predisposed this group to evolve the C_4 pathway. This physiological diversity indicates *Euploca* could be valuable for understanding how C_4 photosynthesis evolved; however, the phylogenetic relationships of the photosynthetic types have been unknown, such that hypotheses of C_4 evolution have not been readily testable in *Euploca*. (Muhaidat *et al.*, 2011).

To understand evolutionary change in *Euploca*, particularly in reference to C_4 evolution, a robust phylogenetic tree with a large number of *Euploca* spp. is required, with a comprehensive identification of the C_4 and non-C_4 species in the genus. Here we present a phylogenic analysis based on plastid (*matK*, *rbcL*) and nuclear (ribosomal nrITS) loci for 74 *Euploca* spp. To evaluate photosynthetic pathway, we determined carbon isotope ratios ($\delta^{13}C$) from *c.* 850 specimens sampled from 230 species (and one variety) of the traditional *Heliotropium s.l.* and a few from the traditional *Tournefortia s.l.* CO_2 assimilation pathway is mapped onto the tree to infer where and how often C_4 photosynthesis arose (or was lost). C_3 plants typically have $\delta^{13}C$ of −22‰ to −32‰, whereas C_4 biomass has $\delta^{13}C$ between −9‰ and −16‰. $\delta^{13}C$ values generally cannot distinguish C_2 species from C_3 species, unless a strong C_4 cycle is present, in which case the species can exhibit $\delta^{13}C$ values between −16 and −22‰ (Monson *et al.*, 1988; Alonso-Cantabrana & von Caemmerer, 2016). Although the phylogenetic and isotope work will clarify specific clades and their photosynthetic pathway, it is not the purpose of this paper to comprehensively resolve taxonomic or nomenclatural questions. To facilitate name comparisons and computerized searches, we list all the studied *Euploca* spp. under both generic names in the Supporting Information, Table S1.

MATERIAL AND METHODS

Phylogenetic studies

Taxon sampling

For initial molecular phylogenetic studies begun in 1989, MWF grew selected Mexican, US and Australian species from seed to extract high-quality DNA for restriction fragment length polymorphism (RFLP) methods. Species were chosen to represent the morphological diversity of *Heliotropium* section *Orthostachys*, but selection was limited by the availability of viable seed, primarily collected by MWF (Mexico and USA) and LAC (Australia).

Recently, we sampled from the worldwide distribution of *Euploca*, focusing on regions of greatest diversity, i.e. North America, Australia and the West Indies. We especially focused on taxa that might be evolutionarily close to the boundary between C_3 and C_4, identified because of previous suggestions of intermediacy (Frohlich, 1978: 65) or because $\delta^{13}C$ values showed the presence of C_3, (or proto-Kranz or C_2) photosynthesis, but flower and inflorescence morphology suggested relationship to C_4 taxa.

Some species from Mexico and the USA were again grown from seed and DNA isolated from fresh tissue. Other samples from Mexico and Australia had been silica gel-dried in the field (Chase & Hills, 1991). Additional samples from these regions, and nearly all samples from elsewhere, were obtained from herbarium specimens. Some important samples were supplied by collaborators who study *Euploca* (often as *Heliotropium*). Samples that gave DNA of sufficient quality to obtain sequences (including DNAs from the 1989 study) are included in the analyses. These comprise 105 accessions of *Euploca*, representing 74

taxonomic species. Samples for the first outgroup, *Myriopus* (four accessions; three species), came from herbarium specimens.

We used only seven accessions to represent the next outgroup, *Heliotropium s.s.*, including a diversity of species from both hemispheres. These sequences mostly come from ongoing parallel projects with collaborators. This small number of samples does not provide useful information on relationships in this outgroup, so these are represented only by a single triangle in the trees presented and are not listed in Table 1; results for *Heliotropium s.s.* will be published elsewhere by H. Akhani and colleagues. We used two accessions of *Hydrophyllum canadense* L. as a further outgroup, also represented by a triangle.

For these analyses, we only included sequences obtained at RBG Kew, as described before, because they provide nearly complete data for all three regions (nrITS, *rbcL* and *matK*). Failure of amplification, despite repeated attempts, occurred in some cases, represented by '?' in the matrix; 3.1% of sequences are missing, counting the excluded nrITS sequences (see below).

DNA EXTRACTIONS, RFLPS, PCR AND SEQUENCING

Most DNA extractions, including early extractions for RFLPs and recent extractions from live, silica-gel-dried and herbarium material were done using the CTAB method (Doyle & Doyle, 1987), with extractions at Kew followed by CsCl gradient purification and dialysis. Some recent extractions from small amounts of herbarium material were done using Machery–Nagel spin columns following the manufacturer's protocols. DNAs are accessioned in the Kew DNA Bank (Table 1): http://apps.kew.org/dnabank/homepage.html. RFLP studies were carried out at the Bailey Hortorium, Cornell, essentially as described in Doyle, Doyle & Brown, (1990), using cloned *Nicotiana* L. and/or *Petunia* Juss. plastid probes, yielding 63 informative RFLP characters.

PCRs were done with ReddyMix PCR master mix (Thermo Scientific) with 1.5 µM $MgCl_2$ for nrITS and 2.5 µM $MgCl_2$ for *rbcL* and *matK*. Reaction volumes were 25 or 50 µL. The larger reactions contained 45 µL of master mix, 1 or 2 µL of genomic DNA solution (of variable concentration), 1 µL of each primer (at 100 ng/µL) and 1 µL of bovine serum albumin (at 0.4%). For nrITS, 0.4 µL of dimethylsulphoxide was also added to destabilize potential secondary structure in nrITS that might impede PCR. Smaller reactions were performed in the same proportions.

Primers for nrITS were 27SE and 17SE (Sun *et al.*, 1994). The amplification program (temperature – time) was 94° 5:00; (94° 1:00; 50° 1:00; 72° 1:00); 72° 7:00; 16° hold, with steps in parentheses repeated for 28 cycles or occasionally, for difficult amplifications, for as many as 40 cycles.

Primers for the 5′ and 3′ parts of *rbcL* were initially 1F (Lledó *et al.*, 1998) with 724(R)d (Lledó *et al.*, 1998; as 724R*), and 636F (Muasya *et al.*, 1998) with 1360R (Reeves *et al.*, 2001), but in *Euploca* the target sequence for 724(R)d differs significantly from that primer, so we substituted a new primer, 724rHT (TCG CAG TTA CCT GCA GTA GC), which proved far superior. The amplification program was 94° 5:00; (94° 1:00; 48° 1:00; 72° 1:00); 72° 7:00; 16° hold, with steps in parentheses repeated 28 cycles or occasionally, for difficult amplifications, for as many as 40 cycles.

For *matK* we initially used forward primer X (TAATTTACGATCAATTCATTC) and reverse primer 5 (GTTCTAGCACAAGAAAGTCG) previously used by others at RBG Kew, but amplifications proved to be difficult. Later, we used primers 1R KIM and 3F KIM (Costion *et al.*, 2011) with much better success. After trying several amplification programs, we settled on a four-step PCR cycle: 94° 3:00; (95° 0:30; 50° 1:00; 65° 0:30; 72° 1:00); 72° 7:00; 18° hold, for 35 or occasionally for as many as 42 cycles for the steps in parentheses. The rationale is that a few complementary nucleotides are added to the primer during the initial low temperature extension at 65°, stabilizing the duplex despite mismatches with the primer. The polymerization is then completed rapidly at the standard higher temperature. This 65° step is probably significant only for the first few cycles before significant product with exact match primer sequences at both ends has appeared.

PCR reactions were purified using NucleoSpin Extraction kit (Macherey-Nagel) following the manufacturer's protocol or on QIAquick 96 PCR Purification kit (Qiagen). Sequencing reactions were done with the PCR primers, and, for the *matK* products amplified with 1R KIM and 3F KIM, also with the internal 5′ primer matKmRht (CCT TAT CAA AGA CTT CTA CAA GAC) and occasionally with 3′ internal primer matKmLht (CKA TTT TTG CTT CAA AAG GGA C). The 3′ primer 1R KIM lies within the PCR product of the primers used initially, so for comparison with the earlier sequences we used the internal primer to obtain sequence all the way up to 1R KIM.

Sequencing reactions were initially done in 10-µL reaction volumes with 0.5 µL of BigDye Terminator Mix (Applied Biosystems), 3 µL 5× cycle sequencing buffer, 0.75 µL of primer (at 10 ng/µL) and *c*. 40–50 ng of purified PCR product. Reactions were cleaned by ethanol precipitation. Subsequently, and with better results, we used 5-µL reaction volumes, containing 0.25 µL of BigDye terminator mix, 1.5 µL 5× cycle

Table 1. Collections used for PCR and DNA sequencing. Authorities for *Euploca* names and synonyms under *Heliotropium* are given in Supporting Information, Table S1.

Name	Kew DNA bank number	Country	Collector and number	Herbarium	GenBank Numbers ITS	Accession *rbcL*	Numbers *matK*
E. aequorea	20588	Australia	*Lepschi 5158*	CANB	MZ578075	MZ611768	MZ617166
E. aequorea	20590	Australia	*Lepschi 5168*	CANB	MZ578093		MZ617184
E. alcyonium	19696	Australia	*Lewis s.n.*	BM or K	MZ578081	MZ611774	MZ617172
E. axillaris	45036	Mexico	*Frohlich 2156*	K	MZ578125	MZ611816	MZ617214
E. axillaris	45037	Mexico	*Frohlich 2512*	K	MZ578124	MZ611815	MZ617213
E. ballii	19707	Australia	*Frazer 240*		MZ578092	MZ611784	MZ617183
E. brachygyne	19698	Australia	*Telford & Butler 9184*	CANB	MZ578072	MZ611765	MZ617163
E. brachygyne	19716	Australia	*Corfield 3096*	BM or K	MZ578071	MZ611764	MZ617162
E. brachythrix	20581	Australia	*Kubien Jesson & Craven 27*	CANB	MZ578063	MZ611756	MZ617155
E. bracteata	45027	Australia	*Fryxell & Craven 4266*	CANB	MZ578079	MZ611772	MZ617170
E. campestris	24410	Bolivia	*De la barra 37*	B	MZ578159	MZ611850	MZ617247
E. chrysantha	24472	Argentina	*Hilger ARG 95.80*	B	MZ578108	MZ611799	MZ617198
E. chrysantha	24474	Argentina	*Hilger ARG 95.81*	B	MZ578107	MZ611798	MZ617197
E. collina	20587	Australia	*Conors 1371*	CANB	MZ578091		MZ617182
E. confertifolia	21309	Mexico	*Frohlich 2523*	K	MZ578122	MZ611813	MZ617211
E. confertifolia	45041	Mexico	*Frohlich 2136*	K	MZ578123	MZ611814	MZ617212
E. conocarpa	45025	Australia	*Fryxell & Craven 3957*	CANB	MZ578066	MZ611759	MZ617158
E. convolvulacea	19734	Texas	*Stephanie Bartel s.n.*	TRT	MZ578143	MZ611834	MZ617231
E. cracens	20576	Australia	*Kubien, Jesson & Craven 3*	CANB	MZ578086	MZ611779	MZ617177
E. cremnogena	33563	Mexico	*Frohlich 2114*	K	MZ578149	MZ611840	MZ617237
E. cunninghamii	19727	Australia	*Craven & Wendel 9288*	CANB	MZ578058	MZ611751	MZ617150
E. cunninghamii	19731	Australia	*Craven 9275*	CANB	MZ578059	MZ611752	MZ617151
E. cupressina	19725	Australia	*Craven & Wendel 9319*	CANB	MZ578095	MZ611786	MZ617186
E. delestangii	19709	Australia	*Cummings 17448*	CANB	MZ578069	MZ611762	
E. dichotoma	45028	Australia	*Craven 8167*	CANB	MZ578060	MZ611753	MZ617152
E. diversifolia	19729	Australia	*Craven & Wendel 9316*	CANB	MZ578088	MZ611781	MZ617179
E. euodes	20583	Australia	*Kubian, Jesson & Craven 30*	CANB	MZ578062	MZ611755	MZ617154
E. fallax	21301	Mexico	*Frohlich 2502*	TRT	MZ578157	MZ611848	MZ617245
E. filiformis	24150	Mexico	*Mexia 9310*	GH	MZ578140	MZ611831	
E. filiformis	24159	Mexico	*Frohlich 2136*	K	MZ578139	MZ611830	MZ617228
E. filiformis	41365	Belize	*Sage s.n.*	TRT	MZ578141	MZ611832	MZ617229
E. flintii	45026	Australia	*Craven 8185*	CANB	MZ578074	MZ611767	MZ617165
E. foliata	19730	Australia	*Craven 9168*	CANB	MZ578089	MZ611782	MZ617180
E. foliata	20580	Australia	*Kunien, Jesson & Craven 22*	CANB	MZ578083	MZ611776	MZ617174
E. foliosissima	21304	Mexico	*Frohlich 2508*	K	MZ578136	MZ611827	MZ617225
E. foliosissima	45029	Mexico	*Frohlich 2105*	K	MZ578135	MZ611826	MZ617224
E. frohlichii	19724	Australia	*Craven et al. 9357*	CANB	MZ578094	MZ611785	MZ617185

Table 1. Continued

Name	Kew DNA bank number	Country	Collector and number	Herbarium	GenBank ITS	Accession *rbcL*	Numbers *matK*
E. fruticosa	45030	Mexico	*Frohlich 2520*	K	MZ578131	MZ611822	MZ617220
E. glabella	19712	Australia	*Cowie & Baker 6401*	BM or K	MZ578055	MZ611748	MZ617147
E. glandulifera	19723	Australia	*Craven & Wendel 9308*	CANB	MZ578158	MZ611849	MZ617246
E. greggii	19732	Texas	*Josh McClung s.n.*	TRT	MZ578105	MZ611796	MZ617195
E. greggii	41554	Texas	*Sage s.n.*	TRT	MZ578106	MZ611797	MZ617196
E. haesum	20585	Australia	*Kubien, Jessun & Craven*	CANB	MZ578057	MZ611750	MZ617149
E. heterantha	19697	Australia	*Wilson & Rowe 1019*	BM or K	MZ578070	MZ611763	MZ617161
E. humifusa	21649	Cayman Islands	*Proctor 35137*	BM	MZ578121	MZ611812	
E. humilis	21608	Jamaica	*Proctor 36510*	BM	MZ578132	MZ611823	MZ617221
E. humilis	41556	Mexico	*Frohlich 2505*	K	MZ578137	MZ611828	MZ617226
E. humilis	45033	Mexico	*Frohlich 2161.2*	K	MZ578138	MZ611829	MZ617227
E. humilis	45034	Mexico	*Frohlich 1980*	K	MZ578133	MZ611824	MZ617222
E. inexplicita	19717	Australia	*Mitchell PRP567*	BM or K	MZ578096	MZ611787	MZ617187
E. karwinskyi	21307	Mexico	*Frohlich 2521*	K	MZ578142	MZ611833	MZ617230
E. lagoensis	21601	Jamaica	*Proctor 33730*	BM	MZ578110	MZ611801	MZ617200
E. lagoensis	21602	Cuba	*Howard et al 289*	BM	MZ578111	MZ611802	MZ617201
E. lagoensis	21704	Argentina	*Pedersen 8758*	K	MZ578112	MZ611803	MZ617202
E. leptalea	19728	Australia	*Craven & Wendel 9287*	CANB	MZ578056	MZ611749	MZ617148
E. limbata	21303	Mexico	*Frohlich 2507*	K	MZ578126	MZ611817	MZ617215
E. mendocina	21703	Argentina	*Maas et al. 8173*	K	MZ578109	MZ611800	MZ617199
E. mexicana	21308	Mexico	*Frohlich 2522*	K	MZ578147	MZ611838	MZ617235
E. mexicana	41553	Mexico	*Frohlich 2522*	K	MZ578148	MZ611839	MZ617236
E. microphylla	21607	Antigua	*Totton 28.2.31*	BM	MZ578115	MZ611806	MZ617205
E. microphylla	21611	Barbados	*Carrington SC4*	BM	MZ578114	MZ611805	MZ617204
E. microsalsoloides	19736	Australia	*Craven & Wendel 9271*	CANB	MZ578068	MZ611761	MZ617160
E. nana	21609	Turks and Caicos Islands	*Gillis 11877*	K	MZ578118	MZ611809	MZ617208
E. nana	21613	Bahamas	*Proctor 30825*	BM	MZ578117	MZ611808	MZ617207
E. nana	21651	Bahamas	*Proctor 30989*	BM	MZ578120	MZ611811	MZ617210
E. nana	21652	Puerto Rico	*Dunbar 394*	BM	MZ578119	MZ611810	MZ617209
E. nashii	21645	Bahamas	*Gillis 11663*	K	MZ578116	MZ611807	MZ617206
E. nesopelyda	19719	Australia	*Craven Stewart & Wendel 9196*	CANB	MZ578076	MZ611769	MZ617167
E. nigricans	42548	Socotra	*Miller 19129*	E	MZ578104	MZ611795	MZ617194
E. ovalifolia	24511	Namibia	*Hilger 93/15*	B	MZ578155	MZ611846	MZ617243
E. ovalifolia	24512	Namibia	*Hilger 93/17*	B	MZ578156	MZ611847	MZ617244
E. paniculata	19706	Australia	*Graham 7*	BM or K	MZ578080	MZ611773	MZ617171
E. peckhamii	19699	Australia	*Stewart Wendel & Edwards 9251*	BM or K	MZ578077	MZ611770	MZ617168
E. personata	31615	Yemen	*Popov 43/2*	BM	MZ578101	MZ611792	MZ617191

Table 1. Continued

Name	Kew DNA bank number	Country	Collector and number	Herbarium	GenBank ITS	Accession *rbcL*	Numbers *matK*
E. personata	42.1	Somalia	*Thulin & Warfa 6138*	UPS	MZ578102	MZ611793	MZ617192
E. plumosa	19726	Australia	*Craven 9147*	CANB	MZ578064	MZ611757	MZ617156
E. plumosa	20579	Australia	*Kubian, Jesson & Craven 18*	CANB	MZ578065	MZ611758	MZ617157
E. polyphella	45035	Florida	*Sanders 3*	K	MZ578134	MZ611825	MZ617223
E. powelliorum	41364	Texas	*Sage s.n.*	TRT	MZ578151	MZ611842	MZ617239
E. pringlei	21605	Arizona	*Soreng & Salazar 1895*	K	MZ578127	MZ611818	MZ617216
E. pringlei	45031	Mexico	*Frohlich 2116-2*	K	MZ578128	MZ611819	MZ617217
E. procumbens	19733	Mexico	*Manzano s.n.*	TRT	MZ578153	MZ611844	MZ617241
E. procumbens	41555	Mexico	*Frohlich 1538 or 1850*	K	MZ578154	MZ611845	MZ617242
E. prostrata	19713	Australia	*Craven 8516*	CANB	MZ578078	MZ611771	MZ617169
E. pulvina	19710	Australia	*Pullen 11243*	BM or K	MZ578073	MZ611766	MZ617164
E. racemosa	41367	Texas	*Sage s.n.*	TRT	MZ578144	MZ611835	MZ617232
E. ramulipatens	19718	Australia	*Craven 8538*	CANB	MZ578087	MZ611780	MZ617178
E. rariflora	24606	Iran	*Wendelbo & Foroughi 15357*	TARI	MZ578103	MZ611794	MZ617193
E. sessei	21306	Mexico	*Frohlich 2519*	K	MZ578152	MZ611843	MZ617240
E. sessilistigma	44.1	Somalia	*Thulin & Abdi Dahir 6475*	UPS	MZ578100	MZ611791	MZ617190
E. strigosa	24678	Yemen	*Radcliffe, Smith & Hetchie 4526*	K	MZ578097	MZ611788	MZ617188
E. strigosa	31614	Iran	*No collector 1793*		MZ578099	MZ611790	
E. strigosa	45.1	Ethiopia	*Thulin et al. 3860*	UPS	MZ578098	MZ611789	MZ617189
E. subreniformis	19714	Australia	*Craven & Stewart 9765*	CANB	MZ578082	MZ611775	MZ617173
E. tenella	29665	Missouri	*Vogan s.n.*	TRT	MZ578145	MZ611836	MZ617233
E. tenella	45043	Texas	*Kearns s.n.*		MZ578146	MZ611837	MZ617234
E. tenuifolia	19715	Australia	*Lewis s.n.*	BM or K	MZ578061	MZ611754	MZ617153
E. texana	45039	Texas	*Frohlich 2300*	K	MZ578129	MZ611820	MZ617218
E. texana	45040	Texas	*Sage s.n.*	TRT	MZ578130	MZ611821	MZ617219
E. torreyi	29666	Texas	*Sage s.n.*	TRT	MZ578150	MZ611841	MZ617238
E. uniflora	19708	Australia	*Chesterfield 204*	BM or K	MZ578067	MZ611760	MZ617159
E. ventricosa	19701	Australia	*Lewis s.n.*	BM or K	MZ578085	MZ611778	MZ617176
E. ventricosa	45024	Australia	*Craven 8160*	CANB	MZ578084	MZ611777	MZ617175
E. vestita	19711	Australia	*Wilson & Rowe 896*	BM or K	MZ578090	MZ611783	MZ617181
E. wigginsii	45038	Mexico	*Frohlich 2601* grown from seed of: *Van de Vender 2006-775*	K	MZ578113	MZ611804	MZ617203
Myriopus psilostachya	41714	Peru	*Hutchinson, P. C. 3463*	K		MZ611854	MZ617251
Myriopus salzmannii	41717	Bolivia	*Wood, Soto, Marrod 23516*	K	MZ578162	MZ611853	MZ617250
Myriopus volubilis	41712	Costa Rica	*Barry Hammel 19401*	K	MZ578161	MZ611852	MZ617249
Myriopus volubilis	41718	Turks and Caicos Islands	*Pollard, B. J. 1347*	K	MZ578160	MZ611851	MZ617248

Table 2. Average $\delta^{13}C$ values of species of *Euploca* and related genera analysed here. Sample size is the number of samples analysed for $\delta^{13}C$. Boldface species are C_4. Black, non-boldface species are confirmed or probable C_3. Red species are C_2; blue species are proto-Kranz. Geographical abbreviations: Afr = Africa; Aus = Australia; Cent Am = Central America; Carib = Caribbean; SAM = South America; cult = from cultivated plant from that region.

Species	$\delta^{13}C$	Sample Size	Region
Euploca			
E. aequorea	−13.6	2	Aus
E. alcyonium	−13.0	3	Aus
E. amnis-edith	−13.1	2	Aus
E. anderssonii	−27.0	4	Galap.
E. antillana	−28.5	3	Carib
E. applanata	−13.0	4	NE Afr
E. arenitensis	−13.2	2	Aus
E. axillaris	−12.7	3	Mex
E. baclei	−31.5	2	Afr
E. ballii	−13.3	2	Aus
E. barbata	−27.0	1	SAM
E. brachygyne	−13.2	2	Aus
E. brachythrix	−13.3	3	Aus
E. bracteata	−14.0	4	Aus
E. bursifera	−13.6	3	Carib
E. campestris	−27.1	5	SAM
E. catamarcensis	−14.3	4	SAM
E. chrysantha	−13.4	7	SAM
E. chrysocarpa	−27.8	2	Aus
E. collina	−13.3	2	Aus
E. confertifolia	−12.6	5	Mex
E. consimilis	−13.5	2	Aus
E. convolvulacea	−28.3	11	NAM
E. cornuta	−15.3	1	India
E. cracens	−14.6	2	Aus
E. cremnogena	−31.0	2	Mex
E. cunninghamii	−13.6	4	Aus
E. cupressina	−13.7	2	Aus
E. delestangii	−14.2	2	Aus
E. dichotoma	−13.3	3	Aus
E. dichroa	−27.8	2	Carib
E. diffusa	−13.7	4	Carib
E. discordea	−13.9	2	Aus
E. distantiflora	−12.2	1	SAM
E. diversifolia	−13.2	3	Aus
E. eggersii	−28.7	2	Carib
E. epacridea	−12.6	2	Aus
E. euodes	−13.8	4	Aus
E. fallax	−26.7	6	Mex Cent Am
E. fasciculata	−12.8	2	Aus
E. ferreyrae	−24.4	2	SAM
E. filagnoides	−12.4	2	Aus
E. filiformis	−29.8	11	New World
E. flintii	−13.7	2	Aus
E. foliata	−14.8	3	Aus
E. foliosissima	−13.1	7	Mex Cent Am
E. foveolata	−13.2	2	Aus
E. frohlichii	−14.2	2	Aus
E. fruticosa	−13.7	11	New World
E. glabella	−12.8	2	Aus
E. glandulifera	−28.5	3	Aus
E. greggii	−26.6	11	NAM
E. haesum	−13.2	3	Aus
E. haitiensis	−12.7	3	Carib
E. hassleriana	−13.4	3	SAM
E. hintonii	−28.0	3	Mex
E. humifusa	−12.5	7	Carib
E. humilis (= E. ternata)	−13.6	21	New World
E. humistrata	−26.4	7	SAM
E. imbricata	−12.2	1	Carib
E. inexplicita	−12.9	2	Aus
E. karwinskyi	−28.5	6	Mex
E. katangensis	−28.7	2	Africa
E. lagoensis	−27.9	8	New World
E. laxa	−13.8	2	NE Afr
E. leptalea	−13.2	5	Aus
E. limbata	−12.1	5	Mex
E. margaretensis	−12.1	3	SAM
E. maxima	−12.4	3	SAM
E. mendocina	−13.1	5	SAM
E. mexicana (= E. calcicola)	−25.6	6	Mex
E. microphylla	−12.9	9	Carib
E. microsalsoloides	−12.6	3	Aus
E. mitchellii	−27.1	2	Aus
E. moorei	−23.8	2	Aus
E. mutica	−13.6	2	Aus
E. myriophylla	−13.8	1	Carib
E. nana	−12.7	13	Carib
E. nashii	−13.7	5	Carib
E. nesopelyda	−13.1	2	Aus
E. nexosa	−13.0	2	Aus
E. nigricans	−29.1	4	Socotra
E. ocellata	−14.6	4	SAM
E. ottonis	−12.3	2	SAM
E. ovalifolia	−28.2	8	Aus Afr
E. oxyloba	−25.8	5	SAM
E. pachyphylla	−26.8	4	Aus
E. pallescens	−14.1	2	SAM
E. paniculata	−13.4	4	Aus
E. paradoxa	−27.4	3	SAM
E. parciflora	−27.8	4	Cent & SAM
E. peckhamii	−13.4	2	Aus
E. pedicellaris	−13.6	2	Carib
E. peninsularis	−13.2	2	Aus

Table 2. Continued

Species	$\delta^{13}C$	Sample Size	Region
E. personata	–13.4	2	NE Afr Yemen
E. pilosa	–26.2	9	SAM
E. plumosa	–13.2	3	Aus
E. polyanthella	–27.4	6	SAM
E. polyphylla	–13.1	5	N&SAM
E. pringlei	–13.3	9	NAM
E. procumbens	–28.1	7	N&SAM
E. prostrata	–12.6	2	Aus
E. protensa	–12.7	2	Aus
E. pulvina	–12.6	4	Aus
E. purdiei	–12.5	2	SAM
E. queretaroana	–27.6	6	Mex
E. racemosa	–29.9	4	NAM
E. ramulipatens	–12.8	2	Aus
E. rariflora	–13.7	7	Afr, Mideast
E. rhadinostachya	–13.4	2	Aus
E. salicioides	–13.6	5	SAM
E. scabra	–13.3	2	India
E. serpylloides	–11.4	2	Cuba
E. sessei	–24.7	5	Mex
E. sessilistigma	–13.2	6	NE Afr
E. skeleton	–14.0	2	Aus
E. sphaerica	–14.3	2	Aus
E. sphaerococca	–12.5	4	Carib
E. strigosa	–13.5	8	Afr
E. styotricha	–13.4	2	Aus
E. subreniformis	–14.9	2	Aus
E. synaimon	–12.1	2	Aus
E. tabuliplagae	–13.0	2	Aus
E. tanythrix	–13.8	2	Aus
E. tenella	–28.2	8	NAM
E. tenuifolia	–13.5	6	Aus
E. texana	–14.3	4	NAM
E. toratensis	–28.0	1	SAM
E. torreyi	–26.4	10	NAM
E. transformis	–13.2	4	Aus
E. tytoides	–12.4	2	Aus
E. uniflora	–13.1	2	Aus
E. uninervis	–28.4	2	Carib
E. vaga	–12.8	2	Aus
E. ventricosa	–12.7	6	Aus
E. vestita	–13.2	4	Aus
E. viator	–12.9	2	Aus
E. wigginsii	–15.0	5	Mex
Heliotropium s.s.			
H. abbreviatum	–25.3	1	SAM
H. adenogynum	–24.0	2	SAM
H. ammophilum	–27.2	3	Aus
H. amplexicaule	–27.9	3	SAM
H. anchusifolium	–27.7	1	SAM
H. angiospermum	–26.1	6	N&SAM
H. anomalum	–26.9	5	Pacific Is.

Table 2. Continued

Species	$\delta^{13}C$	Sample Size	Region
H. arborescens	–29.6	2	NAM (cult)
H. aff. arborescens	–24.4	1	SAM
H. argenteum	–26.4	2	SAM
H. asperrimum	–27.1	2	Aus
H. bacciferum	–26.8	7	South Asia/ Afr
H. bangii	–25.1	2	SAM
H. benadirense	–27.3	2	Africa
H. brahuicum	–26.1	1	South Asia
H. cabulicum	–27.1	2	South Asia
H. cf incanum	–27.3	1	SAM
H. *chenopodiaceum*	–25.4	4	SAM
H. ciliatum	–27.7	4	Afr
H. corymbosum	–25.5	2	SAM
H. crispatum	–27.7	3	Aus
H. curassavicum	–27.8	6	N&SAM, Carib, Aus, Pacific Is.
H. dasycarpum	–26.3	1	S. Asia
H. dasycarpum subsp. transoxanum	–26.1	1	Cent Asia
H. dissitiflorum	–25.6	1	Iran
H. drepanophyllum	–26.9	2	SW Asia/NE Africa
H. elongatum	–28.7	4	SAM
H. eremogenum	–24.3	1	SAM
H. erosum	–28.5	2	NW Afr Canary Is.
H. europaeum	–28.1	9	COS
H. floridum	–26.4	4	SAM
H. geissei	–24.6	2	SAM
H. genovefae	–28.6	2	Carib
H. giessii	–26.4	2	Afr
H. glabriusculum	–26.0	4	Mex
H. glutinosum	–23.7	2	SAM
H. hirsutissimum	–26.6	2	S Europe, Medit.
H. incanum	–26.6	3	SAM
H. indicum	–30.4	3	Americas, Asia, Aus, Afr
H. johnstonii	–25.4	1	SAM
H. krauseanum	–27.4	5	SAM
H. kurtzii	–25.4	6	SAM
H. lanceolatum	–27.7	2	SAM
H. leiocarpum	–27.6	2	SAM
H. lineare	–26.8	4	S Afr
H. linearifolium	–27.2	2	SAM
H. longiflorum	–26.5	6	NE Afr
H. macrostachyum	–27.0	2	Mex
H. mandonii	–26.9	2	SAM
H. messerschmidiodes	–25.1	2	Canary Is.
H. micranthum	–28.9	1	Cent. Asia Iran
H. microstachyum	–25.5	7	SAM

Table 2. Continued

Species	$\delta^{13}C$	Sample Size	Region
H. molle	−24.5	2	Mex, Texas
H. monostachyum	−28.5	1	SAM
H. murinum	−28.0	2	Aus
H. myosotifolium	−23.7	2	SAM
H. nelsonii	−26.0	2	S Afr
H. nicotianaefolium	−29.7	2	SAM
H. oliveranum	−27.3	1	S Afr
H. ophioglossum	−26.1	4	Afr, S Asia
H. oppositifolium	−28.0	1	SAM
H. paronychioides	−26.2	8	SAM
H. patagonicum	−25.6	2	SAM
H. pectinatum	−28.6	4	E Afr
H. peruvianum	−28.9	1	SAM
H. phillipianum	−29.1	2	SAM
H. phylicoides	−28.1	7	SAM
H. pinnatisectum	−27.4	2	SAM
H. pleiopterum	−27.2	2	Aus
H. pterocarpum	−26.2	4	NE Afr
H. pycnophyllum	−25.2	3	SAM
H. ramosissimum	−28.0	4	Afr, S. Asia
H. remotiflorum	−26.0	1	S Asia
H. rufipilum	−28.3	9	Cent Am, SAM
H. sclerocarpum	−28.7	2	SAM
H. scottae	−29.7	2	E Afr
H. sinuatum	−25.2	3	SAM
H. spathulatum	−26.8	1	NAM
H. stenophyllum	−27.0	2	SAM
H. steudneri	−26.7	4	E Afr
H. suaveolens	−26.8	2	Europe
H. submolle	−26.2	5	SAM
H. subspinosum	−28.7	2	NE Afr
H. subulatum	−25.7	2	E Afr
H. supinum	−27.7	5	Europe, Afr, S Asia
H. taltalense	−24.9	2	SAM
H. transalpinum	−29.6	6	SAM
H. tubulosum	−26.9	2	S Afr
H. *undulatum*	−27.1	5	Afr
H. urbanianum	−25.9	1	SAM
H. veronicaefolium	−26.4	6	SAM
H. villosum	−24.6	1	Mideast
H. zeylanicum	−25.7	4	S Asia/Afr
Myriopus			
M. candidulus	−31.3	1	SAM
M. maculatus	−29.3	3	SAM
M. paniculatus	−29.1	1	SAM
M. poliochros	−28.4	1	Carib
M. psilostachya	−24.5	1	SAM
M. salzmanni	−28.0	5	SAM
M. volubilis	−28.3	5	NAM, SAM, Carib

Table 2. Continued

Species	$\delta^{13}C$	Sample Size	Region
Tournefortia			
T. bicolor	−30.1	1	SAM
T. chinchensis	−27.8	3	SAM
T. cuspidata	−31.6	1	SAM
T. gnaphaloides (H. gnaphalodes)	−27.2	1	SAM
T. hirsutissima (H.verdcourtii)	−30.0	1	SAM
T. polystachya	−29.9	1	SAM
T. tarmensis (H. tarmense)	−28.8	1	SAM
T. virgata	−24.4	2	SAM

sequencing buffer, 0.75-μL primer and *c*. 20 ng purified PCR product, with these reactions purified by Magnesil beads using a Beckman–Coulter Biomek NX platform.

Sequences were run on an ABI 3730 sequencer. Sequences were assembled and edited using Sequencher v.4.5 software (Gene Codes; http://genecodes.com). All reads were carefully checked by hand.

Phylogenetic analyses

Alignments of *rbcL* were unambiguous with no insertions/deletions. Alignment of *matK* was done initially at the amino acid level in Muscle (Edgar, 2004) and then converted to aligned DNA sequences by the PAL2NAL web server (http://www.bork.embl.de/pal2nal/#RunP2N). This alignment was unambiguous, requiring no hand correction. Almost all indels in *matK* are multiples of three, as would be expected to maintain the reading frame. We used two alignments of nrITS. Alignment 1 was done with Muscle followed by extensive corrections by hand in PhyDE (http://www.phyde.de). Alignment 2 was done by the MAFFT v.7 web server (http://mafft.cbrc.jp/alignment/server/) without subsequent hand correction. Indels were coded using the 'simple gap scoring' method (Simmons & Ochoterena, 2000) using the SeqState plugin for PhyDE (Müller, 2005). In *matK*, we observed a single base indel that jumped repeatedly between two near-by positions. This jumping indel was not used as an indel character due to the high frequency of its jumps. It was retained for analyses at the DNA level, to avoid the one-base alignment shifts it would cause for the bases between the two indel positions, were it omitted from the DNA sequences. This jumping indel is the subject of further research.

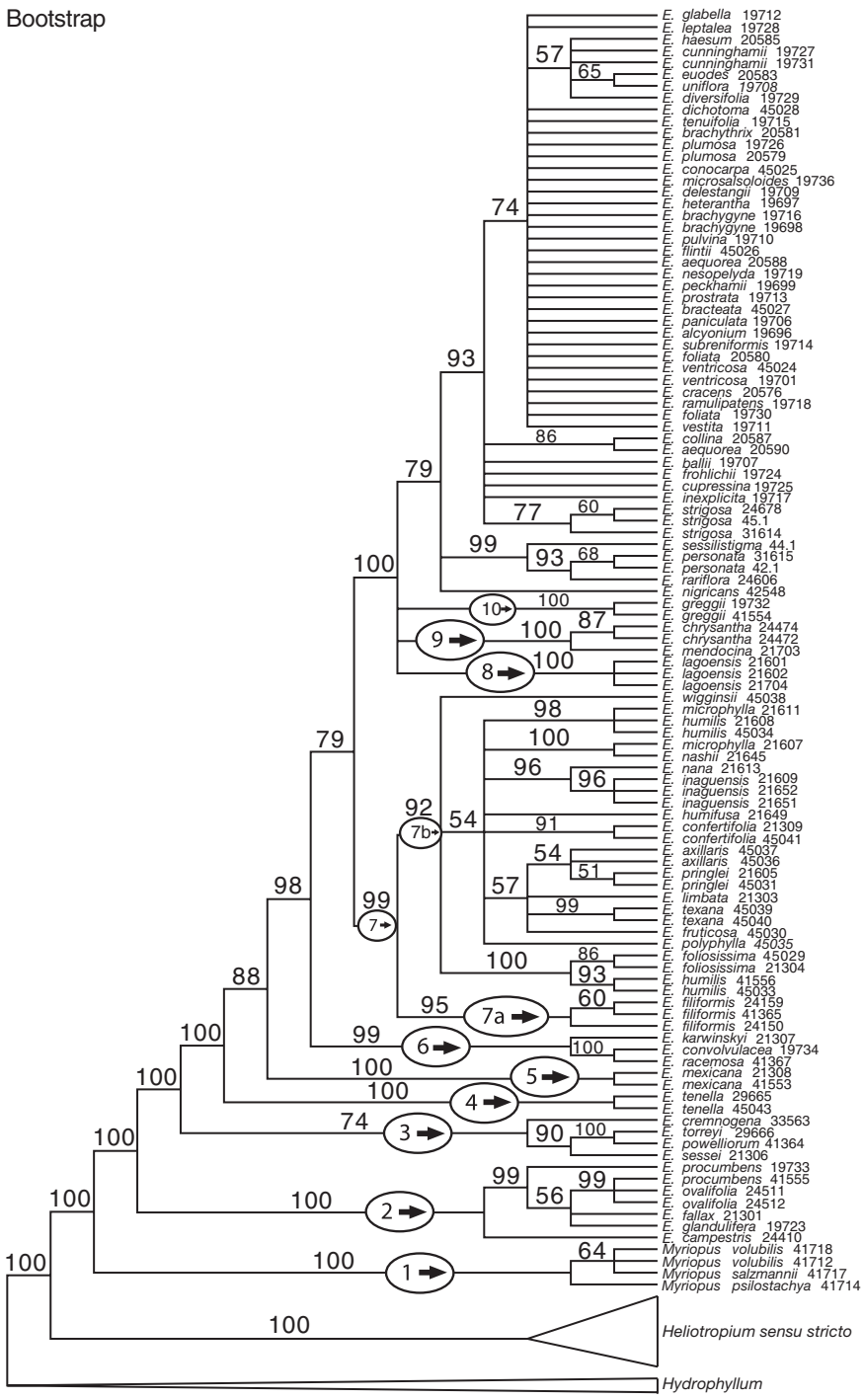

Figure 1. Parsimony bootstrap of the two plastid genes. Numbers following species names are RBG, Kew, DNA bank accession numbers, appended to explicitly and precisely identify the sample. Numbered ovals on the tree identify specific clades, based on their positions in the MrBayes tree (Fig. 7).

Parsimony analyses, with unordered characters of equal weight, were done in PAUP* v.4.0b10 or 4.0a (Swofford, 2003) for all sequence data, with each nrITS alignment, with and without indel characters, and, independently, for each amplified gene segment and for the two compartments. PAUP non-parametric bootstrapping was done using TBR branch swapping, with 15 random-addition starting trees, saving no more than 30 trees for each starting tree, with 380 to 1000 bootstrap replicates. The plastid bootstrap tree

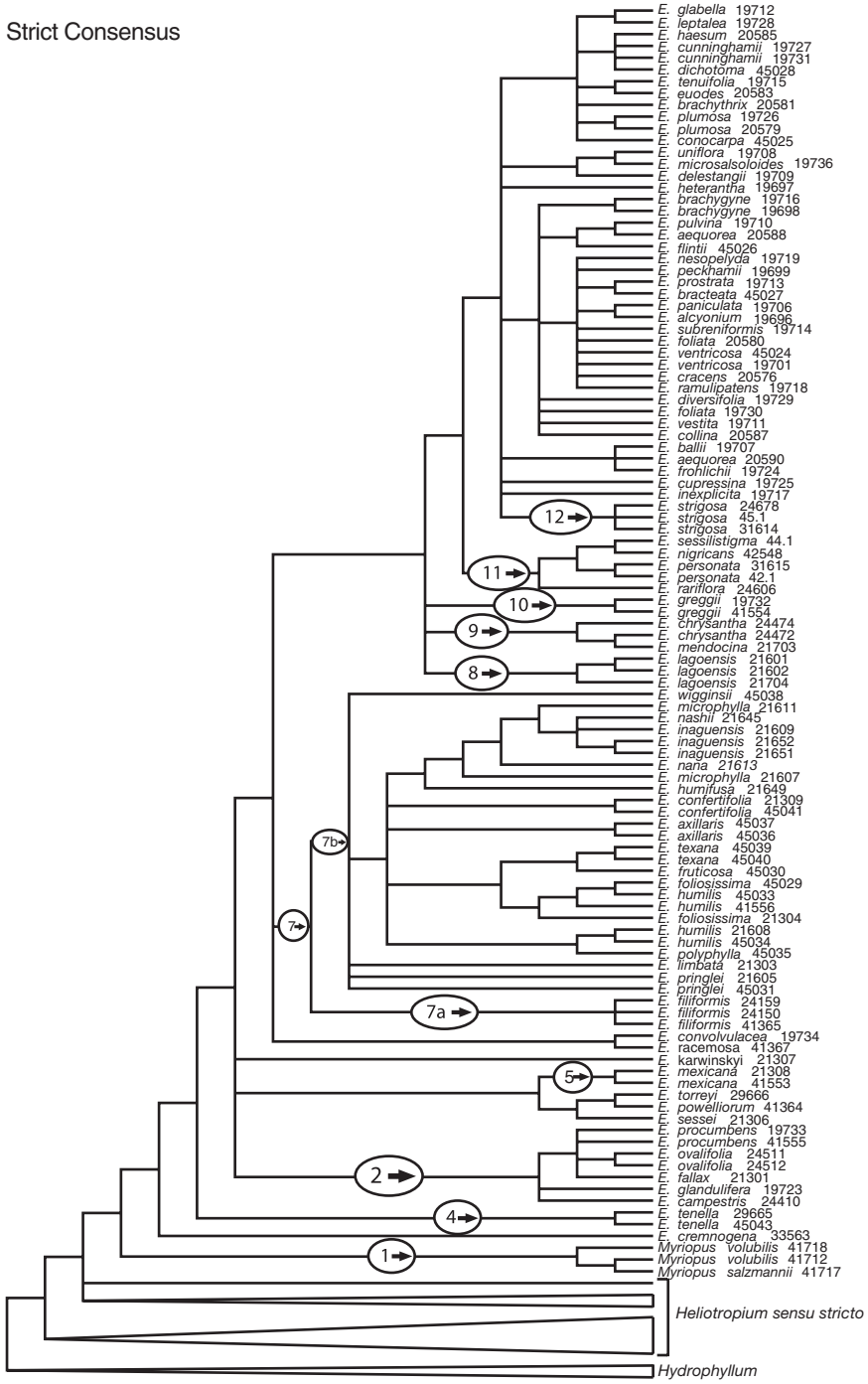

Figure 2. Strict consensus of nrITS loci only (from alignment one). Note placement of *E. tenella*, (clade 4) and *E. cremnogena* compared to Figure 1.

(Fig. 1) shows strong support (≥ 90%) for many nodes, with weak support for others, whereas the nrITS tree has a large polytomy (not shown). When the plastid and nrITS data are combined, support for some basal nodes within *Euploca* is reduced.

The nrITS strict consensus tree (alignment 1) shows placement of *E. cremnogena* and *E. tenella* as two successively diverging sisters to the rest of *Euploca* (Fig. 2). This is unexpected based on morphology and conflicts dramatically with placement in the plastid tree (and, for *E. tenella*, with the old RFLP-based tree; the Supporting Information, Fig. S1). On blastn against GenBank, these sequences do hit *Heliotropium s.l.* sequences.

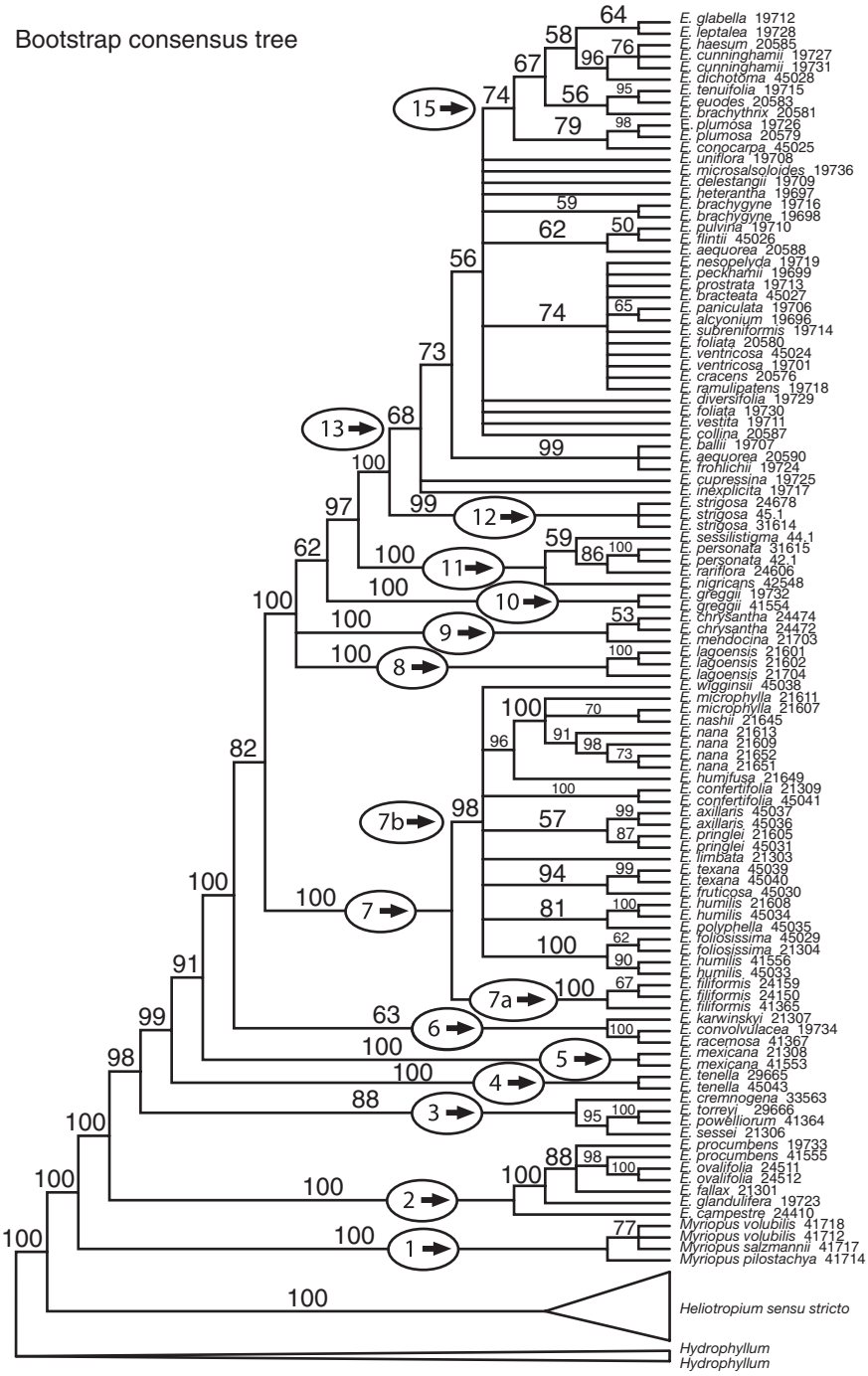

Figure 3. Parsimony bootstrap of the three loci. nrITS sequences for *E. cremnogena* and *E. tenella* were omitted, as in subsequent figures.

The two *E. tenella* sequences are almost identical, and are from different plant sources, and the DNAs were isolated and the PCRs done many months apart, so we do not think they could be spurious. The *E. cremnogena* DNA, from a 1978 herbarium collection, is only of fair quality, and its sequence was difficult to amplify, so we had less confidence in that sequence, but repeat PCRs using specific primers based on the original nrITS sequence gave good amplifications and virtually an identical sequence, so we think that that sequence is also genuine. The nrITS sequences of these two species are consistent with neither the tree supported by the plastid genes nor the expected relationships based

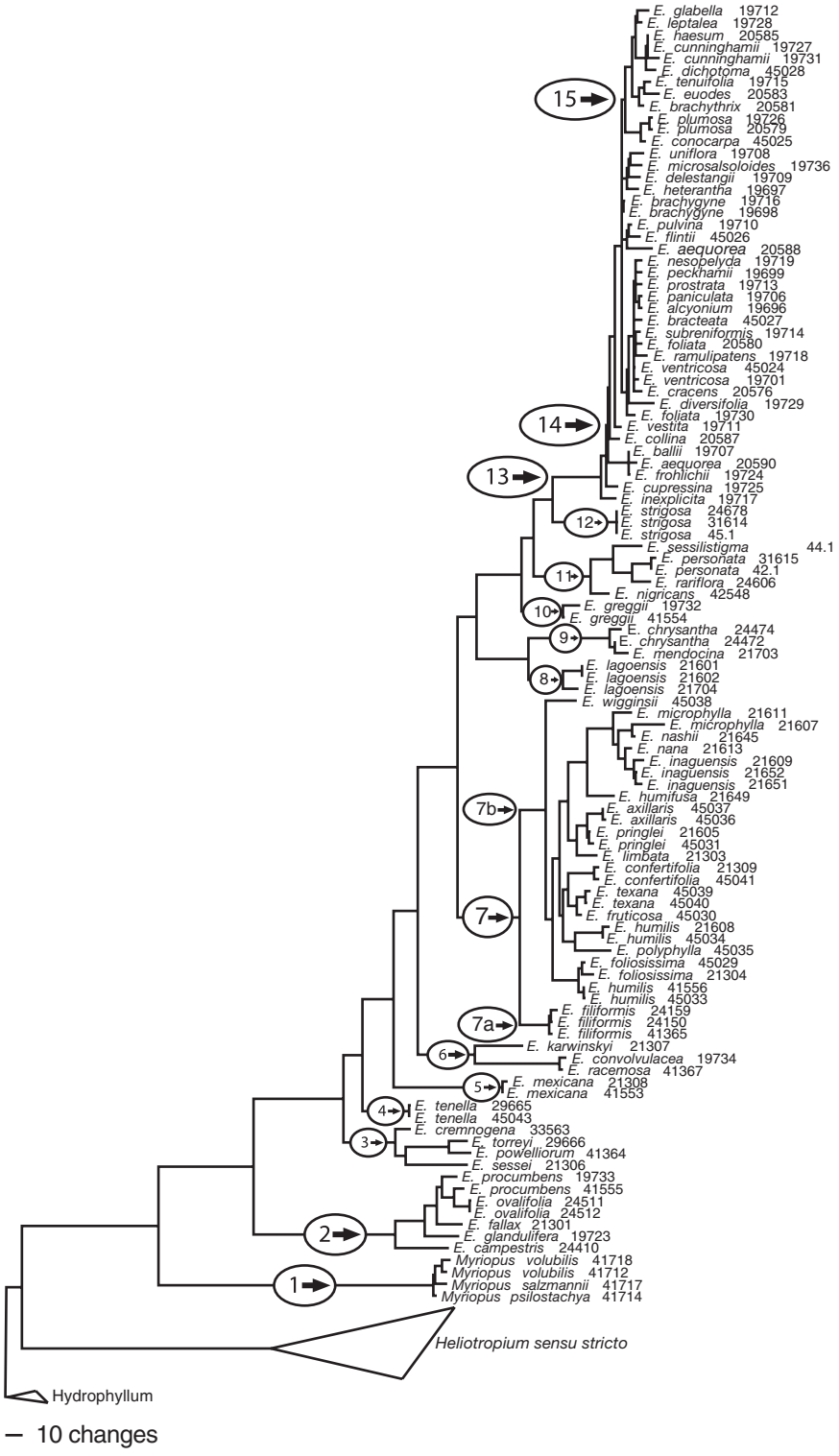

Figure 4. Phylogram of the three loci, corresponding to Figure 3.

on shared morphology, so we removed them from subsequent analyses. Divergent nrITS sequences can be present as small minorities among the many rDNA repeats, rendering them undetectable by PCR. If their copy number increases, they could become the main amplifiable sequences, potentially

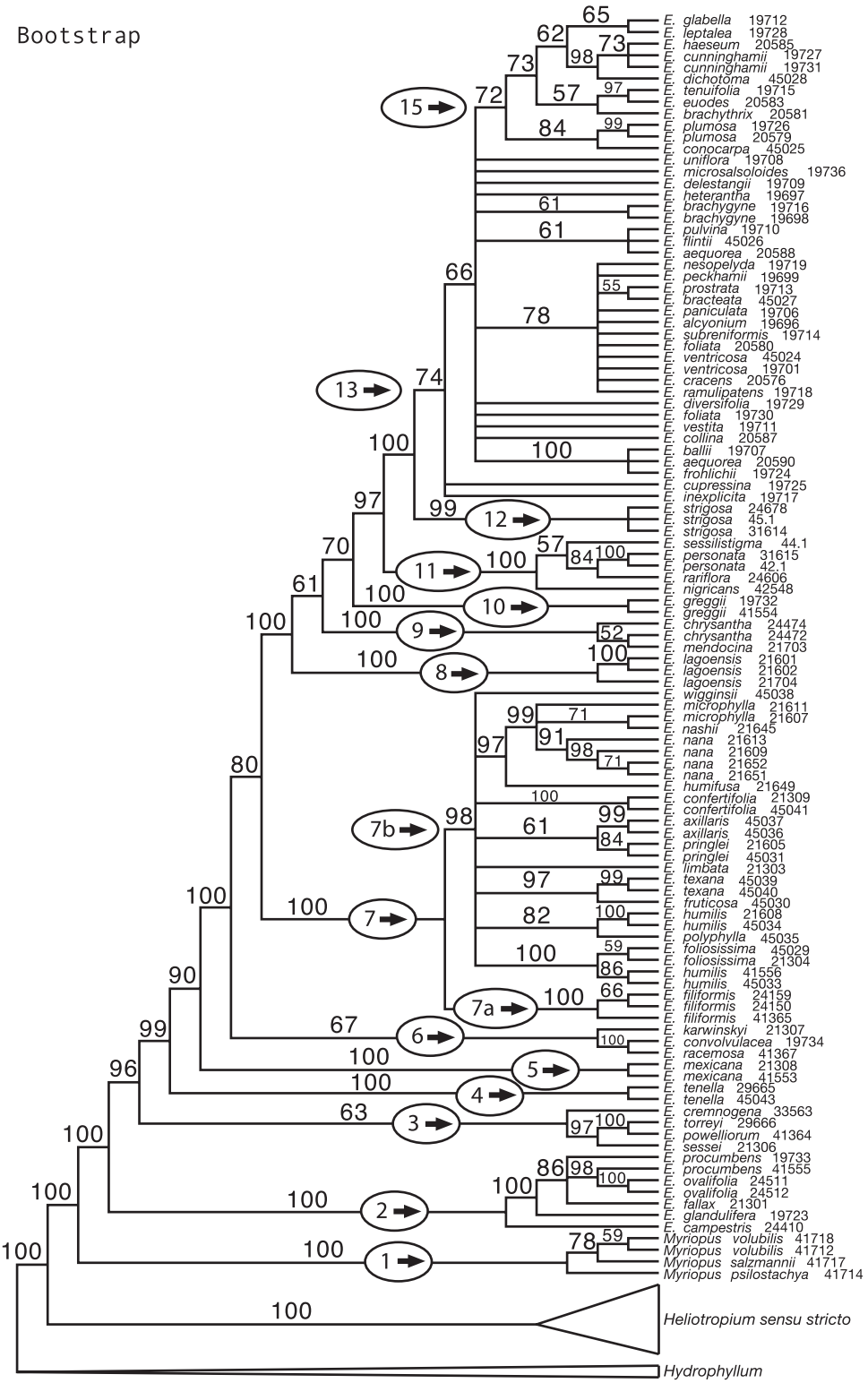

Figure 5. Parsimony bootstrap tree of the combined matrix with indel characters.

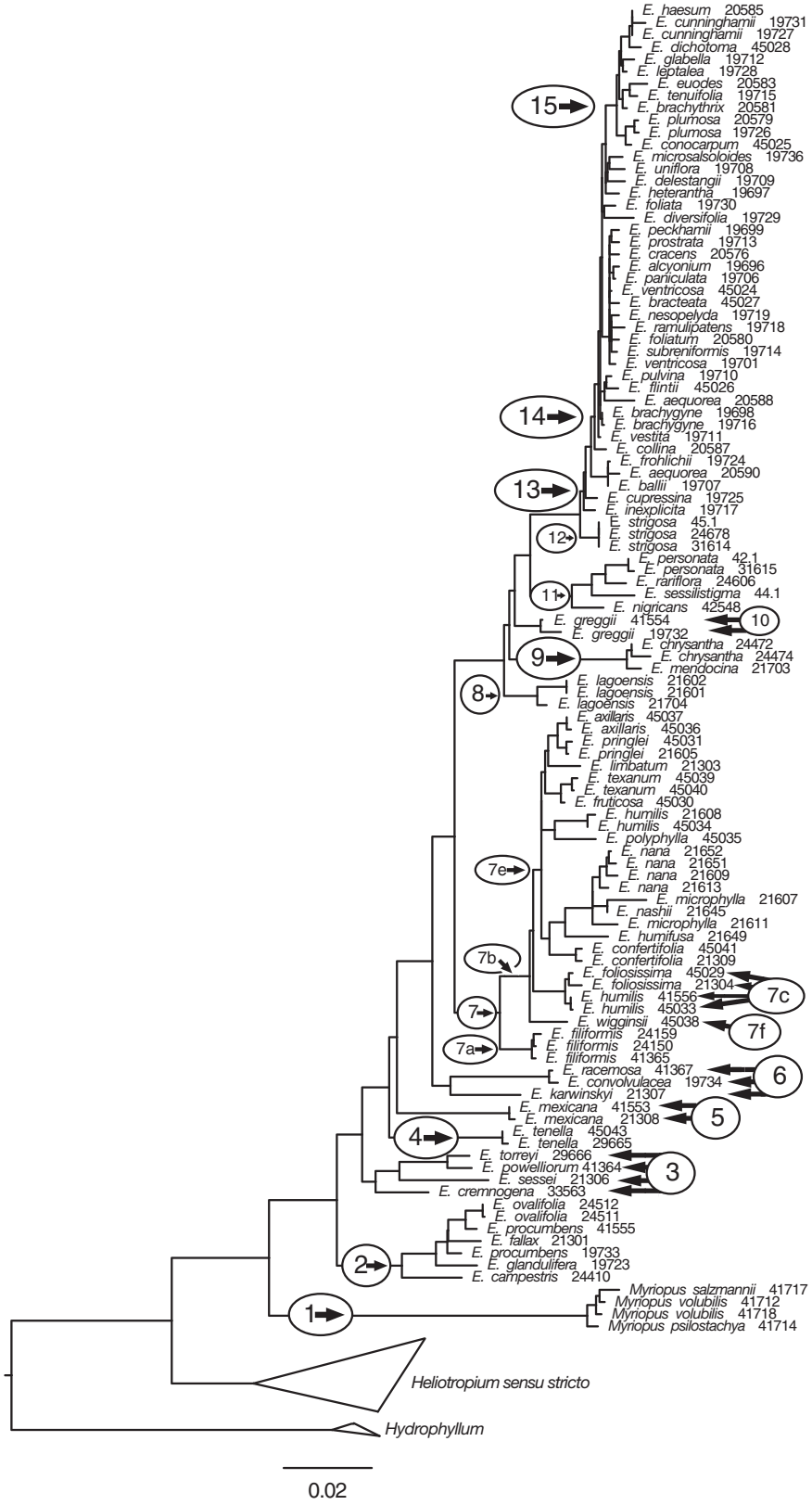

Figure 6. Optimal maximum likelihood tree from GARLI, based on the combined matrix without indels.

confounding phylogenetic estimation. That could be the circumstance here.

When just these three nrITS sequences are excluded, parsimony bootstrap support from all other sequence data gives more strongly supported nodes near the base of the tree than do plastid genes alone, confirming that these three nrITS sequences were the source of the reduced support seen previously. Both nrITS alignments, combined with the plastid data, show similar results; Figure 3 shows the parsimony bootstrap tree for nrITS alignment 2 and plastid data. Figure 4 shows a corresponding tree with branch lengths proportional to the amount of change, with little sequence divergence among the EUs near the top of the tree, corresponding to the large polytomy in the parsimony bootstrap tree.

Indel data alone for all sequences (except *E. cremnogena* and *E. tenella* nrITS) show little resolution. Parsimony analyses using all sequence plus indel data, but excluding *E. cremnogena* and *E. tenella* nrITS, further increase support for many nodes (Fig. 5).

GARLI 2.0 (http://www.phylo.org/tools/obsolete/garli%202.0%20on%20xsede(beta).html) analyses to find the optimal maximum likelihood tree (Fig. 6) were done with data divided into seven partitions, one for nrITS and separate partitions for each of the three codon positions of *rbcL* and *matK*. Each model used the gamma distribution with four rate categories. The rate matrix with six general-time reversible substitution rates (GTR), the base frequencies and the numbers of invariant sites were estimated from the data.

To perform Shimodaira–Hasegawa (SH) tests (Shimodiara & Hasegawa, 1999) and AU tests (Shimodaira, 2002), we used GARLI to find the optimal ML tree under constraints that forced various C_4 taxa together, to the exclusion of C_3, proto-Kranz and C_2 taxa that fell between them in the optimal ML tree, and then used PAUP* 4.0b10 or 4.0a to perform SH and AU tests, with 10 000 RELL bootstrap replicates. For these tests, PAUP* could only use the DNA data, not the indel data, hence these tests do not use all available data.

For MrBayes v.3.2.2 analyses, DNA data were divided into fewer partitions (five) to aid convergence, by combining the first and second codon positions of *rbcL* and also combining the first and second codon positions of *matK*. Models were GTR+gamma, which improved convergence compared to GTR+invgamma. One additional partition was used when including indel data or two if including both indel and RFLP data; these were analysed using the gamma model. Analyses that included indel characters showed more resolution and stronger support than analyses without indel data. Analyses with RFLP data had little effect on resolution, so those data were excluded from final analyses due to their availability for only a few accessions, which might result in long-distance influence (which is not the same as long branch attraction; Maddison, 1993). Final analyses were done with the three loci and indel data: temperature 0.04, with two or four runs and five chains in each analysis; completion criterion Stopval = 0.01; 25% burn-in. Multiple analyses achieved the Stopval completion criterion. All of the individual analyses that were terminated by Stopval and that showed wide, uniform scatter on the plot of log likelihood versus generation (indicating convergence) yielded trees with identical topologies and similar posterior probabilities. All PSRF (potential scale reduction factor) values were close to 1.0, except a single value in one analysis for a bipartition 'not found in all runs' listed as not available. All these analyses were examined in Tracer v.1.7.1, and various individual analyses (mostly with two runs) were combined to achieve all ESS values over 200, yielding six independent converged analyses, as confirmed by the hairy caterpillar plot for all statistics, which also justified the burn-in value. The tree shown in Figure 7 is based on an analysis with four runs, all ESS values > 300, and all PSRF close to 1.0.

Plant material and $\delta^{13}C$ determinations

For $\delta^{13}C$ studies, we sampled leaves or twigs from multiple herbaria (GH, ANUC, AAH, B, K, MICH, BM, NY, TRT, E, listed roughly in order of the number of specimens sampled), attempting to obtain samples of as many species as possible. Identifications of these species followed, to the best of our ability, current authorities in these regions, or we use the most recent names on herbarium sheets, particularly from areas without current workers or floras or from herbarium specimens we cannot re-examine. We sometimes opted to retain species names that have been lumped, to record the diversity of sampling. We only sampled a few of the traditional *Tournefortia s.l.* species. *Tournefortia* has even fewer taxonomic specialists than *Heliotropium s.l.*; for that genus, we generally report names on the herbarium sheets. We also sampled leaves, fruits or stems from collections by RFS and students and MWF, which were included in recent physiological and structural studies (Vogan *et al.*, 2007; Muhaidat *et al.*, 2011) and also many earlier collections by MWF for his PhD studies (1978) and for subsequent alkaloid studies (Catalfamo *et al.*, 1982). To determine $\delta^{13}C$ values, 3–10 mg of plant material was removed from dried specimens and placed in tin sampling cups in 96-well plates and sent to one of two commercial stable isotope laboratories, the University of California at Davis (http://stableisotopefacility.ucdavis.edu; before 2010) or Washington State University

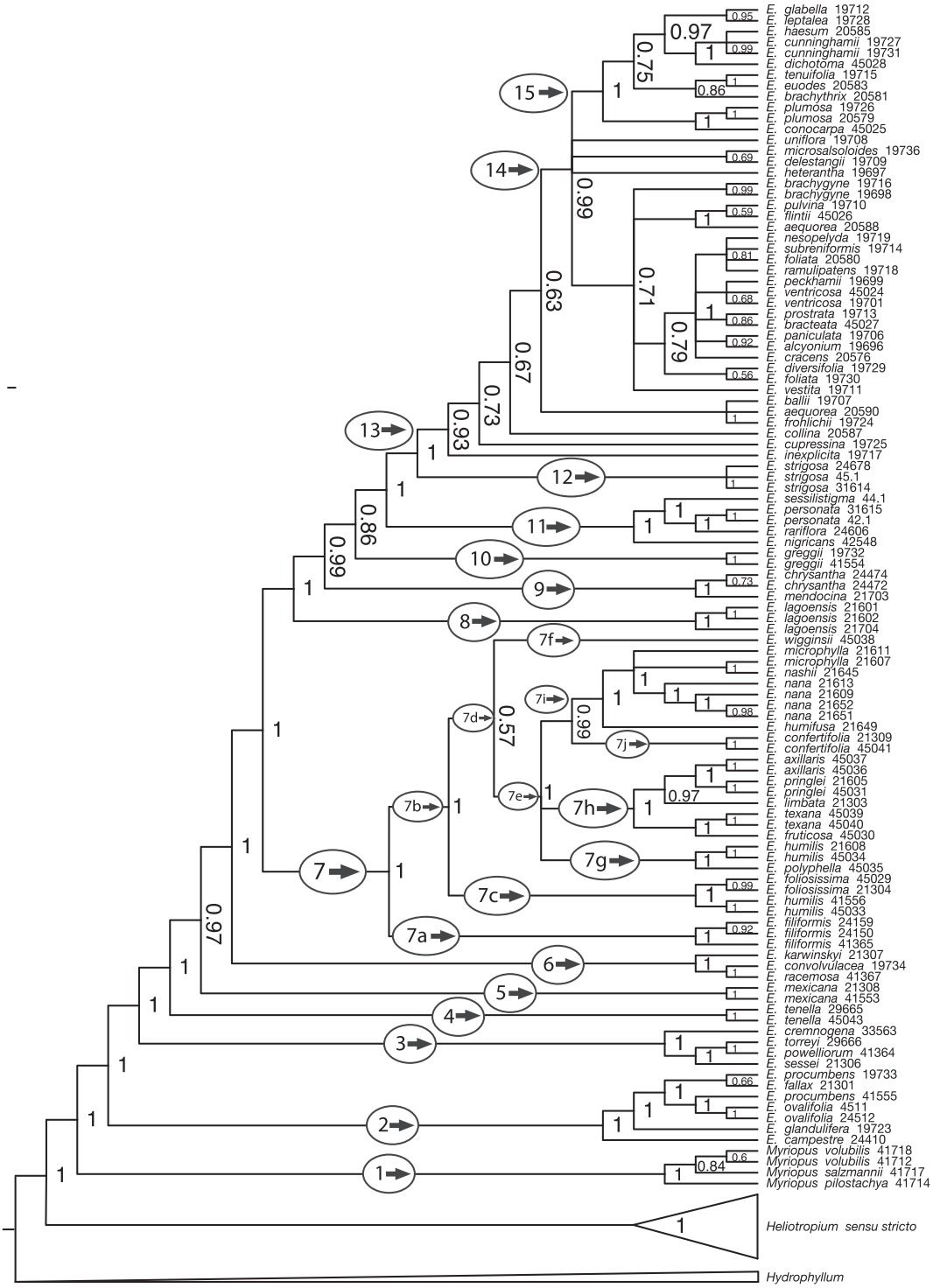

Figure 7. MrBayes tree based on the combined matrix with indel characters. This tree is used for character state history reconstruction.

(http://www.isotopes.wsu.edu; after 2010). All sampled specimens and their $\delta^{13}C$ values are listed in the Supporting Information, Table S1.

OTHER DATA

The morphological features considered here were chosen because they show major variation in *Euploca*,

and/or have been used as features to distinguish species and species groups, and/or are important for the life histories of the plants. Features such as annual versus perennial growth habit and perfect-flowered versus dioecious versus trioecious fall in all three of these categories. Inflorescence features such as the presence/absence of bracts and presence/absence and type of trichomes on the upper corolla surface and in the corolla throat fall in the first and second categories, and were pertinent to the subsectional classification of *Heliotropium* section *Orthostachys* proposed by Johnston (1928).

Morphological information is derived from Frohlich (1978; for North America), Craven (1996, 2005; for Australia), Thulin (2005, 2006; for north-eastern Africa) and herbarium specimens at K and online (by MWF), supplemented by information from various floras. Trichome presence within the corolla throat is often noted in the literature, but only rarely with enough detail to determine trichome type; one exception is the study of Di Fulvio & Ariza Espinar (2016), who provided trichome data for *E. chrysanthum* and *E. mendocinum*. The COVID-19 pandemic, which immediately followed MWF's recovery from being hit by a car, precluded transmitted light microscope analyses of trichome types on the upper corolla surfaces and in the throats of species not previously examined; for these, only data on presence/absence of trichomes, but not trichome type, are available. Primary versus secondary-walled trichomes are distinguished by the birefringence created by the aligned cellulose microfibrils in the discrete layers of the secondary wall, resulting in the wall appearing bright when oriented obliquely between crossed polarizers (Slayter, 1970; Frohlich, 1978). Silica in the trichome wall, if present, is detected by measuring the index of refraction of the trichome, using lines of Becke (Slayter, 1970; Frohlich, 1978), especially before and after treatment with hydrofluoric acid. Hydrofluoric acid treatment removes silica, lowering the index of refraction.

Dioecy is strongly suspected if multiple herbarium specimens show some individual plants with abundant seed set, whereas others set no seed despite having many dead flowers, which was sometimes checked by examining flowers but may have been missed. Trioecy may be suspected in the field if some plants have no seed set, some show moderate seed set and some have abundant seed set, but this is difficult to detect from herbarium collections, so some trioecious species may well not have been detected. For Australia, plants are scored as trioecous if some populations are dioecious, whereas others are perfect-flowered (Craven, 1996); all three forms can be found in populations of *E. torreyi*.

Our characters for geographical distribution are based, first, on regions exhibiting notable diversity of *Euploca* spp., with boundaries between discrete regions based on major separations either by water or by moist habitats unsuitable for most *Euploca* spp. (i.e. in south-eastern Central America and Southeast Asia). Neither Africa nor Asia is a centre of *Euploca* diversity, so they are lumped together. The Horn of Africa and the adjacent Arabian Peninsula have several species, but their distribution boundaries do not closely coincide, so this region is lumped with Africa and Asia. *Euploca nigricans* of Socotra is unusual, so we consider Socotra a separate unit to highlight that species. In the New World some species are widespread, in both North and South America and some also in the Caribbean; for these we create a separate character state 'Widespread in the New World' as an alternative to coding them with polymorphic character states, because such polymorphic characters do not project down the tree, but rather are resolved by any near-by non-polymorphic samples.

RESULTS

Phylogenetic analyses

Figures 5–7 show trees from parsimony (PAUP), maximum likelihood (GARLI) and Bayesian (MrBayes) analyses, respectively. These analyses include all three regions with nrITS alignment 2 (MAFFT). The PAUP and MrBayes analyses included the indel data, but the GARLI analysis did not. All three methods generate similar results.

The trees exhibit a pectinate series of mostly small clades, numbered 1 to 12, starting just above the second outgroup, *Heliotropium s.s.* These same 12 clades appear in all three analyses (Figs 5–7) and are compatible with the RFLP tree (see also Supporting Information, Fig. S1).

Of the 12 numbered clades, ten are individually held together at BS 99–100% by parsimony; the other two are clade 3 (63%) and clade 6 (67%). All 12 have strong support from Bayesian analysis (posterior probability, PP = 1.0). The pectinate arrangement of the 12 clades receives mostly strong support from parsimony percentages (BP 90–100), except for three nodes, and all but one node show strong support from Bayesian analysis (PP 0.95–1.0). The node (clade 7 + sister of clade 7) is separated from clade 6 by only BS = 80, but by PP = 1.0; the node (clade 9 + sister of clade 9) is separated from clade 8 by only BS = 61, but by PP = 0.99. The node (clade 10 + sister of clade 10) is separated from clade 9 by only BS = 70 and by only PP = 0.86, so that node lacks strong support. Therefore, all three possible resolutions of clade 9, clade 10 and sister of clade

10 will be considered in reconstructing character history for carbon assimilation system.

Clade 13 breaks the pectinate numbering by including all remaining accessions sister to clade 12. In clade 13, resolution is poor due to low levels of variation, as indicated by the short internodes in Figure 4 and the ML tree (Fig. 6). In the Bayesian tree there is more resolution in clade 13, but there are still several polytomies and low support at most nodes. Clade 14 (PP 0.99) contains a pentacotomy, with only one of these five (clade 15) held together with strong support (PP 1.0). In the small clades 1–6 and 8–12, the parsimony tree is generally resolved, and the Bayesian tree even more so, with most nodes supported.

Clade 7, with 29 accessions, is the only large clade (BP 100, PP 1.0) along the pectinate backbone of the tree. In it, *E. filiformis* (7a) is strongly supported (BP 98, PP 1.0) as sister to the rest (7b), for which parsimony shows a large polytomy. The Bayesian tree shows somewhat more resolution; in 7b, clade 7c, containing *E. foliosissima* and two accessions of *E. humilis*, diverges first and is sister to 7d, but the subclades in 7d, 7e and 7f, are barely held together, by only PP = 0.57. Indeed, the GARLI ML tree shows an alternative resolution in 7b: 7f diverges first, paired with (7c + 7e). Clade 7e contains a trichotomy, with its three components each strongly supported, as are their constituents.

Note that the same species, *E. humilis* (formerly *E. ternata* and *H. ternatum*) falls in two clades: clades 7c (paired with *E. foliosissima*) and clade 7g (paired with *E. polyphylla*). Removing *E. foliosissima* and/or *E. polyphella* (individually or both) in MrBayes analyses does not bring these *E. humilis* accessions together.

Our SH and AU tests compared the preferred ML tree (Fig. 6) to constrained trees that forced rearrangements of the C_4 and non-C_4 taxa. Forcing all the C_4 taxa together was rejected at $P = 0.0001$ by the SH test and $P = 0$ by the AU test. Forcing *E. nigricans* (a non-C_4 Old World species) out from a forced clade of only Old World C_4 species was rejected by SH at $P = 0.0307$, which is significant ($P < 0.05$) and much more strongly rejected by the AU test at $P = 0$. Note that the indel data, which adds significant support to multiple nodes in the tree, could not be used for these tests. Shimodiara (2002) regards the AU test as the more reliable.

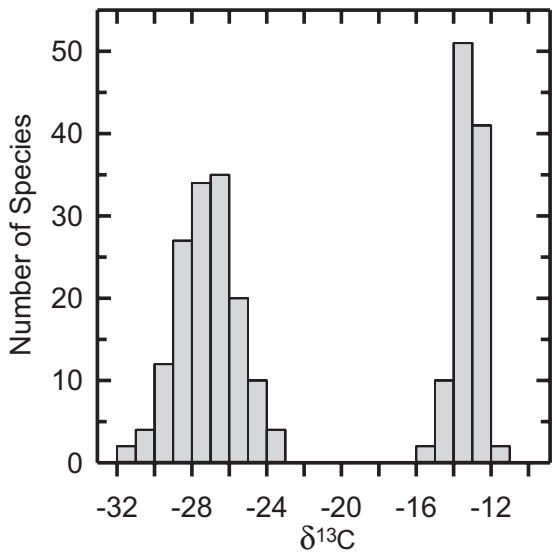

Figure 8. The distribution of carbon isotope ratios in the species sampled in this study. See Table 2 for a species list and their respective mean $\delta^{13}C$ values (and see Supporting Information, Table S1 for a detailed list of samples).

Carbon isotopes

$\delta^{13}C$ ratios were determined on *c.* 850 specimens representing 245 species (see also Supporting Information, Table S1). The $\delta^{13}C$ values occur in recognized C_3 or C_4 ranges (Fig. 8), indicating an absence of intermediate species operating a strong C_4 cycle (Fig. 8). Of these 245 species, 99 had C_4 isotopic ratios—all were *Euploca* spp. (Table 2). All *Myriopus* and *Heliotropium s.s.* (which will include all other former *Tournefortia* spp.) exhibited C_3-like isotopic ratios. In *Euploca*, 39 species exhibited C_3-like $\delta^{13}C$ values (28% of the species in the genus). In this group, five C_2 (*E. convolvulacea*, *E. cremnogena*, *E. greggii*, *E.lagoensis* and *E. racemosa*) and three proto-Kranz species (*E. karwinskyi*, *E. ovalifolia* and *E. procumbens*) have been previously identified, whereas six [*E. mexicana* (formerly *H. calcicola* and *E. calcicola*, Frohlich *et al.*, 2020), *E. fallax, E. filiformis, E. tenella, E. torreyi,* and *E. sessei*] are confirmed C_3 species based on gas exchange or microscopy studies (Frohlich, 1978: 65–70; Vogan *et al.*, 2007; Muhaidat *et al.*, 2011; Sage, unpublished).

Other data

Results for inflorescence bract presence/absence, trichome presence/absence/type on the upper corolla surface and geographical distribution are shown in Figures 9–11, respectively. Character state history was estimated by parsimony in MacClade 4.08a (Maddison & Maddison, 2000), based on the MrBayes tree (Fig. 7), but with the large Australian clade reduced to a single terminal for some figures. For bracts and trichome type, unordered character states were used. For geography, character state history was based on a step matrix (Fig. 11), so the cost of character state change could reflect the distance between geographical regions. MacClade cannot estimate character state history using a step matrix if the tree contains polytomies, so for

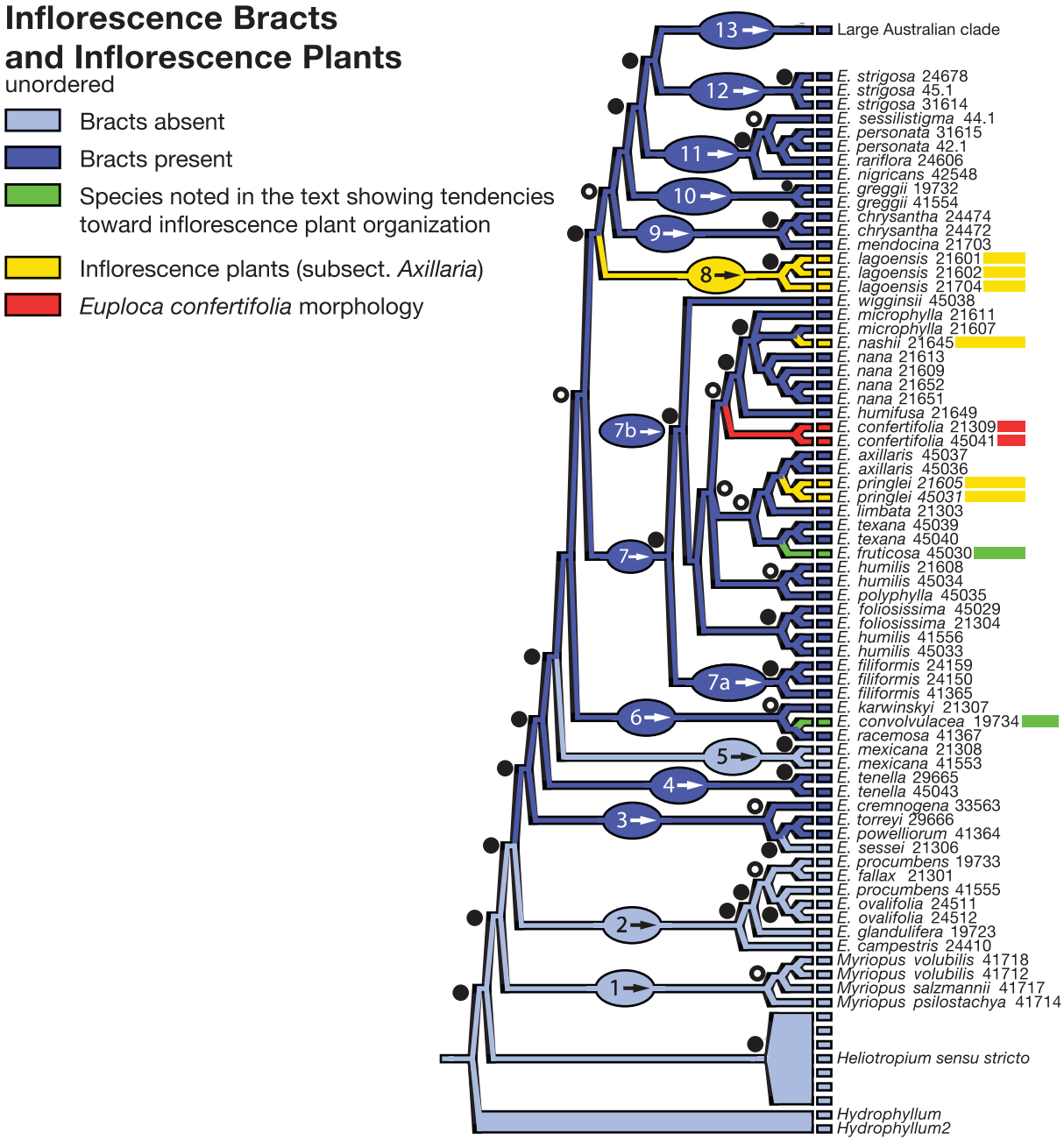

Figure 9. Inflorescence bracts and inflorescence plants. Character state history is reconstructed by parsimony on the MrBayes tree (Fig. 7). The large Australian clade is represented by a single terminal. Solid black dots identify major nodes with strong support from both parsimony (BS 90–100%) and Bayesian analysis (PP 0.95–1.0). Dots with white centres identify nodes with strong support from Bayesian analysis only. Support is not marked at nodes near branch tips, due to space limitation, unless significant for character history. Status of some accessions is highlighted by colour bars adjacent to or behind names.

Figure 11, polytomies were randomly resolved, but this has no effect on biogeographic inferences. Results for perennial/annual growth habit and for breeding system are shown in Figures 12 and 13, respectively, with character state history estimated by parsimony.

DISCUSSION

Our results resolve many issues concerning *Euploca*. These include its infrageneric taxonomy (Johnston, 1928), which must be abandoned due to the extensive parallelism/convergence and reversals in characters on which that was based. Parallelism and convergence (and perhaps reversals)

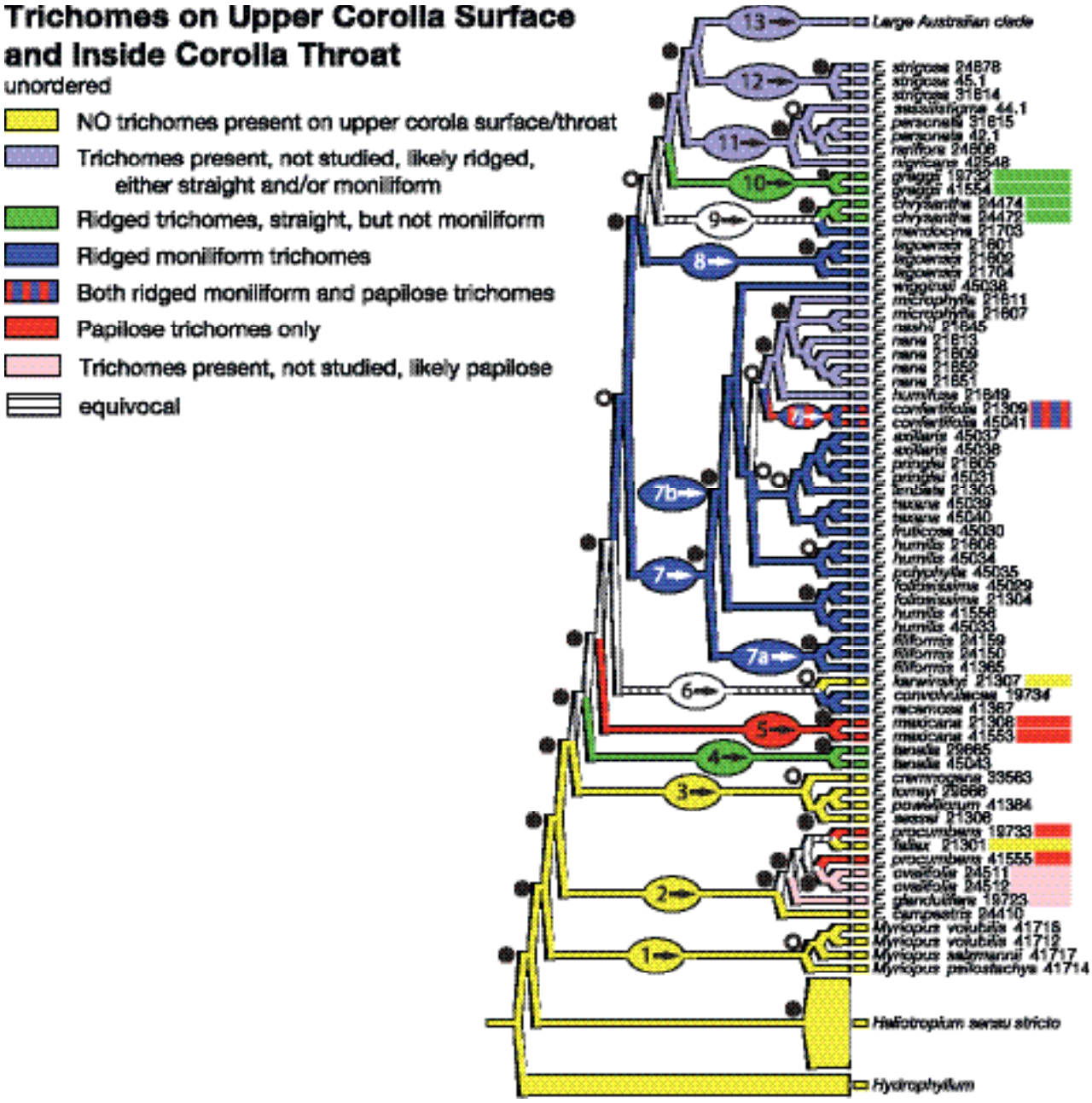

Figure 10. Trichomes on upper corolla surface and inside corolla throat. Character state history and node support dots are as in Figure 9.

are inferred for many morphological features, and notably for carbon assimilation systems. Next, the characters used for infrageneric taxonomy are discussed, then other morphological and life-history features are considered, followed by carbon assimilation systems.

PHYLOGENETICS

Our results are consistent with earlier molecular phylogenetic studies that included only a few *Euploca* spp. (Diane *et al.*, 2002, Hilger & Diane, 2003; Luebert *et al.*, 2011a). In particular, the traditional circumscriptions of *Heliotropium s.l.* and *Tournefortia s.l.* are non-monophyletic, consistent with recognition of *Heliotropium* section *Orthostachys* as the genus *Euploca* and *Tournefortia* section *Cyphocyema* as *Myriopus*, which are monophyletic. Our phylogenetic focus is on *Euploca*, so we consider relatively few accessions outside *Euploca*, which only serve as outgroups for *Euploca*.

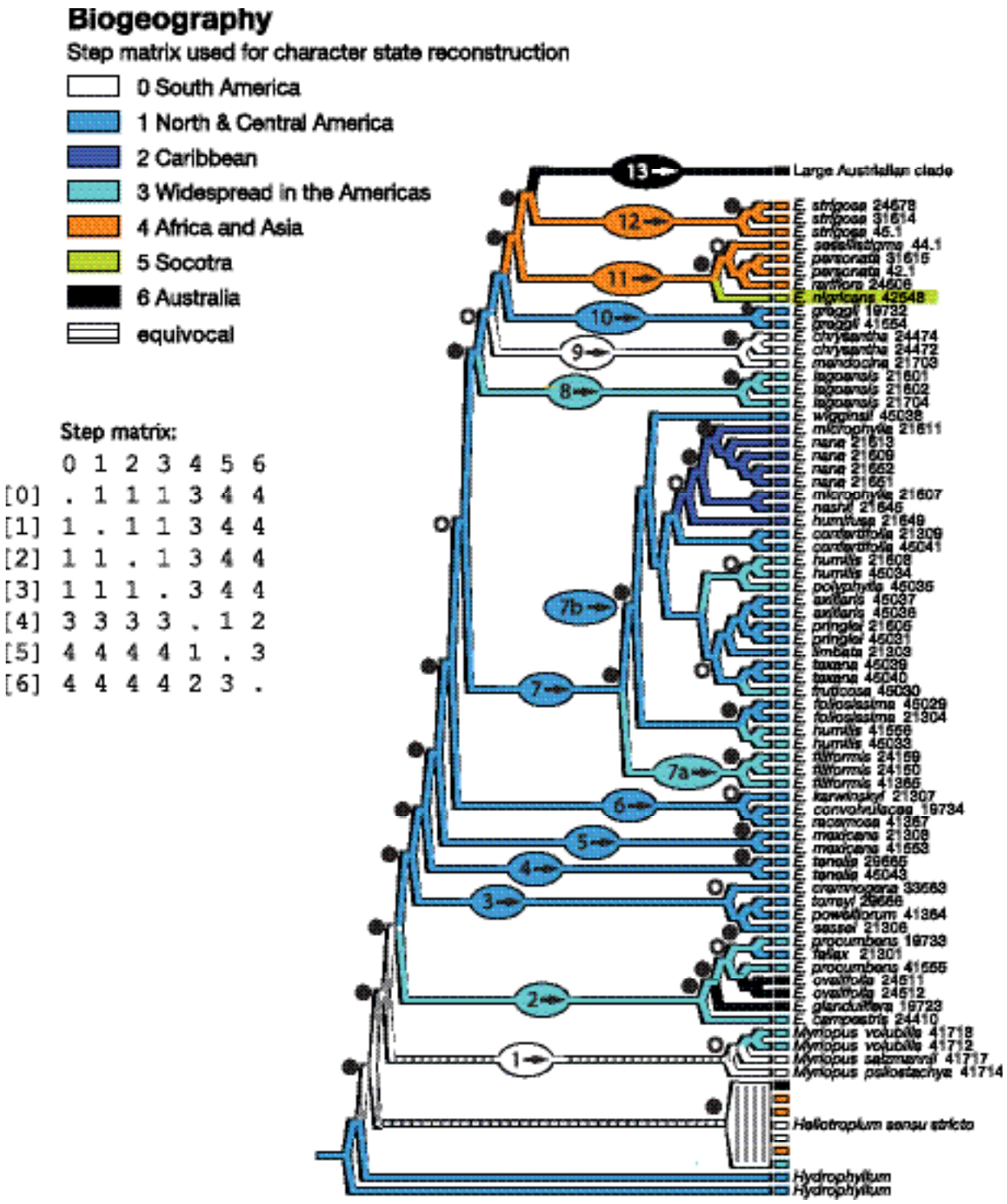

Figure 11. Biogeography. Biogeographic character state history is reconstructed by parsimony, using the step matrix shown, which requires a purely bifurcating tree, so polytomies in the MrBayes tree (Fig. 7) are randomly resolved. Node support is indicated as in Figure 9. See text for explanation of character states.

Johnston's subsections and defining characters

None of the three subsections of *H.* section *Orthostachys* (*Euploca*) erected by Johnston (1928) is monophyletic. He separated *Ebracteata* and *Bracteata* by bract presence, but these two states are intermixed in the tree (Fig. 9). Subsection *Ebracteata* contains

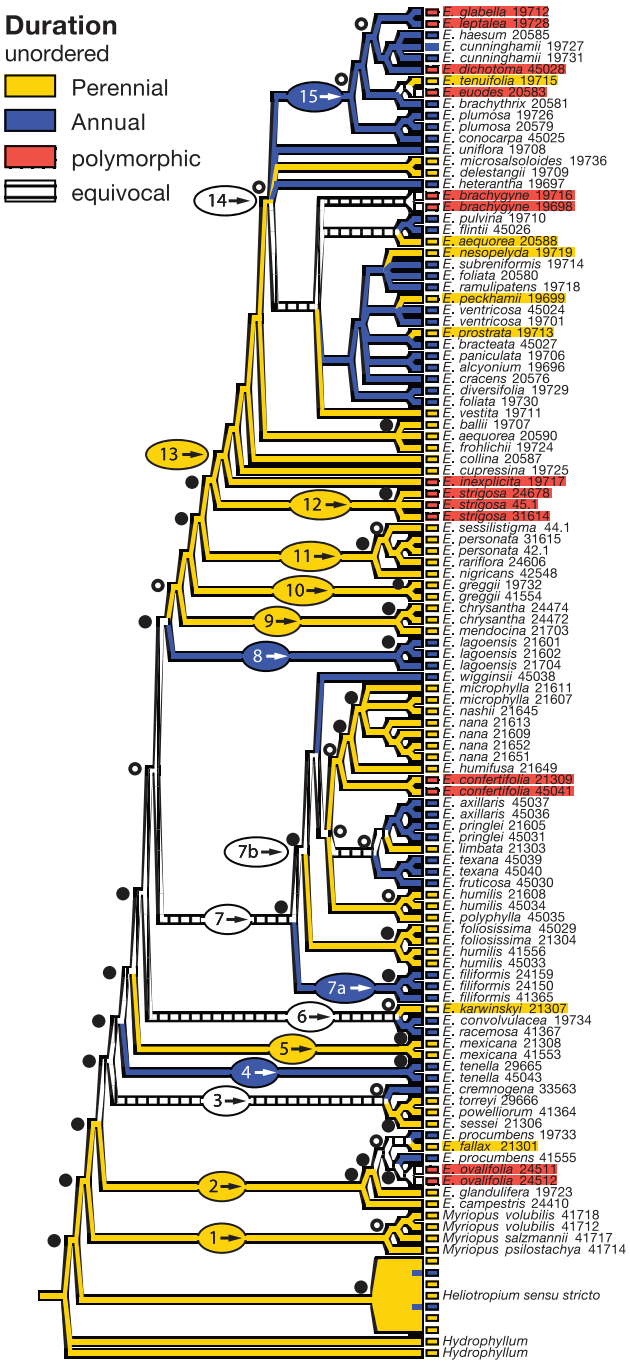

Figure 12. Duration. Character state history for perennial versus annual duration is reconstructed, and node support is indicated, as in Figure 9, with the large Australian clade shown in full.

the earliest-diverging *Euploca* species (clade 2) and some species in clade 3 and also *E. mexicana* (formerly *E. calcicola*, clade 5) but not *E. tenella* (clade 4). In clade 3, one species, *E. sessei*, completely lacks bracts, but *E. torreyi* has some prominent inflorescence bracts, yet Johnston nevertheless placed it in *Ebracteata* (as *H. angustifolium* Torr.), remarking that it is 'predominantly bractless' (Johnston, 1937: 19). *Euploca cremnogena* (clade 3) also has occasional prominent bracts (Johnston, 1939), and Johnston placed it in *Bracteata*. *Euploca tenella* (clade 4) has numerous bracts resembling small leaves and is typical for subsection *Bracteata*. All species of clade 6 and above are members of *Bracteata* or belong to

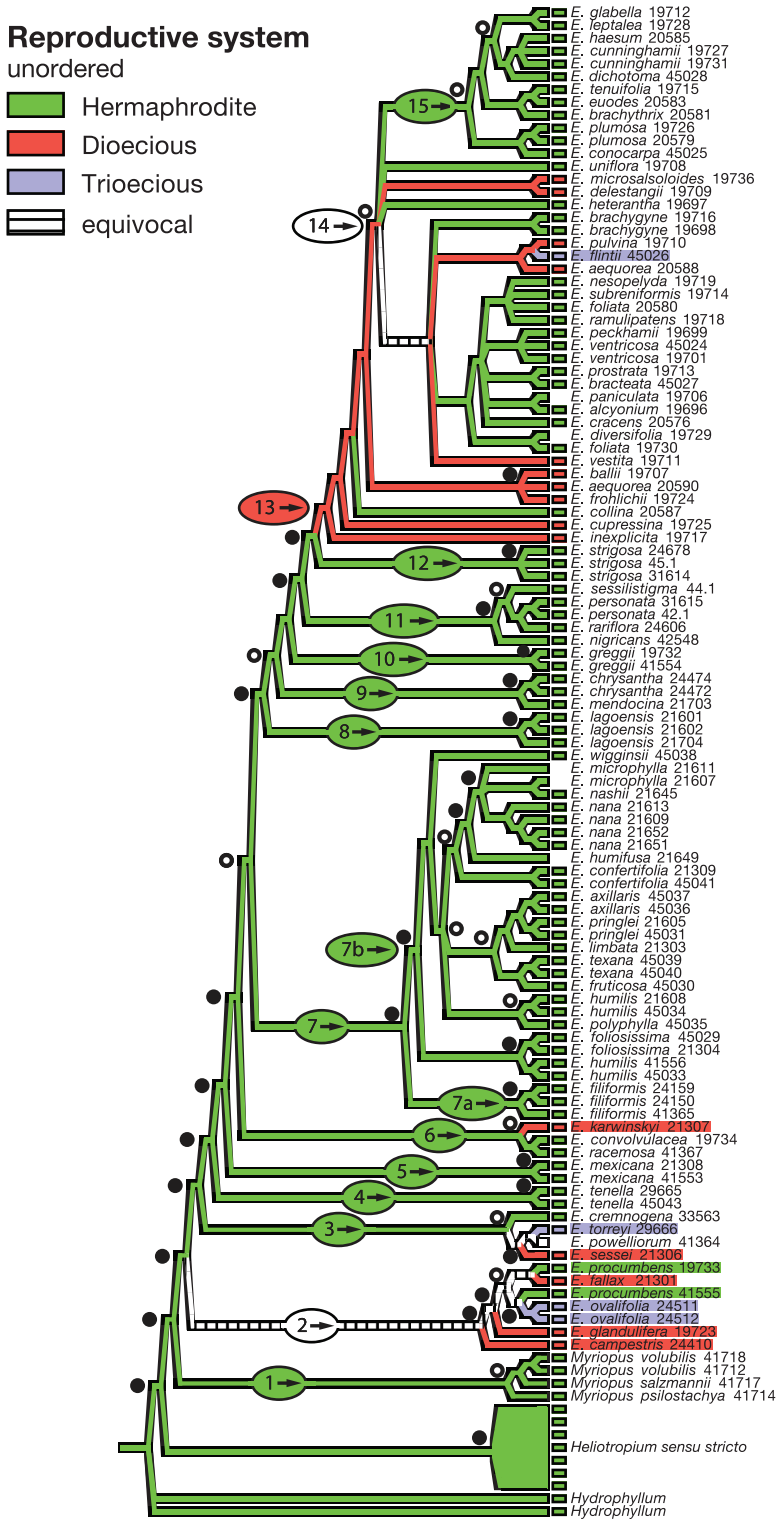

Figure 13. Reproductive System. Character state history and node support are as in Figure 9, with the large Australian clade shown in full.

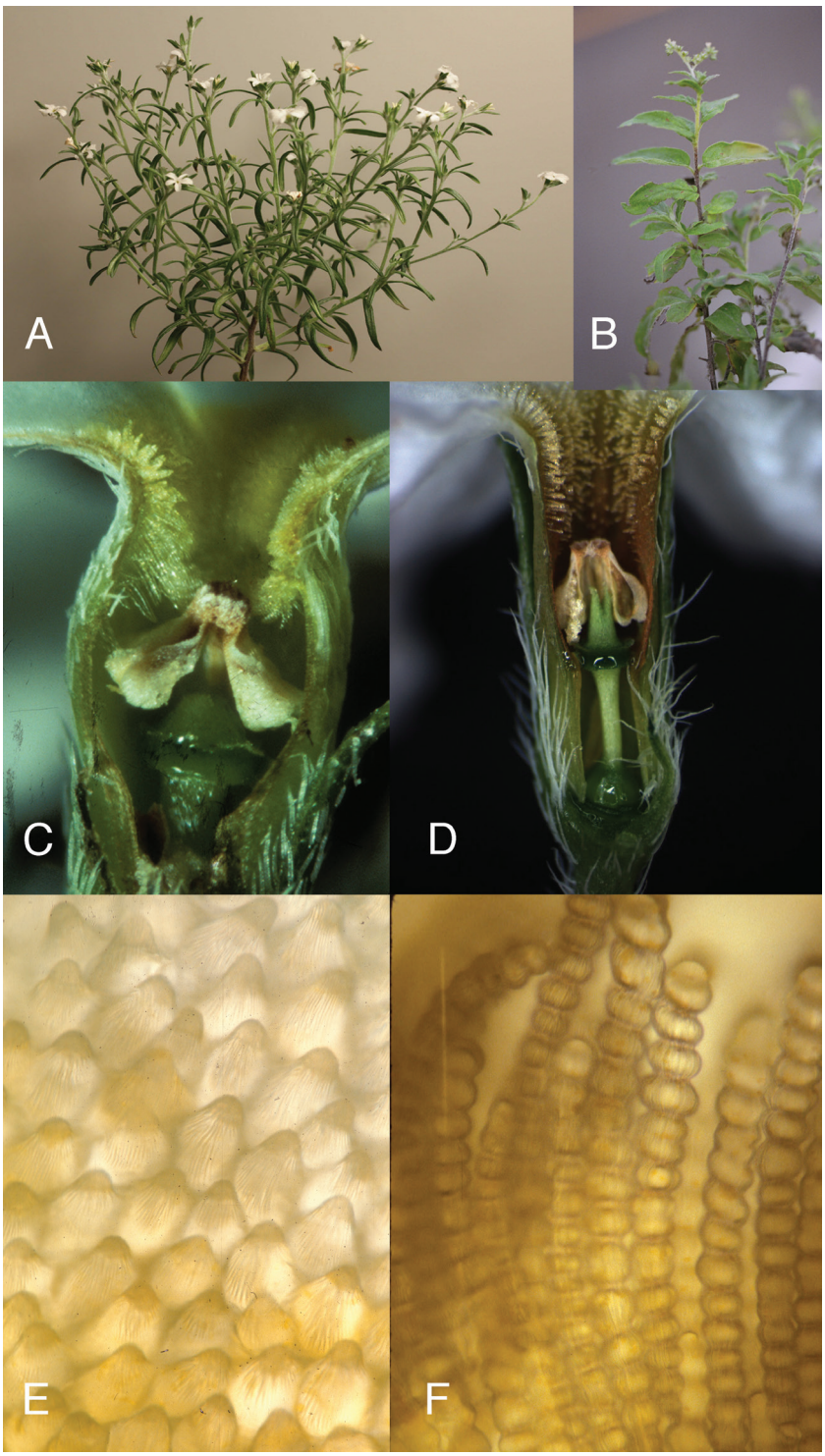

Figure 14. *Euploca* gross morphology and ridged/moniliform trichomes in corolla throat. A, *E. tenella* in flower and fruit. The vegetative and inflorescence shoots look similar, as bracts are similar in size and spacing to the leaves, and the stems are similarly robust. Such inflorescences are termed 'anthoclades', *sensu* Luebert *et al.* (2016). Open flowers are near the inflorescence tips. Developing fruits are visible further down the inflorescences. B, *E. fallax* in flower. The vegetative shoots are very different from the slender, crowded inflorescence shoots at the top of the leafy shoot. C, *E. humilis* flower, living, cut open lengthwise, with ridged cells grading into ridged cylindrical and then moniliform trichomes at the corolla throat. D,

Axillaria. Species in subsection *Axillaria*, (marked yellow in Fig. 9), which Johnston (1928) described as having 'flowers borne along the leafy stem', include *E. lagoensis* (clade 8), *E. nashii*, clade 7i (as labelled in Fig. 7) and *E. pringlei*, clade 7h. Hence, *Axillaria* is also polyphyletic. The 'leaves' of subsection *Axillaria* were shown to be the bracts of these inflorescence plants (Frohlich, 1978; see below).

More or less accompanying the transition to bracteate inflorescences, the mature inflorescence axes become more similar to the vegetative shoots in thickness, often in the spacing of appendages (i.e. flowers or leaves) and in general appearance (Frohlich, 1978). In the bractless species, inflorescences typically fork basally in two or three cymes, whereas in the bracteate species inflorescences either fork only once, to make two cymes or fail to fork, so the single cyme is collinear with the stem that bears it, augmenting the resemblance between them. For example, the inflorescences of *E. tenella* (clade 4) have a gross aspect so similar to the vegetative shoots that one may easily overlook some inflorescences when observing the whole plant (Fig. 14A). Such inflorescences are termed anthoclades (*sensu* Luebert *et al.*, 2016). By contrast, in the bractless species, inflorescence axes are typically significantly thinner than vegetative shoots, with flowers much more closely spaced than are the leaves of vegetative shoots (Fig. 14B). The bractless species all have cymes that are tightly coiled at the tip, which are termed scorpioid, but most bracteate species have cymes that are only somewhat bent at the tip, which are termed cincinni (Buys & Hilger, 2003), which further augments the resemblance of these cincinni to vegetative shoots.

Another character that transitions approximately where bracts appear in the tree involves the primary-walled corolla trichomes found in some species [Frohlich, 1978: 71, **72,** 73–76 (pages in boldface have figures); Förther, 1998]. These must not be confused with the abundant trichomes on leaves, stems, abaxial sides of sepals and the exterior of the corolla in *Euploca* (and other members of Boraginaceae); those trichomes have thick, silicified secondary walls covered with large blunt papillae. The largest have cystoliths at their bases.

Two kinds of primary-walled trichomes are borne on the corolla of some *Euploca* species (Frohlich, 1978: 71, **72,** 73–76). These primary-walled trichomes lack silica, unlike the abundant trichomes on the exterior of the corolla and on sepals, leaves and stems of most Boraginaceae. The most frequent kind of these primary-walled trichomes intergrades with the ridged papillose epidermal cells found on all *Euploca* petals (and on many other flowers). In these species, each epidermal cell on the upper limb surface bears a prominent papilla with numerous, closely spaced, sharp-edged cuticular ridges. As one scans towards the corolla throat, these papillae become progressively longer until they must be called trichomes. In many species, within the corolla throat, these cylindrical trichomes transition to become fully moniliform (Figs 14C–F). Figure 10 places the characters ridged cylindrical or moniliform trichomes, either confirmed or likely, on the tree and marked dark blue or light blue, respectively. Presence of these trichomes correlates almost perfectly with presence of bracts, except that *E. torreyi* (clade 3), with bracts, lacks these trichomes; perhaps this was an unstated reason why Johnston placed that species in subsection *Ebracteata*. Also, *E. karwinskyi* (clade 6), with bracts, lacks these trichomes, but this may be related to its urceolate corolla shape, unique in *Euploca*, which otherwise has salverform or funnelform corollas. We are not aware of such ridged cylindrical or moniliform trichomes in *Heliotropium s.s.*

A second sort of primary-walled trichome has a straight shaft with a moderately thickened wall with numerous tiny papillae on the surface (Frohlich, 1978: **184**). These trichomes do not intergrade with other types of cells. This sort of trichome is found only in scattered species, most notably in *E. procumbens* (clade 2, red in Fig. 10; Fig. 15A) and *E. mexicana* (clade 5, red in Fig. 10), which has so many that they form a dense yellow tuft in the corolla throat (Fig. 15B) and in *E. confertifolia* (clade 7j, red and blue striped in Fig. 10), which has moniliform trichomes as well. These trichomes are easily distinguished from secondary-walled papillae-bearing trichomes by observation between crossed polarizers. Figure 15C, D shows a hand section of live *E. procumbens* corolla tube in ordinary transmitted light and between crossed polarizers; only the secondary walls are bright between the polarizers, showing that only they have a significant secondary wall. Presence of this type of primary-walled trichome in widely separated species is a remarkable example of homoplasy. Furthermore, these trichomes also occur in some *Heliotropium s.s.* (e.g. in *H. angiospermum* Murray), demonstrating even greater homoplasy and almost certainly parallel acquisition of this trichome type on the corolla. Some other species in *Heliotropium s.s.* have dense groups of secondary-walled trichomes

E. wigginsii prepared similarly, also has ridged cells grading into ridged cylindrical and moniliform trichomes in the corolla throat. Numerous species-level differences are obvious between this and C. E, Each epidermal cell on the corolla limb has a big papilla with longitudinal ridges; this live, air mount makes ridges visible. F, Moniliform trichomes, air mount, with ridges visible in places. Photos by MWF.

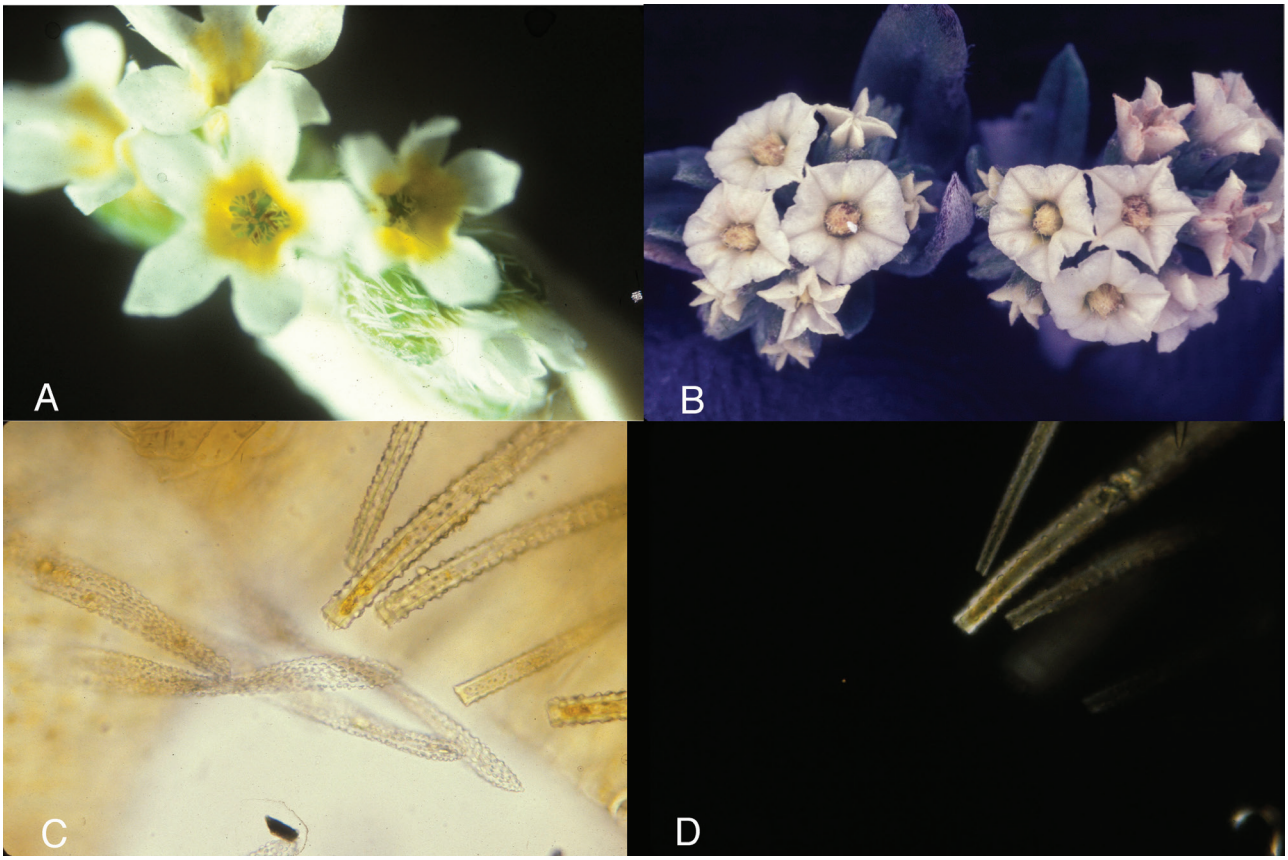

Figure 15. Flowers with primary-walled papillose trichomes. A, *E. procumbens* flower with sparse, primary-walled papillose trichomes in the corolla throat. B, *E. mexicana* (formerly *E. calcicola* and *H. calcicola*) with a tuft of such trichomes in the corolla throat. (The colour distortion is an artefact; leaves were green and the trichomes bright yellow.) C, *E. procumbens* corolla throat hand section, live, with primary-walled papillose trichomes from the inner surface visible on the left, angling down towards the right, and secondary-walled papillose trichomes from the outer corolla surface, on the right, angling down towards the left. The latter trichomes had their ends cut off in making the section. D, The same view as in C, but between crossed polarizers to detect the birefringence of secondary wall (Slayter, 1970), so only the trichomes on the right, with secondary wall, are bright. Photos by MWF.

within the corolla throat (e.g. *H. biannulatiforme* Popov); we have not found such secondary-walled trichomes in flowers of *Euploca*.

That *E. mexicana* (clade 5) falls above *E. tenella* (clade 4) was a great surprise because *E. tenella* has the bracts, elongate ridged trichomes and inflorescence axes similar in size to vegetative shoots, as is typical of clades 6 and up, whereas *E. mexicana* lacks these features. *Euploca mexicana* uniquely has opposite or whorled leaves. These two taxa have appeared in this order in all analyses, and this arrangement has strong support in the final analyses, implying that these features change relatively easily. Inclusion of bracteate and bractless taxa in clade 3 supports high mutability of that character.

Förther (1998) created the genus *Hilgeria* for three species from the West Indies, for which we were not able to obtain PCR products. The three differ greatly in morphology, and Förther's description of this new genus does not really distinguish his proposed genus from other species included in our analysis (such as *E. confertifolia*, clade 7j, that fall well within *Euploca*). Some species of Förther's *Hilgeria* were included in other studies and fell within what is now recognized as *Euploca* (Hilger & Diane, 2003; Nazaire & Hufford, 2012). We do not recognize *Hilgeria*.

Inflorescence plants

Johnston (1928: 47) described members of subsection *Axillaria* as having 'flowers borne along the leafy stem and not aggregated into a definite spicate or racemose inflorescence', which is a strange morphology for a member of Boraginaceae. These flowers are inserted individually among the leaves along the shoot but not in leaf axils. To clarify this strange morphology, Frohlich

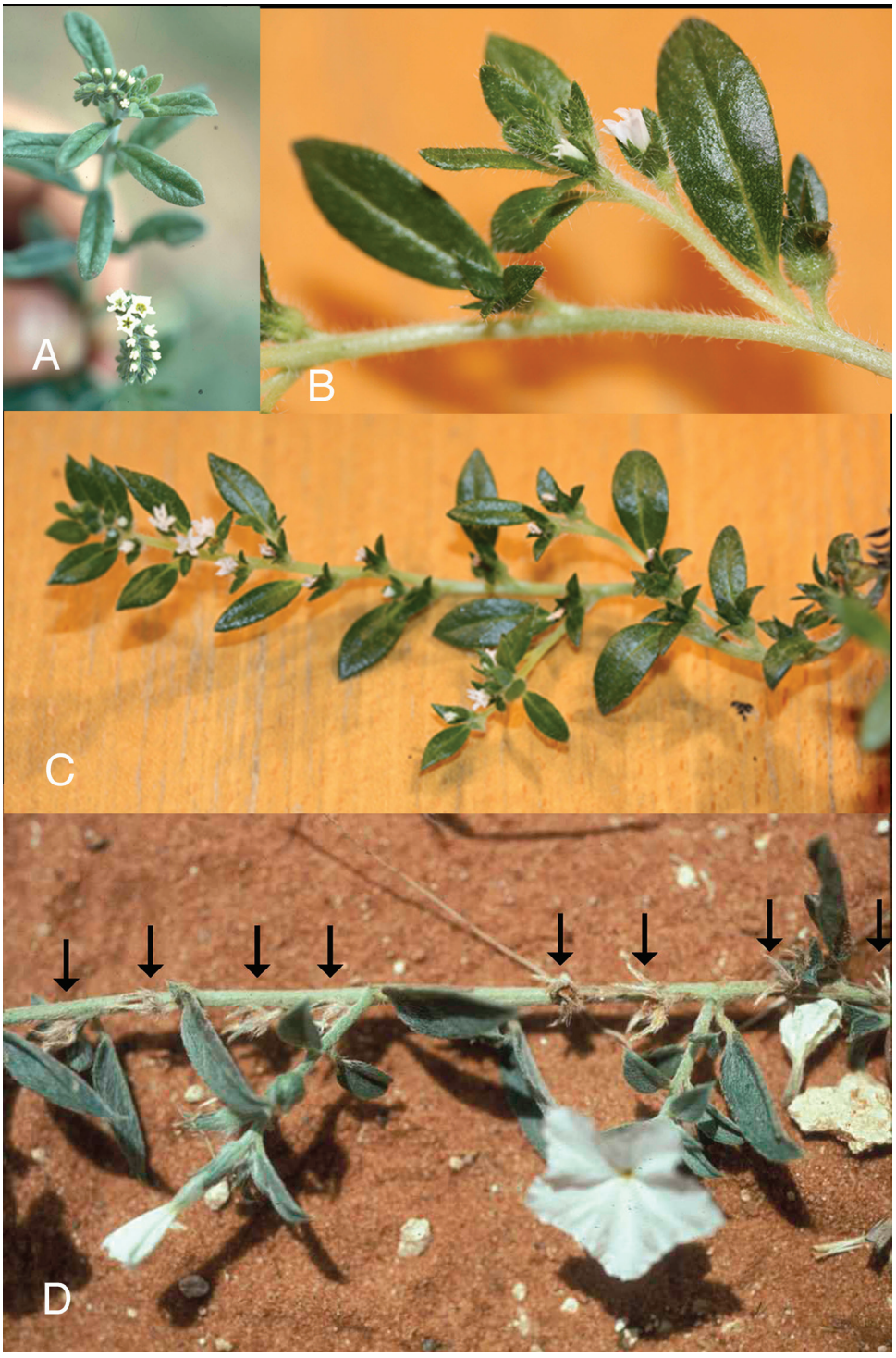

Figure 16. Cymes and inflorescence plants. A, *E. humilis* shows the common plant architecture in *Euploca*. The prominent vegetative shoot with spiral phyllotaxy terminates in the scorpioid cyme near the top of the image (with flower buds just starting to open) and a small vegetative renewal shoot is visible, just to the right of that cyme, growing from a leaf axil just below the inflorescence. The cyme near the bottom of the image is from the previous iteration of this pattern; the prominent vegetative shoot shown here is the renewal shoot from that previous iteration, having grown from a leaf axil just below the lower cyme. *E. humilis* has inflorescence bracts, but they are typically appressed to, and shorter than, the flowers, so are not visible here. B, Close-up of the inflorescence plant *E. pringlei*. The horizontal stem is morphologically a cyme. It bears

(1978: 18–51) studied overall morphology and apical development of vegetative and inflorescence shoots in all three of Johnston's subsections. Plants in *Ebracteata* and *Bracteata* have unremarkable vegetative shoots bearing terminal cymose inflorescences and renewal shoots arising from the uppermost leaf axils below these inflorescences (Fig. 16A; Frohlich, 1978: **19**, 20–29, 37–38). Vegetative axis development is typical for eudicots, with leaf primordia arising in a spiral phyllotaxy (Frohlich, 1978: **43**). The only oddity is the appearance of a precocious axillary renewal shoot in some collections of *E. procumbens*; subsequent elongation of these apical regions often separates the renewal shoot from its subtending leaf and stretches the insertion zone to make the renewal shoot appear to gradually diverge from the shoot subtending the cymes, resulting in what was termed a 'displaced branch' (Frohlich, 1978: 27–28, 46–47, **184:** Fig. 3). At maturity, this renewal shoot only gradually becomes distinct from the main shoot in the zone between the subtending leaf and complete separation of the renewal shoot from the main shoot, as is strikingly apparent from their vascular tissues (Frohlich, 1978: 32, **33**, 34, **184:** Fig. 8). Although commonly described as the renewal shoot being 'fused' with the main shoot, this has the unfortunate implication that the renewal shoot and the main shoot are initially distinct structures secondarily conjoined. Within the inflorescence, Endress (2010) termed this feature 'metatopy' and explicitly stated the two structures are 'congenitally fused', but describing this as a 'stretched attachment' would avoid unfortunate connotations and reflects the genesis of this feature.

Development of cymose inflorescences is stereotyped: flower primordia arise in two adjacent rows by bifurcation of the inflorescence apex. Sepal primordia in these flowers arise in a 3/5 spiral phyllotaxy that has mirror-symmetric handedness between the two rows of flowers (Frohlich, 1978: 39, **40**, 41–42, **43**, 45–46). Bracts, if present, arise on the opposite side of the inflorescence apex from the two rows of flowers, not between the flower primordia, so the flowers are not axillary to the bracts (Frohlich, 1978: **44:** Figs 8–10). The number of bracts varies greatly and is typically many fewer than the number of flowers. Developing bracts curve around the inflorescence axis and become appressed against developing flowers and inflorescence tip, protecting them, notably in the species with bent cymes (cincinni) rather than tightly coiled (scorpioid) cymes. In bractless species, the tightly coiled inflorescence tip is protected by the back of the older part of the inflorescence axis against which it is appressed.

The ordinary shoots of subsection *Axillaria* were studied most carefully in *E. pringlei*, and also in *E. nashii* and *E. antillana* (then lumped with *H. lagoense*). These shoots show the same development as the bracteate cymes of subsection *Bracteata*: the flower primordia arise in two adjacent rows by bifurcation of the apex, with sepals arising in mirror-symmetric spiral patterns in the two rows of flowers, and the 'leaves' arising just like bracts in *Bracteata*, on the opposite side of the inflorescence apex from the flowers (Frohlich, 1978: 47–51, **44**: Figs 11, 12). Hence, the ordinary *Axillaria* shoots are cincinni, with the bracts expanding to serve as leaves. These leaves bear axillary buds that develop directly into more inflorescence shoots (Figs 16B, C). Seedlings of *E. pringlei* initially make about six leaves in a spiral phyllotaxy and then convert to inflorescence growth for the rest of the life of the plant. Seedlings of *E. nashii* produce only two or three leaves before the axis begins making flowers, converting permanently to inflorescence growth. These are 'inflorescence plants' (yellow in Fig. 9). These inflorescences fall within the definition of 'anthoclades' (Luebert *et al.*, 2016), but inflorescence plants have the additional features that (1) inflorescence bracts make axillary branches that are wholly inflorescences and (2) the plant body almost completely lacks any non-inflorescence shoots, except perhaps in the first-formed, short, seedling axis, that bears fewer than ten leaves arranged in a spiral phyllotaxy.

This highly stereotyped inflorescence development allows such inflorescence plants to be recognized without detailed study of apical development: it is clear on herbarium specimens if flowers are borne in two rows on a shoot. The direction of the spiral phyllotactic origin of the sepals is reflected by their relative sizes

a developing fruit at the right (covered by the accrescent sepals) and a second fruit near the middle and a third fruit at the left edge of the photo (though the latter two are not so obvious). The bract/leaf at the right subtends an axillary shoot that immediately makes a flower, so it is morphologically a cyme from its inception. C, *E. pringlei* shoot system is all cymes (cincinni, because the tip is not coiled). The developing fruits along the main shoot lie in a zig-zag pattern, because the flowers form in two rows (by bifurcation of the shoot apex). The leaves are morphologically bracts, and grow axillary cymes, but the flowers/fruit are not axillary to the leaves/bracts. D, *E. convolvulacea* has the typical architecture: vegetative shoots that terminate in cymes, with, in this species, widely spaced large bracts. Here, the horizontal stem crossing the image is an old cyme; the black arrows mark its dead flowers. Unusually, the bracts on this old cyme have grown axillary branches, which start out as vegetative shoots, with spiral phyllotaxy, but soon convert to cymes and make the flowers in this image. Photos by MWF.

and slight basal imbrication. The last critical features, axillary branches directly forming more inflorescence shoots and no ordinary vegetative shoots, are easily determined. Cuban *Euploca serpylloides* (Griseb.) Diane & Hilger may well be another inflorescence plant; we were unable to amplify DNA sequences from our herbarium samples. Other likely inflorescence species occur in Brazil.

In our cladogram, *E. lagoensis* is also a confirmed inflorescence plant. *Euploca lagoensis*, *E. pringlei* and *E. nashii* are the only inflorescence plants in our cladogram (Fig. 9, yellow). *Euploca axillaris*, sister to *E. pringlei*, is not an inflorescence plant, despite its name. It has the gross appearance of an inflorescence plant due to the similar sizes, spacings and morphology of the vegetative leaves and the bracts, with long inflorescences. However, on close observation it has ordinary vegetative shoots with terminal cymes and renewal vegetative shoots formed near the tops of the older vegetative shoots.

There are, however, some other species with both standard vegetative shoots as well as ordinary cymes that nevertheless show a tendency towards becoming an inflorescence plant. In particular, some annual species, like *E. fruticosa* (green in Fig. 9), stop making vegetative renewal shoots late in the growing season but continue extending their existing inflorescences, making more flowers and occasional large bracts. By the end of the season, there may be more photosynthetic tissue in the bracts than in the remaining live leaves on the true vegetative shoots, as older leaves gradually senesce. *Euploca convolvulacea* (clade 6, green in Fig. 9) also has large bracts in its inflorescences, and sometimes those bracts produce axillary branches. These axillary branches begin as vegetative shoots, with multiple leaves born in a spiral phyllotaxy, but they soon convert to inflorescence growth (Fig. 16D). Hence, axillary buds associated with bracts can arise without full conversion to an inflorescence plant. African *E. katangensis* (of the *E. baclei* group; Simons & Wieringa, 2019), not in our phylogenetic analyses, also has numerous large inflorescence bracts, similar to leaves in size and spacing, and these bracts can produce axillary branches that typically make a few tiny leaves in a spiral phyllotaxy before switching from vegetative to inflorescence growth. This short bit of vegetative growth may not be obvious, so the branch cyme appears inserted directly in the axil of the bract. The general appearance so resembles an inflorescence plant that Simmons & Wieringa (2019) placed *E. madagascariensis* (another segregate of *E. baclei*), not in our phylogenetic analyses, in Johnston's former subsection *Axillaria*, but these plants also have multiple purely vegetative shoots before making inflorescences, as well as making short vegetative growth from bract axils, so they are not inflorescence plants.

Euploca confertifolia (red in Fig. 9), shows yet another strange inflorescence morphology. Its vegetative shoots have small linear leaves that are crowded near the shoot tips before the internodes elongate. The vegetative shoots bear terminal cymes and form standard vegetative renewal shoots. The cymes bear bracts that are about the same size and shape as vegetative leaves, and, after making one or a few flowers, the cymes cease making flowers but continue to make bracts/leaves, initially in an irregular phyllotaxy that soon settles down to an ordinary spiral phyllotaxy, so after making a few flowers the cyme converts into a vegetative shoot. These reconverted vegetative shoots then bear further terminal cymes, and the process repeats multiple times (Frohlich, 1978: 222–224). This suggests that inflorescence morphology is remarkably plastic in this part of *Euploca*, which is probably related to the observation that inflorescence axes in the bracteate species are less distinct from vegetative shoots than in the bractless species. Of the two varieties of *E. confertifola* (Turner, 2016) we have examined the erect variety more carefully. The congenitally procumbent variety has more crowded appendages at shoot tips, complicating observations, but it appears to behave similarly. A confusing feature for the recognition of these two varieties is that shoots of the erect form will lie down in age and make tufts of leaves at their tips, simulating a congenitally procumbent plant.

Craven's descriptions (1996) of Australian species suggest there are more examples of unusual inflorescence morphology among those species, which we have not examined carefully. Craven (1996) listed 11 species as always or sometimes having 'solitary' flowers, which could reflect an inflorescence-plant growth habit or *E. confertifolium*-like inflorescences.

Note that the three inflorescence plants are widely separated on the cladogram (Fig. 9, yellow), with multiple non-inflorescence plants between them, suggesting three independent origins. The placement of leaves (i.e. morphologically the bracts) in *E. pringlei*, which are inserted close to flowers, unlike in the other two, supports the separate evolutionary origin of its inflorescence-plant architecture. Furthermore, the two species with a tendency towards inflorescence-plant morphology (green in Fig. 9) are not closely related to the real inflorescence plants, suggesting a widespread capacity to evolve inflorescence-plant morphology in this part of the *Euploca* tree.

This inflorescence-plant architecture may have adaptive value for plants that live in habitats with brief and, especially, growing conditions of unpredictable duration (Frohlich, 1978: 105). These plants automatically produce both flowers and leaves concurrently as each shoot grows, so they are continuously able to balance reproductive demand

for photosynthate with the growth of leaves to satisfy that demand. *Euploca* plants that produce extensive vegetative growth on a shoot, which then switches to make only flowers in scorpioid cymes/cincinni, with renewal shoots that take some time to develop, do not have such an automatic balance; the ratio of leaves to developing fruit oscillates on connected shoots as first only leaves and then only flowers and fruit are produced. For small, annual plants this might be an especially significant problem. *Euploca pringlei* plants are notably small.

When a plant of *E. pringlei* suffers drought, it immediately stops new growth and only matures the fruits (and perhaps some flowers) that were already formed, thus taking advantage of the remaining available water to mature a maximum number of seeds. For this strategy to be efficient there needs to be a continuous balance between demand and production of photosynthate. *Euploca nashii* also grows in zones that have unpredictable water availability. *Euploca lagoensis*, growing on exposed riverbanks and periodically flooded sites (Melo & Semir, 2010; Costa & Melo, 2019), may suffer unexpected inundation and so also experiences unpredictability.

Of the species with the standard architecture of distinct vegetative shoots versus cymes, the precocious renewal shoots of *E. procumbens* (normally an annual) could also maintain a balance between leaves producing photosynthate and flowers/fruit requiring it (Frohlich, 1978: 104–105). Vigorous plants often have two to four iterations of shoots, and their renewal shoots growing simultaneously. The younger shoots produce leaves at the same time as the older shoots are making flowers and fruit. Such simultaneous growth could maintain the balance between photosynthate demand and production. *Euploca procumbens* is a widespread weed that also grows on natural disturbed sites along riverbanks, so it may also experience growing conditions of unpredictable duration. In contrast, many non-inflorescence species are perennial shrubs, probably with deeper roots providing a more stable water supply, and perennials can store excess photosynthate, remaining at the end of the growing season, for future use.

Plant duration and reproductive system

Longevity and breeding system both show extensive homoplasy (Figs 12, 13). Annual and perennial small clades (and some individual species) intermingle throughout the tree. Longevity is a continuum, especially in climates without severe winter or an intense dry season; in such climates, a normally annual plant may occasionally survive a moderately inclement season to reproduce in additional seasons. That does happen: e.g. *E. procumbens* is normally an annual, but will survive indefinitely if conditions are not extreme.

Some perennial *Euploca* spp. are, in contrast, long-lived shrubs of stable habitats, and some of these are sisters to annuals that apparently never survive more than one growing season. For example, the tall shrub *E. karwinskyi* is sister to *E. convolvulacea* + *E. racemosa*, which are annuals. The long-lived perennial clade (*E. torreyi* + *E. powelliorum* + *E. sessei*) is sister to the small annual *E. cremnogena*. Furthermore, some annuals (e.g. *E. pringlei*) even exhibit apparent programmed death: if a plant once suffers from significant drought, the plant stops growing, matures its developing fruit and dies. In cultivation, such plants cannot be saved after a brief drought, even though subsequently given much water and fertilizer. The outgroup *Myriopus* is perennial, as are most early-diverging *Euploca* spp., so *Euploca* is probably ancestrally perennial.

Euploca breeding systems show wide variation. Most species have perfect flowers, but there are a number of fully dioecious species and a few that are trioecious (Fig. 13). The dioecious and trioecious species do not all group together, but instead are scattered across the tree. The mechanisms that achieve dioecy have not been studied in detail, but, at least for *E. fallax*, *E. sessei* and *E. karwinskyi*, the morphology is similar (Frohlich, 1978: **78**, 166, 177, 199). Female flowers produce thin anthers containing aborted tissue, but male flowers produce ovaries that are not obviously defective. This is despite these species all having perfect-flowered sister species, which is the optimized ancestral condition of the genus and the inferred state along the backbone of the tree. If dioecy here really did arise multiple times, in (at least) superficially similar ways, from perfect-flowered ancestors, this would be a remarkable example of parallel/convergent evolution. *Heliotropium s.s.* is perfect-flowered, except for a report of gynodioecy or leaky dioecy in *Tournefortia argentea* L.f. (Wang *et al.*, 2020).

Most perfect-flowered *Euploca* spp. are insect pollinated and in Mexico are visited by small Lepidoptera and tiny Hymenoptera (Frohlich, unpublished). Frohlich (1978: 113–150, 266–277) modelled seed set to recreate the observed ratios of fruits with zero to four seeds and found that the contagious distribution best fits the data, interpreted as each flower being visited by a Poisson number of insects that each brings a Poisson number of effective pollen grains. *Euploca procumbens*, however, has a low pollen-ovule ratio and should be obligately self-pollinated; this species was cited in Cruden (1977) as *Heliotropium* sp. Its flowers are visited by Lepidoptera. Almost all *E. procumbens* fruits contain four seeds. *Euploca folliosissima* is a putative obligate apomict: the many observed pollen mother cells all have one to

seven stray chromosomes at the late stages of meiosis, and these stray chromosomes form microcyte pollen grains that typically have a single pore, whereas its large grains usually have five pores, although sometimes fewer (Frohlich, 1978: 88–91, **93**, 94, **95**, 96). Frohlich (1978) examined maturing anthers from multiple populations and found such microcytes in every anther, but this species sets abundant seed in the wild. It apparently requires pollination because plants cultivated in growth chambers rarely set seed (and then only when hosting aphids). It has an extensive range in the central highlands of Mexico and is morphologically uniform and distinct from other *Euploca* spp. *Euploca foliosissima* may be of hybrid origin; its chromosome number is $2n = 48$, compared to its likely sister *E. humilis* that has $2n = 24$ (Frohlich, 1978: 91 [as *H. ternatum*]). A hybrid origin of *E. foliosissima* could be the cause of the two positions of *E. humilis* in the cladogram. Polyploidy may be common in *Euploca* (Frohlich, 1978: 90). Both duration of growth and breeding systems are subject to much homoplasy, including reversals.

Biogeography

Geographical distribution characters are placed on the MrBayes tree in Figure 11. *Euploca* clearly originated and diversified in the Americas. Its sister group, *Myriopus* (clade 1) is exclusively American (scored as equivocal between North and South America in Figure 11). *Euploca* clades 3–10 are also exclusively American. Clade 2, which includes both Old and New World species, must have originated in the Americas with one (or conceivably two) migrations to the Old World because *E. campestris*, sister to the rest of clade 2, is American, as are the clades above and below clade 2. *Euploca procumbens* is an abundant, widespread weed in the Americas and is scored as American in Figure 11, although MWF has seen specimens from Pacific islands that fit the concept of *E. procumbens*, except for having longer trichomes on the sterile tip of the stigmatic apparatus above the stigmatic ring. These could be recent inadvertent introductions, especially as there are some 19th century specimens of *Heliotropium* growing on discarded ballast in various port cities, but natural migration to distant islands also clearly occurs, as there are endemic species of *Euploca* and *Heliotropium s.s.* (and former *Tournefortia*) on oceanic islands, including the Galapagos Islands, Hawaii and other Polynesian islands. In clade 2, *E. glandulifera* is found in Australia, whereas *E. ovalifolia* is widespread in Africa, South Asia and Australia. Strictly interpreting the Bayesian tree would suggest two migrations to the Old World, but it would also suggest two origins of the distinctive *E. procumbens* (PP 1.00 for both nodes). The parsimony tree is less resolved. We think it more likely that there was one migration to the Old World in this clade. Except for the weedy *E. procumbens* and some populations of *E. ovalifolium* (Craven, 1996), these species in both hemispheres are dioecious (Fig. 9). Dioecy impedes long-distance dispersal (Jordan, 2001), but it clearly did occur, but perhaps dioecy is 'leaky', and some plants produce some perfect flowers. In *E. fallax*, MWF examined hundreds of old inflorescences on male plants and did find one normal-looking seed. Although resulting from far fewer than 0.1% of 'male' flowers, this might have allowed successful propagation after long-distance dispersal, even of a single seed. Comparable seed set without pollination in female plants could not be estimated.

In the American clades, clades 3–10, many of the listed species occur in North and Central America. Clade 7 is most diverse in Mexico and adjacent USA and Central America, with some species extending into South America (lightest blue) and one subclade restricted to the West Indies (dark blue). There are several additional South American *Euploca* spp. not in our phylogenetic analyses. Several of the included Mexican species are highly restricted (e.g. *E. karwinskyi* and *E. cremnogena*). One of us (MW) has observed diversity in dry Andean valleys hinting at undescribed species there. Clade 8 is widespread in tropical America in wet conditions, e.g. mud exposed on receding shores of lakes and streams and on wet ground. Clade 9 is found only in southern South America, and clade 10 is only in North America (northern Mexico and Texas). These two are found in dry habitats, even in semi-deserts.

Above clade 10, all taxa are found exclusively in the Old World. Clade 11 includes one widespread species, but most are restricted to dry habitats in the Horn of Africa and Arabia, with one endemic to Socotra. Clade 12 is a widespread weedy species. The large clade 13 is exclusively Australian, which is probably a recent radiation, given the limited sequence divergence among these species (Figs 4, 6). *Euploca* nutlets are distinctive. D. Steart (pers. comm.) searched the Australian fossil seed and fruit literature, but found no records of *Euploca* or of *Heliotropium s.s.* Clades 9, 10, 11 and (at least predominantly) 13 are found in dry, sunny habitats.

Clades 9, 10 and 11 are hermaphrodite (Fig. 13), which would seem to facilitate long-distance dispersal (Jordan, 2001) to the Old World and possibly also between dry zones of southern North America (clade 10) and South America (clade 9). Other such long-distance dispersals between these regions have been inferred, most famously for *Larrea* Cav. (Lia *et al.*, 2001) and even other Boraginaceae, e.g. *Tiquilia* (Moore,

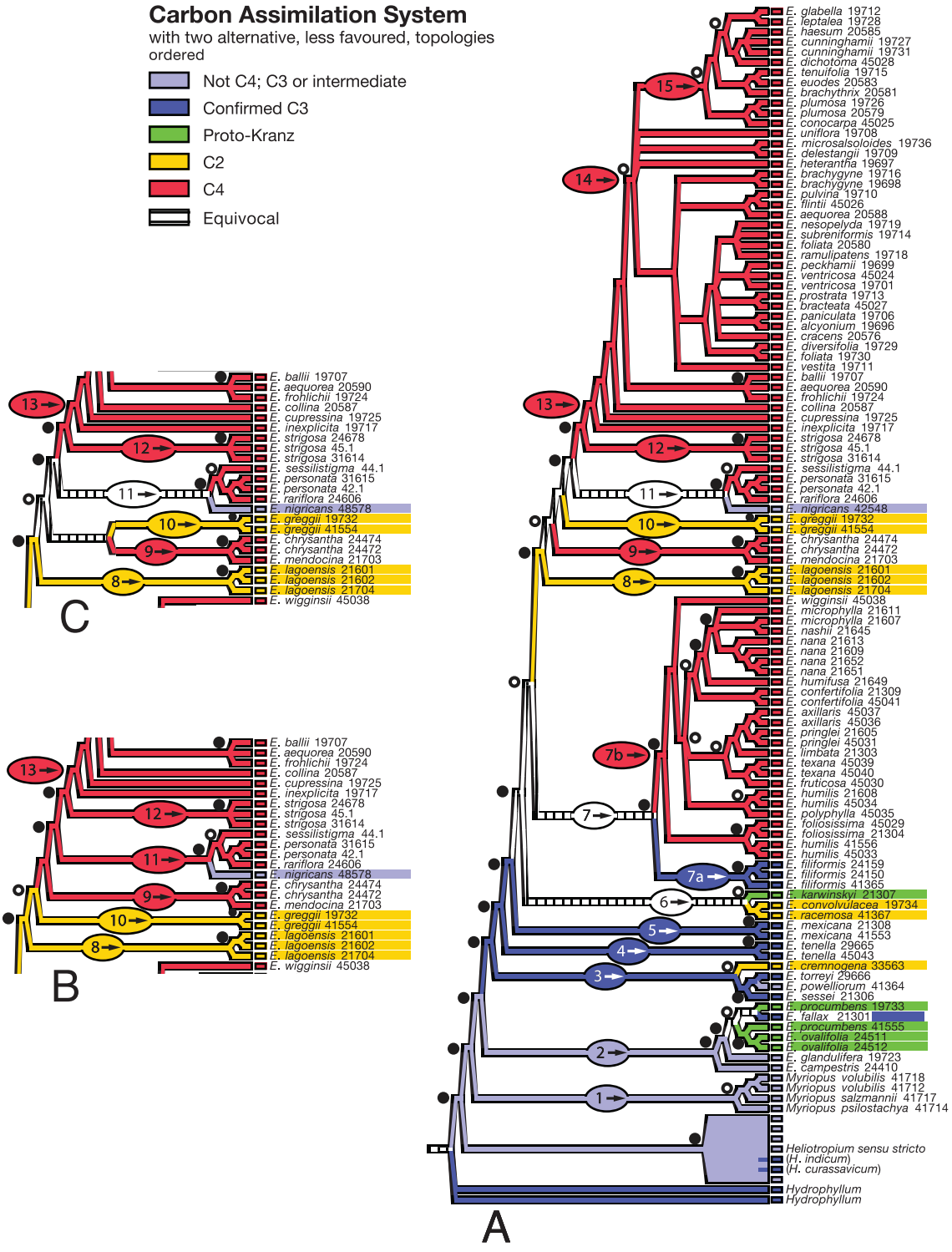

Figure 17. Carbon Assimilation System. Character history reconstruction and node support are as in Figure 9. MrBayes only weakly unites clade 10 with its sister (PP 0.86), separating them from clade 9, so all three resolutions of that potential trichotomy are shown. A, Character state history on the MrBayes tree. B, Alternative less favoured topology, swapping the order of clades 9 and 10 along the pectinate backbone. C, Alternative less favoured topology, uniting clades 9 and 10. Carbon isotope ratio ($\delta^{13}C$) is diagnostic for C_4, but cannot distinguish C_3 from C_2 or proto-Kranz. Taxa with C_3-like $\delta^{13}C$, but lacking

Tye & Jansen, 2006) and Amsinckiinae (Guilliams *et al.*, 2017). The tragic death of Lyn Craven, who monographed the Australian species (Craven, 1996), precludes discussion of biogeography in Australia.

Carbon fixation

In Boraginaceae, C_3 photosynthesis is the ancestral condition with C_4 occurring in at least ninety nine *Euploca* spp. In the related non-*Euploca* clades, i.e. *Heliotropium s.s.* (including most of *Tournefortia*) and *Myriopus*, only species with C_3-like $\delta^{13}C$ values have been found. Our results are in agreement with Akhani (2007), who also found no C_4 species in *Heliotropium s.s.* Limited examination of these outgroup species (Muhaidat *et al.*, 2011; Sage, unpublished) showed no evidence of non-C_3 states, leading us to tentatively identify them as purely C_3. It is a puzzle why C_4 has evolved in *Euploca* but not in its sister groups, even though some grow in habitats that would seem to favour C_4 evolution. *Heliotropium* section *Cochranea* is found primarily in the Atacama Desert of Chile (Luebert, 2013). In south-western Asia, *Heliotropium s.s.* is abundant and diverse in dry areas, in particular in the Irano-Turanian and Saharo-Sindian regions, which are centres of evolution and diversification of C_4 taxa (Akhani, 2007; Rudov *et al.*, 2020).

In *Euploca* (Fig. 17), probable C_3 (pale blue) and confirmed C_3 (dark blue) dominate the early-diverging clades 2–5. Notably, clades 2 and 3 also contain proto-Kranz (green) and a putative C_2 species, *E. cremnogena* (yellow-orange), based on its Kranz-like anatomy (Frohlich, 1978: 65) coupled with a $\delta^{13}C$ value showing they are not C_4. This suggests the earliest-diverging branches were independently evolving a variety of intermediate physiologies well before the C_4 clades diverged. In clades 4 and 5, C_3 species *E. tenella* and *E. mexicana* (formerly *H. calcicola*) exhibit enlarged, organelle-enriched bundle sheath cells, which have been hypothesized to represent a physiological activation of the bundle sheath cells and, as such, may precede the formation of the proto-Kranz condition that arises in clades 2 (*E. procumbens*) and 6 (*E. karwinskyi*) (Muhaidat *et al.*, 2011). The sister position of *E. karwinskyi* with *E. convolvulacea* + *E. racemosa* in clade 6 supports the hypothesis that proto-Kranz facilitates the origin of the C_2 condition in *Euploca*.

In *Euploca*, placement of three C_2 species (*E. convolvulacea* in clade 6, *E. lagoensis* in clade 8, and *E greggii* in clade 10) supports the hypothesis that the C_2 condition facilitates the rise of C_4 photosynthesis, as they occur in a sister position to C_4 clades. We specifically interpret the phylogenetic pattern (Fig. 17A) as indicating the backbone of the tree between clades 5 and 10 was probably predominantly C_2 in nature, with multiple origins of C_4 clades: in clades 7b and 9 and perhaps independently in clades 11 and/or 12. The less favoured alternative topologies (Fig. 17B, C) are consistent with this inference. Two species complicate this assessment, generating the equivocal portion of the backbone around clades 6 and 7. In clade 7A, the C_3 species *E. filiformis* sits above the backbone and is sister to a C_4 clade, and in clade 6 the proto-Kranz *E. karwinskyi* is sister to a pair of C_2 species; this, coupled with C_3 projected from the base of the cladogram and C_2 from above, renders a section of the backbone equivocal. Note, though, that of the six character state inputs to this equivocal zone, two each are C_3 and C_2, one is proto-Kranz and only one is C_4, suggesting that an intermediate state characterized this backbone section. If *E. karwinskyi* is re-assigned as C_2, then this section of backbone becomes fully C_2 (not shown). If *E. karwinskyi* were C_3 then this backbone section would remain equivocal. The confirmed C_3 species *E. filiformis* could be a revertant from C_2, or perhaps the backbone up to clade 7 includes C_3 species.

In clade 11, *E. nigricans* is isotopically C_3 or intermediate (confirmed by four $\delta^{13}C$ determinations on three collections, including the type, with permission). Living material is not available to determine whether it is C_2, proto-Kranz or confirmed C_3. The sister and the next outgroup of *E. nigricans* are both C_4, but the subsequent outgroup is C_2 (Fig. 17A, C) or C_4 (Fig. 17B). Alternative arrangements, that would join all of the C_4 taxa of the sister group of *E. nigricans* and the next outgroup, to the exclusion of *E. nigricans* itself, were strongly rejected by the AU test and also rejected by the SH test. The alternative arrangement shown in Figure 17B implies that C_4 arose once (not twice) in a clade containing both the South American C_4 species (clade 9) and the Old World C_4 taxa (clades 11, 12 and 13), (hence three times total in *Euploca*) but for that to be true, C_4 must have been lost by *E. nigricans*. It is possible that *E. nigricans* is a revertant from a C_4 ancestor, although reversion from C_4 is considered difficult and thus less likely (Oakley *et al.*, 2014; Sage *et al.*, 2014). If *E. nigricans* retains an ancestral C_2 condition, then C_4 arose four times from a likely C_2 backbone in *Euploca*, even if the topology of Figure 14B were correct: for *E. nigricans* to have retained C_2 would require the backbone around clade 9

further evidence to distinguish C_3 from an intermediate carbon assimilation system, are labelled pale blue. Dark blue taxa have been confirmed as C_3 by physiological studies. Status of some accessions, notably intermediates, is highlighted by colour bars behind or adjacent to their names, e.g., *E. nigricans*, which would otherwise be hard to see.

(Fig. 17B), as well as the branch leading to clade 11, to have been C_2, despite the parsimony-based character state reconstruction in that part of the tree. Note that, regardless of the above inferences, proto-Kranz must have arisen twice (in clades 2 and 6), and C_2 must have arisen at least twice (in clades 6 and 8, and perhaps separately in clade 10).

Detailed anatomical and enzyme activity studies on two C_4 taxa in clade 7b demonstrate presence of the NADP-ME C_4 system, with both chloroplasts and mitochondria located centripetally in the bundle sheaths. The C_2 species *E. greggii* (clade 10) shows elevated NADP-malic enzyme, and all intermediates studied show a tendency for centripetal chloroplasts and mitochondria in bundle sheaths (Muhaidat, Sage & Dengler, 2007; Muhaidat *et al.*, 2011). All Kranz-anatomy *Euploca* spp. sectioned by Frohlich (1978: 65) showed centripetal chloroplasts, including at least one representative of each C_4 clade. The C_4 *Euploca* spp. may all be of the NADP-ME subtype.

If it becomes possible to identify specific genetic changes that result in C_4, C_2 and/or proto-Kranz physiologies, then the above question of carbon fixation system along the backbone may well be resolved by identifying instances of parallelism, convergence and reversion among these *Euploca* species: diverse DNA base changes could probably have generated similar physiologies, so complex, identical changes would suggest synapomorphy, whereas different base changes resulting in similar physiologies suggest parallelism/convergence. Reversion from C_4 would be unlikely to undo all the changes by which C_4 had arisen; retention of some C_4 sequence substitutions would identify revertants.

Sage *et al.* (2018) interpreted the habitat distributions of these C_3 and C_2 species as indications of the ecological drivers of C_4 evolution, namely, those promoting increased photorespiration induced by drought and heat stress. The C_3 species *E. fallax*, *E. sessei*, *E. mexicana* and *E. torreyi* are long-lived, evergreen shrubs of old growth communities, presumably photosynthetic throughout the year, with the first three on dry limestone in the central highlands of Mexico, whereas the last is found further north in Mexico and Texas on rocky, semi-arid scrublands. *Euploca tenella* is a summer annual of the middle of North America; it is the northern-most *Euploca* sp., but it grows on exposed, shallow soils where the bedrock is at or close to the surface; these soils are prone to episodic drought. *Euploca filiformis* is a small annual herb that grows in wet (often flooded) soil in the hot tropics, often in disturbed habitats, even in roadside ditches. A preliminary investigation of this species indicates it is fully C_3 (Sage & Busch, unpublished). The proto-Kranz *E. karwinskyi* is an evergreen shrub on limestone that is a local endemic at high elevations in Tamaulipas, Mexico, where it grows with *E. mexicana*. (and the C_4 summer annual *E. confertifolia*). *Euploca procumbens* is a common, widespread annual weed of diverse habitats.

The C_2 species *E. convolvulacea* is a summer annual on exposed and often hot sand dunes from North America (Texas to California and northern Mexico; Sage *et al.*, 2018), and *E. racemosa* is a summer annual from southern Texas that is also found on sandy soils. *Euploca lagoensis* is a low-latitude ephemeral that grows on wet soils and mud along receding shores of lakes and streams in the hot dry season. *Euploca greggii* is a perennial geophyte of sandy soils in northern Mexico and Texas. The short-lived summer annual herb *E. cremnogena* has been collected only twice; the only habitat information records it in the partial shade of trees at 250 m elevation in the Tierra Caliente region.

The distinct C_4 origin in clade 7b most probably took place in the semi-arid to arid landscapes of North and Central America, as indicated by the distribution of most clade 7 species. The C_4 species in clade 9 both occur in arid South America (Argentina), indicating a C_4 origin there, probably in the sandy, arid habitat where these species now occur; these are habitats that favour C_4 plants (Sage *et al.*, 2018; Mahdavi & Bergmeier, 2018). The two C_4 species of clade 9 (*E. mendocina* and *E. chrysantha*) share an unusual feature with the C_2 species *E. greggii* (clade 10): they all spread by underground structures, shown to be stems in *E. greggii*, and form large clones, with the South American species forming small, spherical storage structures. They all have linear leaves and a low, much-branched habit. *Euploca greggii* is widespread in northern Mexico and adjacent Texas, a region with notable disjunctions to South America (e.g. *Larrea*, Zygophyllaceae and *Tiquilia* and Amsinckiinae, Boraginaceae).

The exclusively Old World distributions of all species in clades 11 and higher suggest the C_4 origin(s) inferred at the base of clade 12, and possibly in clade 11, occurred in the Old World, possibly Africa, where the C_4 species of clades 11 and 12 occur. *Euploca baclei* and *E. katangensis* of Africa are not C_4 (not available for our phylogenetic studies). Alternatively, there could be a single origin on the backbone below clade 9, with radiation to the Old World beginning with the C_4 species of clades 11 and 12, and possible reversion from C_4 to C_2 in clade 10, and from C_4 to C_3 (*E. nigricans*) in clade 11. *Euploca nigricans* is a clear candidate for reversion because it is restricted to Socotra, with all reasonable source populations on the African or Asian mainland being C_4 *Euploca* species from clades 11 and higher. *Euploca nigricans* has been observed by one of us (AF); it is found in two habitats on Socotra: as an understory shrub in *Dracaena* woodland and on the

south-west escarpment in semi-deciduous woodland on limestone. Both areas harvest fog and mist of the south-west monsoon.

The phylogenetic tree also clarifies how *Euploca* spp. spread across the globe, and the role C_4 may have played. In southern North America, where C_4 arose in clade 7b, there is a marked diversification of C_4 species across the dry landscapes of Mexico, the USA and the Caribbean basin. The innovation of C_4 is correlated with and may have promoted the radiation of clade 7b, possibly by rapid evolution into new ecological niches as observed in other C_4 clades such as the African grass *Alloteropsis semialata* (Lundgren *et al.*, 2015). The burst of species diversification in clades 13 and higher represent diversification following dispersal to Australia. Multiple polytomies, short-branch lengths and little change in the DNA sequences in the Australian clades are evidence for rapid species radiation following the arrival of the first C_4 *Euploca* species in Australia. As the driest continent, Australia is surprisingly lacking in taxa that evolved C_4 in Australia, with only a few species-poor clades identified (Sage, 2016). These observations support a scenario where Australia has many dry niches that were poorly filled by the native flora, allowing the founding population of a C_4 *Euploca* spp. to rapidly radiate upon arrival from a source population in either Africa or Asia. A similar scenario is also evident in C_4 species of *Gomphrena* L. (Amaranthaceae), which rapidly radiated following a founding event (Sage *et al.*, 2007). In contrast, C_3 *Euploca* spp., relatives of *E. glandulifera* and *E. ovalifolia* of clade 2, did not radiate extensively following introduction to Australia.

CONCLUSIONS

Our tree resolves patterns of evolutionary innovation in *Euploca* because we included numerous species, especially species showing divergent physiological and morphological attributes of interest, supplemented by species from the worldwide range of the genus. The existence of four distinct C_4 *Euploca* clades separated by non-C_4 taxa implies four acquisitions of C_4, or fewer acquisitions with some losses. The scattered placements of confirmed C_3, C_2 and proto-Kranz taxa also implies multiple origins/losses of these carbon fixation systems as well. This documents a large amount of parallel/convergent evolution in the carbon fixation systems of *Euploca*, making *Euploca* an ideal evo-devo system to unravel how these carbon fixation systems evolve. In spite of living in dry environments similar to those of *Euploca*, C_4 photosynthesis has not evolved in the species-rich and highly diverse *Heliotropium s.s.*, even in the deserts of south-western Asia, and it has not so far been found among the many species of the traditional *Tournefortia* or other genera of Boraginaceae. It is mysterious why such extensive parallel and convergent evolution on carbon fixation pathways should have happened only in *Euploca*.

The strange inflorescence plants provide another example of remarkable parallel evolution: the three confirmed inflorescence plants are widely separated on the cladogram, clearly showing independent origins of this unusual morphology (and there are hints of more instances among the Australian taxa). Also, plants that show attributes tending towards this morphology are not phylogenetically close to the inflorescence plants, and furthermore, there is *E. confertifolia*, which has yet another peculiar inflorescence morphology. Other features discussed here, including breeding system, plant duration and a special corolla trichome type, also show much parallelism/convergence.

Such instances of frequent parallel/convergent evolution of some features (but not of others) in a small group support a hypothesis from the era of the Modern Synthesis, that some groups of organisms have a 'tendency to evolve' particular novel features (Frohlich, 2006). The idea of a tendency to evolve was largely rejected as cladistics became widely accepted because it runs counter to the doctrine of parsimony, but the evidence here of extensive parallelism/convergences in *Euploca*, but not in related genera, of C_4 photosynthesis and inflorescence-plant morphology, suggests *Euploca* does have a 'tendency to evolve' those features. Tendencies to evolve apply only to specific, limited attributes of some groups and do not imply rapid evolution of all attributes, or of similar attributes in all groups. For example, fruit structure in *Euploca* is uniform, whereas fruit structure is highly variable in *Heliotropium s.s.* Presumably, such tendencies to evolve reflect both similar adaptive pressure on multiple members of the affected group and, more importantly, a degree of evolutionary flexibility that allows members of the group to respond to evolutionary pressure in similar ways, repeatedly, in different clades in the group. This sort of 'tendency to evolve' is not the same as the nineteenth/early twentieth century idea of orthogenesis, which posits an inherent, near-universal tendency for organisms to evolve in similar ways, (i.e. 'directed evolution') due to inherent attributes of organisms that cause morphological change along a particular trajectory, without the agency of Darwinian natural selection (but perhaps with the inheritance of acquired characteristics) (Ulett, 2014, and references therein).

Examining the gene changes behind these repeated instances of parallelism/convergence in *Euploca* should reveal why such tendencies to evolve exist (Frohlich, 2006). The carbon fixation systems in *Euploca* are an ideal system for this because *Euploca* has so many intermediates scattered across the tree and because

so much is already known about the biochemical, physiological and morphological requirements for the C_3, C_4, C_2 and proto-Kranz carbon assimilation systems.

ACKNOWLEDGEMENTS

We thank the herbaria listed here for samples for $\delta^{13}C$ determinations and for DNA extraction. We thank Tom Van Devender, Ana Lilia Reina-Guerrero, Jonathan Wendel, George Yatskievych and Denis Kearns for providing samples, Edith Kapinos and Laslo Csiba for DNA extractions, Bryn T. M. Dentinger for assistance with phylogenetic analysis programs, David Steart for searching for records of fossil *Euploca* and *Heliotropium* fruits from Australia, David Steart for searching for records of fossil *Euploca* and *Heliotropium* fruits from Australia, Saba Rokni for photographs of specimens in spite of COVID-19, Jerzy Rzedowski and especially Thomas Boone Hallberg Barlow for much assistance with fieldwork in Mexico. Funding was provided by an National Science Foundation (USA) Mid-Career fellowship BSR-8909933 to MWF, by Discovery Grant number RGPIN-2017-06476 from the Natural Sciences and Engineering Research Council of Canada to RFS, and by RBG Kew.

DATA AVAILABILITY

The data produced for these analyses are available in the GenBank Nucleotide Database at https://www.ncbi.nlm.nih.gov/nucleotide/ and can be located using the accession numbers in Table 1.

REFERENCES

Akhani H. 2007. Diversity, biogeography, and photosynthetic pathways of *Argusia* and *Heliotropium* (Boraginaceae) in South-West Asia with an analysis of phytogeographical units. *Botanical Journal of the Linnean Society* **155:** 401–425.

Akhani H, **Förther H. 1994.** The genus *Heliotropium* L. (Boraginaceae) in *Flora Iranica* area. *Sendtnera* **2:** 187–276.

Alonso-Cantabrana H, **von Caemmerer S. 2016.** Carbon isotope discrimination as a diagnostic tool for C_4 photosynthesis in C_3-C_4 intermediate species. *Journal of Experimental Botany* **67:** 3109–3121.

APG III. **2009.** An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society* **161:** 105–121.

APG IV. **2016.** An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society* **181:** 1–20.

Birecka H, **Frohlich MW**, **Glickman LM. 1983.** Free and esterified necines in *Heliotropium* species from Mexico and Texas. *Phytochemistry* **22:** 1167–1171.

Brown R. 1810. *Prodromus Florae Novae Hollandae et Insulae Van-Diemen*, Vol. 1. London: Taylor.

Buys MH, **Hilger HH. 2003.** Boraginaceae cymes are exclusively scorpioid and not helicoid. *Taxon* **52:** 719–724.

de Candolle A. 1845. *Prodromus [systematis naturalis regni vegetabilis, sive, Enumeratio contracta ordinum generum specierumque plantarum huc usque cognitarium, juxta methodi naturalis, normas digesta]*. Vol. 9. Paris: Treuttel et Würtz.

Catalfamo JL, **Frohlich MW**, **Martin WB Jr**, **Birecka H. 1982.** Necines of alkaloids in *Heliotropium* species from Mexico and the U.S.A. *Phytochemistry* **21:** 2677–2682.

Chase MW, **Hills HG. 1991.** Silica gel: an ideal material for field preservation of leaf samples for DNA studies. *Taxon* **40:** 215–220.

Costa FCP, **Melo JIM. 2019.** Boraginales (Boraginaceae *s.l.*) and Lamiales (Lamiaceae and Verbenaceae) in a conservation area in the semiarid region of northeastern Brazil. *Rodriguésia* **70:** e01472017.

Costion C, **Ford A**, **Cross H**, **Crayn D**, **Harrington M**, **Lowe A. 2011.** Plant DNA barcodes can accurately estimate species richness in poorly known floras. *PLoS One* **6:** e26841.

Craven LA. 1996. A taxonomic revision of *Heliotropium* (Boraginaceae) in Australia. *Australian Systematic Botany* **9:** 521–657.

Craven LA. 2005. Seven new species of *Heliotropium* (Boraginaceae) from the monsoon and arid zones of Australia. *The Beagle* **21:** 11–25.

Cruden RW. 1977. Pollen-ovule ratios: a conservative indicator of breeding systems in flowering plants. *Evolution* **31:** 32–46.

Di Fulvio TE, **Ariza Espinar L. 2016.** Las especies argentinas de *Heliotropium* (Boraginaceae). *Boletin de la Sociedad Argentina de Botanica* **51:** 745–787.

Diane N, **Förther H**, **Hilger HH. 2002.** A systematic analysis of *Heliotropium*, *Tournefortia* and allied taxa of the Heliotropiaceae (Boraginales) based on ITS1 sequences and morphological data. *American Journal of Botany* **89:** 287–295.

Diane N, **Hilger HH**, **Förther H**, **Weigend M**, **Luebert F. 2016.** Heliotropiaceae. In: Kadereit JW, Bittrich V, eds. *The families and genera of vascular plants*. Cham: Springer International Publishing, 203–211.

Doyle JJ, **Doyle JL. 1987.** A rapid isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin, Botanical Society of America* **19:** 11–15.

Doyle JJ, **Doyle JL**, **Brown AHD. 1990.** A chloroplast-DNA phylogeny of the wild perennial relatives of soybean (*Glycine* subgenus *Glycine*): congruence with morphological and crossing groups. *Evolution* **44:** 371–389.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32:** 1792–1797.

Endress P. 2010. Disentangling confusions in inflorescence morphology: patterns and diversity of reproductive shoot

ramification in angiosperms. *Journal of Systematics and Evolution* **48:** 225–239.

Förther H. 1998. Die infragenerische Gliederung der Gattung *Heliotropium* L. und ihre Stellung innerhalb der subfam. Heliotropioideae (Schrad.) Arn. (Boraginaceae). *Sendtnera* **5:** 35–241.

Frohlich MW. 1978. *Systematics of* Heliotropium *section* Orthostachys *in Mexico*. Unpublished PhD thesis, Harvard University. Available from ResearchGate.

Frohlich MW. 2006. Recommendations and goals for evo-devo research: scenarios, genetic constraint, and developmental homeostasis. In: Columbus JT, Friar EA, Porter JM, Prince LM, Simpson MG, eds. *Monocots comparative biology and evolution excluding Poales*. Claremont: Rancho Santa Ana.

Frohlich MW, **Thulin M**, **Chase MW. 2020.** Ninety-three new combinations in *Euploca* for species of *Heliotropium* section *Orthostachys* (Boraginaceae *sensu* APG). *Phytotaxa* **434:** 13–21.

GPWG: Grass Phylogeny Working Group II. **2012.** New grass phylogeny resolves deep evolutionary relationships and discovers C_4 origins. *New Phytologist* **193:** 304–312.

Guilliams CM, **Hasenstab-Lehman KE**, **Mabry ME**, **Simpson MG. 2017.** Memoirs of a frequent flier: phylogenomics reveals 18 long-distance dispersals between North America and South America in the popcorn flowers (Amsinckiinae). *American Journal of Botany* **104:** 1717–1728.

Hilger HH, **Diane N. 2003.** A systematic analysis of Heliotropiaceae (Boraginales) based on *trnL* and ITS1 sequence data. *Botanische Jahrbücher für Systematik, Pflanzengeschichte und Pflanzengeographie* **125:** 19–51.

Johnston IM. 1928. Studies in the Boraginaceae VII: the South American species of *Heliotropium*. *Contributions from Gray Herbarium of Harvard University* **81:** 14–73.

Johnston IM. 1937. Studies in the Boraginaceae XII: novelties and critical notes. *Journal of the Arnold Arboretum* **18:** 10–25.

Johnston IM. 1939. *Studies in the Boraginaceae, XIII, new or otherwise noteworthy species, chiefly from Western United States. Journal of the Arnold Arboretum* **20:** 376–377.

Jordan GJ. 2001. An investigation of long-distance dispersal based on species native to both Tasmania and New Zealand. *Australian Journal of Botany* **49:** 333–340.

Kadereit G, **Ackerly D**, **Pirie MD. 2012.** A broader model for C_4 photosynthesis evolution in plants inferred from the goosefoot family (Chenopodiaceae *s.s.*). *Proceedings of the Royal Society of London B: Biological Sciences* **279:** 3304–3311.

Lia VV, **Confalonieri VA**, **Comas CI**, **Hunziker JH. 2001.** Molecular phylogeny of *Larrea* and its allies (Zygophyllaceae): reticulate evolution and the probable time of creosote bush arrival to North America. *Molecular Phylogenetics and Evolution* **21:** 309–320.

Linnaeus C. 1753. *Species plantarum*. Stockholm: L. Salvius.

Lledó MD, **Crespo MB**, **Cameron KM**, **Fay MF**, **Chase MW. 1998.** Systematics of Plumbaginaceae based on cladistic analysis of *rbcL* sequence data. *Systematic Botany* **23:** 21–29.

Luebert F. 2013. A revision of *Heliotropium* sect. *Cochranea* (Heliotropiaceae). *Kew Bulletin* **68:** 1–54.

Luebert F, **Brokamp G**, **Wen J**, **Weigend M**, **Hilger HH. 2011a.** Phylogenetic relationships and morphological diversity in Neotropical *Heliotropium* (Heliotropiaceae). *Taxon* **60:** 663–680.

Luebert F, **Cecchi L**, **Frohlich MW**, **Gottschling M**, **Guilliams CM**, **Hasenstab-Lehman KE**, **Hilger HH**, **Miller JS**, **Mittelbach M**, **Nazaire M**, **Nepi M**, **Nocentini D**, **Ober D**, **Olmstead RG**, **Selvi F**, **Simpson MG**, **Sutorý K**, **Valdés B**, **Walden GK**, **Weigend M. 2016.** Familial classification of the Boraginales. *Taxon* **65:** 502–522.

Luebert F, **Hilger HH**, **Weigend M. 2011b.** Diversification in the Andes: age and origins of South American *Heliotropium* lineages (Heliotropiaceae, Boraginales). *Molecular Phylogenetics and Evolution* **61:** 90–102.

Lundgren MR, **Besnard G**, **Ripley BS**, **Lehmann CER**, **Chatelet DS**, **Kynest RG**, **Namaganda M**, **Vorontsova MS**, **Hall RC**, **Elia J**, **Osborne CP**, **Christin PA. 2015.** Photosynthetic innovation broadens the niche within a single species. *Ecology Letters* **18:** 1021–1029.

Maddison WP. 1993. Missing data versus missing characters in phylogenetic analysis. *Systematic Biology* **42:** 576–581.

Maddison DR, **Maddison WP. 2000.** *MacClade 4: analysis of phylogeny and character evolution. Version 4.0*. Sunderland: Sinauer Associates.

Mahdavi P, **Bergmeier E. 2018.** Distribution of C_4 plants in sand habitats of different climatic regions. *Folia Geobotanica* **53:** 201–211.

McKown AD, **Moncalvo JM**, **Dengler NG. 2005.** Phylogeny of *Flaveria* (Asteraceae) and inference of C_4 photosynthesis evolution. *American Journal of Botany* **92:** 1911–1928.

Melo JIM, **Semir J. 2010.** Taxonomia do gênero *Euploca* Nutt. (Heliotropiaceae) no Brasil. *Acta Botanica Brasilica* **24:** 111–132.

Monson RK, **Teeri JA**, **Ku MSB**, **Gurevitch J**, **Mets LJ**, **Dudley S. 1988.** Carbon-isotope discrimination by leaves of *Flaveria* species exhibiting different amounts of C_3-cycle and C4-cycle co-function. *Planta* **174:** 145–151.

Moore MJ, **Tye A**, **Jansen RK. 2006.** Patterns of long-distance dispersal in *Tiquilia* subg. *Tiquilia* (Boraginaceae): implications for the origins of amphitropical disjuncts and Galápagos Islands endemics. *American Journal of Botany* **93:** 1163–1177.

Muasya AM, **Simpson DA**, **Chase MW**, **Culham A. 1998.** An assessment of suprageneric phylogeny in Cyperaceae using *rbcL* DNA sequences. *Plant Systematics and Evolution* **211:** 257–271.

Muhaidat R, **Sage RF**, **Dengler NG. 2007.** Diversity of Kranz anatomy and biochemistry in C_4 eudicots. *American Journal of Botany* **94:** 362–381.

Muhaidat R, **Sage TL**, **Frohlich MW**, **Dengler NG**, **Sage RF. 2011.** Characterization of C_3–C_4 intermediate species in the genus *Heliotropium* L. (Boraginaceae): anatomy, ultrastructure and enzyme activity. *Plant, Cell & Environment* **34:** 1723–1736.

Müller K. 2005. SeqState: primer design and sequence statistics for phylogenetic DNA datasets. *Applied Bioinformatics* **4:** 65–69.

Nazaire M, **Hufford L. 2012.** A broad phylogenetic analysis of Boraginaceae: implications for the relationships of *Mertensia*. *Systematic Botany* **37:** 758–783.

Nuttall T. 1837. Collections toward a flora of the territory of Arkansas. *Transactions of the American Philosophical Society, series 2* **5:** 189–190.

Oakley JC, **Sultmanis S**, **Stinson CR**, **Sage TL**, **Sage RF. 2014.** Comparative studies of C_3 and C_4 *Atriplex* hybrids in the genomics era: physiological assessments. *Journal of Experimental Botany* **65:** 3637–3647.

Rawsthorne S. 1992. C_3-C_4 intermediate photosynthesis: linking physiology to gene expression. *Plant Journal* **2:** 267–274.

Reeves G, **Chase MW**, **Goldblatt P**, **Rudall P**, **Fay MF**, **Cox AV**, **Lejeune B**, **Souza-Chies T. 2001.** Molecular systematics of Iridaceae: evidence from four plastid DNA regions. *American Journal of Botany* **88:** 2074–2087.

Reveal JL, **Chase MW. 2011.** APG III: bibliographical information and synonymy of Magnoliidae. *Phytotaxa* **19:** 71–134.

Rudov A, **Mashkour M**, **Djamali M**, **Akhani H. 2020.** A review of C_4 plants in southwest Asia: an ecological, geographical and taxonomical analysis of a region with high diversity of C_4 eudicots. *Frontiers in Plant Science* **11:** 546518.

Sage RF. 2016. A portrait of the C_4 photosynthetic family on the 50th anniversary of its discovery: species number, evolutionary lineages, and hall of fame. *Journal of Experimental Botany* **67:** 4039–4056.

Sage RF, **Khoshravesh R**, **Sage TL. 2014.** From proto-Kranz to C_4 Kranz: building the bridge to C_4 photosynthesis. *Journal of Experimental Botany* **65:** 3341–3356.

Sage RF, **Monson RK**, **Ehleringer JR**, **Adachi S**, **Pearcy RW. 2018.** Some like it hot: the physiological ecology of C_4 plant evolution. *Oecologia* **187:** 941–966.

Sage RF, **Sage TL**, **Kocacinar F. 2012.** Photorespiration and the evolution of C_4 photosynthesis. *Annual Review of Plant Biology* **63:** 19–47.

Sage RF, **Sage TL**, **Pearcy RW**, **Borsch T. 2007.** The taxonomic distribution of C_4 photosynthesis in Amaranthaceae *sensu stricto*. *American Journal of Botany* **94:** 1992–2003.

Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. *Systematic Biology* **5:** 492–508.

Shimodiara H, **Hasegawa M. 1999.** Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution* **16:** 1114–1116.

Simmons MP, **Ochoterena H. 2000.** Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology* **49:** 369–381.

Simons ELAN, **Wieringa JJ. 2019.** The *Euploca baclei* complex (Boraginaceae subfam. Heliotropioideae). *Blumea* **64:** 92–95.

Slayter EM. 1970. *Optical methods in biology*. New York: Wiley.

Small JK. 1933. *Manual of the southeastern flora*. Chapel Hill: University of North Carolina Press.

Stata M, **Sage TL**, **Sage RF. 2019.** Mind the gap: the evolutionary engagement of the C_4 metabolic cycle in support of net carbon assimilation. *Current Opinion in Plant Biology* **49:** 27–34.

Sun Y, **Skinner DZ**, **Liang GH**, **Hulbert SH. 1994.** Phylogenetic analysis of *Sorghum* and related taxa using internal transcribed spacers of nuclear ribosomal DNA. *Theoretical and Applied Genetics* **89:** 26–32.

Swofford DL. 2003. *PAUP* Phylogenetic analysis using parsimony (*and other methods), v. 4.0 beta 10*. Sunderland: Sinauer Associates.

Thulin M. 2005. Three new species of *Heliotropium* (Boraginaceae) from the Horn of Africa region. *Nordic Journal of Botany* **23:** 527–532.

Thulin M. 2006. *Flora of Somalia*. Vol. 3. Kew: Royal Botanic Gardens.

Turner BL. 2016. Distribution of *Euploca confertifolia* (Boraginaceae), including its two varietal components. *Phytologia* **98:** 311–312.

Ulett MA. 2014. Making the case for orthogenesis: the popularization of definitely directed evolution (1890–1926). *Studies in History and Philosophy of Biological and Biomedical Sciences* **45:** 124–132.

Vogan PJ, **Frohlich MW**, **Sage RF. 2007.** The functional significance of C_3-C_4 intermediate traits in *Heliotropium* L. (Boraginaceae): gas exchange perspectives. *Plant, Cell and Environment* **30:** 1337–1345.

Wang X, **Wen M**, **Wu M**, **Xu W**, **Zhang K**, **Zhang D. 2020.** Gynodioecy or leaky dioecy? The unusual sexual system of a coral dune-habitant *Tournefortia argentea* (Boraginaceae). *Phytologia* **98:** 311–312.

Weigend M, **Luebert F**, **Gottschling M**, **Couvreur TLP**, **Hilger HH**, **Miller JS. 2014.** From capsules to nutlets – phylogenetic relationships in the Boraginales. *Cladistics* **30:** 508–518.

Yang Y, **Berry PE. 2011.** Phylogenetics of the *Chamaesyce* clade (*Euphorbia*, Euphorbiaceae): reticulate evolution and long-distance dispersal in a prominent C_4 lineage. *American Journal of Botany* **98:** 1486–1503.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site.

Table S1. List of samples analysed for $\delta^{13}C$.

Figure S1. Parsimony bootstrap of RFLP data from 1989. The *Euploca* portion is fully compatible with results from sequence data; clades 2, 4, 7 and 13 fall in the same order on the *Euploca* backbone. Sample numbers are MWF collection numbers.